Wiley Series in Discrete Mathematics and Optimization

Fourth Edition

THE PROBABILISTIC METHOD

NOGA ALON • JOEL H. SPENCER

WILEY

THE PROBABILISTIC METHOD

WILEY SERIES IN
DISCRETE MATHEMATICS AND OPTIMIZATION

A complete list of titles in this series appears at the end of this volume.

THE PROBABILISTIC METHOD

Fourth edition, July 2015, Tel Aviv and New York

NOGA ALON
School of Mathematics,
Raymond and Beverly Sackler Faculty of Exact Sciences,
Tel Aviv University,
Tel Aviv, Israel.

JOEL H. SPENCER
Courant Institute of Mathematical Sciences,
New York University,
New York, USA

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Alon, Noga.
The probabilistic method / Noga Alon, Joel H. Spencer. – Fourth edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-119-06195-3 (cloth)
1. Combinatorial analysis. 2. Probabilities. I. Spencer, Joel H. II. Title.
QA164.A46 2016
511′.6–dc23

2015021599

Typeset in 10/12pt TimesLtStd by SPi Global, Chennai, India.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

4 2016

To Nurit and Mary Ann

Contents

Preface

The probabilistic method is one of the most powerful and widely used tools applied in combinatorics. One of the major reasons for its rapid development is the important role of randomness in theoretical computer science and in statistical physics.

The interplay between discrete mathematics and computer science suggests an algorithmic point of view in the study of the probabilistic method in combinatorics, and this is the approach we tried to adopt in this book. The book thus includes a discussion of algorithmic techniques together with a study of the classical method as well as the modern tools applied in it. The first part of the book contains a description of the tools applied in probabilistic arguments, including the basic techniques that use expectation and variance, as well as the more recent applications of martingales and correlation inequalities. The second part includes a study of various topics in which probabilistic techniques have been successful. This part contains chapters on discrepancy and random graphs, as well as on several areas in theoretical computer science: Circuit Complexity, Computational Geometry, Graph Property Testing and, Derandomization of randomized algorithms. Scattered between the chapters are gems described under the heading "The Probabilistic Lens." These are elegant proofs that are not necessarily related to the chapters after which they appear and can usually be read separately.

The basic probabilistic method can be described as follows: In order to prove the existence of a combinatorial structure with certain properties, we construct an appropriate probability space and show that a randomly chosen element in this space has the desired properties with positive probability. This method was initiated by Paul Erdős, who contributed so much to its development over a 50-year period that it seems appropriate to call it "The Erdős Method." His contribution can be measured not only by his numerous deep results in the subject but also by the many intriguing problems and conjectures posed by him that stimulated a big portion of the research in the area.

It seems impossible to write an encyclopedic book on the probabilistic method; too many recent interesting results apply probabilistic arguments, and we do not even try

to mention all of them. Our emphasis is on methodology, and we thus try to describe the ideas, and not always to give the best possible results if these are too technical to allow a clear presentation. Many of the results are asymptotic, and we use the standard asymptotic notation: for two functions f and g, we write $f = O(g)$ if $f \leq cg$ for all sufficiently large values of the variables of the two functions, where c is an absolute positive constant. We write $f = \Omega(g)$ if $g = O(f)$ and $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$. If the limit of the ratio f/g tends to zero as the variables of the functions tend to infinity we write $f = o(g)$. Finally, $f \sim g$ denotes that $f = (1 + o(1))g$; that is, f/g tends to 1 when the variables tend to infinity. Each chapter ends with a list of exercises. The more difficult ones are marked by $(*)$. The exercises enable readers to check their understanding of the material and also provide the possibility of using the book as a textbook.

This is the fourth edition of the book; it contains several improved results and covers various additional topics that developed extensively during the last few years. A breakthrough approach to the Local Lemma is described in Chapter 5. A new algorithmic approach to the "six standard deviations" result in discrepancy theory is presented in Chapter 12. A novel proof for the study of Property B, based on a random greedy coloring, appears in Chapter 3. In all the above cases, the algorithmic proofs provide essentially new arguments for the existence of the desired objects. A new, short section on graph limits has been added to Chapter 9. A technique for counting independent sets in graphs and its application in a graph coloring problem is described in Chapter 1. Further additions include a new Probabilistic Lens, several additional exercises, and a new appendix with hints to selected exercises.

As in the previous editions, it is a special pleasure to thank our wives, Nurit and Mary Ann. Their patience, understanding, and encouragement have been key ingredients in the success of this enterprise.

Noga Alon
Joel H. Spencer
Tel Aviv and New York, 2015

Acknowledgments

We are very grateful to all our students and colleagues who contributed to the creation of this fourth edition through joint research, helpful discussions, and useful comments. These include Simon Blackburn, Miklós Bóna, Steve Cook, Ehud Friedgut, Oded Goldreich, Omri Ben-Eliezer, Krzysztof Choromanski, Oliver Cooley, Ohad Feldheim, Naomi Feldheim, Asaf Ferber, Laura Florescu, Lior Gishbboliner, Matan Harel, Danny Hefetz, Timo Hirscher, Rani Hod, Mihyun Kang, Joel Lewis, Nati Linial, Guy Moshkovitz, Dhruv Mubayi, Tahl Nowik, Roberto Oliveira, Ron Peled, Will Perkins, Oliver Riordan, Guy Rutenberg, Jeffrey Shallit, Asaf Shapira, Clara Shikhelman, Philipp Sprüssel, Aravind Srinivasan, John Steinberger, Elmar Teufl, Shai Vardi, Amit Weinstein, Jed Yang, Mariano Zelke and Yufei Zhao, who pointed out various inaccuracies and misprints, and suggested improvements in the presentation as well as in the results. Needless to say, the responsibility for the remaining mistakes, as well as the responsibility for the (hopefully not many) new ones, is solely ours.

PART I

METHODS

1

The Basic Method

What you need is that your brain is open.
–Paul Erdős

1.1 THE PROBABILISTIC METHOD

The probabilistic method is a powerful tool for tackling many problems in discrete mathematics. Roughly speaking, the method works as follows: trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in these structures with positive probability. The method is best illustrated by examples. Here is a simple one. The *Ramsey number* $R(k, \ell)$ is the smallest integer n such that in any two-coloring of the edges of a complete graph on n vertices K_n by red and blue, either there is a red K_k (i.e., a complete subgraph on k vertices all of whose edges are colored red) or there is a blue K_ℓ. Ramsey (1929) showed that $R(k, \ell)$ is finite for any two integers k and ℓ. Let us obtain a lower bound for the diagonal Ramsey numbers $R(k, k)$.

Proposition 1.1.1 *If* $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$, *then* $R(k, k) > n$. *Thus* $R(k, k) > \lfloor 2^{k/2} \rfloor$ *for all* $k \geq 3$.

Proof. Consider a random two-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equally likely. For any

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

fixed set R of k vertices, let A_R be the event that the induced subgraph of K_n on R is *monochromatic* (i.e., that either all its edges are red or they are all blue). Clearly, $\Pr[A_R] = 2^{1-\binom{k}{2}}$. Since there are $\binom{n}{k}$ possible choices for R, the probability that at least one of the events A_R occurs is at most $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$. Thus, with positive probability, no event A_R occurs and there is a two-coloring of K_n without a monochromatic K_k; that is, $R(k,k) > n$. Note that if $k \geq 3$ and we take $n = \lfloor 2^{k/2} \rfloor$, then

$$\binom{n}{k} 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \cdot \frac{n^k}{2^{k^2/2}} < 1$$

and hence $R(k,k) > \lfloor 2^{k/2} \rfloor$ for all $k \geq 3$. ■

This simple example demonstrates the essence of the probabilistic method. To prove the existence of a good coloring, we do not present one explicitly, but rather show, in a nonconstructive way, that it exists. This example appeared in a paper of P. Erdős from 1947. Although Szele had applied the probabilistic method to another combinatorial problem, mentioned in Chapter 2, already in 1943, Erdős was certainly the first to understand the full power of this method and apply it successfully over the years to numerous problems. One can, of course, claim that the probability is not essential in the proof given above. An equally simple proof can be described by counting; we just check that the total number of two-colorings of K_n is larger than the number of those containing a monochromatic K_k.

Moreover, since the vast majority of the probability spaces considered in the study of combinatorial problems are finite, this claim applies to most of the applications of the probabilistic method in discrete mathematics. Theoretically, this is indeed the case. However, in practice the probability is essential. It would be hopeless to replace the applications of many of the tools appearing in this book, including, for example, the second moment method, the Lovász Local Lemma and the concentration via martingales by counting arguments, even when these are applied to finite probability spaces.

The probabilistic method has an interesting algorithmic aspect. Consider, for example, the proof of Proposition 1.1.1, which shows that there is an edge two-coloring of K_n without a monochromatic $K_{2\log_2 n}$. Can we actually find such a coloring? This question, as asked, may sound ridiculous; the total number of possible colorings is finite, so we can try them all until we find the desired one. However, such a procedure may require $2^{\binom{n}{2}}$ steps; an amount of time that is exponential in the size $\left[= \binom{n}{2}\right]$ of the problem. Algorithms whose running time is more than polynomial in the size of the problem are usually considered impractical. The class of problems that can be solved in polynomial time, usually denoted by **P** (see, e.g., Aho, Hopcroft and Ullman (1974)), is, in a sense, the class of all solvable problems. In this sense, the exhaustive search approach suggested above for finding a good coloring of K_n is not acceptable, and this is the reason for our remark that the proof of Proposition 1.1.1 is nonconstructive; it does not supply a constructive, efficient,

and deterministic way of producing a coloring with the desired properties. However, a closer look at the proof shows that, in fact, it can be used to produce, effectively, a coloring that is very likely to be good. This is because, for large k, if $n = \lfloor 2^{k/2} \rfloor$, then

$$\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!}\left(\frac{n}{2^{k/2}}\right)^k \leq \frac{2^{1+\frac{k}{2}}}{k!} \ll 1.$$

Hence, a random coloring of K_n is very likely not to contain a monochromatic $K_{2\log n}$. This means that if, for some reason, we *must* present a two-coloring of the edges of K_{1024} without a monochromatic K_{20}, we can simply produce a random two-coloring by flipping a fair coin $\binom{1024}{2}$ times. We can then deliver the resulting coloring safely; the probability that it contains a monochromatic K_{20} is less than $2^{11}/20!$, probably much smaller than our chances of making a mistake in any rigorous proof that a certain coloring is good! Therefore, in some cases the probabilistic, nonconstructive method does supply effective probabilistic algorithms. Moreover, these algorithms can sometimes be converted into deterministic ones. This topic is discussed in some detail in Chapter 16.

The probabilistic method is a powerful tool in combinatorics and graph theory. It is also extremely useful in number theory and in combinatorial geometry. More recently, it has been applied in the development of efficient algorithmic techniques and in the study of various computational problems. In the rest of this chapter, we present several simple examples that demonstrate some of the broad spectrum of topics in which this method is helpful. More complicated examples, involving various more delicate probabilistic arguments, appear in the rest of the book.

1.2 GRAPH THEORY

A *tournament* on a set V of n players is an orientation $T = (V, E)$ of the edges of the complete graph on the set of vertices V. Thus for every two distinct elements x and y of V, either (x, y) or (y, x) is in E, but not both. The name "tournament" is natural, since one can think of the set V as a set of players in which each pair participates in a single match, where (x, y) is in the tournament iff x beats y. We say that T has the property S_k if, for every set of k Players, there is one that beats them all. For example, a directed triangle $T_3 = (V, E)$, where $V = \{1, 2, 3\}$ and $E = \{(1, 2), (2, 3), (3, 1)\}$, has S_1. Is it true that for every finite k there is a tournament T (on more than k vertices) with the property S_k? As shown by Erdős (1963b), this problem, raised by Schütte, can be solved almost trivially by applying probabilistic arguments. Moreover, these arguments even supply a rather sharp estimate for the minimum possible number of vertices in such a tournament. The basic (and natural) idea is that, if n is sufficiently large as a function of k, then a *random* tournament on the set $V = \{1, \ldots, n\}$ of n players is very likely to have the property S_k. By a random tournament we mean here a tournament T on V obtained by choosing, for each $1 \leq i < j \leq n$, independently, either the edge (i, j) or the edge (j, i), where each of these two choices is equally

likely. Observe that in this manner, all the $2^{\binom{n}{2}}$ possible tournaments on V are equally likely; that is, the probability space considered is symmetric. It is worth noting that we often use in applications symmetric probability spaces. In these cases, we shall sometimes refer to an element of the space as a *random element*, without describing explicitly the probability distribution . Thus, for example, in the proof of Proposition 1.1.1 random two-colorings of K_n were considered; that is, all possible colorings were equally likely. Similarly, in the proof of the next simple result we study random tournaments on V.

Theorem 1.2.1 *If* $\binom{n}{k}(1-2^{-k})^{n-k} < 1$*, then there is a tournament on n vertices that has the property* S_k.

Proof. Consider a random tournament on the set $V = \{1, \ldots, n\}$. For every fixed subset K of size k of V, let A_K be the event that there is no vertex that beats all the members of K. Clearly, $\Pr[A_K] = (1-2^{-k})^{n-k}$. This is because, for each fixed vertex $v \in V - K$, the probability that v does not beat all the members of K is $1-2^{-k}$, and all these $n-k$ events corresponding to the various possible choices of v are independent. It follows that

$$\Pr\left[\bigvee_{\substack{K\subset V\\ |K|=k}} A_K\right] \le \sum_{\substack{K\subset V\\ |K|=k}} \Pr[A_K] = \binom{n}{k}(1-2^{-k})^{n-k} < 1.$$

Therefore, with positive probability, no event A_K occurs; that is, there is a tournament on n vertices that has the property S_k. ■

Let $f(k)$ denote the minimum possible number of vertices of a tournament that has the property S_k. Since $\binom{n}{k} < \left(\frac{en}{k}\right)^k$ and $(1-2^{-k})^{n-k} < e^{-(n-k)/2^k}$, Theorem 1.2.1 implies that $f(k) \le k^2 \cdot 2^k \cdot (\ln 2)(1+o(1))$. It is not too difficult to check that $f(1) = 3$ and $f(2) = 7$. As proved by Szekeres (cf. Moon (1968)), $f(k) \ge c_1 \cdot k \cdot 2^k$.

Can one find an explicit construction of tournaments with at most c_2^k vertices having property S_k? Such a construction is known but is not trivial; it is described in Chapter 9.

A *dominating set* of an undirected graph $G = (V, E)$ is a set $U \subseteq V$ such that every vertex $v \in V - U$ has at least one neighbor in U.

Theorem 1.2.2 *Let* $G = (V, E)$ *be a graph on n vertices, with minimum degree* $\delta > 1$. *Then G has a dominating set of at most* $n\dfrac{1+\ln(\delta+1)}{\delta+1}$ *vertices.*

Proof. Let $p \in [0, 1]$ be, for the moment, arbitrary. Let us pick, randomly and independently, each vertex of V with probability p. Let X be the (random) set of all vertices picked and let $Y = Y_X$ be the random set of all vertices in $V - X$ that do not have any neighbor in X. The expected value of $|X|$ is clearly np. For each fixed vertex $v \in V$,

$\Pr[v \in Y] = \Pr[v$ and its neighbors are not in $X] \leq (1-p)^{\delta+1}$. Since the expected value of a sum of random variables is the sum of their expectations (even if they are not independent) and since the random variable $|Y|$ can be written as a sum of n indicator random variables χ_v ($v \in V$), where $\chi_v = 1$ if $v \in Y$ and $\chi_v = 0$ otherwise, we conclude that the expected value of $|X| + |Y|$ is at most $np + n(1-p)^{\delta+1}$. Consequently, there is at least one choice of $X \subseteq V$ such that $|X| + |Y_X| \leq np + n(1-p)^{\delta+1}$. The set $U = X \cup Y_X$ is clearly a dominating set of G whose cardinality is at most this size.

The above argument works for any $p \in [0,1]$. To optimize the result we use elementary calculus. For convenience, we bound $1 - p \leq e^{-p}$ (this holds for all nonnegative p and is a fairly close bound when p is small) to give the simpler bound

$$|U| \leq np + ne^{-p(\delta+1)}.$$

Take the derivative of the right-hand side with respect to p and set it equal to zero. The right-hand side is minimized at

$$p = \frac{\ln(\delta+1)}{\delta+1}.$$

Formally, we set p equal to this value in the first line of the proof. We now have $|U| \leq n\dfrac{1+\ln(\delta+1)}{\delta+1}$, as claimed. ∎

Three simple but important ideas are incorporated in the last proof. The first is the linearity of expectation; many applications of this simple, yet powerful principle appear in Chapter 2. The second is perhaps more subtle and is an example of the "alteration" principle that is discussed in Chapter 3. The random choice did not supply the required dominating set U immediately; it only supplied the set X, which has to be altered a little (by adding to it the set Y_X) to provide the required dominating set. The third involves the optimal choice of p. One often wants to make a random choice but is not certain what probability p should be used. The idea is to carry out the proof with p as a parameter giving a result that is a function of p. At the end, that p is selected which gives the optimal result. Here, there is yet a fourth idea that might be called asymptotic calculus. We want the asymptotics of $\min np + n(1-p)^{\delta+1}$, where p ranges over $[0,1]$. The actual minimum $p = 1 - (\delta+1)^{-1/\delta}$ is difficult to deal with, and in many similar cases precise minima are impossible to find in a closed form. Rather, we give away a little bit, bounding $1 - p \leq e^{-p}$, yielding a clean bound. A good part of the *art* of the probabilistic method lies in finding suboptimal but clean bounds. Did we give away too much in this case? The answer depends on the emphasis for the original question. For $\delta = 3$, our rough bound gives $|U| \leq 0.596n$, while the more precise calculation gives $|U| \leq 0.496n$, perhaps a substantial difference. For δ large, both methods give asymptotically $n \ln \delta/\delta$.

It can easily be deduced from the results in Alon (1990b) that the bound in Theorem 1.2.2 is nearly optimal. A non-probabilistic, algorithmic proof of this theorem can be obtained by choosing the vertices for the dominating set one by

one, when in each step a vertex that covers the maximum number of yet-uncovered vertices is picked. Indeed, for each vertex v, denote by $C(v)$ the set consisting of v together with all its neighbors. Suppose that during the process of picking vertices the number of vertices u that do not lie in the union of the sets $C(v)$ of the vertices chosen so far is r. By the assumption, the sum of the cardinalities of the sets $C(u)$ over all such uncovered vertices u is at least $r(\delta + 1)$, and, hence by averaging, there is a vertex v that belongs to at least $r(\delta + 1)/n$ such sets $C(u)$. Adding this v to the set of chosen vertices, we observe that the number of uncovered vertices is now at most $r(1 - (\delta + 1)/n)$. It follows that in each iteration of the above procedure the number of uncovered vertices decreases by a factor of $1 - (\delta + 1)/n$ and, hence after $n \ln(\delta + 1)/(\delta + 1)$ steps, there will be at most $n/(\delta + 1)$ yet uncovered vertices that can now be added to the set of chosen vertices to form a dominating set of size at most equal to the one in the conclusion of Theorem 1.2.2.

Combining this with some ideas of Podderyugin and Matula, we can obtain a very efficient algorithm to decide whether a given undirected graph on n vertices is, say, $n/3$ edge-connected. A *cut* in a graph $G = (V, E)$ is a partition of the set of vertices V into two nonempty disjoint sets $V = V_1 \cup V_2$. If $v_1 \in V_1$ and $v_2 \in V_2$, we say that the cut *separates* v_1 and v_2. The *size* of the cut is the number of edges of G having one end in V_1 and the other end in V_2. In fact, we sometimes identify the cut with the set of these edges. The *edge connectivity* of G is the minimum size of a cut of G. The following lemma is due to Podderyugin and Matula (independently).

Lemma 1.2.3 *Let $G = (V, E)$ be a graph with minimum degree δ, and let $V = V_1 \cup V_2$ be a cut of size smaller than δ in G. Then every dominating set U of G has vertices in V_1 and in V_2.*

Proof. Suppose this is false and $U \subseteq V_1$. Choose, arbitrarily, a vertex $v \in V_2$, and let $v_1, v_2, \ldots, v_\delta$ be δ of its neighbors. For each i, $1 \leq i \leq \delta$, define an edge e_i of the given cut as follows: if $v_i \in V_1$, then $e_i = \{v, v_i\}$, otherwise $v_i \in V_2$, and since U is dominating, there is at least one vertex $u \in U$ such that $\{u, v_i\}$ is an edge; take such a u and put $e_i = \{u, v_i\}$. The δ edges $e_1, \ldots, e_\delta$ are all distinct and all lie in the given cut, contradicting the assumption that its size is less than δ. This completes the proof. ∎

Let $G = (V, E)$ be a graph on n vertices, and suppose we wish to decide whether G is $n/3$ edge-connected; that is, whether its edge connectivity is at least $n/3$. Matula showed, by applying Lemma 1.2.3, that this can be done in time $O(n^3)$. By the remark following the proof of Theorem 1.2.2, we can slightly improve it and get an $O(n^{8/3} \log n)$ algorithm as follows. We first check if the minimum degree δ of G is at least $n/3$. If not, G is not $n/3$ edge-connected, and the algorithm ends. Otherwise, by Theorem 1.2.2, there is a dominating set $U = \{u_1, \ldots, u_k\}$ of G, where $k = O(\log n)$, and it can in fact be found in time $O(n^2)$. We now find, for each i, $2 \leq i \leq k$, the minimum size s_i of a cut that separates u_1 from u_i. Each of these problems can be solved by solving a standard network flow problem in time $O(n^{8/3})$ (see, e.g., Tarjan (1983)). By Lemma 1.2.3, the edge connectivity of G is simply the minimum between δ and $\min_{2 \leq i \leq k} s_i$. The total time of the algorithm is $O(n^{8/3} \log n)$, as claimed.

1.3 COMBINATORICS

A *hypergraph* is a pair $H = (V, E)$, where V is a finite set whose elements are called *vertices*, and E is a family of subsets of V, called *edges*. It is *n-uniform* if each of its edges contains precisely n vertices. We say that H has *property B*, or that it is two-*colorable*, if there is a two-coloring of V such that no edge is monochromatic. Let $m(n)$ denote the minimum possible number of edges of an n-uniform hypergraph that does not have property B.

Proposition 1.3.1 [Erdős (1963a)] *Every n-uniform hypergraph with less than* 2^{n-1} *edges has property B. Therefore* $m(n) \geq 2^{n-1}$.

Proof. Let $H = (V, E)$ be an n-uniform hypergraph with less than 2^{n-1} edges. Color V randomly by two colors. For each edge $e \in E$, let A_e be the event such that e is monochromatic. Clearly, $\Pr[A_e] = 2^{1-n}$. Therefore,

$$\Pr\left[\bigvee_{e\in E} A_e\right] \leq \sum_{e\in E} \Pr[A_e] < 1$$

and there is a two-coloring without monochromatic edges. ∎

In Section 3.6 we present a more delicate argument, due to Cherkashin and Kozik (2015), which shows that

$$m(n) \geq \Omega\left(\left(\frac{n}{\ln n}\right)^{1/2} 2^n\right).$$

The best known upper bound to $m(n)$ is found by turning the probabilistic argument "on its head." Basically, the sets become random and each coloring defines an event. Fix V with v points, where we shall later optimize v. Let χ be a coloring of V with a points in one color, $b = v - a$ points in the other. Let $S \subset V$ be a uniformly selected n-set. Then

$$\Pr[S \text{ is monochromatic under } \chi] = \frac{\binom{a}{n} + \binom{b}{n}}{\binom{v}{n}}.$$

Let us assume v is even for convenience. As $\binom{y}{n}$ is convex, this expression is minimized when $a = b$. Thus

$$\Pr[S \text{ is monochromatic under} \chi] \geq p,$$

where we set

$$p = \frac{2\binom{v/2}{n}}{\binom{v}{n}}$$

for notational convenience. Now let $S_1, \ldots, S_m$ be uniformly and independently chosen n-sets, with m to be determined. For each coloring χ, let A_χ be the event in which none of the S_i is monochromatic. By the independence of the S_i

$$\Pr[A_\chi] \le (1-p)^m.$$

There are 2^v colorings, so

$$\Pr\left[\bigvee_\chi A_\chi\right] \le 2^v(1-p)^m.$$

When this quantity is less than 1, there exist $S_1, \ldots, S_m$ so that no A_χ holds; that is, $S_1, \ldots, S_m$ is not two-colorable and hence $m(n) \le m$.

The asymptotics provide a fairly typical example of those encountered when employing the probabilistic method. We first use the inequality $1 - p \le e^{-p}$. This is valid for all positive p, and the terms are quite close when p is small. When

$$m = \left\lceil \frac{v \ln 2}{p} \right\rceil,$$

then $2^v(1-p)^m < 2^v e^{-pm} \le 1$ so $m(n) \le m$. Now we need to find v to minimize v/p. We may interpret p as twice the probability of picking n white balls from an urn with $v/2$ white and $v/2$ black balls, sampling without replacement. It is tempting to estimate p by 2^{-n+1}, the probability for sampling with replacement. This approximation would yield $m \sim v2^{n-1}(\ln 2)$. As v gets smaller, however, the approximation becomes less accurate and, as we wish to minimize m, the tradeoff becomes essential. We use a second-order approximation

$$p = \frac{2\binom{v/2}{n}}{\binom{v}{n}} = 2^{1-n}\prod_{i=0}^{n-1}\frac{v-2i}{v-i} \sim 2^{1-n}e^{-n^2/2v}$$

as long as $v \gg n^{3/2}$, estimating

$$\frac{v-2i}{v-i} = 1 - \frac{i}{v} + O\left(\frac{i^2}{v^2}\right) = e^{-i/v+O(i^2/v^2)}.$$

Elementary calculus gives $v = n^2/2$ for the optimal value. The evenness of v may require a change of at most 2, which turns out to be asymptotically negligible. This yields the following result of Erdős (1964):

Theorem 1.3.2 $m(n) < (1+o(1))\dfrac{e \ln 2}{4} n^2 2^n$.

Let $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^h$ be a family of pairs of subsets of an arbitrary set. We call $\mathcal{F}$ a (k, ℓ)-*system* if $|A_i| = k$ and $|B_i| = \ell$ for all $1 \le i \le h$, $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \ne \emptyset$ for

all distinct i, j, with $1 \le i, j \le h$. Bollobás (1965) proved the following result, which has many interesting extensions and applications:

Theorem 1.3.3 *If $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^{h}$ is a (k, ℓ)-system then $h \le \binom{k+\ell}{k}$.*

Proof. Put $X = \bigcup_{i=1}^{h}(A_i \cup B_i)$ and consider a random order π of X. For each i, $1 \le i \le h$, let X_i be the event that all the elements of A_i precede all those of B_i in this order. Clearly, $\Pr[X_i] = 1/\binom{k+\ell}{k}$. It is also easy to check that the events X_i are pairwise disjoint. Indeed, assume this is false, and let π be an order in which all the elements of A_i precede those of B_i and all the elements of A_j precede those of B_j. Without loss of generality, we may assume that the last element of A_i does not appear after the last element of A_j. But in this case, all elements of A_i precede all those of B_j, contradicting the fact that $A_i \cap B_j \neq \emptyset$. Therefore, all the events X_i are pairwise disjoint, as claimed. It follows that

$$1 \ge \Pr\left[\bigvee_{i=1}^{h} X_i\right] = \sum_{i=1}^{h} \Pr[X_i] = h/\binom{k+\ell}{k},$$

completing the proof. ∎

Theorem 1.3.3 is sharp, as shown by the family $\mathcal{F} = \{(A, X \setminus A) : A \subset X, |A| = k\}$, where $X = \{1, 2, \ldots, k+\ell\}$.

1.4 COMBINATORIAL NUMBER THEORY

A subset A of an abelian group G is called *sum-free* if $(A + A) \cap A = \emptyset$, that is, if there are no $a_1, a_2, a_3 \in A$ such that $a_1 + a_2 = a_3$.

Theorem 1.4.1 [Erdős (1965a)] *Every set $B = \{b_1, \ldots, b_n\}$ of n nonzero integers contains a sum-free subset A of size $|A| > \frac{1}{3}n$.*

Proof. Let $p = 3k + 2$ be a prime that satisfies $p > 2\max_{1 \le i \le n}|b_i|$, and put $C = \{k + 1, k + 2, \ldots, 2k + 1\}$. Observe that C is a sum-free subset of the cyclic group Z_p and that

$$\frac{|C|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.$$

Let us choose at random an integer x, $1 \le x < p$, according to a uniform distribution on $\{1, 2, \ldots, p-1\}$, and define $d_1, \ldots, d_n$ by $d_i \equiv xb_i (\mathrm{mod}\ p)$, $0 \le d_i < p$. Trivially, for every fixed i, $1 \le i \le n$, as x ranges over all numbers $1, 2, \ldots, p-1$, d_i ranges over all nonzero elements of Z_p, and hence $\Pr[d_i \in C] = |C|/(p-1) > \frac{1}{3}$. Therefore, the expected number of elements b_i such that $d_i \in C$ is more than $n/3$. Consequently, there is an x, $1 \le x < p$, and a subsequence A of B of cardinality $|A| > n/3$, such that $xa (\mathrm{mod}\ p) \in C$ for all $a \in A$. This A is clearly sum-free, since, if $a_1 + a_2 = a_3$ for some $a_1, a_2, a_3 \in A$, then $xa_1 + xa_2 \equiv xa_3 (\mathrm{mod}\ p)$, contradicting the fact that C is a sum-free subset of Z_p. This completes the proof. ∎

Remark. The above proof works whenever p is a prime that does not divide any of the numbers b_i. This can be used to design an efficient deterministic algorithm for finding a sum-free subset A of size bigger than $|B|/3$ in a given set B as above. In Alon and Kleitman (1990), it is shown that every set of n nonzero elements of an arbitrary abelian group contains a sum-free subset of more than $2n/7$ elements, and that the constant $2/7$ is the best possible. For quite some time it was not clear whether or not the constant $1/3$ in Theorem 1.4.1 can be replaced by a larger constant, until Eberhard, Green and Manners (2013) proved that the constant $1/3$ is tight. The problem of deciding whether or not every set of n nonzero integers contains a sum-free subset of cardinality at least $n/3 + w(n)$, where $w(n)$ tends to infinity with n, remains open. It will be very surprising if there is no such $w(n)$.

1.5 DISJOINT PAIRS

The probabilistic method is most striking when it is applied to prove theorems whose statement does not seem to suggest at all the need for probability. Most of the examples given in the previous sections are simple instances of such statements. In this section we describe a (slightly) more complicated result, due to Alon and Frankl (1985), which solves a conjecture of Daykin and Erdős.

Let $\mathcal{F}$ be a family of m distinct subsets of $X = \{1, 2, \ldots, n\}$. Let $d(\mathcal{F})$ denote the number of disjoint pairs in $\mathcal{F}$, that is

$$d(\mathcal{F}) = |\{\{F, F'\} : F, F' \in \mathcal{F}, \quad F \cap F' = \emptyset\}|.$$

Daykin and [Erdős] conjectured that, if $m = 2^{(1/2+\delta)n}$, then for every fixed $\delta > 0$, $d(\mathcal{F}) = o(m^2)$, as n tends to infinity. This result follows from the following theorem, which is a special case of a more general result:

Theorem 1.5.1 *Let $\mathcal{F}$ be a family of $m = 2^{(1/2+\delta)n}$ subsets of $X = \{1, 2, \ldots, n\}$, where $\delta > 0$. Then*

$$d(\mathcal{F}) < m^{2-\delta^2/2}. \tag{1.1}$$

Proof. Suppose (1.1) is false; pick independently t members $A_1, A_2, \ldots, A_t$ of $\mathcal{F}$ with repetitions at random, where t is a large positive integer, to be chosen later. We will show that with positive probability $|A_1 \cup A_2 \cup \cdots \cup A_t| > n/2$ and still this union is disjoint to more than $2^{n/2}$ distinct subsets of X. This contradiction will establish (1.1).

In fact,

$$\begin{aligned} &\Pr[|A_1 \cup A_2 \cup \cdots \cup A_t| \le n/2] \\ &\le \sum_{S \subset X, |S| = n/2} \Pr[A_i \subset S, i = 1, \ldots, t] \\ &\le 2^n (2^{n/2}/2^{(1/2+\delta)n})^t = 2^{n(1-\delta t)} . \end{aligned} \tag{1.2}$$

Define

$$v(B) = |\{A \in \mathcal{F} : B \cap A = \emptyset\}|.$$

Clearly,

$$\sum_{B \in \mathcal{F}} v(B) = 2d(\mathcal{F}) \geq 2m^{2-\delta^2/2}.$$

Let Y be a random variable whose value is the number of members $B \in \mathcal{F}$ that are disjoint to all the A_i $(1 \leq i \leq t)$. By the convexity of z^t, the expected value of Y satisfies

$$E[Y] = \sum_{B \in \mathcal{F}} \left(\frac{v(B)}{m}\right)^t = \frac{1}{m^t} \cdot m \left(\frac{\sum v(B)^t}{m}\right)$$
$$\geq \frac{1}{m^t} \cdot m \left(\frac{2d(\mathcal{F})}{m}\right)^t \geq 2m^{1-t\delta^2/2}.$$

Since $Y \leq m$, we conclude that

$$\Pr[Y \geq m^{1-t\delta^2/2}] \geq m^{-t\delta^2/2}. \tag{1.3}$$

One can check that, for $t = \lceil 1 + 1/\delta \rceil$, $m^{1-t\delta^2/2} > 2^{n/2}$ and the right-hand side of (1.3) is greater than the right-hand side of (1.2). Thus, with positive probability, $|A_1 \cup A_2 \cup \cdots \cup A_t| > n/2$ and still this union is disjoint to more than $2^{n/2}$ members of F. This contradiction implies inequality (1.1). ■

1.6 INDEPENDENT SETS AND LIST COLORING

Containers

A recent powerful method has been developed independently by Saxton and Thomason (2012) and by Balogh, Morris and Samotij (2014). This method supplies a structural characterization of the independent sets in uniform hypergraphs satisfying certain natural conditions, by showing that in such hypergraphs every independent set is almost fully contained in one of a small number of sparse sets (called containers). This general result leads to many interesting consequences including sparse random analogs of several classical results like Szemerédi's theorem and Turán's theorem. The method is elementary but somewhat technical; here we only present the basic approach dealing with independent sets in regular graphs, and describe one interesting application to a seemingly unrelated graph coloring problem. Many additional applications can be found in Saxton and Thomason (2012) and in Balogh et al. (2014).

The basic approach for regular graphs has been discovered earlier by several researchers, most notably by Sapozhenko (2001). We proceed with the statement and its short proof.

Theorem 1.6.1 *Let $G = (V, E)$ be a d-regular graph on n vertices, and let $\epsilon > 0$ be a positive real. Then there is a collection $\mathcal{C}$ of subsets of V, so that*

$$|\mathcal{C}| \leq \sum_{i \leq n/(\epsilon d)} \binom{n}{i}$$

each $C \in \mathcal{C}$ is of size at most $\frac{n}{\epsilon d} + \frac{n}{2-\epsilon}$, and every independent set in G is fully contained in a member $C \in \mathcal{C}$. Moreover, for each $C \in \mathcal{C}$, the degree of each vertex $v \in C$ in the induced subgraph of G on C is at most ϵd.

Proof. Let S be an independent set in G. Define a set C containing S as follows: Starting with $T = \emptyset$, as long as there is a vertex $v \in S$ so that $|N(v) - N(T)| \geq \epsilon d$, add it to T. Here, $N(v)$ is the set of all neighbors of v, and $N(T)$ is the set of all neighbors of vertices in T. Note that T may depend on the order in which the vertices of S are inspected, but for our purpose here any order will do. This process clearly ends with a subset $T \subset S$, where $|T| \leq \frac{n}{\epsilon d}$ as each addition of a vertex to T increases $|N(T)|$ by at least ϵd. Moreover, each vertex $v \in S - T$ has at least $(1 - \epsilon)d$ neighbors in $N(T)$. Let $B(T)$ denote the set of all vertices $v \in V - (T \cup N(T))$ that have at least $(1 - \epsilon)d$ neighbors in $N(T)$. Note that, crucially, $B(T)$ is determined by T. Define $C = T \cup B(T)$. By the discussion above $S \subset C$, every vertex of T has no neighbors in C, and every vertex in $C - T$ has at most ϵd neighbors in C. Since $C - T = B(T)$ is contained in $V - N(T)$, its size is at most $n - |N(T)|$, and as each of its vertices has at least $(1 - \epsilon)d$ neighbors in $N(T)$, it follows that $|B(T)| \leq \frac{|N(T)d|}{(1-\epsilon)d} = \frac{|N(T)|}{1-\epsilon}$. Taking a convex combination of the above two bounds, we conclude that

$$|B(T)| \leq \frac{1}{2-\epsilon}(n - |N(T)|) + \frac{1-\epsilon}{2-\epsilon}\frac{|N(T)|}{1-\epsilon} = \frac{n}{2-\epsilon}.$$

The set of containers $\mathcal{C}$ can thus be defined as the collection of all sets $T \cup B(T)$, where T is an independent set of size at most $\frac{n}{\epsilon d}$ in G. ■

List Coloring

The *list chromatic number* (or *choice number*) $\chi_\ell(G)$ of a graph $G = (V, E)$ is the minimum integer k such that, for every assignment of a list of k colors to each vertex v of G, there is a proper vertex coloring of G in which the color of each vertex is in its list. This notion was introduced independently by Vizing (1976) and by Erdős, Rubin and Taylor (1980). In both papers, the authors realized that this is a variant of usual coloring that exhibits several new interesting properties, and that in general $\chi_\ell(G)$, which is always at least as large as the chromatic number of G, may be arbitrarily large even for graphs G of chromatic number 2.

An intriguing property of list coloring of graphs, which is not shared by ordinary vertex coloring, is the fact that the list chromatic number of any graph with a large average degree is large. Indeed, it is shown in Alon (2000) that the list chromatic number of any graph with average degree d is at least $\Omega(\log d)$. Here we present a

short proof of this result for regular graphs, using the notion of containers. This proof appears in Saxton and Thomason (2012) and provides an asymptotically sharp lower bound for the choice number in terms of the degree of regularity. It can be extended, with some additional work, to nonregular graphs as well, but for simplicity we restrict the description to regular graphs.

Theorem 1.6.2 *Let $d > k > 2$ be integers satisfying*

$$k^2 \cdot H\left(\frac{\log d}{d}\right) < \left[1 - \left(\frac{1}{2} + \frac{1}{\log d}\right)\left(\frac{k}{k-1}\right)\right]^k \log e, \tag{1.4}$$

where $H(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function and all logarithms are in base 2. Then the choice number of any d-regular graph exceeds k. Therefore, there exists an absolute positive constant c so that, if $d \geq ck^4 2^k$, then the choice number of any d-regular graph exceeds k.

Proof. Let $G = (V, E)$ be a d-regular graph on n vertices, and let k be an integer so that (1.4) holds. Fix a set $K = \{1, 2, \ldots, k^2\}$ of k^2 colors and assign to each vertex $v \in V$, randomly and independently, a subset L_v of cardinality k chosen uniformly among all k-subsets of K. We claim that, with positive probability, there is no proper coloring of G assigning to each vertex v a color from its list L_v. To prove this claim using the union bound, it suffices to show that the probability that there are k^2 independent sets $S_1, S_2, \ldots, S_{k^2}$ in G so that for each vertex v there is an independent set S_i satisfying $v \in S_i$ and $i \in L_v$ is smaller than 1. Indeed, in any proper coloring, the set S_i of all vertices colored i forms an independent set, and if the color i of a vertex v belongs to its list L_v, then we must have $v \in S_i$ and $i \in L_v$. However, the number of independent sets in G may well be too large for using the union bound, hence we replace the independent sets by the containers described above. By Theorem 1.6.1 with $\epsilon = 1/\log d$, there is a family $\mathcal{C}$ of at most

$$\sum_{i \leq n \log d/d} \binom{n}{i} \leq 2^{H(\log d/d)n}$$

subsets C of V, each of size at most $n\left(\frac{\log d}{d} + \frac{1}{2-1/\log d}\right) < n\left(\frac{1}{2} + \frac{1}{\log d}\right)$ so that any independent set is fully contained in at least one of them. It suffices to show that, with positive probability, for any choice of k^2 containers $C_1, C_2, \ldots, C_{k^2}$, there is a vertex v so that v is not contained in C_i for any $i \in L_v$. As the number of containers is much smaller than the total number of independent sets, this can be proved by the union bound. The details follow. There are $|\mathcal{C}|^{k^2}$ ways to choose the containers $C_1, \ldots, C_{k^2}$. Fix such a choice and note that, since each container is small, so is their average size, implying that the average, over the vertices v, number of containers C_i that contain v is at most $k^2\left(\frac{1}{2} + \frac{1}{\log d}\right)$. Let k_v denote the number of containers C_i such that $v \in C_i$, and let $\overline{k} = \frac{1}{n}\sum_v k_v$ be its average over the vertices v. The probability that the list L_v of v does not contain any index i so that $v \in C_i$ is exactly

$$\frac{\binom{k^2-k_v}{k}}{\binom{k^2}{k}} \geq g(k_v),$$

where the function $g(z)$ is defined by

$$g(z) = \left[\frac{k^2-k-z}{k^2-k}\right]^k = \left(1-\frac{z}{k^2-k}\right)^k$$

for $0 \leq z < k^2-k$ and by $g(z)=0$ for $z \geq k^2-k$. It follows that the probability that, for the above fixed choice of containers, for each vertex v there is an $i \in L_v$ with $v \in C_i$, is at most

$$\prod_v [1-g(k_v)] \leq e^{-\sum_v g(k_v)}.$$

Since the function $g(z)$ is convex for all $z \geq 0$, it follows by Jensen's inequality that $\sum_v g(k_v) \geq ng(\overline{k})$, and thus the probability that the random lists do yield a proper coloring by color classes contained in the fixed set of containers above is at most $e^{-ng(\overline{k})}$. Since $g(z)$ is non-increasing and $\overline{k} \leq k^2(\frac{1}{2}+\frac{1}{\log d})$, it follows that

$$g(\overline{k}) \geq \left[\frac{k^2-k-k^2\left(\frac{1}{2}+\frac{1}{\log d}\right)}{k^2-k}\right]^k = \left[1-\left(\frac{1}{2}+\frac{1}{\log d}\right)\frac{k}{k-1}\right]^k$$

and the above probability is at most

$$e^{-n\left[1-\left(\frac{1}{2}+\frac{1}{\log d}\right)\frac{k}{k-1}\right]^k}.$$

By (1.4), this probability multiplied by the number of choices of a sequence of k^2 containers, which is at most

$$2^{k^2 H\left(\frac{\log d}{d}\right)},$$

is smaller than 1, and the union bound completes the proof. ■

1.7 EXERCISES

1. Prove that, if there is a real p, $0 \leq p \leq 1$ such that

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}} < 1,$$

then the Ramsey number $R(k,t)$ satisfies $R(k,t) > n$. Using this, show that

$$R(4,t) \geq \Omega(t^{3/2}/(\ln\, t)^{3/2}).$$

2. Suppose $n \geq 4$, and let H be an n-uniform hypergraph with at most $4^{n-1}/3^n$ edges. Prove that there is a coloring of the vertices of H by four colors so that in every edge all four colors are represented.
3. (*) Prove that for every two independent and identically distributed real random variables X and Y,

$$\Pr[|X - Y| \leq 2] \leq 3 \Pr[|X - Y| \leq 1].$$

4. (*) Let $G = (V, E)$ be a graph with n vertices and minimum degree $\delta > 10$. Prove that there is a partition of V into two disjoint subsets A and B so that $|A| \leq O(n \ln \delta/\delta)$, and each vertex of B has at least one neighbor in A and at least one neighbor in B.
5. (*) Let $G = (V, E)$ be a graph on $n \geq 10$ vertices, and suppose that if we add to G any edge not in G, then the number of copies of a complete graph on 10 vertices in it increases. Show that the number of edges of G is at least $8n - 36$.
6. (*) Theorem 1.2.1 asserts that for every integer $k > 0$ there is a tournament $T_k = (V, E)$ with $|V| > k$ such that for every set U of at most k vertices of T_k there is a vertex v so that all directed arcs $\{(v, u) : u \in U\}$ are in E.
Show that each such tournament contains at least $\Omega(k2^k)$ vertices.
7. Let $\{(A_i, B_i), 1 \leq i \leq h\}$ be a family of pairs of subsets of the set of integers such that $|A_i| = k$ for all i and $|B_i| = l$ for all i, $A_i \cap B_i = \emptyset$, and $(A_i \cap B_j) \cup (A_j \cap B_i) \neq \emptyset$ for all $i \neq j$. Prove that $h \leq (k + l)^{k+l}/(k^k l^l)$.
8. (Prefix-free codes; Kraft inequality). Let F be a finite collection of binary strings of finite lengths, and assume that no member of F is a prefix of another one. Let N_i denote the number of strings of length i in F. Prove that

$$\sum_i \frac{N_i}{2^i} \leq 1.$$

9. (*) (Uniquely decipherable codes; Kraft–McMillan inequality). Let F be a finite collection of binary strings of finite lengths, and assume that no two distinct concatenations of two finite sequences of codewords result in the same binary sequence. Let N_i denote the number of strings of length i in F. Prove that

$$\sum_i \frac{N_i}{2^i} \leq 1.$$

10. Prove that there is an absolute constant $c > 0$ with the following property: let A be an $n \times n$ matrix with pairwise distinct entries. Then there is a permutation of the rows of A so that no column in the permuted matrix contains an increasing subsequence of length at least $c\sqrt{n}$.

THE PROBABILISTIC LENS:
The Erdős–Ko–Rado Theorem

A family $\mathcal{F}$ of sets is called intersecting if $A, B \in \mathcal{F}$ implies $A \cap B \neq \emptyset$. Suppose $n \geq 2k$, and let $\mathcal{F}$ be an intersecting family of k-element subsets of an n-set, for definiteness $\{0, \ldots, n-1\}$. The Erdős–Ko–Rado theorem is that $|\mathcal{F}| \leq \binom{n-1}{k-1}$. This is achievable by taking the family of k-sets containing a particular point. We give a short proof due to Katona (1972).

Lemma 1 *For $0 \leq s \leq n-1$, set $A_s = \{s, s+1, \ldots, s+k-1\}$, where addition is modulo n. Then $\mathcal{F}$ can contain at most k of the sets A_s.*

Proof. Fix some $A_s \in \mathcal{F}$. All other sets A_t that intersect A_s can be partitioned into $k-1$ pairs $\{A_{s-i}, A_{s+k-i}\}$, $(1 \leq i \leq k-1)$, and the members of each such pair are disjoint. The result follows, since $\mathcal{F}$ can contain at most one member of each pair. ■

Now we prove the Erdős–Ko–Rado theorem. Let a permutation σ of $\{0, \ldots, n-1\}$ and $i \in \{0, \ldots, n-1\}$ be chosen randomly, uniformly and independently and set $A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+k-1)\}$, addition again modulo n. Conditioning on any choice of σ, the lemma gives $\Pr[A \in \mathcal{F}] \leq k/n$. Hence $\Pr[A \in \mathcal{F}] \leq k/n$. But A is uniformly chosen from all k-sets so

$$\frac{k}{n} \geq \Pr[A \in \mathcal{F}] = \frac{|\mathcal{F}|}{\binom{n}{k}}$$

and

$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}.$$

2

Linearity of Expectation

The search for truth is more precious than its possession.
–Albert Einstein

2.1 BASICS

Let $X_1, \ldots, X_n$ be random variables, $X = c_1X_1 + \cdots + c_nX_n$. Linearity of Expectation states that

$$\mathrm{E}[X] = c_1\mathrm{E}[X_1] + \cdots + c_n\mathrm{E}[X_n] \ .$$

The power of this principle comes from there being no restrictions on the dependence or independence of X_i. In many instances, $\mathrm{E}[X]$ can easily be calculated by a judicious decomposition into simple (often indicator) random variables X_i.

Let σ be a random permutation on $\{1, \ldots, n\}$, uniformly chosen. Let $X(\sigma)$ be the number of fixed points of σ. To find $\mathrm{E}[X]$, we decompose $X = X_1 + \cdots + X_n$, where X_i is the indicator random variable of the event $\sigma(i) = i$. Then

$$\mathrm{E}[X_i] = \Pr[\sigma(i) = i] = \frac{1}{n}$$

so that

$$\mathrm{E}[X] = \frac{1}{n} + \cdots + \frac{1}{n} = 1 \ .$$

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

In applications, we often use that there is a point in the probability space for which $X \geq \mathrm{E}[X]$ and a point for which $X \leq \mathrm{E}[X]$. We have selected results with a purpose of describing this basic methodology. The following result of Szele (1943) is oftentimes considered the first use of the probabilistic method.

Theorem 2.1.1 *There is a tournament T with n players and at least $n!2^{-(n-1)}$ Hamiltonian paths.*

Proof. In the random tournament, let X be the number of Hamiltonian paths. For each permutation σ, let X_σ be the indicator random variable for σ giving a Hamiltonian path, that is, satisfying $(\sigma(i), \sigma(i+1)) \in T$ for $1 \leq i < n$. Then $X = \sum X_\sigma$ and

$$\mathrm{E}[X] = \sum \mathrm{E}[X_\sigma] = n!2^{-(n-1)}$$

Thus some tournament has at least $\mathrm{E}[X]$ Hamiltonian paths. ■

Szele conjectured that the maximum possible number of Hamiltonian paths in a tournament on n players is at most $n!/(2 - o(1))^n$. This was proved in Alon (1990a) and is presented in "The Probabilistic Lens: Hamiltonian Paths" (following Chapter 4).

2.2 SPLITTING GRAPHS

Theorem 2.2.1 *Let $G = (V, E)$ be a graph with n vertices and e edges. Then G contains a bipartite subgraph with at least $e/2$ edges.*

Proof. Let $T \subseteq V$ be a random subset given by $\Pr[x \in T] = 1/2$, these choices being mutually independent. Set $B = V - T$. Call an edge $\{x, y\}$ crossing if exactly one of x, y is in T. Let X be the number of crossing edges. We decompose

$$X = \sum_{\{x,y\} \in E} X_{xy} \, ,$$

where X_{xy} is the indicator random variable for $\{x, y\}$ being crossing. Then

$$\mathrm{E}[X_{xy}] = \frac{1}{2}$$

as two fair coin flips have probability $1/2$ of being different. Then

$$\mathrm{E}[X] = \sum_{\{x,y\} \in E} \mathrm{E}[X_{xy}] = \frac{e}{2} \, .$$

Thus $X \geq e/2$ for some choice of T, and the set of those crossing edges forms a bipartite graph. ■

A more subtle probability space gives a small improvement (which is tight for complete graphs).

Theorem 2.2.2 *If G has $2n$ vertices and e edges, then it contains a bipartite subgraph with at least $en/(2n-1)$ edges. If G has $2n+1$ vertices and e edges, then it contains a bipartite subgraph with at least $e(n+1)/2n+1$ edges.*

Proof. When G has $2n$ vertices, let T be chosen uniformly from among all n-element subsets of V. Any edge $\{x,y\}$ now has probability $n/(2n-1)$ of being crossing, and the proof concludes as before. When G has $2n+1$ vertices, choose T uniformly from among all n-element subsets of V, and the proof is similar. ■

Here is a more complicated example in which the choice of distribution requires a preliminary lemma. Let $V = V_1 \cup \cdots \cup V_k$, where the V_i are disjoint sets of size n. Let $h : V^k \to \{\pm 1\}$ be a two-coloring of the k-sets. A k-set E is crossing if it contains precisely one point from each V_i. For $S \subseteq V$ set $h(S) = \sum h(E)$, the sum over all k-sets $E \subseteq S$.

Theorem 2.2.3 *Suppose $h(E) = +1$ for all crossing k-sets E. Then there is an $S \subseteq V$ for which*

$$|h(S)| \geq c_k n^k \ .$$

Here, c_k is a positive constant, independent of n.

Lemma 2.2.4 *Let P_k denote the set of all homogeneous polynomials $f(p_1, \ldots, p_k)$ of degree k with all coefficients having absolute value at most 1, and $p_1 p_2 \cdots p_k$ having coefficient 1. Then for all $f \in P_k$, there exist $p_1, \ldots, p_k \in [0,1]$ with*

$$|f(p_1, \ldots, p_k)| \geq c_k \ .$$

Here, c_k is positive and independent of f.

Proof. Set

$$M(f) = \max_{p_1,\ldots,p_k \in [0,1]} |f(p_1, \ldots, p_k)| \ .$$

For $f \in P_k$, $M(f) > 0$ as f is not the zero polynomial. As P_k is compact and $M : P_k \to R$ is continuous, M must assume its minimum c_k. ■

Proof [Theorem 2.2.3]. Define a random $S \subseteq V$ by setting

$$\Pr[x \in S] = p_i, \quad x \in V_i \ ,$$

these choices being mutually independent, with p_i to be determined. Set $X = h(S)$. For each k-set E, set

$$X_E = \begin{cases} h(E) & \text{if } E \subseteq S, \\ 0 & \text{otherwise.} \end{cases}$$

Say, E has type $(a_1, \ldots, a_k)$ if $|E \cap V_i| = a_i$, $1 \le i \le k$. For these E,

$$\mathrm{E}[X_E] = h(E)\Pr[E \subseteq S] = h(E)p_1^{a_1} \cdots p_k^{a_k} .$$

Combining terms by type

$$\mathrm{E}[X] = \sum_{a_1+\cdots+a_k=k} p_1^{a_1} \cdots p_k^{a_k} \sum_{E \text{ of type } (a_1,\cdots,a_k)} h(E) .$$

When $a_1 = \cdots = a_k = 1$, all $h(E) = 1$ by assumption, so

$$\sum_{E \text{ of type } (1,\ldots,1)} h(E) = n^k .$$

For any other type, there are fewer than n^k terms, each ± 1, so

$$\left| \sum_{E \text{ of type } (a_1,\ldots,a_k)} h(E) \right| \le n^k .$$

Thus

$$\mathrm{E}[X] = n^k f(p_1, \ldots, p_k) ,$$

where $f \in P_k$, as defined by Lemma 2.2.4.

Now select $p_1, \ldots, p_k \in [0, 1]$ with $|f(p_1, \ldots, p_k)| \ge c_k$. Then

$$\mathrm{E}[|X|] \ge |\mathrm{E}[X]| \ge c_k n^k .$$

Some particular value of $|X|$ must exceed or equal its expectation. Hence there is a particular set $S \subseteq V$ with $|X| = |h(S)| \ge c_k n^k$. ■

Theorem 2.2.3 has an interesting application to Ramsey Theory. It is known (see Erdős (1965b)) that, given any coloring with two colors of the k-sets of an n-set, there exist k disjoint m-sets, $m = \Theta((\ln\ n)^{1/(k-1)})$, so that all crossing k-sets are the same color. From Theorem 2.2.3, there then exists a set of size $\Theta((\ln\ n)^{1/(k-1)})$, at least $\frac{1}{2} + \epsilon_k$ of whose k-sets are the same color. This is somewhat surprising since it is known that there are colorings in which the largest monochromatic set has size at most the $k - 2$-fold logarithm of n.

2.3 TWO QUICKIES

Linearity of Expectation sometimes gives very quick results.

Theorem 2.3.1 *There is a two-coloring of K_n with at most*

$$\binom{n}{a} 2^{1-\binom{a}{2}}$$

monochromatic K_a.

Proof [Outline]. Take a random coloring . Let X be the number of monochromatic K_a and find E[X]. For some coloring, the value of X is at most this expectation. ■

In Chapter 16, it is shown how such a coloring can be found deterministically and efficiently.

Theorem 2.3.2 *There is a two-coloring of $K_{m,n}$ with at most*

$$\binom{m}{a}\binom{n}{b} 2^{1-ab}$$

monochromatic $K_{a,b}$.

Proof [Outline]. Take a random coloring. Let X be the number of monochromatic $K_{a,b}$ and find E[X]. For some coloring, the value of X is at most this expectation. ■

2.4 BALANCING VECTORS

The next result has an elegant *non*-probabilistic proof, which we defer to the end of this chapter. Here, $|v|$ is the usual Euclidean norm.

Theorem 2.4.1 *Let $v_1, \ldots, v_n \in R^n$, all $|v_i| = 1$. Then there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that*

$$|\epsilon_1 v_1 + \cdots + \epsilon_n v_n| \leq \sqrt{n}\,,$$

and also there exist $\epsilon_1, \ldots, \epsilon_n = \pm 1$ so that

$$|\epsilon_1 v_1 + \cdots + \epsilon_n v_n| \geq \sqrt{n}\,.$$

Proof. Let $\epsilon_1, \ldots, \epsilon_n$ be selected uniformly and independently from $\{-1, +1\}$. Set

$$X = |\epsilon_1 v_1 + \cdots + \epsilon_n v_n|^2\,.$$

Then

$$X = \sum_{i=1}^{n}\sum_{j=1}^{n} \epsilon_i \epsilon_j v_i \cdot v_j\,.$$

Thus

$$\mathrm{E}[X] = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j \mathrm{E}[\epsilon_i \epsilon_j] \ .$$

When $i \neq j$, $\mathrm{E}[\epsilon_i \epsilon_j] = \mathrm{E}[\epsilon_i]\mathrm{E}[\epsilon_j] = 0$. When $i = j$, $\epsilon_i^2 = 1$ so $\mathrm{E}[\epsilon_i^2] = 1$. Thus

$$\mathrm{E}[X] = \sum_{i=1}^{n} v_i \cdot v_i = n \ .$$

Hence there exist specific $\epsilon_1, \ldots, \epsilon_n = \pm 1$ with $X \geq n$ and with $X \leq n$. Taking square roots gives the theorem. ■

The next result includes part of Theorem 2.4.1 as a linear translation of the $p_1 = \cdots = p_n = 1/2$ case.

Theorem 2.4.2 *Let $v_1, \ldots, v_n \in R^n$, all $|v_i| \leq 1$. Let $p_1, \ldots, p_n \in [0, 1]$ be arbitrary, and set $w = p_1 v_1 + \cdots + p_n v_n$. Then there exist $\epsilon_1, \ldots, \epsilon_n \in \{0, 1\}$ so that, setting $v = \epsilon_1 v_1 + \cdots + \epsilon_n v_n$,*

$$|w - v| \leq \frac{\sqrt{n}}{2} \ .$$

Proof. Pick ϵ_i independently with

$$\Pr[\epsilon_i = 1] = p_i, \quad \Pr[\epsilon_i = 0] = 1 - p_i \ .$$

The random choice of ϵ_i gives a random v and a random variable

$$X = |w - v|^2 \ .$$

We expand

$$X = \left| \sum_{i=1}^{n} (p_i - \epsilon_i) v_i \right|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j (p_i - \epsilon_i)(p_j - \epsilon_j)$$

so that

$$\mathrm{E}[X] = \sum_{i=1}^{n} \sum_{j=1}^{n} v_i \cdot v_j \mathrm{E}[(p_i - \epsilon_i)(p_j - \epsilon_j)] \ .$$

For $i \neq j$,

$$\mathrm{E}[(p_i - \epsilon_i)(p_j - \epsilon_j)] = \mathrm{E}[p_i - \epsilon_i]\mathrm{E}[p_j - \epsilon_j] = 0 \ .$$

For $i = j$,

$$\mathrm{E}[(p_i - \epsilon_i)^2] = p_i(p_i - 1)^2 + (1 - p_i)p_i^2 = p_i(1 - p_i) \leq \frac{1}{4} \ ,$$

$(\mathrm{E}[(p_i - \epsilon_i)^2] = \mathrm{Var}[\epsilon_i]$, the *variance* to be discussed in Chapter 4.) Thus

$$\mathrm{E}[X] = \sum_{i=1}^{n} p_i(1-p_i)|v_i|^2 \le \frac{1}{4}\sum_{i=1}^{n}|v_i|^2 \le \frac{n}{4}$$

and the proof concludes as in that of Theorem 2.4.1. ■

2.5 UNBALANCING LIGHTS

Theorem 2.5.1 *Let $a_{ij} = \pm 1$ for $1 \le i,j \le n$. Then there exist $x_i, y_j = \pm 1$, $1 \le i,j \le n$ so that*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_iy_j \ge \left(\sqrt{\frac{2}{\pi}} + o(1)\right)n^{3/2} .$$

This result has an amusing interpretation. Let an $n \times n$ array of lights be given, each either on ($a_{ij} = +1$) or off ($a_{ij} = -1$). Suppose for each row and each column there is a switch so that if the switch is pulled ($x_i = -1$ for row i and $y_j = -1$ for column j) all of the lights in that line will be "switched" on to off or off to on. Then for any initial configuration it is possible to perform switchings so that the number of lights on minus the number of lights off is at least $(\sqrt{2/\pi} + o(1))n^{3/2}$.

Proof [Theorem 2.5.1]. Forget the x's. Let $y_1, \ldots, y_n = \pm 1$ be selected independently and uniformly and set

$$R_i = \sum_{j=1}^{n} a_{ij}y_j \ , R = \sum_{i=1}^{n} |R_i| \ .$$

Fix i. Regardless of a_{ij}, $a_{ij}y_j$ is ± 1 with probability $1/2$, and their values (over j) are independent; that is, whatever the ith row is initially after random switching, it becomes a uniformly distributed row, with all 2^n possibilities equally likely. Thus R_i has distribution S_n – the distribution of the sum of n independent uniform $\{-1, 1\}$ random variables – and so

$$\mathrm{E}[|R_i|] = \mathrm{E}[|S_n|] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right)\sqrt{n} \ .$$

These asymptotics may be found by estimating S_n by $\sqrt{n}N$, where N is standard normal and using elementary calculus. Alternatively, a closed form

$$\mathrm{E}[|S_n|] = n2^{1-n}\binom{n-1}{\lfloor (n-1)/2 \rfloor}$$

may be derived combinatorially (a problem in the 1974 Putnam competition!) and the asymptotics follows from Stirling's formula.

Now apply the Linearity of Expectation to R:

$$\mathrm{E}[R] = \sum_{i=1}^{n} \mathrm{E}[|R_i|] = \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2} .$$

There exist $y_1, \ldots, y_n = \pm 1$ with R at least this value. Finally, pick x_i with the same sign as R_i so that

$$\sum_{i=1}^{n} x_i \sum_{j=1}^{n} a_{ij} y_j = \sum_{i=1}^{n} x_i R_i = \sum_{i=1}^{n} |R_i| = R \geq \left(\sqrt{\frac{2}{\pi}} + o(1)\right) n^{3/2} . \qquad \blacksquare$$

Another result on unbalancing lights appears in "The Probabilistic Lens: Unbalancing Lights" (following Chapter 13). The existence of Hadamard matrices and the discussion in Section 9.1 show that the estimate in the last theorem cannot be improved to anything bigger than $n^{3/2}$.

2.6 WITHOUT COIN FLIPS

A non-probabilistic proof of Theorem 2.2.1 may be given by placing each vertex in either T or B sequentially. At each stage, place x in either T or B so that at least half of the edges from x to previous vertices are crossing. With this effective algorithm, at least half the edges will be crossing.

There is also a simple sequential algorithm for choosing signs in Theorem 2.4.1. When the sign for v_i is to be chosen, a partial sum $w = \epsilon_1 v_1 + \cdots + \epsilon_{i-1} v_{i-1}$ has been calculated. Now if it is desired that the sum be small, select $\epsilon_i = \pm 1$ so that $\epsilon_i v_i$ makes an obtuse (or right) angle with w. If the sum need be big, make the angle acute or right. In the extreme case when all angles are right angles, Pythagoras and induction give that the final w has norm $\sqrt{n}$; otherwise, it is either less than $\sqrt{n}$ or greater than $\sqrt{n}$, as desired.

For Theorem 2.4.2, a greedy algorithm produces the desired ϵ_i. Given $v_1, \ldots, v_n \in R^n$, $p_1, \ldots, p_n \in [0, 1]$, suppose $\epsilon_1, \ldots, \epsilon_{s-1} \in \{0, 1\}$ have already been chosen. Set $w_{s-1} = \sum_{i=1}^{s-1} (p_i - \epsilon_i) v_i$, the partial sum. Select ϵ_s so that

$$w_s = w_{s-1} + (p_s - \epsilon_s) v_s = \sum_{i=1}^{s} (p_i - \epsilon_i) v_i$$

has minimal norm. A random $\epsilon_s \in \{0, 1\}$ chosen with $\Pr[\epsilon_s = 1] = p_s$ gives

$$\begin{aligned} \mathrm{E}[|w_s|^2] &= |w_{s-1}|^2 + 2 w_{s-1} \cdot v_s \mathrm{E}[p_s - \epsilon_s] + |v_s|^2 \mathrm{E}[p_s - \epsilon_s]^2 \\ &= |w_{s-1}|^2 + p_s(1 - p_s)|v_s|^2 \end{aligned}$$

so for some choice of $\epsilon_s \in \{0, 1\}$,

$$|w_s|^2 \leq |w_{s-1}|^2 + p_s(1-p_s)|v_s|^2 \ .$$

As this holds for all $1 \leq s \leq n$ (taking $w_0 = 0$), the final

$$|w_n|^2 \leq \sum_{i=1}^{n} p_i(1-p_i)|v_i|^2 \ .$$

While the proofs appear similar, a direct implementation of the proof of Theorem 2.4.2 to find $\epsilon_1, \ldots, \epsilon_n$ might take an exhaustive search with exponential time. In applying the greedy algorithm at the sth stage, one makes two calculations of $|w_s|^2$, depending on whether $\epsilon_s = 0$ or 1, and picks that ϵ_s giving the smaller value. Hence there are only a linear number of calculations of norms to be made, and the entire algorithm takes only quadratic time. In Chapter 16, we discuss several similar examples in a more general setting.

2.7 EXERCISES

1. Suppose $n \geq 2$ and let $H = (V, E)$ be an n-uniform hypergraph with $|E| = 4^{n-1}$ edges. Show that there is a coloring of V by four colors so that no edge is monochromatic.
2. Prove that there is a positive constant c so that every set A of n nonzero reals contains a subset $B \subset A$ of size $|B| \geq cn$, so that there are no $b_1, b_2, b_3, b_4 \in B$ satisfying
$$b_1 + 2b_2 = 2b_3 + 2b_4 \ .$$
3. Prove that every set of n nonzero *real* numbers contains a subset A of *strictly* more than $n/3$ numbers such that there are no $a_1, a_2, a_3 \in A$ satisfying $a_1 + a_2 = a_3$.
4. Suppose $p > n > 10m^2$, with p prime, and let $0 < a_1 < a_2, < \cdots < a_m < p$ be integers. Prove that there is an integer x, $0 < x < p$, for which the m numbers
$$(xa_i \pmod p) \bmod n, \quad 1 \leq i \leq m$$
are pairwise distinct.
5. Let H be a graph, and let $n > |V(H)|$ be an integer. Suppose there is a graph on n vertices and t edges containing no copy of H, and suppose that $tk > n^2 \log_e n$. Show that there is a coloring of the edges of the complete graph on n vertices by k colors with no monochromatic copy of H.
6. (*) Prove, using the technique in the probabilistic lens on Hamiltonian paths, that there is a constant $c > 0$ such that for every even $n \geq 4$ the following holds: for every undirected complete graph K on n vertices whose edges are colored red and

blue, the number of alternating Hamilton cycles in K (i.e., properly edge-colored cycles of length n) is at most

$$n^c \frac{n!}{2^n} .$$

7. Let $\mathcal{F}$ be a family of subsets of $N = \{1, 2, \ldots, n\}$, and suppose there are no $A, B \in \mathcal{F}$ satisfying $A \subset B$. Let $\sigma \in S_n$ be a random permutation of the elements of N, and consider the random variable

$$X = |\{i : \{\sigma(1), \sigma(2), \ldots, \sigma(i)\} \in \mathcal{F}\}| .$$

   By considering the expectation of X, prove that $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

8. (*) Let X be a collection of pairwise orthogonal unit vectors in R^n, and suppose the projection of each of these vectors on the first k coordinates is of Euclidean norm at least ϵ. Show that $|X| \leq k/\epsilon^2$, and this is tight for all $\epsilon^2 = k/2^r < 1$.

9. Let $G = (V, E)$ be a bipartite graph with n vertices and a list $S(v)$ of more than $\log_2 n$ colors associated with each vertex $v \in V$. Prove that there is a proper coloring of G assigning to each vertex v a color from its list $S(v)$.

THE PROBABILISTIC LENS:
Brégman's Theorem

Let $A = [a_{ij}]$ be an $n \times n$ matrix with all $a_{ij} \in \{0, 1\}$. Let $r_i = \sum_{1 \le j \le n} a_{ij}$ be the number of 1's in the ith row. Let S be the set of permutations $\sigma \in S_n$ with $a_{i,\sigma i} = 1$ for $1 \le i \le n$. Then the permanent per(A) is simply $|S|$. The following result was conjectured by Minc and proved by Brégman (1973). The proof presented here is similar to that of Schrijver (1978).

Theorem 1 [Brégman's Theorem] $\quad \text{per}(A) \le \prod_{1 \le i \le n} (r_i!)^{1/r_i}$.

Pick $\sigma \in S$ and $\tau \in S_n$ independently and uniformly. Set $A^{(1)} = A$. Let $R_{\tau 1}$ be the number of 1's in row $\tau 1$ in $A^{(1)}$. Delete row $\tau 1$ and column $\sigma\tau 1$ from $A^{(1)}$ to give $A^{(2)}$. In general, let $A^{(i)}$ denote A with rows $\tau 1, \ldots, \tau(i-1)$ and columns $\sigma\tau 1, \ldots, \sigma\tau(i-1)$ deleted and let $R_{\tau i}$ denote the number of 1's of row τi in $A^{(i)}$. (This is nonzero as the $\sigma\tau i$th column has a 1.) Set

$$L = L(\sigma, \tau) = \prod_{1 \le i \le n} R_{\tau i} \ .$$

We think, roughly, of L as Lazyman's permanent calculation. There are $R_{\tau 1}$ choices for a 1 in row $\tau 1$, each of which leads to a different subpermanent calculation. Instead, Lazyman takes the factor $R_{\tau 1}$, takes the one from permutation σ, and examines $A^{(2)}$. As $\sigma \in S$ is chosen uniformly, Lazyman tends toward the high subpermanents and so it should not be surprising that he tends to overestimate the permanent. To make this precise, we define the geometric mean $G[Y]$. If $Y > 0$ takes values $a_1, \ldots, a_s$ with probabilities $p_1, \ldots, p_s$, respectively, then $G[Y] = \prod a_i^{p_i}$. Equivalently, $G[Y] = e^{\text{E}[\ln Y]}$. Linearity of Expectation translates into the geometric mean of a product being the product of the geometric means.

Claim 1 *per*(A) $\le G[L]$.

Proof. We show this for any fixed τ. Set $\tau 1 = 1$ for convenience of notation. We use induction on the size of the matrix. Reorder, for convenience, so that the first row has 1's in the first r columns, where $r = r_1$. For $1 \le j \le r$, let t_j be the permanent of A with the first row and jth column removed or, equivalently, the number of $\sigma \in S$ with $\sigma 1 = j$. Set

$$t = \frac{t_1 + \cdots + t_r}{r}$$

so that $\operatorname{per}(A) = rt$. Conditioning on $\sigma 1 = j$, $R_2 \cdots R_n$ is Lazyman's calculation of $\operatorname{per}(A^{(2)})$, where $A^{(2)}$ is A with the first row and jth column removed. By induction,

$$G[R_2 \cdots R_n | \sigma 1 = j] \ge t_j$$

and so

$$G[L] \ge \prod_{j=1}^{r} (rt_j)^{t_j/\operatorname{per}(A)} = r \prod_{j=1}^{r} t_j^{t_j/rt} . \qquad \blacksquare$$

Lemma 2 $\left(\prod_{j=1}^{r} t_j^{t_j} \right)^{1/r} \ge t^t.$

Proof. Taking logarithms, this is equivalent to

$$\frac{1}{r} \sum_{j=1}^{r} t_j \ln t_j \ge t \ln t ,$$

which follows from the convexity of the function $f(x) = x \ln x$. $\blacksquare$

Applying the lemma,

$$G[L] \ge r \prod_{j=1}^{r} t_j^{t_j/rt} \ge r(t^t)^{1/t} = rt = \operatorname{per}(A) . \qquad \blacksquare$$

Now we calculate $G[L]$ conditional on a fixed σ. For convenience of notation, reorder so that $\sigma i = i$, all i, and assume that the first row has 1's in precisely the first r_1 columns. With τ selected uniformly, the columns $1, \ldots, r_1$ are deleted in a random order uniform over all $r_1!$ possibilities. R_1 is the number of those columns remaining when the first column is to be deleted. As the first column is equally likely to be in any position among those r_1 columns, R_1 is uniformly distributed from 1 to r_1 and $G[R_1] = (r_1!)^{1/r_1}$. "Linearity" then gives

$$G[L] = G\left[\prod_{i=1}^{n} R_i \right] = \prod_{i=1}^{n} G[R_i] = \prod_{i=1}^{n} (r_i!)^{1/r_i} .$$

The overall $G[L]$ is the geometric mean of the conditional $G[L]$ and hence has the same value. That is,

$$\operatorname{per}(A) \le G[L] = \prod_{i=1}^{n} (r_i!)^{1/r_i} .$$

3

Alterations

Beauty is the first test: there is no permanent place in the world for ugly mathematics.
–G. H. Hardy

The basic probabilistic method was described in Chapter 1 as follows: trying to prove that a structure with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this space with positive probability. In this chapter we consider situations where the "random" structure does not have all the desired properties but may have a few "blemishes." With a small alteration, we remove the blemishes, giving the desired structure.

3.1 RAMSEY NUMBERS

Recall from Section 1.1 that $R(k, l) > n$ means there exists a two-coloring of the edges of K_n by red and blue so that there is neither a red K_k nor a blue K_l.

Theorem 3.1.1 *For any integer n,* $R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$.

Proof. Consider a random two-coloring of the edges of K_n obtained by coloring each edge independently either red or blue, where each color is equally likely. For any set R of k vertices, let X_R be the indicator random variable for the event for which the induced subgraph of K_n on R is monochromatic. Set $X = \sum X_R$, the sum over all

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

such R. By Linearity of Expectation, $\mathrm{E}[X] = \sum \mathrm{E}[X_R] = m$ with $m = \binom{n}{k} 2^{1-\binom{k}{2}}$. Thus there exists a two-coloring for which $X \leq m$. Fix such a coloring. Remove from K_n one vertex from each monochromatic k-set. At most m vertices have been removed (we may have "removed" the same vertex more than once but this only helps), so s vertices remain with $s \geq n - m$. This coloring on these s points has no monochromatic k-set. ∎

We are left with the "calculus" problem of finding that n which will optimize the inequality. Some analysis shows that we should take $n \sim e^{-1} k 2^{k/2}(1 - o(1))$, giving

$$R(k,k) > \frac{1}{e}(1 + o(1))k2^{k/2}.$$

A careful examination of Proposition 1.1.1 gives the lower bound

$$R(k,k) > \frac{1}{e\sqrt{2}}(1 + o(1))k2^{k/2}.$$

The more powerful Lovász Local Lemma (see Chapter 5) gives

$$R(k,k) > \frac{\sqrt{2}}{e}(1 + o(1))k2^{k/2}.$$

The distinctions between these bounds may be considered inconsequential since the best known upper bound for $R(k,k)$ is $(4 + o(1))^k$. The upper bounds do not involve probabilistic methods and may be found, for example, in Graham, Rothschild and Spencer (1990). We give all three lower bounds in following our philosophy of emphasizing *methodologies* rather than results.

In dealing with the off-diagonal Ramsey numbers, the distinction between the basic method and the alteration is given in the following two results.

Theorem 3.1.2 *If there exists $p \in [0,1]$ with*

$$\binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l}(1-p)^{\binom{l}{2}} < 1,$$

then $R(k,l) > n$.

Theorem 3.1.3 *For all integers n and $p \in [0,1]$,*

$$R(k,l) > n - \binom{n}{k} p^{\binom{k}{2}} - \binom{n}{l}(1-p)^{\binom{l}{2}}.$$

Proof. In both cases, we consider a random two-coloring of K_n obtained by coloring each edge independently either red or blue, where each edge is red with probability p.

Let X be the number of red k-sets plus the number of blue l-sets. Linearity of Expectation gives

$$\mathrm{E}[X] = \binom{n}{k} p^{\binom{k}{2}} + \binom{n}{l} (1-p)^{\binom{l}{2}}.$$

For Theorem 3.1.2, $\mathrm{E}[X] < 1$, so there exists a two-coloring with $X = 0$. For Theorem 3.1.3, there exists a two-coloring with s "bad" sets (either red k-sets or blue l-sets), $s \leq \mathrm{E}[X]$. Removing one point from each bad set gives a coloring of at least $n - s$ points with no bad sets. ■

The asymptotics of Theorems 3.1.2 and 3.1.3 can get fairly complex. Oftentimes, Theorem 3.1.3 gives a substantial improvement on Theorem 3.1.2. Even further improvements may be found using the Lovász Local Lemma. These bounds have been analyzed in Spencer (1977).

3.2 INDEPENDENT SETS

Here is a short and sweet argument that gives roughly half of the celebrated Turán's theorem. $\alpha(G)$ is the independence number of a graph G; $\alpha(G) \geq t$ means there exist t vertices with no edges between them.

Theorem 3.2.1 *Let $G = (V, E)$ have n vertices and $nd/2$ edges, $d \geq 1$. Then $\alpha(G) \geq n/2d$.*

Proof. Let $S \subseteq V$ be a random subset defined by

$$\Pr[v \in S] = p,$$

p to be determined, the events $v \in S$ being mutually independent. Let $X = |S|$, and let Y be the number of edges in $G|_S$. For each $e = \{i, j\} \in E$, let Y_e be the indicator random variable for the event $i, j \in S$ so that $Y = \sum_{e \in E} Y_e$. For any such e,

$$\mathrm{E}[Y_e] = \Pr[i, j \in S] = p^2,$$

so by Linearity of Expectation,

$$\mathrm{E}[Y] = \sum_{e \in E} \mathrm{E}[Y_e] = \frac{nd}{2} p^2.$$

Clearly, $\mathrm{E}[X] = np$, so, again by Linearity of Expectation

$$\mathrm{E}[X - Y] = np - \frac{nd}{2} p^2.$$

We set $p = 1/d$ (here using $d \geq 1$) to maximize this quantity, giving

$$\mathrm{E}[X - Y] = \frac{n}{2d}.$$

Thus there exists a specific S for which the number of vertices of S minus the number of edges in S is at least $n/2d$. Select one vertex from each edge of S and delete it. This leaves a set S^* with at least $n/2d$ vertices. All edges having been destroyed, S^* is an independent set. ■

The full result of Turán is given in "The Probabilistic Lens: Turán's Theorem" (following Chapter 6).

3.3 COMBINATORIAL GEOMETRY

For a set S of n points in the unit square U, let $T(S)$ be the minimum area of a triangle whose vertices are three distinct points of S. Put $T(n) = \max T(S)$, where S ranges over all sets of n points in U. Heilbronn conjectured that $T(n) = O(1/n^2)$. This conjecture was disproved by Komlós, Pintz and Szemerédi (1982) who showed, by a rather involved probabilistic construction, that there is a set S of n points in U such that $T(S) = \Omega(\log n/n^2)$. As this argument is rather complicated, we only present here a simpler one showing that $T(n) = \Omega(1/n^2)$.

Theorem 3.3.1 *There is a set S of n points in the unit square U such that $T(S) \geq 1/(100n^2)$.*

Proof. We first make a calculation. Let P, Q, R be independently and uniformly selected from U, and let $\mu = \mu(PQR)$ denote the area of the triangle PQR. We bound $\Pr[\mu \leq \epsilon]$ as follows: Let x be the distance from P to Q so that

$$\Pr[b \leq x \leq b + \Delta b] \leq \pi(b + \Delta b)^2 - \pi b^2$$

and in the limit $\Pr[b \leq x \leq b + db] \leq 2\pi b\, db$. Given P, Q at distance b, the altitude from R to the line PQ must have height $h \leq 2\epsilon/b$, and so R must lie in a strip of width $4\epsilon/b$ and length at most $\sqrt{2}$. This occurs with probability at most $4\sqrt{2}\epsilon/b$. As $0 \leq b \leq \sqrt{2}$, the total probability is bounded by

$$\int_0^{\sqrt{2}} (2\pi b)(4\sqrt{2}\epsilon/b)db = 16\pi\epsilon.$$

Now let $P_1, \ldots, P_{2n}$ be selected uniformly and independently in U, and let X denote the number of triangles $P_iP_jP_k$ with area less than $1/(100n^2)$. For each particular i, j, k, the probability of this occurring is less than $0.6n^{-2}$, and so

$$\mathrm{E}[X] \leq \binom{2n}{3}(0.6n^{-2}) < n.$$

Thus there exists a specific set of $2n$ vertices with fewer than n triangles of area less than $1/(100n^2)$. Delete one vertex from the set from each such triangle. This leaves at least n vertices, and now no triangle has area less than $1/(100n^2)$. ■

We note the following construction of Erdős showing $T(n) \geq 1/(2(n-1)^2)$ with n prime: On $[0, n-1] \times [0, n-1]$, consider the n points (x, x^2), where x^2 is reduced mod n (more formally, (x, y) where $y \equiv x^2 \bmod n$ and $0 \leq y < n$). If some three points of this set were collinear, they would lie on a line $y = mx + b$, and m would be a rational number with denominator less than n. But then in Z_n^2, the parabola $y = x^2$ would intersect the line $y = mx + b$ at three points, so that the quadratic $x^2 - mx - b$ would have three distinct roots, an impossibility. Triangles between lattice points in the plane have as their areas either half-integers or integers, hence the areas must be at least $1/2$. Contracting the plane by an $n - 1$ factor in both coordinates gives the desired set. While this gem does better than Theorem 3.3.1, it does not lead to the improvements of Komlós, Pintz, and Szemerédi.

3.4 PACKING

Let C be a bounded measurable subset of R^d, and let $B(x)$ denote the cube $[0, x]^d$ of side x. A *packing* of C into $B(x)$ is a family of mutually disjoint copies of C, all lying inside $B(x)$. Let $f(x)$ denote the largest size of such a family. The packing constant $\delta = \delta(C)$ is defined by

$$\delta(C) = \mu(C) \lim_{x \to \infty} f(x) x^{-d},$$

where $\mu(C)$ is the measure of C. This is the maximal proportion of space that may be packed by copies of C (this limit can be proved always to exist but even without that result the following result holds with lim replaced by lim inf).

Theorem 3.4.1 *Let C be bounded, convex, and centrally symmetric around the origin. Then $\delta(C) \geq 2^{-d-1}$.*

Proof. Let P, Q be selected independently and uniformly from $B(x)$, and consider the event $(C + P) \cap (C + Q) \neq \emptyset$. For this to occur, we must have, for some $c_1, c_2 \in C$,

$$P - Q = c_1 - c_2 = 2\frac{c_1 - c_2}{2} \in 2C$$

by central symmetry and convexity. The event $P \in Q + 2C$ has probability at most $\mu(2C)x^{-d}$ for each given Q, hence

$$\Pr[(C + P) \cap (C + Q) \neq \emptyset] \leq \mu(2C)x^{-d} = 2^d x^{-d} \mu(C).$$

Now let $P_1, \ldots, P_n$ be selected independently and uniformly from $B(x)$, and let X be the number of $i < j$ with $(C + P_i) \cap (C + P_j) \neq \emptyset$. From Linearity of Expectation,

$$\mathrm{E}[X] \leq \frac{n^2}{2} 2^d x^{-d} \mu(C).$$

Hence there exists a specific choice of n points with fewer than that many intersecting copies of C. For each P_i, P_j with $(C + P_i) \cap (C + P_j) \neq \emptyset$, remove either P_i or P_j from the set. This leaves at least $n - (n^2/2)2^d x^{-d}\mu(C)$ nonintersecting copies of C. Set $n = x^d 2^{-d}/\mu(C)$ to maximize this quantity, so that there are at least $x^d 2^{-d-1}/\mu(C)$ nonintersecting copies of C. These do not all lie inside $B(x)$ but, letting w denote an upper bound on the absolute values of the coordinates of the points of C, they do all lie inside a cube of side $x + 2w$. Hence

$$f(x + 2w) \geq x^d 2^{-d-1}/\mu(C)$$

and so $\delta(C) \geq \lim_{x\to\infty} \mu(C)f(x + 2w)(x + 2w)^{-d} \geq 2^{-d-1}$. ■

A simple greedy algorithm does somewhat better. Let $P_1, \ldots, P_m$ be *any* maximal subset of $[0, x]^d$ with the property that the sets $C + P_i$ are disjoint. We have seen that $C + P_i$ overlaps $C + P$ if and only if $P \in 2C + P_i$. Hence the sets $2C + P_i$ must cover $[0, x]^d$. As each such set has measure $\mu(2C) = 2^d\mu(C)$, we must have $m \geq x^d 2^{-d}/\mu(C)$. As before, all sets $C + P_i$ lie in a cube of side $x + 2w$, with w a constant, so that

$$f(x + 2w) \geq m \geq x^d 2^{-d}/\mu(C)$$

and so

$$\delta(C) \geq 2^{-d}.$$

A still further improvement appears in "The Probabilistic Lens: Efficient Packing" (following Chapter 14).

3.5 GREEDY COLORING

When a random coloring does not suffice for solving a problem, there are several ways to proceed. One of them is to apply a random recoloring to fix the blemishes left by the original random procedure. This has been useful in the study of Property B discussed in Section 1.3. Here we use the same notation: $m(n) > m$ means that, given any n-uniform hypergraph $H = (V, E)$ with m edges, there exists a two-coloring of V so that no edge is monochromatic. Beck (1978) improved Erdős' 1963 bound to $m(n) = \Omega(2^n n^{1/3})$. Building on his methods, Radhakrishnan and Srinivasan (2000) proved $m(n) = \Omega(2^n(n/\ln\ n)^{1/2})$. Both proofs applied random recoloring.

A simpler beautiful proof, based on a random greedy coloring, has been found by Cherkashin and Kozik (2015), following an approach of Pluhár (2009), and it is that proof that we describe below. The first edition of the book includes Beck's proof, the previous edition includes that of Radhakrishnan and Srinivasan (2000), and this one contains the new proof of Cherkashin and Kozik (2015). Note that the natural normalized quantity of interest here is $m(n)/2^{n-1}$, since the expected number of monochromatic edges in a random two-coloring of a hypergraph with m edges, each of size n, is $m/2^{n-1}$. Thus, the expression $m(n)/2^{n-1}$ measures the ratio between

$m(n)$ and the trivial lower bound for it. This means that, after more than 50 years of study of this problem, which led to lots of elegant ideas, the upper and lower bounds on $m(n)/2^{n-1}$, which are $O(n^2)$ and $\Omega(\sqrt{n/\ln n})$, remain quite far apart!

Theorem 3.5.1 *If there exists $p \in [0,1]$ with $k(1-p)^n + k^2 p < 1$, then $m(n) > 2^{n-1}k$.*

Corollary 3.5.2 $m(n) = \Omega(2^n(n/\ln n)^{1/2})$.

Proof. Bound $1 - p \le e^{-p}$. The function $ke^{-pn} + k^2 p$ is minimized at $p = \ln(n/k)/n$. Substituting back in, if

$$\frac{k^2}{n}(1 + \ln(n/k)) < 1,$$

then the condition of Theorem 3.5.1 holds. This inequality is true when $k = c(n/\ln n)^{1/2}$ for any $c < \sqrt{2}$, with n sufficiently large. ■

Proof. [Theorem 3.5.1] Fix $H = (V, E)$ with $m = 2^{n-1}k$ edges and p satisfying the condition. We describe a randomized algorithm that yields a coloring of V.

For each vertex $v \in V$, let x_v be a uniform random label in $[0,1]$, where all choices are independent. Note that, with probability 1, all labels are distinct. The algorithm goes over the vertices by an increasing order of their labels, coloring each vertex blue, unless it is the last vertex in an edge in which all previous vertices have been colored blue. In this case, the vertex is colored red. By definition, there are no blue edges. Also, a red edge can appear only if each of its vertices, and in particular its first vertex, is the last vertex of another edge all of whose previous vertices are colored blue. Call an ordered pair of edges (e,f) a conflicting pair if the last vertex of e is the first vertex of f. By the discussion above, if there are no conflicting pairs, then the coloring will produce no red edge. We proceed to show that with positive probability there are no such pairs.

To prove this fact, split the interval $[0,1]$ into three subintervals, L, M, and R (denoting left, middle, and right) as follows: $L = [0, \frac{1-p}{2}), M = [\frac{1-p}{2}, \frac{1+p}{2})$, and $R = [\frac{1+p}{2}, 1]$. The probability that there is a conflicting pair (e,f) in which $e \subset L$ or $f \subset R$ is clearly bounded by the probability that there is an edge of H contained in L or in R, which is at most $2k2^{n-1}(\frac{1-p}{2})^n = k(1-p)^n$.

For any other conflicting pair (e,f), the unique vertex $v = e \cap f$ must satisfy $x_v \in M$. In addition, for any vertex $u \in e - v$, $x_u < x_v$, and for any vertex $w \in f - v$, $x_w > x_v$. For each ordered pair of edges (e,f) of H whose intersection is a single vertex v, we bound the probability that (e,f) forms a conflicting pair as above as follows: The probability that $x_v \in M$ is p. Given that this is the case, and given the value of x_v, the probability that $x_u < x_v$ for all $u \in e - v$ and $x_w > x_v$ for all $w \in f - v$ is $x_v^{n-1}(1 - x_v)^{n-1} \le (\frac{1}{4})^{n-1}$. Therefore, the probability that (e,f) is such a conflicting pair is at most $p(\frac{1}{4})^{n-1}$. As there are less than $k^2 4^{n-1}$ ordered pairs of edges of H, the probability of having a conflicting pair of any kind is smaller than

$$k(1-p)^n + k^2 4^{n-1} p\left(\frac{1}{4}\right)^{n-1} = k(1-p)^n + k^2 p$$

which is less than 1, by assumption. This shows that the algorithm produces a coloring with no monochromatic edges with positive probability, completing the proof. ■

Note that the algorithm in the proof above can also be formulated as a recoloring procedure, justifying the inclusion of this section in the present chapter. Indeed, after selecting the random labels x_v, one can color all vertices v with $x_v \in L \cup M$ blue and all vertices u with $x_u \in R$ red. We can then go over the vertices of M according to the order of their labels, and recolor any blue vertex which is the last one in a blue edge red.

3.6 CONTINUOUS TIME

Discrete random processes can sometimes be analyzed by placing them in a continuous time framework. This allows powerful methods of analysis (such as integration!) to be applied. The approach seems most effective when dealing with random orderings. The proof of Theorem 3.5.1 is of this form. We here give a second example, involving *Random Greedy Packing.*

Let H be a $(k+1)$-uniform hypergraph on a vertex set V of size N. The $e \in H$, which we call edges, are simply subsets of V of size $k+1$. We assume the following conditions:

Degree condition: that is, every $v \in V$ is in precisely D edges.

Codegree condition: that is, every distinct pair $v, v' \in V$ have only $o(D)$ edges in common.

We think of k fixed ($k = 2$ being an illustrative example) and the asymptotics as $N, D \to \infty$, with no set relationship between N and D.

A packing is a family P of vertex disjoint edges $e \in H$. Clearly $|P| \leq N/(k+1)$. We define a randomized algorithm to produce a (not necessarily optimal) packing. Assign to each $e \in H$ uniformly and independently a birth time $x_e \in [0, D)$. [The choice of $[0, D)$ rather than $[0, 1]$ proves to be a technical convenience. Note that, as the x_e are real variables with probability 1, there are no ties.] At time zero, $P \leftarrow \emptyset$. As time progresses from 0 to D, when an edge e is born, it is added to P if possible—that is, unless there is already some $e' \in P$ that overlaps e. Let P_c denote the value of P just before time c—when all e with birth times $t_e < c$ have been examined. Set $P^{\text{FINAL}} = P_D$. Note that by time D all edges have been born and their births were in random order. Thus P^{FINAL} is identical to the discrete process—often called the random greedy algorithm—in which H is first randomly ordered and then the $e \in H$ are considered sequentially.

Theorem 3.6.1 Spencer (1995) *The expected value of* $|P^{\text{FINAL}}|$ *is asymptotic to* $N/(k+1)$.

We say $v \in V$ survives at time c if no $e \in P_c$ contains v, and we let S_c denote the set of $v \in V$ so surviving. Rather than looking at P^{FINAL}, we shall examine P_c, where c is an arbitrary fixed nonnegative real. Let

$$f(c) = \lim^* \Pr[v \in S_c],$$

where, formally, we mean here that for all $\epsilon > 0$ there exist D_0, N_0, and $\delta > 0$ so that, if H is $(k+1)$-uniform on $N > N_0$ vertices with each v in $D > D_0$ edges and every distinct pair $v, v' \in V$ has less than δD common edges, then $|f(c) - \Pr[v \in S_c]| < \epsilon$ for all $v \in V$.

The heart of the argument lies in showing that $f(c)$ exists by defining a continuous time birth process yielding that value. We now describe the birth process, omitting some of the epsilondeltamanship needed to formally show the limit.

Our birth process starts at time c, and time goes backwards to 0. It begins with root Eve, our anthropomorphized v. Eve has births in time interval $[0, c)$. The number of births is given by a Poisson distribution with mean c and, given their number, their times are uniformly and independently distributed. [This is a standard Poisson process with intensity 1. Equivalently, on any infinitesimal time interval $[x, x + dx)$, Eve has probability dx of giving birth and these events are independent over disjoint intervals.] Our fertile Eve always gives birth to k-tuplets. Each child is born fertile under the same rules, so if Alice in born at time x, she (in our unisexual model) has a Poisson distribution with mean x of births, uniformly distributed in $[0, x)$.

The resulting random tree $T = T_c$ can be shown to be finite (note the time interval is finite) with probability 1. Given a finite T, we say for each vertex Alice that Alice survives or dies according to the following scheme:

Menendez Rule: If Alice has given birth to a set (or possibly several sets) of k-tuplets all of whom survived, then she dies; otherwise she survives.

In particular, if Alice is childless, she survives. We can then work our way up the tree to determine of each vertex whether she survives or dies.

Example. $c = 10, k = 2$. Eve gives birth to Alice, Barbara at time 8.3, and then to Rachel, Siena at time 4.3. Alice gives birth to Nancy, Olive at time 5.7 and Rachel gives birth to Layla, Mayavati at time 0.4. There are no other births. The leaves Nancy, Olive, Layla, Mayavati, Barbara, and Siena then survive. Working up the tree, Alice and Rachel die. In neither of Eve's births did both children Survive, and therefore Eve survives.

We define $f(c)$ to be the probability that the root Eve survives in the random birth tree $T = T_c$.

We outline the equivalence by defining a tree $T = T_c(v)$ for $v \in H$. For each edge e containing v with birth time $t = t_e < c$, we say that $e - \{v\}$ is a set of k-tuplets born to v at time t. We work recursively; if w is born at time t, then for each e' containing w with birth time $t' = t_{e'} < t$ we say that $e' - \{w\}$ is a set of k-tuplets born to w at time t'. Possibly, this process does not give a tree since the same vertex w may be reached in more than one way—the simplest example is if $v \in e, e'$, where both have birth times less than c and e, e' share another common vertex w. Then the process is stillborn, and $T_c(v)$ is not defined. We will argue that, for any particular tree T

$$\lim^* \Pr[T_c(v) \cong T] = \Pr[T_c = T]. \tag{3.1}$$

As $\sum_T \Pr[T_c = T] = 1$, this gives a rather roundabout argument that the process defining $T_c(v)$ is almost never stillborn.

We find $T_c(v)$ in stages. First consider the D edges e containing v. The number of them with birth time $t_e < c$ has binomial distribution BIN$[D, c/D]$, which approaches (critically) the Poisson distribution with mean c. Given that there are l such e, their birth times t_e are uniformly distributed. There are (by the codegree condition) $o(D^2)$ pairs e, e' containing v and also some other vertex so there is probability $o(1)$ that two such e, e' have birth time less than c. Now suppose $T_c(v)$ has been built out to a certain level and a vertex w has been born at time t. There are only $o(D)$ common edges between w and any of the finite number of w' already born, so there are still about D edges e containing w and no other such w'. We now examine their birth times, the number with $t_e < x$ has binomial distribution BIN$[D - o(D), x/D]$ which approaches the Poisson distribution with mean x. As above, almost surely no two such e, e' will have a common vertex other than w itself. For any fixed T, the calculation of $\Pr[T_c(v) \cong T]$ involves a finite number of these limits, which allows us to conclude (3.1).

With $c < d$, the random tree T_d includes T_c as a subtree by considering only those births of Eve occurring in $[0, c)$. If Eve survives in T_d, she must survive in T_c. Hence, $f(d) \le f(c)$. We now claim

$$\lim_{c\to\infty} f(c) = 0.$$

If not, the nondecreasing f would have a limit $L > 0$, and all $f(x) \ge L$. Suppose in T_c Eve had i births. In each birth there would be probability at least L^k that all k children survived. The probability that Eve survived would then be at most $(1 - L^k)^i$. Since the number of Eve's births is Poisson with mean c

$$f(c) \le \sum_{i=0}^{\infty} e^{-c}\frac{c^i}{i!}(1 - L^k)^i = e^{-L^k c}$$

but then $\lim_{c\to\infty} f(c) = 0$, a contradiction.

By Linearity of Expectation, $\mathrm{E}[|S_c|] \to f(c)n$. As $(k+1)|P_c| + |S_c| = n$, $\mathrm{E}[|P_c|] \to (1 - f(c))n/(k+1)$. But $\mathrm{E}[|P^{\mathrm{FINAL}}|] \ge \mathrm{E}[|P_c|]$. We make $f(c)$ arbitrarily small by taking c appropriately big, so that $\mathrm{E}[|P^{\mathrm{FINAL}}|] \ge (1 - o(1))n/(k+1)$. As $|P^{\mathrm{FINAL}}| \le n/(k+1)$ always, the theorem follows.

Remark. We can actually say more about $f(c)$. For Δc small, $f(c + \Delta c) - f(c) \sim -(\Delta c) f(c)^{k+1}$ as, roughly, an Eve starting at time $c + \Delta c$ might have a birth in time interval $[c, c + \Delta c)$, all of whose children survive, while Eve has no births in $[0, c)$, all of whose children survive. Letting $\Delta c \to 0$ yields the differential equation $f'(c) = -f(c)^{k+1}$. The initial value $f(0) = 1$ gives a unique solution $f(c) = (1 + ck)^{-1/k}$. It is intriguing to plug in $c = D$. This is not justified, as our limit arguments were for c fixed and $N, D \to \infty$. Nonetheless, that *would* yield $\mathrm{E}[|S_D|] = O(ND^{-1/k})$, that is, the random greedy algorithm would leave $O(ND^{-1/k})$ vertices uncovered. Suppose we replace the codegree condition by the stronger condition that every distinct pair $v, v' \in V$ has at most one edge in common. There is computer simulation data which

show that in those cases the random greedy algorithm does leave $O(ND^{-1/k})$ vertices uncovered. This remains an open question, though it is shown in Alon, Kim and Spencer (1997) that this is the case for a modified version of the greedy algorithm.

Corollary 3.6.2 *Under the assumptions of the theorem, there exists a packing P of size $\sim N/(k+1)$.*

Proof. We have defined a random process that gives a packing with expected size $\sim N/(k+1)$, and our usual magic implies that such a P must exist. ■

In particular, this gives an alternate proof to the Erdős–Hanani conjecture, first proved by Rödl (1985) as given in Section 4.7. We use the notation of that section and define the packing number $m(n,k,l)$ as the maximal size of a family F of k-element subsets of $[n] = \{1,\ldots,n\}$ such that no l-set is contained in more than one k-set. Define a hypergraph $H = H(n,k,l)$ as follows: The vertices of H are the l-element subsets of $[n]$. For each k-element $A \subset [n]$, we define an edge e_A as the set of l-element subsets of A. A family F satisfying the above conditions then corresponds to a packing $P = \{e_A : A \in F\}$ in H. H has $N = \binom{n}{l}$ vertices. Each edge e_A has size $K+1 = \binom{k}{l}$. Each vertex is in $D = \binom{n-l}{k-l}$ edges. The number of edges containing two vertices v, v' depends on their intersection. It is largest (given $v \neq v'$) when v, v' (considered as l-sets) overlap in $l-1$ points, and then it is $\binom{n-l-1}{k-l-1}$. We assume (as in Section 4.7) that k, l are fixed and $n \to \infty$ so this number of common edges is $o(D)$. The assumptions of Section 4.7 give $K+1$ fixed, $N, D \to \infty$, so that there exists P with

$$m(n,k,l) = |P| \sim N/(K+1) \sim \binom{n}{l} \Big/ \binom{k}{l}.$$

3.7 EXERCISES

1. As shown in Section 3.1, the Ramsey number $R(k,k)$ satisfies

$$R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$$

for every integer n. Conclude that

$$R(k,k) \geq (1-o(1))\frac{k}{e}2^{k/2}.$$

2. Prove that the Ramsey number $R(4,k)$ satisfies

$$R(4,k) \geq \Omega((k/\ln\, k)^2).$$

3. Prove that every three-uniform hypergraph with n vertices and $m \geq n/3$ edges contains an independent set (i.e., a set of vertices containing no edges) of size at least
$$\frac{2n^{3/2}}{3\sqrt{3}\sqrt{m}}.$$

4. (*) Show that there is a finite n_0 such that any directed graph on $n > n_0$ vertices in which each outdegree is at least $\log_2 n - \frac{1}{10}\log_2 \log_2 n$ contains an even simple directed cycle.

THE PROBABILISTIC LENS:
High Girth and High Chromatic Number

Many consider this one of the most pleasing uses of the probabilistic method, as the result is surprising and does not appear to call for nonconstructive techniques. The *girth* of a graph G is the size of its shortest cycle, $\alpha(G)$ is the size of the largest independent set in G, and $\chi(G)$ denotes its chromatic number.

Theorem 1 Erdős (1959) *For all k, l, there exists a graph G with girth$(G) > l$ and $\chi(G) > k$.*

Proof. Fix $\theta < 1/l$ and let $G \sim G(n,p)$ with $p = n^{\theta-1}$; that is, G is a random graph on n vertices chosen by picking each pair of vertices as an edge randomly and independently with probability p. Let X be the number of cycles of size at most l. Then

$$\mathrm{E}[X] = \sum_{i=3}^{l} \frac{(n)_i}{2i} p^i \le \sum_{i=3}^{l} \frac{n^{\theta i}}{2i} = o(n)$$

as $\theta l < 1$. In particular,

$$\Pr[X \ge n/2] = o(1).$$

Set $x = \lceil (3/p) \ln n \rceil$ so that

$$\Pr[\alpha(G) \ge x] \le \binom{n}{x} (1-p)^{\binom{x}{2}} < \left[ne^{-p(x-1)/2}\right]^x = o(1).$$

Let n be sufficiently large so that both these events have probability less than 0.5. Then there is a specific G with less than $n/2$ cycles of length at most l and

with $\alpha(G) < 3n^{1-\theta} \ln n$. Remove from G a vertex from each cycle of length at most l. This gives a graph G^* with at least $n/2$ vertices. G^* has girth greater than l and $\alpha(G^*) \leq \alpha(G)$. Thus

$$\chi(G^*) \geq \frac{|G^*|}{\alpha(G^*)} \geq \frac{n/2}{3n^{1-\theta} \ln n} = \frac{n^\theta}{6 \ln n}.$$

To complete the proof, let n be sufficiently large so that this is greater than k. ■

4

The Second Moment

Many persons who have not studied mathematics confuse it with arithmetic and consider it a dry and arid science. Actually, however, this science requires great fantasy.
–Sophia Kovalevsky

4.1 BASICS

After expectation, the most vital statistic for a random variable X is the *variance*. We denote it $\mathrm{Var}[X]$. It is defined by

$$\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2]$$

and measures how spread out X is from its expectation. We shall generally, following standard practice, let μ denote expectation and σ^2 denote variance. The positive square root σ of the variance is called the *standard deviation*. With this notation, here is our basic tool.

Theorem 4.1.1 [Chebyshev's Inequality] *For any positive λ,*

$$\Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

Proof. $\sigma^2 = \mathrm{Var}[X] = \mathrm{E}[(X - \mu)^2] \geq \lambda^2\sigma^2 \Pr[|X - \mu| \geq \lambda\sigma]$. ■

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

The use of Chebyshev's Inequality is called the *second moment method*.

Chebyshev's Inequality is best possible when no additional restrictions are placed on X as X may be $\mu + \lambda\sigma$ and $\mu - \lambda\sigma$ with probability $1/2\lambda^2$ and otherwise μ. Note, however, that when X is a normal distribution with mean μ and standard deviation σ, then

$$\Pr[|X - \mu| \geq \lambda\sigma] = 2\int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

and for λ large, this quantity is asymptotically $\sqrt{2/\pi}e^{-\lambda^2/2}/\lambda$, which is significantly smaller than $1/\lambda^2$. In Chapters 7 and 8, we shall see examples where X is the sum of "nearly independent" random variables and these better bounds can apply.

Suppose we have a decomposition

$$X = X_1 + \cdots + X_m.$$

Then $\mathrm{Var}[X]$ may be computed by the formula

$$\mathrm{Var}[X] = \sum_{i=1}^{m} \mathrm{Var}[X_i] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j].$$

Here the second sum is over ordered pairs, and the *covariance* $\mathrm{Cov}[Y, Z]$ is defined by

$$\mathrm{Cov}[Y, Z] = \mathrm{E}[YZ] - \mathrm{E}[Y]\mathrm{E}[Z].$$

In general, if Y, Z are independent, then $\mathrm{Cov}[Y, Z] = 0$. This often simplifies variance calculations considerably. Now suppose further, as will generally be the case in our applications, that the X_i are indicator random variables: that is, $X_i = 1$ if a certain event A_i holds, and $X_i = 0$ otherwise. If X_i is one with probability $p_i = \Pr[A_i]$, then

$$\mathrm{Var}[X_i] = p_i(1 - p_i) \leq p_i = \mathrm{E}[X_i],$$

and so

$$\mathrm{Var}[X] \leq \mathrm{E}[X] + \sum_{i \neq j} \mathrm{Cov}[X_i, X_j].$$

4.2 NUMBER THEORY

The second moment method is an effective tool in number theory. Let $\nu(n)$ denote the number of primes p dividing n. (We do not count multiplicity though it would make little difference.) The following result says, roughly, that "almost all" n have "very close to" $\ln \ln n$ prime factors. This was first shown by Hardy and Ramanujan (1920) by a quite complicated argument. We give a remarkably simple proof of Turán (1934), a proof that played a key role in the development of probabilistic methods in number theory.

Theorem 4.2.1 *Let $\omega(n) \to \infty$ arbitrarily slowly. Then the number of x in $\{1, \dots, n\}$ such that*

$$|\nu(x) - \ln \ln n| > \omega(n)\sqrt{\ln \ln n}$$

is $o(n)$.

Proof. Let x be randomly chosen from $\{1, \dots, n\}$. For p prime, set

$$X_p = \begin{cases} 1 & \text{if } p|x, \\ 0 & \text{otherwise.} \end{cases}$$

Set $M = n^{1/10}$ and set $X = \sum X_p$, with the summation being over all primes $p \le M$. As no $x \le n$ can have more than 10 prime factors larger than M, we have $\nu(x) - 10 \le X(x) \le \nu(x)$ so that large deviation bounds on X will translate into asymptotically similar bounds for ν. [Here 10 could be any (large) constant.] Now

$$\mathrm{E}[X_p] = \frac{\lfloor n/p \rfloor}{n}.$$

As $y - 1 < \lfloor y \rfloor \le y$,

$$\mathrm{E}[X_p] = 1/p + O(1/n).$$

By Linearity of Expectation,

$$\mathrm{E}[X] = \sum_{p \le M} \left(\frac{1}{p} + O\left(\frac{1}{n}\right) \right) = \ln \ln n + O(1),$$

where we used the well-known fact that $\sum_{p \le x}(1/p) = \ln \ln x + O(1)$, which can be proved by combining (in a clever way) Stirling's formula with Abel summation.

Now we find an asymptotic expression for

$$\mathrm{Var}[X] = \sum_{p \le M} \mathrm{Var}[X_p] + \sum_{p \ne q} \mathrm{Cov}[X_p, X_q].$$

As $\mathrm{Var}[X_p] = (1/p)(1 - 1/p) + O(1/n)$,

$$\sum_{p \le M} \mathrm{Var}[X_p] = \left(\sum_{p \le M} \frac{1}{p} \right) + O(1) = \ln \ln n + O(1).$$

With p, q distinct primes, $X_p X_q = 1$ if and only if $p|x$ and $q|x$, which occurs if and only if $pq|x$. Hence

$$\begin{aligned} \mathrm{Cov}[X_p, X_q] &= \mathrm{E}[X_p X_q] - \mathrm{E}[X_p]\mathrm{E}[X_q] \\ &= \frac{\lfloor n/pq \rfloor}{n} - \frac{\lfloor n/p \rfloor}{n} \frac{\lfloor n/q \rfloor}{n} \end{aligned}$$

$$\le \frac{1}{pq} - \left(\frac{1}{p} - \frac{1}{n}\right)\left(\frac{1}{q} - \frac{1}{n}\right)$$
$$\le \frac{1}{n}\left(\frac{1}{p} + \frac{1}{q}\right).$$

Thus

$$\sum_{p\neq q} \mathrm{Cov}[X_p, X_q] \le \frac{1}{n}\sum_{p\neq q}\left(\frac{1}{p} + \frac{1}{q}\right) \le \frac{2M}{n}\sum \frac{1}{p}.$$

Thus

$$\sum_{p\neq q} \mathrm{Cov}[X_p, X_q] \le O(n^{-9/10} \ln\ln n) = o(1)$$

and similarly

$$\sum_{p\neq q} \mathrm{Cov}[X_p, X_q] \ge -o(1).$$

That is, the covariances do not affect the variance $\mathrm{Var}[X] = \ln\ln n + O(1)$, and Chebyshev's inequality actually gives

$$\Pr\left[|X - \ln\ln n| > \lambda\sqrt{\ln\ln n}\right] < \lambda^{-2} + o(1)$$

for any constant $\lambda > 0$. As $|X - \nu| \le 10$, the same holds for ν. ■

In a classic paper, Erdős and Kac (1940) showed, essentially, that ν does behave like a normal distribution with mean and variance $\ln\ln n$. Here is their precise result:

Theorem 4.2.2 *Let λ be fixed, positive, negative, or zero. Then*

$$\lim_{n\to\infty} \frac{1}{n}\left|\left\{x : 1 \le x \le n, \nu(x) \ge \ln\ln n + \lambda\sqrt{\ln\ln n}\right\}\right| = \int_\lambda^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Proof. We outline the argument, emphasizing the similarities to Turán's proof. Fix a function $s(n)$ with $s(n) \to \infty$ and $s(n) = o((\ln\ln n)^{1/2})$—for example, $s(n) = \ln\ln\ln n$. Set $M = n^{1/s(n)}$. Set $X = \sum X_p$, the summation over all primes $p \le M$. As no $x \le n$ can have more than $s(n)$ prime factors greater than M, we have $\nu(x) - s(n) \le X(x) \le \nu(x)$ so that it suffices to show Theorem 4.2.2 with ν replaced by X. Let Y_p be independent random variables with $\Pr[Y_p = 1] = p^{-1}$, $\Pr[Y_p = 0] = 1 - p^{-1}$, and set $Y = \sum Y_p$, the summation over all primes $p \le M$. This Y represents an idealized version of X. Set

$$\mu = \mathrm{E}[Y] = \sum_{p\le M} p^{-1} = \ln\ln n + o\left((\ln\ln n)^{1/2}\right)$$

and

$$\sigma^2 = \mathrm{Var}[Y] = \sum_{p \le M} p^{-1}(1 - p^{-1}) \sim \ln \ln n$$

and define the normalized $\widetilde{Y} = (Y - \mu)/\sigma$. From the central limit theorem, $\widetilde{Y}$ approaches the standard normal N and $\mathrm{E}[\widetilde{Y}^k] \to \mathrm{E}[N^k]$ for every positive integer k. Set $\widetilde{X} = (X - \mu)/\sigma$. We compare $\widetilde{X}, \widetilde{Y}$.

For any distinct primes $p_1, \ldots, p_s \le M$,

$$\mathrm{E}[X_{p_1} \cdots X_{p_s}] - \mathrm{E}[Y_{p_1} \cdots Y_{p_s}] = \frac{1}{n} \left\lfloor \frac{n}{p_1 \cdots p_s} \right\rfloor - \frac{1}{p_1 \cdots p_s} = O(n^{-1}).$$

We let k be an arbitrary fixed positive integer and compare $\mathrm{E}[\widetilde{X}^k]$ and $\mathrm{E}[\widetilde{Y}^k]$. Expanding, $\widetilde{X}^k$ is a polynomial in X with coefficients $n^{o(1)}$. Further expanding each $X^j = (\sum X_p)^j$—always reducing X_p^a to X_p when $a \ge 2$—gives the sum of $O(M^k) = n^{o(1)}$ terms of the form $X_{p_1} \cdots X_{p_s}$. The same expansion applies to $\widetilde{Y}$. As the corresponding terms have expectations within $O(n^{-1})$, the total difference

$$\mathrm{E}[\widetilde{X}^k] - \mathrm{E}[\widetilde{Y}^k] = n^{-1+o(1)} = o(1).$$

Hence, each moment of $\widetilde{X}$ approaches that of the standard normal N. A standard, though nontrivial, theorem in probability theory gives that $\widetilde{X}$ must therefore approach N in distribution. ■

We recall the famous quotation of G. H. Hardy:

> 317 is a prime, not because we think so, or because our minds are shaped in one way rather than another, but *because it is so*, because mathematical reality is built that way.

How ironic—though not contradictory—that the methods of probability theory can lead to a greater understanding of the prime factorization of integers! Additional results applying information about the moments of a distribution in order to determine it appear in Chapter 8; see also Billingsley (1995).

4.3 MORE BASICS

Let X be a nonnegative integral-valued random variable, and suppose we want to bound $\Pr[X = 0]$ given the value $\mu = \mathrm{E}[X]$. If $\mu < 1$, we may use the inequality

$$\Pr[X > 0] \le \mathrm{E}[X]$$

so that, if $\mathrm{E}[X] \to 0$, then $X = 0$ almost always. (Here we are imagining an infinite sequence of X dependent on some parameter n going to infinity.) But now suppose $\mathrm{E}[X] \to \infty$. It does *not* necessarily follow that $X > 0$ almost always. For example,

let X be the number of deaths due to nuclear war in the 12 months after reading this paragraph. Calculation of $\mathrm{E}[X]$ can make for lively debate, but few would deny that it is quite large. Yet we may believe—or hope—that $\Pr[X \neq 0]$ is very close to zero. We can sometimes deduce $X > 0$ almost always if we have further information about $\mathrm{Var}[X]$.

Theorem 4.3.1 $\Pr[X = 0] \leq \dfrac{\mathrm{Var}[X]}{\mathrm{E}[X]^2}$.

Proof. Set $\lambda = \mu/\sigma$ in Chebyshev's Inequality. Then

$$\Pr[X = 0] \leq \Pr[|X - \mu| \geq \lambda\sigma] \leq \frac{1}{\lambda^2} = \frac{\sigma^2}{\mu^2}.$$

■

We generally apply this result in asymptotic terms.

Corollary 4.3.2 *If* $\mathrm{Var}[X] = o(\mathrm{E}[X]^2)$*, then* $X > 0$ *almost always.*

The proof of Theorem 4.3.1 actually gives that, for any $\epsilon > 0$,

$$\Pr[|X - \mathrm{E}[X]| \geq \epsilon \mathrm{E}[X]] \leq \frac{\mathrm{Var}[X]}{\epsilon^2 \mathrm{E}[X]^2}$$

and thus in asymptotic terms we actually have the following stronger assertion:

Corollary 4.3.3 *If* $\mathrm{Var}[X] = o(\mathrm{E}[X]^2)$*, then* $X \sim \mathrm{E}[X]$ *almost always.*

Suppose, again, $X = X_1 + \cdots + X_m$, where X_i is the indicator random variable for event A_i. For indices i, j, write $i \sim j$ if $i \neq j$ and the events A_i, A_j are not independent. We set (the sum over ordered pairs)

$$\Delta = \sum_{i \sim j} \Pr[A_i \wedge A_j].$$

Note that, when $i \sim j$

$$\mathrm{Cov}[X_i, X_j] = \mathrm{E}[X_i X_j] - \mathrm{E}[X_i]\mathrm{E}[X_j] \leq \mathrm{E}[X_i X_j] = \Pr[A_i \wedge A_j]$$

and that when $i \neq j$ and not $i \sim j$, then $\mathrm{Cov}[X_i, X_j] = 0$. Thus

$$\mathrm{Var}[X] \leq \mathrm{E}[X] + \Delta.$$

Corollary 4.3.4 *If* $\mathrm{E}[X] \to \infty$ *and* $\Delta = o(\mathrm{E}[X]^2)$*, then* $X > 0$ *almost always. Furthermore* $X \sim \mathrm{E}[X]$ *almost always.*

Let us say $X_1, \ldots, X_m$ are *symmetric* if for every $i \neq j$ there is a measure preserving mapping of the underlying probability space that permutes the m events and sends

event A_i to event A_j. Examples will appear in the next section. In this instance, we write

$$\Delta = \sum_{i\sim j} \Pr[A_i \wedge A_j] = \sum_i \Pr[A_i] \sum_{j\sim i} \Pr[A_j \mid A_i]$$

and note that the inner summation is independent of i. We set

$$\Delta^* = \sum_{j\sim i} \Pr[A_j \mid A_i],$$

where i is any fixed index. Then

$$\Delta = \sum_i \Pr[A_i]\Delta^* = \Delta^* \sum_i \Pr[A_i] = \Delta^* \mathrm{E}[X].$$

Corollary 4.3.5 *If* $\mathrm{E}[X] \to \infty$ *and* $\Delta^* = o(\mathrm{E}[X])$*, then* $X > 0$ *almost always. Furthermore,* $X \sim \mathrm{E}[X]$ *almost always.*

The condition of Corollary 4.3.5 has the intuitive sense that conditioning on any specific A_i holding does not substantially increase the expected number $\mathrm{E}[X]$ of events holding.

4.4 RANDOM GRAPHS

The random graph $G(n,p)$ is, informally, the graph on n labeled vertices, obtained by selecting each pair of vertices to be an edge, randomly and independently, with probability p. A property of graphs is a family of graphs closed under isomorphism. A function $r(n)$ is a *threshold function* for some property P, if whenever $p = p(n) \ll r(n)$, then $G(n,p)$ does not satisfy P almost always, and whenever $p \gg r(n)$, then $G(n,p)$ satisfies P almost always. For more precise definitions of the random graph $G(n,p)$ and of threshold functions, see Section 10.1.

The results of this section are generally surpassed by those of Chapter 10 but they were historically the first results and provide a good illustration of the second moment. We begin with a particular example. By $\omega(G)$ we denote here and in the rest of the book the number of vertices in the maximum clique of the graph G.

Theorem 4.4.1 *The property* $\omega(G) \geq 4$ *has threshold function* $n^{-2/3}$.

Proof. For every 4-set S of vertices in $G(n,p)$, let A_S be the event "S is a clique" and X_S its indicator random variable. Then

$$\mathrm{E}[X_S] = \Pr[A_S] = p^6$$

as six different edges must all lie in $G(n,p)$. Set

$$X = \sum_{|S|=4} X_S$$

so that X is the number of 4-cliques in G, and $\omega(G) \geq 4$ if and only if $X > 0$. Linearity of Expectation gives

$$\mathrm{E}[X] = \sum_{|S|=4} \mathrm{E}[X_S] = \binom{n}{4} p^6 \sim \frac{n^4 p^6}{24}.$$

When $p(n) \ll n^{-2/3}$, $\mathrm{E}[X] = o(1)$ and so $X = 0$, almost surely.

Now suppose $p(n) \gg n^{-2/3}$ so that $\mathrm{E}[X] \to \infty$, and consider the Δ^* of Corollary 4.3.5. (All 4-sets "look the same" so that the X_S are symmetric.) Here $S \sim T$ if and only if $S \neq T$ and S, T have common edges, that is, if and only if $|S \cap T| = 2$ or 3. Fix S. There are $O(n^2)$ sets T with $|S \cap T| = 2$, and for each of these, $\Pr[A_T \mid A_S] = p^5$. There are $O(n)$ sets T with $|S \cap T| = 3$, and for each of these $\Pr[A_T \mid A_S] = p^3$. Thus

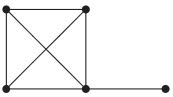
$$\Delta^* = O(n^2 p^5) + O(np^3) = o(n^4 p^6) = o(\mathrm{E}[X])$$

since $p \gg n^{-2/3}$. Corollary 4.3.5 therefore applies and $X > 0$; that is, there *does* exist a clique of size 4, almost always. ■

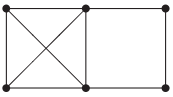
The proof of Theorem 4.4.1 appears to require a fortuitous calculation of Δ^*. The following definitions pave the way for the more general Theorem 4.4.2:

Definition 1 *Let H be a graph with v vertices and e edges. We call $\rho(H) = e/v$ the* density *of H. We call H* balanced *if every subgraph H' has $\rho(H') \leq \rho(H)$. We call H* strictly balanced *if every proper subgraph H' has $\rho(H') < \rho(H)$.*

Example. K_4 and, in general, K_k are strictly balanced. The graph



is not balanced, as it has density $7/5$, while the subgraph K_4 has density $3/2$. The graph



is balanced but not strictly balanced as it and its subgraph K_4 have density $3/2$.

Theorem 4.4.2 *Let H be a balanced graph with v vertices and e edges. Let $A(G)$ be the event that H is a subgraph (not necessarily induced) of G. Then $p = n^{-v/e}$ is the threshold function for A.*

Proof. We follow the argument of Theorem 4.4.1. For each v-set S, let A_S be the event that $G|_S$ contains H as a subgraph. Then

$$p^e \leq \Pr[A_S] \leq v! p^e.$$

(Any particular placement of H has probability p^e of occurring and there are at most $v!$ possible placements. The precise calculation of $\Pr[A_S]$ is, in general, complicated due to the overlapping of potential copies of H.) Let X_S be the indicator random variable for A_S and

$$X = \sum_{|S|=v} X_S$$

so that A holds if and only if $X > 0$. Linearity of expectation gives

$$\mathrm{E}[X] = \sum_{|S|=v} \mathrm{E}[X_S] = \binom{n}{v} \Pr[A_S] = \Theta(n^v p^e).$$

If $p \ll n^{-v/e}$, then $\mathrm{E}[X] = o(1)$, so $X = 0$ almost always.

Now assume $p \gg n^{-v/e}$ so that $\mathrm{E}[X] \to \infty$ and consider the Δ^* of Corollary 4.3.5. (All v-sets look the same, so the X_S are symmetric.) Here $S \sim T$ if and only if $S \neq T$ and S, T have common edges, that is, if and only if $|S \cap T| = i$ with $2 \le i \le v-1$. Let S be fixed. We split

$$\Delta^* = \sum_{T \sim S} \Pr[A_T \mid A_S] = \sum_{i=2}^{v-1} \sum_{|T \cap S|=i} \Pr[A_T \mid A_S].$$

For each i, there are $O(n^{v-i})$ choices of T. Fix S, T and consider $\Pr[A_T \mid A_S]$. There are $O(1)$ possible copies of H on T. Each has—since, critically, H is balanced—at most ie/v edges with both vertices in S and thus at least $e - (ie/v)$ other edges. Hence

$$\Pr[A_T \mid A_S] = O\left(p^{e-(ie/v)}\right)$$

and

$$\begin{aligned}
\Delta^* &= \sum_{i=2}^{v-1} O\left(n^{v-i} p^{e-(ie/v)}\right) \\
&= \sum_{i=2}^{v-1} O\left((n^v p^e)^{1-i/v}\right) \\
&= \sum_{i=2}^{v-1} o(n^v p^e) \\
&= o(\mathrm{E}[X])
\end{aligned}$$

since $n^v p^e \to \infty$. Hence Corollary 4.3.5 applies. ∎

Theorem 4.4.3 *In the notation of Theorem 4.4.2, if H is* not *balanced, then $p = n^{-v/e}$ is* not *the threshold function for A.*

Proof. Let H_1 be a subgraph of H with v_1 vertices, e_1 edges, and $e_1/v_1 > e/v$. Let α satisfy $v_1/e_1 < \alpha < v/e$, and set $p = n^{-\alpha}$. The expected number of copies of H_1 is then $o(1)$, so almost always $G(n,p)$ contains no copy of H_1. But if it contains no copy of H_1, then it surely can contain no copy of H. ■

The threshold function for the property of containing a copy of H, for general H, was examined in the original papers of Erdős and Rényi (1960). It still provides an excellent introduction to the theory of random graphs. Let H_1 be that subgraph with maximal density $\rho(H_1) = e_1/v_1$. (When H is balanced, we may take $H_1 = H$.) They showed that $p = n^{-v_1/e_1}$ is the threshold function. We do not show this here though it follows fairly straightforwardly from these methods.

We finish this section with two strengthenings of Theorem 4.4.2.

Theorem 4.4.4 *Let H be balanced with v vertices, e edges and a automorphisms. Let X be the number of copies of H in $G(n,p)$. Assume $p \gg n^{-v/e}$. Then almost always*

$$X \sim \frac{n^v p^e}{a}.$$

Proof. Label the vertices of H by $1, \dots, v$. For each ordered $x_1, \dots, x_v$, let $A_{x_1,\dots,x_v}$ be the event in which $x_1, \dots, x_v$ provides a copy of H in that order. Specifically we define

$$A_{x_1,\dots,x_v} : \{i,j\} \in E(H) \Rightarrow \{x_i, x_j\} \in E(G).$$

We let $I_{x_1,\dots,x_v}$ be the corresponding indicator random variable. We define an equivalence class on v-tuples by setting $(x_1, \dots, x_v) \equiv (y_1, \dots, y_v)$ if there is an automorphism σ of $V(H)$ so that $y_{\sigma(i)} = x_i$ for $1 \le i \le v$. Then

$$X = \sum I_{x_1,\dots,x_v}$$

gives the number of copies of H in G, where the sum is taken over one entry from each equivalence class. As there are $(n)_v/a$ terms,

$$\mathrm{E}[X] = \frac{(n)_v}{a}\mathrm{E}[I_{x_1,\dots,x_v}] = \frac{(n)_v p^e}{a} \sim \frac{n^v p^e}{a}.$$

Our assumption $p \gg n^{-v/e}$ implies $\mathrm{E}[X] \to \infty$. It suffices therefore to show $\Delta^* = o(\mathrm{E}[X])$. Fixing $x_1, \dots, x_v$,

$$\Delta^* = \sum_{(y_1,\dots,y_v)\sim(x_1,\dots,x_v)} \Pr[A_{(y_1,\dots,y_v)} \mid A_{(x_1,\dots,x_v)}].$$

There are $v!/a = O(1)$ terms with $\{y_1, \dots, y_v\} = \{x_1, \dots, x_v\}$, and for each, the conditional probability is at most 1 (actually, at most p), thus contributing $O(1) = o(\mathrm{E}[X])$ to Δ^*. When $\{y_1, \dots, y_v\} \cap \{x_1, \dots, x_v\}$ has i elements, $2 \le i \le v-1$, the argument of

Theorem 4.4.2 gives that the contribution to Δ^* is $o(\mathrm{E}[X])$. Altogether, $\Delta^* = o(\mathrm{E}[X])$ and we apply Corollary 4.3.5. ■

Theorem 4.4.5 *Let H be* any *fixed graph. For every subgraph H' of H (including H itself), let $X_{H'}$ denote the number of copies of H' in $G(n,p)$. Assume p is such that $\mathrm{E}[X_{H'}] \to \infty$ for every H'. Then*

$$X_H \sim \mathrm{E}[X_H]$$

almost always.

Proof. Let H have v vertices and e edges. As in Theorem 4.4.4, it suffices to show $\Delta^* = o(\mathrm{E}[X])$. We split Δ^* into a finite number of terms. For each H' with w vertices and f edges, we have those $(y_1, \ldots, y_v)$ that overlap with the fixed $(x_1, \ldots, x_v)$ in a copy of H'. These terms contribute, up to constants,

$$n^{v-w}p^{e-f} = \Theta\left(\frac{\mathrm{E}\left[X_H\right]}{\mathrm{E}[X_{H'}]}\right) = o(\mathrm{E}[X_H])$$

to Δ^*. Hence Corollary 4.3.5 does apply. ■

4.5 CLIQUE NUMBER

Now we fix the edge probability $p = \frac{1}{2}$ and consider the clique number $\omega(G)$. We set

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}},$$

the expected number of k-cliques. The function $f(k)$ drops under one at $k \sim 2\log_2 n$. (Very roughly, $f(k)$ is like $n^k 2^{-k^2/2}$.)

Theorem 4.5.1 *Let $k = k(n)$ satisfy $k \sim 2\log_2 n$ and $f(k) \to \infty$. Then almost always $\omega(G) \geq k$.*

Proof. For each k-set S, let A_S be the event "S is a clique" and X_S the corresponding indicator random variable. We set

$$X = \sum_{|S|=k} X_S$$

so that $\omega(G) \geq k$ if and only if $X > 0$. Then $\mathrm{E}[X] = f(k) \to \infty$, and we examine the Δ^* of Corollary 4.3.5. Fix S and note that $T \sim S$ if and only if $|T \cap S| = i$, where $2 \leq i \leq k-1$. Hence

$$\Delta^* = \sum_{i=2}^{k-1} \binom{k}{i} \binom{n-k}{k-i} 2^{\binom{i}{2}-\binom{k}{2}}$$

and so

$$\frac{\Delta^*}{\mathrm{E}[X]} = \sum_{i=2}^{k-1} g(i),$$

where we set

$$g(i) = \frac{\binom{k}{i}\binom{n-k}{k-i}}{\binom{n}{k}} 2^{\binom{i}{2}}.$$

Observe that $g(i)$ may be thought of as the probability that a randomly chosen T will intersect a fixed S in i points times the factor increase in $\Pr[A_T]$ when it does. Setting $i = 2$,

$$g(2) = 2\frac{\binom{k}{2}\binom{n-k}{k-2}}{\binom{n}{k}} \sim \frac{k^4}{n^2} \leq o(n^{-1}).$$

At the other extreme $i = k - 1$,

$$g(k-1) = \frac{k(n-k)2^{-(k-1)}}{\binom{n}{k}2^{-\binom{k}{2}}} \sim \frac{2kn2^{-k}}{\mathrm{E}[X]}.$$

As $k \sim 2\log_2 n$, the numerator is $n^{-1+o(1)}$. The denominator approaches infinity and so $g(k-1) \leq o(n^{-1})$. Some detailed calculation (which we omit) gives that the remaining $g(i)$ and their sum are also negligible so that Corollary 4.3.5 applies. ■

Theorem 4.5.1 leads to a strong concentration result for $\omega(G)$. For $k \sim 2\log_2 n$,

$$\frac{f(k+1)}{f(k)} = \frac{n-k}{k+1}2^{-k} = n^{-1+o(1)} = o(1).$$

Let $k_0 = k_0(n)$ be that value with $f(k_0) \geq 1 > f(k_0+1)$. For "most" n, the function $f(k)$ will jump from a large $f(k_0)$ to a small $f(k_0+1)$. The probability that G contains a clique of size $k_0 + 1$ is at most $f(k_0+1)$, which will be very small. When $f(k_0)$ is large, Theorem 4.5.1 implies that G contains a clique of size k_0 with probability nearly 1. Together, with very high probability $\omega(G) = k_0$. For some n, one of the values $f(k_0), f(k_0+1)$ may be of moderate size so this argument does not apply. Still one may show a strong concentration of the result found independently by Bollobás and Erdős (1976) and Matula (1976).

Corollary 4.5.2 *There exists $k = k(n)$ so that*

$$\Pr[\omega(G) = k \text{ or } k+1] \to 1.$$

We give yet stronger results on the distribution of $\omega(G)$ in Section 10.2.

4.6 DISTINCT SUMS

A set $x_1, \dots, x_k$ of positive integers is said to have distinct sums if all sums

$$\sum_{i \in S} x_i, S \subseteq \{1, \dots, k\}$$

are distinct. Let $f(n)$ denote the maximal k for which there exists a set

$$\{x_1, \dots, x_k\} \subset \{1, \dots, n\}$$

with distinct sums. The simplest example of a set with distinct sums is $\{2^i : i \leq \log_2 n\}$. This example shows

$$f(n) \geq 1 + \lfloor \log_2 n \rfloor .$$

Erdős offered \$300 for a proof or disproof that

$$f(n) \leq \log_2 n + C$$

for some constant C. From above, as all $2^{f(n)}$ sums are distinct and less than nk,

$$2^{f(n)} < nk = nf(n),$$

and so

$$f(n) < \log_2 n + \log_2 \log_2 n + O(1).$$

Examination of the second moment gives a modest improvement. Fix $\{x_1, \dots, x_k\} \subset \{1, \dots, n\}$ with distinct sums. Let $\epsilon_1, \dots, \epsilon_k$ be independent with

$$\Pr[\epsilon_i = 1] = \Pr[\epsilon_i = 0] = 1/2$$

and set

$$X = \epsilon_1 x_1 + \cdots + \epsilon_k x_k .$$

(We may think of X as a random sum.) Set

$$\mu = \mathrm{E}[X] = \frac{x_1 + \cdots + x_k}{2}$$

and $\sigma^2 = \mathrm{Var}[X]$. We bound

$$\sigma^2 = \frac{x_1^2 + \cdots + x_k^2}{4} \leq \frac{n^2 k}{4}$$

so that $\sigma \leq n\sqrt{k}/2$. By Chebyshev's Inequality, for any $\lambda > 1$,

$$\Pr\left[|X - \mu| \geq \lambda n\sqrt{k}/2\right] \leq \lambda^{-2}.$$

Reversing,

$$1 - \frac{1}{\lambda^2} \leq \Pr\left[|X - \mu| < \lambda n\sqrt{k}/2\right].$$

But X has any particular value with probability either zero or 2^{-k} since, critically, a sum can be achieved in at most one way. Thus

$$\Pr\left[|X - \mu| < \lambda n\sqrt{k}/2\right] \leq 2^{-k}(\lambda n\sqrt{k} + 1)$$

and

$$n \geq \frac{2^k(1 - \lambda^{-2}) - 1}{\sqrt{k}\lambda}.$$

While $\lambda = \sqrt{3}$ gives optimal results, any choice of $\lambda > 1$ gives the following:

Theorem 4.6.1 $f(n) \leq \log_2 n + (1/2)\log_2 \log_2 n + O(1)$.

4.7 THE RÖDL NIBBLE

For $2 \leq l < k < n$, let $M(n, k, l)$, the covering number, denote the minimal size of a family $\mathcal{K}$ of k-element subsets of $\{1, \ldots, n\}$ having the property that every l-element set is contained in at least one $A \in \mathcal{K}$. Clearly, $M(n, k, l) \geq \binom{n}{l} / \binom{k}{l}$ since each k-set covers $\binom{k}{l}$ l-sets and every l-set must be covered. Equality holds if and only if the family $\mathcal{K}$ has the property that every l-set is contained in exactly one $A \in \mathcal{K}$. This is called an (n, k, l) tactical configuration (or block design). For example, $(n, 3, 2)$ tactical configurations are better known as Steiner Triple Systems. The question of the existence of tactical configurations, raised by Steiner in 1853, has been a central one for combinatorics. In the previous editions of this book, we have written that this question is one for which probabilistic methods (at least so far!) played little role. Very recently, Keevash (2014) settled this question using a sophisticated combination of probabilistic and algebraic arguments. He proved that whenever the necessary divisibility conditions hold, and n is sufficiently large as a function of k and l, then an (n, k, l) block design exists. The full argument is complicated, and we will not give it here. Instead, we present a simpler result, proving the existence of asymptotic near designs.

In 1963, Paul Erdős and Haim Hanani conjectured that for fixed $2 \leq l < k$,

$$\lim_{n\to\infty} \frac{M(n, k, l)}{\binom{n}{l} / \binom{k}{l}} = 1.$$

Their conjecture was, roughly, that one can get asymptotically close to a tactical configuration. While this conjecture seemed ideal for a probabilistic analysis, it was a full generation before Rödl (1985) found the proof, which we describe in this section. [One may similarly define the packing number $m(n,k,l)$ as the maximal size of a family $\mathcal{K}$ of k-element subsets of $\{1,\dots,n\}$ having the property that every l-element set is contained in at most one $A \in \mathcal{K}$. Erdős and Hanani noticed from elementary arguments that

$$\lim_{n\to\infty} \frac{M(n,k,l)}{\binom{n}{l}/\binom{k}{l}} = 1 \quad \Longleftrightarrow \quad \lim_{n\to\infty} \frac{m(n,k,l)}{\binom{n}{l}/\binom{k}{l}} = 1\,.$$

While the Rödl result may be formulated in terms of either packing or covering, here we deal only with the covering problem.]

Several researchers realized that the Rödl method applies in a much more general setting, dealing with covers in uniform hypergraphs. This was first observed by Frankl and Rödl and has been simplified and extended by Pippenger and Spencer (1989) as well as by Kahn (1996). Our treatment here follows the one in Pippenger and Spencer (1989) and is based on the description of Füredi (1988), where the main tool is the second moment method.

For an r-uniform hypergraph $H = (V,E)$ and for a vertex $x \in V$, we let $d_H(x)$ [or simply $d(x)$, when there is no danger of confusion] denote the degree of x in H, that is, the number of edges containing x. Similarly, for $x, y \in V$, $d(x,y) = d_H(x,y)$ is the number of edges of H containing both x and y. A covering of H is a set of edges whose union contains all vertices. In what follows, whenever we write $\pm\delta$ we mean a quantity between $-\delta$ and δ. The following theorem is due to Pippenger, following Frankl and Rödl.

Theorem 4.7.1 *For every integer $r \geq 2$ and reals $k \geq 1$ and $a > 0$, there are $\gamma = \gamma(r,k,a) > 0$ and $d_0 = d_0(r,k,a)$ such that for every n and $D \geq d_0$ the following holds:*

Every r-uniform hypergraph $H = (V,E)$ on a set V of n vertices in which all vertices have positive degrees and which satisfies the following conditions:

(1) For all vertices $x \in V$ but at most γn of them, $d(x) = (1 \pm \gamma)D$;

(2) For all $x \in V$, $d(x) < kD$;

(3) For any two distinct $x, y \in V$, $d(x,y) < \gamma D$;

contains a cover of at most $(1+a)(n/r)$ edges.

The basic idea in the proof is simple. Fixing a small $\epsilon > 0$, one shows that a random set of roughly $\epsilon n/r$ edges has, with high probability, only some $O(\epsilon^2 n)$ vertices covered more than once, and hence covers at least $\epsilon n - O(\epsilon^2 n)$ vertices. Moreover, after deleting the vertices covered, the induced hypergraph on the remaining vertices still satisfies the properties described in (1), (2), and (3) above (for some other values of n, γ, k, and D). Therefore, one can choose again a random set of edges of this

hypergraph, covering roughly an ϵ-fraction of its vertices with nearly no overlaps. Proceeding in this way for a large number of times, we are finally left with at most ϵn uncovered vertices, and we then cover them trivially by taking for each of them an arbitrarily chosen edge containing it. Since ϵ is sufficiently small, although this last step is very inefficient, it can be tolerated.

The technical details require a careful application of the second moment method, used several times in the proof of the following lemma:

Lemma 4.7.2 *For every integer $r \geq 2$ and reals $K \geq 1$ and $\epsilon > 0$, and for every real $\delta' > 0$, there are $\delta = \delta(r, K, \epsilon, \delta') > 0$ and $D_0 = D_0(r, K, \epsilon, \delta')$ such that for every n and $D \geq D_0$ the following holds:*

Every r-uniform hypergraph $H = (V, E)$ on a set V of n vertices which satisfies the following conditions:

(i) For all vertices $x \in V$ but at most δn of them, $d(x) = (1 \pm \delta)D$;
(ii) For all $x \in V$, $d(x) < KD$;
(iii) For any two distinct $x, y \in V$, $d(x, y) < \delta D$;

contains a set E' of edges with the following properties:

(iv) $|E'| = (\epsilon n/r)(1 \pm \delta')$.
(v) The set $V' = V - \cup_{e \in E'} e$ is of cardinality $|V'| = ne^{-\epsilon}(1 \pm \delta')$.
(vi) For all vertices $x \in V'$ but at most $\delta'|V'|$ of them, the degree $d'(x)$ of x in the induced hypergraph of H on V' satisfies $d'(x) = De^{-\epsilon(r-1)}(1 \pm \delta')$.

Proof. Throughout the proof we assume, whenever this is needed, that D (and hence n) is sufficiently large. We denote by $\delta_1, \delta_2, \ldots$ positive constants (that can be explicitly estimated) that tend to 0 when δ tends to 0 and D tends to infinity (for fixed r, K, ϵ). Therefore, by choosing δ and D_0 appropriately, we can ensure that each of those will be smaller than δ'.

Let E' be a random subset of E obtained by picking, randomly and independently, each edge in E to be a member of E' with probability $p = \epsilon/D$. We have to show that with positive probability, the properties *(iv)*, *(v)*, and *(vi)* hold.

The proof that *(iv)* holds is easy. Note that, by the assumptions, H has at least $(1 - \delta)n$ vertices of degree at least $(1 - \delta)D$, showing that its number of edges is at least $(1 - \delta)^2 nD/r$. Similarly, the number of edges of H does not exceed $[(1 + \delta)Dn + \delta nKD]/r$. Therefore $|E| = (1 \pm \delta_1)Dn/r$. It follows that the expected value of the size of E' satisfies $\mathrm{E}[|E'|] = |E|p = (1 \pm \delta_1)(\epsilon n/r)$ and its variance is $\mathrm{Var}[|E'|] = |E|p(1 - p) \leq (1 \pm \delta_1)(\epsilon n/r)$. Therefore, by Chebyshev's inequality, for an appropriately chosen $\delta_2 > 0$,

$$\Pr\left[|E'| = (1 \pm \delta_2)\frac{\epsilon n}{r}\right] > 0.99,$$

say, giving *(iv)*.

To prove *(v)*, define for each vertex $x \in V$ an indicator random variable I_x, where $I_x = 1$ if $x \notin \cup_{e \in E'} e$ and $I_x = 0$ otherwise. Note that $|V'| = \sum_{x \in V} I_x$. Call a vertex $x \in V$ *good* if $d(x) = (1 \pm \delta)D$; otherwise call it *bad*. If x is good, then

$$\mathrm{E}[I_x] = \Pr[I_x = 1] = (1-p)^{d(x)} = \left(1 - \frac{\epsilon}{D}\right)^{(1\pm\delta)D} = e^{-\epsilon}(1 \pm \delta_3).$$

If x is bad, then, clearly, $0 \le \mathrm{E}[I_x] \le 1$. Since there are at most δn bad vertices, it follows, by Linearity of Expectation, that the expected value of $|V'|$ is $ne^{-\epsilon}(1 \pm \delta_4)$.

To compute the variance of $|V'| = \sum_{x \in V} I_x$, note that

$$\begin{aligned}\mathrm{Var}[|V'|] &= \sum_{x \in V} \mathrm{Var}[I_x] + \sum_{x,y \in V, x \neq y} \mathrm{Cov}[I_x, I_y] \\ &\le \mathrm{E}[|V'|] + \sum_{x,y \in V, x \neq y} \mathrm{Cov}[I_x, I_y].\end{aligned}$$

However,

$$\begin{aligned}\mathrm{Cov}[I_x, I_y] &= \mathrm{E}[I_x I_y] - \mathrm{E}[I_x]\mathrm{E}[I_y] \\ &= (1-p)^{d(x)+d(y)-d(x,y)} - (1-p)^{d(x)+d(y)} \\ &\le (1-p)^{-d(x,y)} - 1 \le \left(1 - \frac{\epsilon}{D}\right)^{-\delta D} - 1 \le \delta_5.\end{aligned}$$

It follows that

$$\mathrm{Var}[|V'|] \le \mathrm{E}[|V'|] + \delta_5 n^2 \le \delta_6 (\mathrm{E}[|V'|])^2,$$

which, by Chebyshev, implies that with probability at least 0.99

$$|V'| = (1 \pm \delta_7)\mathrm{E}[|V'|] = (1 \pm \delta_8) ne^{-\epsilon},$$

as claimed in *(v)*.

It remains to prove *(vi)*. To do so, note, first, that all but at most $\delta_9 n$ vertices x satisfy the following two conditions:

(A) $d(x) = (1 \pm \delta)D$, and

(B) all but at most $\delta_{10} D$ edges $e \in E$ with $x \in e$ satisfy

$$|\{f \in E \,:\, x \notin f, f \cap e \neq \emptyset\}| = (1 \pm \delta_{11})(r-1)D. \tag{4.1}$$

Indeed, (A) holds for all but $\delta n < \delta_9 n/2$ vertices, by assumption. Moreover, the total number of edges containing vertices whose degrees are not $(1 \pm \delta)D$ is at most δnKD and hence the number of vertices contained in more than $\delta_{10} D$ such edges is at most $\delta nKDr/(\delta_{10} D) \le \delta_9 n/2$ for an appropriate choice of δ_9, δ_{10}. Note, next,

that if $x \in e$ and e contains no vertex of degree which is not $(1 \pm \delta)D$, then, since $d(y,z) < \delta D$ for all y, z, the number of edges f not containing x that intersect e is at most $(r-1)(1 \pm \delta)D$ and at least $(r-1)(1 \pm \delta)D - \binom{r-1}{2} \delta D$, and hence e satisfies (4.1).

It thus suffices to show that, for most of the vertices x satisfying (A) and (B), $d'(x)$ satisfies (vi). Fix such a vertex x. Call an edge e with $x \in e$ *good* if it satisfies (4.1). Conditioning on $x \in V'$, the probability that a good edge containing x stays in the hypergraph on V' is $(1-p)^{(1 \pm \delta_{11})(r-1)D}$. Therefore the expected value of $d'(x)$ is

$$\mathrm{E}[d'(x)] = (1 \pm \delta_{10} \pm \delta)D(1-p)^{(1 \pm \delta_{11})(r-1)D} \pm \delta_{10} D = e^{-\epsilon(r-1)} D(1 \pm \delta_{12}).$$

For each edge e containing x, let I_e denote the indicator random variable whose value is 1 iff e is contained in V'. Then, the degree $d'(x)$ is simply the sum of these indicator random variables, conditioned on $x \in V'$. It follows that

$$\begin{aligned} \mathrm{Var}[d'(x)] &\leq \mathrm{E}[d'(x)] + \sum_{x \in e, x \in f} \mathrm{Cov}[I_e, I_f] \\ &\leq \mathrm{E}[d'(x)] + 2\delta_{10} D^2 (1 \pm \delta) + \sum_{x \in e, x \in f, e, f \text{ good}} \mathrm{Cov}[I_e, I_f]. \end{aligned}$$

It remains to bound the sum $\sum_{x \in e, x \in f, e, f \text{ good}} \mathrm{Cov}[I_e, I_f]$. For each fixed good e, this sum is a sum of the form $\sum_{x \in f, f \text{ good}} \mathrm{Cov}[I_e, I_f]$. There are at most $(r-1)\delta D$ edges f in the last sum for which $|e \cap f| > 1$, and their contribution to the sum cannot exceed $(r-1)\delta D$. If $e \cap f = \{x\}$, then let $t(e,f)$ denote the number of edges of H that intersect both e and f and do not contain x. Clearly, in this case, $t(e,f) \leq (r-1)^2 \delta D$. It follows that for such e and f, $\mathrm{Cov}[I_e, I_f] \leq (1-p)^{-t(e,f)} - 1 \leq \delta_{13}$, implying that for each fixed good edge e,

$$\sum_{x \in f, f \text{ good}} \mathrm{Cov}[I_e, I_f] \leq (r-1)\delta D + D(1+\delta)\delta_{13} \leq \delta_{14} D.$$

As the sum $\sum_{x \in e, x \in f, e, f \text{ good}} \mathrm{Cov}[I_e, I_f]$ is the sum of at most $D(1+\delta)$ such quantities, we conclude that

$$\mathrm{Var}[d'(x)] \leq \mathrm{E}[d'(x)] + \delta_{15} D^2 \leq \delta_{16} (\mathrm{E}[d'(x)])^2.$$

It thus follows, by Chebyshev, that with probability at most δ_{17}, $d'(x)$ is not $(1 \pm \delta_{18}) D e^{-\epsilon(r-1)}$, and therefore, by Markov, that with probability at least, say, 0.99, for all but at most $\delta_{19} n$ vertices, $d'(x) = (1 \pm \delta_{18}) D e^{-\epsilon(r-1)}$. This completes the proof of the lemma. ■

Proof. [Theorem 4.7.1] Fix $\epsilon > 0$ such that

$$\frac{\epsilon}{1 - e^{-\epsilon}} + r\epsilon < 1 + a,$$

and fix $1/10 > \delta > 0$ such that

$$(1+4\delta)\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon < 1 + a.$$

Fix an integer t so that $e^{-\epsilon t} < \epsilon$. The theorem is proved by applying the lemma t times. Put $\delta = \delta_t$ and then define, by reverse induction, $\delta_t > \delta_{t-1} > \cdots > \delta_0$ such that $\delta_i \le \delta_{i+1} e^{-\epsilon(r-1)}$, $\prod_{i=0}^{t}(1+\delta_i) < 1 + 2\delta$, and for $n \ge D \ge R_i$ one can apply the lemma with r, $K = ke^{\epsilon i(r-1)}$, ϵ, $\delta' = \delta_{i+1}$ and $\delta = \delta_i$. This will give the assertion of the theorem with $\gamma = \delta_0$, $d_0 = \max R_i$. Indeed, by applying the lemma repeatedly we obtain a decreasing sequence of sets of vertices $V = V_0, V_1, \ldots, V_t$, each contained in the previous one, and a sequence of sets of edges $E_1, E_2, \ldots, E_t$, where E_i is the set of edges E' obtained in the application of the lemma to the hypergraph induced on V_{i-1}. Here

$$|V_i| = |V_{i-1}|e^{-\epsilon}(1 \pm \delta_i) \qquad (= |V_0|e^{-i\epsilon}(1 \pm 2\delta)),$$
$$|E_i| = \frac{\epsilon|V_{i-1}|}{r}(1 \pm \delta_i) \le (1+4\delta)\frac{\epsilon n}{r}e^{-(i-1)\epsilon},$$

and

$$D_i = D_{i-1}e^{-\epsilon(r-1)} = De^{-\epsilon i(r-1)}.$$

By covering each vertex of V_t separately by an edge containing it, we conclude that the total number of edges in the cover obtained is at most

$$\begin{aligned}(1+4\delta)\sum_{i=0}^{t-1}\frac{\epsilon n}{r}e^{-i\epsilon} + |V_t| &\le (1+4\delta)\frac{\epsilon n}{r}\frac{1}{1-e^{-\epsilon}} + (1+2\delta)ne^{-\epsilon t}\\ &\le \frac{n}{r}(1+4\delta)\left(\frac{\epsilon}{1-e^{-\epsilon}} + r\epsilon\right)\\ &< (1+a)\frac{n}{r}.\end{aligned}$$

This completes the proof. ∎

We conclude the section by showing how the theorem quickly implies Rödl solution of the Erdős–Hanani problem mentioned at the beginning of the section.

Theorem 4.7.3 [Rödl] *For k, l fixed,*

$$M(n,k,l) \le (1+o(1))\binom{n}{l}\Big/\binom{k}{l}$$

where the $o(1)$ term tends to zero as n tends to infinity.

Proof. Put $r = \binom{k}{l}$, and let H be the r-uniform hypergraph whose vertices are all l-subsets of $\{1, 2, \ldots, n\}$, and whose edges are all collections of $\binom{k}{l}$ l-tuples that lie

in a k-set. H has $\binom{n}{l}$ vertices, each of its vertices has degree $D = \binom{n-l}{k-l}$, and every two distinct vertices lie in at most $\binom{n-l-1}{k-l-1} = o(D)$ common edges. Therefore, by Theorem 4.7.1, H has a cover of size at most $(1 + o(1))\binom{n}{l}\binom{k}{l}$, as needed. ■

4.8 EXERCISES

1. Let X be a random variable taking integral nonnegative values, let $\mathrm{E}[X^2]$ denote the expectation of its square, and let $\mathrm{Var}[X]$ denote its variance. Prove that
$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X^2]}.$$

2. (*) Show that there is a positive constant c such that the following holds: For any n reals $a_1, a_2, \ldots, a_n$ satisfying $\sum_{i=1}^{n} a_i^2 = 1$, if $(\epsilon_1, \ldots, \epsilon_n)$ is a $\{-1, 1\}$-random vector obtained by choosing each ϵ_i randomly and independently with uniform distribution to be either -1 or 1, then
$$\Pr\left[\left|\sum_{i=1}^{n} \epsilon_i a_i\right| \leq 1\right] \geq c.$$

3. (*) Show that there is a positive constant c such that the following holds: For any n vectors $a_1, a_2, \ldots, a_n \in R^2$ satisfying $\sum_{i=1}^{n} \|a_i\|^2 = 1$ and $\|a_i\| \leq 1/10$, where $\|\cdot\|$ denotes the usual Euclidean norm, if $(\epsilon_1, \ldots, \epsilon_n)$ is a $\{-1, 1\}$-random vector obtained by choosing each ϵ_i randomly and independently with uniform distribution
$$\Pr\left[\left\|\sum_{i=1}^{n} \epsilon_i a_i\right\| \leq 1/3\right] \geq c.$$

4. Let X be a random variable with expectation $\mathrm{E}[X] = 0$ and variance σ^2. Prove that for all $\lambda > 0$,
$$\Pr[X \geq \lambda] \leq \frac{\sigma^2}{\sigma^2 + \lambda^2}.$$

5. Let $v_1 = (x_1, y_1), \ldots, v_n = (x_n, y_n)$ be n two-dimensional vectors, where each x_i and each y_i is an integer whose absolute value does not exceed $2^{n/2}/(100\sqrt{n})$. Show that there are two disjoint sets $I, J \subset \{1, 2, \ldots, n\}$ such that
$$\sum_{i \in I} v_i = \sum_{j \in J} v_j.$$

6. (*) Prove that for every set X of at least $4k^2$ distinct residue classes modulo a prime p, there is an integer a such that the set $\{ax \pmod p : x \in X\}$ intersects every interval in $\{0, 1, \ldots, p-1\}$ of length at least p/k.

THE PROBABILISTIC LENS:
Hamiltonian Paths

What is the maximum possible number of directed Hamiltonian paths in a tournament on n vertices? Denote this number by $P(n)$. The first application of the probabilistic method in combinatorics is the result of Szele (1943) described in Chapter 2, which states that $P(n) \geq n!/2^{n-1}$. This bound follows immediately from the observation that the right-hand side is the expected number of such paths in a random tournament on n vertices. In the same paper, Szele shows that

$$\frac{1}{2} \leq \lim_{n\to\infty} \left(\frac{P(n)}{n!}\right)^{1/n} \leq \frac{1}{2^{3/4}},$$

proves that this limit does exist, and conjectures that its correct value is $1/2$.

This conjecture is proved in Alon (1990a). The proof is given below. The main tool is the Brégman proof of the Minc Conjecture for the permanent of a (0, 1)-matrix, described in "The Probabilistic Lens: Brégman Theorem" (following Chapter 2).

Theorem 1 *There exists a positive constant c such that, for every n,*

$$P(n) \leq cn^{3/2}\frac{n!}{2^{n-1}}.$$

Proof. For a tournament T, denote by $P(T)$ the number of directed Hamiltonian paths of T. Similarly, $C(T)$ denotes the number of directed Hamiltonian cycles of T, and $F(T)$ denotes the number of spanning subgraphs of T in which the indegree and the outdegree of every vertex is exactly 1. Clearly,

$$C(T) \leq F(T). \tag{1}$$

If $T = (V, E)$ is a tournament on a set $V = \{1, 2, \ldots, n\}$ of n vertices, the *adjacency matrix* of T is the n by n $(0, 1)$-matrix $A_T = (a_{ij})$ defined by $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise. Let r_i denote the number of ones in row i. Clearly,

$$\sum_{i=1}^{n} r_i = \binom{n}{2}. \tag{2}$$

By interpreting combinatorially the terms in the expansion of the permanent $\mathrm{per}(A_T)$, it follows that

$$\mathrm{per}(A_T) = F(T). \tag{3}$$

We need the following technical lemma.

Lemma 2 *For every two integers a, b satisfying $b \geq a + 2 > a \geq 1$, the inequality*

$$(a!)^{1/a} \cdot (b!)^{1/b} < ((a+1)!)^{1/(a+1)} \cdot ((b-1)!)^{1/(b-1)}$$

holds.

Proof. The assertion is simply that $f(a) < f(b-1)$, for the function f defined by $f(a) = (a!)^{1/a}/((a+1)!)^{1/(a+1)}$. Thus it suffices to show that for every integer $x \geq 2$, $f(x-1) < f(x)$. Substituting the expression for f and raising both sides to the power $x(x-1)(x+1)$, it follows that it suffices to show that for all $x \geq 2$,

$$((x-1)!)^{x(x+1)} \cdot ((x+1)!)^{x(x-1)} < (x!)^{2(x^2-1)},$$

that is,

$$\left(\frac{x^x}{x!}\right)^2 > \left(\frac{x+1}{x}\right)^{x(x-1)}.$$

This is certainly true for $x = 2$. For $x \geq 3$, it follows from the facts that $4^x > e^{x+1}$, that $x! < ((x+1)/2)^x$ and that $e^{x-1} > ((x+1)/x)^{x(x-1)}$. ■

Corollary 3 *Define $g(x) = (x!)^{1/x}$. For every integer $S \geq n$, the maximum of the function $\prod_{i=1}^{n} g(x_i)$, subject to the constraints $\sum_{i=1}^{n} x_i = S$ and $x_i \geq 1$ are integers, is obtained iff the variables x_i are as equal as possible (i.e., iff each x_i is either $\lfloor S/n \rfloor$ or $\lceil S/n \rceil$.)*

Proof. If there are two indices i and j such that $x_i \geq x_j + 2$, then by Lemma 2 the value of the product would increase once we add 1 to x_j and subtract 1 from x_i. ■

Returning to our tournament T, we observe that the numbers r_i defined above are precisely the outdegrees of the vertices of T. If at least one of these is 0, then clearly $C(T) = F(T) = 0$. Otherwise, by Brégman's Theorem, by Corollary 3 and by (2) and (3), $F(T)$ is at most the value of the function $\prod_{i=1}^{n} (r_i!)^{1/r_i}$, where the integral

variables r_i satisfy (2) and are as equal as possible. By a straightforward (though somewhat tedious) derivation of the asymptotics using Stirling's formula this gives the following:

Proposition 4 *For every tournament T on n vertices,*

$$C(T) \leq F(T) \leq (1+o(1))\frac{\sqrt{\pi}}{\sqrt{2}e} n^{3/2} \frac{(n-1)!}{2^n}.$$

To complete the proof of the theorem, we have to derive a bound for the number of Hamiltonian paths in a tournament from the above result. Given a tournament S on n vertices, let T be the random tournament obtained from S by adding to it a new vertex y and by orienting each edge connecting y with one of the vertices of S, randomly and independently. For every fixed Hamiltonian path in S, the probability that it can be extended to a Hamiltonian cycle in T is precisely $1/4$. Thus the expected number of Hamiltonian cycles in T is $\frac{1}{4}P(S)$, and hence there is a specific T for which $C(T) \geq \frac{1}{4}P(S)$. However, by Proposition 4

$$C(T) \leq (1+o(1))\frac{\sqrt{\pi}}{\sqrt{2}e}(n+1)^{3/2}\frac{n!}{2^{n+1}},$$

and thus

$$P(S) \leq O\left(n^{3/2}\frac{n!}{2^{n-1}}\right),$$

completing the proof of Theorem 1. ■

5

The Local Lemma

It's a thing that non-mathematicians don't realize. Mathematics is actually an esthetic subject almost entirely.
–John Conway

5.1 THE LEMMA

In a typical probabilistic proof of a combinatorial result, one usually has to show that the probability of a certain event is positive. However, many of these proofs actually give more and show that the probability of the event considered is not only positive but is also large. In fact, most probabilistic proofs deal with events that hold with high probability: that is, a probability that tends to 1 as the dimensions of the problem grow. For example, consider the proof given in Chapter 1 that for each $k \geq 1$ there are tournaments in which for every set of k players there is one who beats them all. The proof actually shows that for every fixed k, if the number n of players is sufficiently large, then almost all tournaments with n players satisfy this property; that is, the probability that a random tournament with n players has the desired property tends to 1 as n tends to infinity.

On the other hand, there is a trivial case in which one can show that a certain event holds with positive, though very small, probability. Indeed, if we have n mutually independent events and each of them holds with probability at least $p > 0$, then the probability that all events hold simultaneously is at least p^n, which is positive, although it may be exponentially small in n.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

It is natural to expect that the case of mutual independence can be generalized to that of rare dependencies and provide a more general way of proving that certain events hold with positive, though small, probability. Such a generalization is, indeed, possible and is stated in the following lemma, known as the Lovász Local Lemma. This simple lemma, first proved in Erdős and Lovász (1975), is an extremely powerful tool, as it offers a way for dealing with rare events.

Lemma 5.1.1 [The Local Lemma; General Case] *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. A directed graph $D = (V, E)$ on the set of vertices $V = \{1, 2, \ldots, n\}$ is called a* dependency digraph *for the events $A_1, \ldots, A_n$ if for each i, $1 \leq i \leq n$, the event A_i is mutually independent of all the events $\{A_j : (i,j) \notin E\}$. Suppose that $D = (V, E)$ is a dependency digraph for the above events and suppose there are real numbers $x_1, \ldots, x_n$ such that $0 \leq x_i < 1$ and $\Pr[A_i] \leq x_i \prod_{(i,j)\in E}(1 - x_j)$ for all $1 \leq i \leq n$. Then*

$$\Pr\left[\bigwedge_{i=1}^{n} \overline{A_i}\right] \geq \prod_{i=1}^{n}(1 - x_i).$$

In particular, with positive probability, no event A_i holds.

Proof. We first prove, by induction on s, that for any $S \subset \{1, \ldots, n\}$, $|S| = s < n$, and any $i \notin S$,

$$\Pr\left[A_i | \bigwedge_{j\in S} \overline{A_j}\right] \leq x_i. \tag{5.1}$$

This is certainly true for $s = 0$. Assuming it holds for all $s' < s$, we prove it for s. Put $S_1 = \{j \in S : (i,j) \in E\}$, $S_2 = S \setminus S_1$. Then

$$\Pr\left[A_i | \bigwedge_{j\in S} \overline{A_j}\right] = \frac{\Pr\left[A_i \wedge \left(\bigwedge_{j\in S_1} \overline{A_j}\right) \middle| \bigwedge_{\ell\in S_2} \overline{A_\ell}\right]}{\Pr\left[\bigwedge_{j\in S_1} \overline{A_j} \middle| \bigwedge_{\ell\in S_2} \overline{A_\ell}\right]}. \tag{5.2}$$

To bound the numerator, observe that, since A_i is mutually independent of the events $\{A_\ell : \ell \in S_2\}$,

$$\begin{aligned}\Pr\left[A_i \wedge \left(\bigwedge_{j\in S_1} \overline{A_j}\right) \middle| \bigwedge_{\ell\in S_2} \overline{A_\ell}\right] &\leq \Pr\left[A_i \middle| \bigwedge_{\ell\in S_2} \overline{A_\ell}\right] \\ &= \Pr[A_i] \leq x_i \prod_{(i,j)\in E} (1 - x_j)\,. \end{aligned} \tag{5.3}$$

The denominator, on the other hand, can be bounded by the induction hypothesis. Indeed, suppose $S_1 = \{j_1, j_2, \ldots, j_r\}$. If $r = 0$, then the denominator is 1, and (5.1)

follows. Otherwise

$$\begin{aligned}
&\Pr\left[\overline{A_{j_1}} \wedge \overline{A_{j_2}} \wedge \cdots \wedge \overline{A_{j_r}} \mid \bigwedge_{\ell \in S_2} \overline{A_\ell}\right] \\
&= \left(1 - \Pr\left[A_{j_1} \mid \bigwedge_{\ell \in S_2} \overline{A_\ell}\right]\right) \cdot \left(1 - \Pr\left[A_{j_2} \mid \overline{A_{j_1}} \wedge \bigwedge_{\ell \in S_2} \overline{A_\ell}\right]\right) \cdot \cdots \\
&\quad \cdot \cdots \left(1 - \Pr\left[A_{j_r} \mid \overline{A_{j_1}} \wedge \cdots \wedge \overline{A_{j_{r-1}}} \wedge \bigwedge_{\ell \in S_2} \overline{A_\ell}\right]\right) \\
&\geq (1 - x_{j_1})(1 - x_{j_2}) \cdots (1 - x_{j_r}) \geq \prod_{(i,j) \in E} (1 - x_j) \,. \qquad (5.4)
\end{aligned}$$

Substituting (5.3) and (5.4) into (5.2), we conclude that $\Pr[A_i | \bigwedge_{j \in S} \overline{A_j}] \leq x_i$, completing the proof of the induction.

The assertion of Lemma 5.1.1 now follows easily, as

$$\begin{aligned}
\Pr\left[\bigwedge_{i=1}^{n} \overline{A_i}\right] &= (1 - \Pr[A_1]) \cdot (1 - \Pr[A_2 | \overline{A_1}]) \cdot \cdots \\
&\quad \cdot \cdots \left(1 - \Pr\left[A_n \mid \bigwedge_{i=1}^{n-1} \overline{A_i}\right]\right) \geq \prod_{i=1}^{n} (1 - x_i) \,,
\end{aligned}$$

completing the proof. ∎

Corollary 5.1.2 [The Local Lemma; Symmetric Case] *Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. Suppose that each event A_i is mutually independent of a set of all the other events A_j but at most d, and that $\Pr[A_i] \leq p$ for all $1 \leq i \leq n$. If*

$$ep(d + 1) \leq 1, \qquad (5.5)$$

then $\Pr[\wedge_{i=1}^{n} \overline{A_i}] > 0$.

Proof. If $d = 0$, the result is trivial. Otherwise, by the assumption there is a dependency digraph $D = (V, E)$ for the events $A_1, \ldots, A_n$ in which, for each i, $|\{j : (i,j) \in E\}| \leq d$. The result now follows from Lemma 5.1.1 by taking $x_i = 1/(d + 1) (< 1)$ for all i and using the fact that, for any $d \geq 1$, $(1 - 1/(d + 1))^d > 1/e$. ∎

It is worth noting that as shown by Shearer (1985), the constant "e" is the best possible constant in inequality (5.5). Note also that the proof of Lemma 5.1.1 indicates that the conclusion remains true even when we replace the two assumptions that each A_i is mutually independent of $\{A_j : (i,j) \notin E\}$ and that for each i

$$\Pr[A_i] \leq x_i \prod_{(ij) \in E} (1 - x_j)$$

by the weaker assumption that for each i and each $S_2 \subset \{1, \ldots, n\} \setminus \{j : (i,j) \in E\}$,

$$\Pr\left[A_i | \bigwedge_{j \in S_2} \overline{A_j}\right] \leq x_i \prod_{(i,j) \in E} (1 - x_j).$$

This turns out to be useful in certain applications.

In the next few sections we present various applications of the Local Lemma for obtaining combinatorial results. There is no known proof of any of these results, which does not use the Local Lemma. Additional applications of the Local Lemma for coloring problems, and many more, can be found in Molloy and Reed (2002).

5.2 PROPERTY B AND MULTICOLORED SETS OF REAL NUMBERS

Recall that a hypergraph $H = (V, E)$ has property B (i.e., is two-colorable) if there is a coloring of V by two colors so that no edge $f \in E$ is monochromatic.

Theorem 5.2.1 *Let $H = (V, E)$ be a hypergraph in which every edge has at least k elements, and suppose that each edge of H intersects at most d other edges. If $e(d + 1) \leq 2^{k-1}$, then H has property B.*

Proof. Color each vertex v of H, randomly and independently, either blue or red (with equal probability). For each edge $f \in E$, let A_f be the event that f is monochromatic. Clearly, $\Pr[A_f] = 2/2^{|f|} \leq 1/2^{k-1}$. Moreover, each event A_f is clearly mutually independent of all the other events $A_{f'}$ for all edges f' that do not intersect f. The result now follows from Corollary 5.1.2. ■

A special case of Theorem 5.2.1 is that, for any $k \geq 9$, any k-uniform k-regular hypergraph H has property B. Indeed, since any edge f of such an H contains k vertices, each of which is incident with k edges (including f), it follows that f intersects at most $d = k(k-1)$ other edges. The desired result follows, since $e(k(k-1)+1) < 2^{k-1}$ for each $k \geq 9$.

The next result we consider, which appeared in the original paper of Erdős and Lovász, deals with k-colorings of the real numbers. For a k-coloring $c : \mathbb{R} \to \{1, 2, \ldots, k\}$ of the real numbers by the k colors $1, 2, \ldots, k$, and for a subset $T \subset \mathbb{R}$, we say that T is *multicolored* (with respect to c) if $c(T) = \{1, 2, \ldots, k\}$, that is, if T contains elements of all colors.

Theorem 5.2.2 *Let m and k be two positive integers satisfying*

$$e(m(m-1)+1)k\left(1 - \frac{1}{k}\right)^m \leq 1. \tag{5.6}$$

Then, for any set S of m real numbers, there is a k-coloring so that each translation $x + S$ (for $x \in \mathbb{R}$) is multicolored.

Note that (5.6) holds whenever $m > (3 + o(1))k \log k$.

Proof. We first fix a *finite* subset $X \subseteq \mathbb{R}$ and show the existence of a k-coloring so that each translation $x + S$ (for $x \in X$) is multicolored. This is an easy consequence of the Local Lemma. Indeed, put $Y = \bigcup_{x \in X}(x + S)$ and let $c : Y \to \{1, 2, \ldots, k\}$ be a random k-coloring of Y obtained by choosing, for each $y \in Y$, randomly and independently, $c(y) \in \{1, 2, \ldots, k\}$ according to a uniform distribution on $\{1, 2, \ldots, k\}$. For each $x \in X$, let A_x be the event in which $x + S$ is not multicolored (with respect to c). Clearly, $\Pr[A_x] \leq k(1 - 1/k)^m$. Moreover, each event A_x is mutually independent of all the other events $A_{x'}$ but those for which $(x + S) \cap (x' + S) \neq \emptyset$. As there are at most $m(m - 1)$ such events, the desired result follows from Corollary 5.1.2.

We can now prove the existence of a coloring of the set of all reals with the desired properties, by a standard compactness argument. Since the discrete space with k points is (trivially) compact, Tikhonov's Theorem (which is equivalent to the axiom of choice) implies that an arbitrary product of such spaces is compact. In particular, the space of all functions from $\mathbb{R}$ to $\{1, 2, \ldots, k\}$, with the usual product topology, is compact. In this space, for every fixed $x \in \mathbb{R}$, the set C_x of all colorings c, such that $x + S$ is multicolored, is closed. (In fact, it is both open and closed, since a basis to the open sets is the set of all colorings whose values are prescribed in a finite number of places). As we proved above, the intersection of any finite number of sets C_x is nonempty. It thus follows, by compactness, that the intersection of all sets C_x is nonempty. Any coloring in this intersection has the properties in the conclusion of Theorem 5.2.2. ■

Note that it is impossible, in general, to apply the Local Lemma to an infinite number of events and conclude that in some point of the probability space none of them holds. In fact, there are trivial examples of countably many mutually independent events A_i, satisfying $\Pr[A_i] = 1/2$ and $\bigwedge_{i \geq 1} \overline{A_i} = \emptyset$. Thus the compactness argument is essential in the above proof.

5.3 LOWER BOUNDS FOR RAMSEY NUMBERS

The derivation of lower bounds for Ramsey numbers by Erdős in 1947 was one of the first applications of the probabilistic method. The Local Lemma provides a simple way of improving these bounds. Let us obtain, first, a lower bound for the diagonal Ramsey number $R(k, k)$. Consider a random two-coloring of the edges of K_n. For each set S of k vertices of K_n, let A_S be the event where the complete graph on S is monochromatic. Clearly, $\Pr[A_S] = 2^{1-\binom{k}{2}}$. It is obvious that each event A_S is mutually independent of all the events A_T but those that satisfy $|S \cap T| \geq 2$, since this is the only case in which the corresponding complete graphs share an edge. We can therefore apply Corollary 5.1.2 with $p = 2^{1-\binom{k}{2}}$ and $d < \binom{k}{2}\binom{n-2}{k-2}$ to conclude the following:

Proposition 5.3.1 *If* $e\binom{k}{2}\binom{n-2}{k-2} \cdot 2^{1-\binom{k}{2}} < 1$, *then* $R(k, k) > n$.

A short computation shows that this gives $R(k,k) > (\sqrt{2}/e)(1+o(1))k2^{k/2}$, only a factor 2 improvement on the bound obtained by the straightforward probabilistic method. Although this minor improvement is somewhat disappointing, it is certainly not surprising; the Local Lemma is most powerful when the dependencies between events are rare, and this is not the case here. Indeed, there is a total number of $K = \binom{n}{k}$ events considered, and the maximum outdegree d in the dependency digraph is roughly $\binom{k}{2}\binom{n}{k-2}$. For large k and much larger n (which is the case of interest for us), we have $d > K^{1-O(1/k)}$, that is, quite a lot of dependencies. On the other hand, if we consider small sets S (e.g., sets of size 3), we observe that out of the total $K = \binom{n}{3}$ of them, each shares an edge with only $3(n-3) \approx K^{1/3}$. This suggests that the Local Lemma may be much more significant in improving the off-diagonal Ramsey numbers $R(k,\ell)$, especially if one of the parameters, say, ℓ, is small. Let us consider, for example, following Spencer (1977), the Ramsey number $R(k,3)$. Here, of course, we have to apply the nonsymmetric form of the Local Lemma. Let us two-color the edges of K_n randomly and independently, where each edge is colored blue with probability p. For each set of three vertices T, let A_T be the event where the triangle on T is blue. Similarly, for each set of k vertices S, let B_S be the event where the complete graph on S is red. Clearly, $\Pr[A_T] = p^3$ and $\Pr[B_S] = (1-p)^{\binom{k}{2}}$. Construct a dependency digraph for the events A_T and B_S by joining two vertices by edges (in both directions) if the corresponding complete graphs share an edge. Clearly, each A_T-node of the dependency graph is adjacent to $3(n-3) < 3n$ $A_{T'}$-nodes and to at most $\binom{n}{k}$ $B_{S'}$-nodes. Similarly, each B_S-node is adjacent to at most $\binom{k}{2}(n-2) < k^2 n/2$ $A_{T'}$-nodes and to at most $\binom{n}{k}$ $B_{S'}$-nodes. It follows from the general case of the Local Lemma (Lemma 5.1.1) that, if we can find a $0 < p < 1$ and two real numbers $0 \le x < 1$ and $0 \le y < 1$ such that

$$p^3 \le x(1-x)^{3n}(1-y)^{\binom{n}{k}}$$

and

$$(1-p)^{\binom{k}{2}} \le y(1-x)^{k^2 n/2}(1-y)^{\binom{n}{k}},$$

then $R(k,3) > n$.

Our objective is to find the smallest possible $k = k(n)$ for which there is such a choice of p, x, and y. An elementary (but tedious) computation shows that the best choice is when $p = c_1 n^{-1/2}$, $k = c_2 n^{1/2}\log n$, $x = c_3/n^{3/2}$, and y so that $\binom{n}{k} y = c_4$. This gives $R(k,3) > c_5 k^2/\log^2 k$. A similar argument gives $R(k,4) > k^{5/2+o(1)}$. In both cases, the amount of computation required is considerable. However, the hard work does pay; the bound $R(k,3) > c_5 k^2/\log^2 k$ matches a lower bound of Erdős proved in 1961 by a highly complicated probabilistic argument. This was improved to $R(k,3) > c_6 k^2/\log k$ by Kim (1995). The bound above for $R(k,4)$ is better than any bound for $R(k,4)$ known to be proved without the Local Lemma.

5.4 A GEOMETRIC RESULT

A family of open unit balls F in the three-dimensional Euclidean space $\mathbb{R}^3$ is called a k-fold covering of $\mathbb{R}^3$ if any point $x \in \mathbb{R}^3$ belongs to at least k balls. In particular, a onefold covering is simply called a *covering*. A k-fold covering $\mathcal{F}$ is called *decomposable* if there is a partition of $\mathcal{F}$ into two pairwise disjoint families $\mathcal{F}_1$ and $\mathcal{F}_2$, each being a covering of $\mathbb{R}^3$. Mani-Levitska and Pach (1988) constructed, for any integer $k \geq 1$, a nondecomposable k-fold covering of $\mathbb{R}^3$ by open unit balls. On the other hand, they proved that any k-fold covering of $\mathbb{R}^3$ in which no point is covered by more than $c2^{k/3}$ balls is decomposable. This reveals a somewhat surprising phenomenon: it is more difficult to decompose coverings that cover some of the points of $\mathbb{R}^3$ too often than to decompose coverings that cover every point about the same number of times. The exact statement of the Mani-Levitska–Pach Theorem is the Following:

Theorem 5.4.1 *Let $\mathcal{F} = \{B_i\}_{i\in I}$ be a k-fold covering of the three-dimensional Euclidean space by open unit balls. Suppose, further, that no point of $\mathbb{R}^3$ is contained in more than t members of $\mathcal{F}$. If*

$$e \cdot t^3 2^{18}/2^{k-1} \leq 1,$$

then $\mathcal{F}$ is decomposable.

Proof. Define an infinite hypergraph $H = (V(H), E(H))$ as follows: The set of vertices of H, $V(H)$, is simply $\mathcal{F} = \{B_i\}_{i\in I}$. For each $x \in \mathbb{R}^3$, let E_x be the set of balls $B_i \in \mathcal{F}$ that contain x. The set of edges of H, $E(H)$, is simply the set of E_x, with the understanding that when $E_x = E_y$ the edge is taken only once. We claim each edge E_x intersects less than $t^3 2^{18}$ other edges E_y of H. If $x \in B_i$, the center of B_i is within distance 1 of x. If now $B_j \cap B_i \neq \emptyset$ the center of B_j is within distance three of x and so B_j lies entirely inside the ball of radius four centered at x. Such a B_j covers precisely $4^{-3} = 2^{-6}$ of the volume of that ball. As no vertex is covered more than t times, there can be at most $2^6 t$ such balls. It is not too difficult to check that m balls in $\mathbb{R}^3$ cut $\mathbb{R}^3$ into less than m^3 connected components so that there are at most $(2^6 t)^3$ distinct E_y overlapping E_x.

Consider, now, any finite subhypergraph L of H. Each edge of L has at least k vertices, and it intersects at most $d < t^3 2^{18}$ other edges of L. Since, by assumption, $e(d + 1) \leq 2^{k-1}$, Theorem 5.2.1 (which is a simple corollary of the Local Lemma) implies that L is two-colorable. This means that one can color the vertices of L blue and red so that no edge of L is monochromatic. Since this holds for any finite L, a compactness argument, analogous to the one used in the proof of Theorem 5.2.2, shows that H is two-colorable. Given a two-coloring of H with no monochromatic edges, we simply let $\mathcal{F}_1$ be the set of all blue balls, and $\mathcal{F}_2$ be the set of all red ones. Clearly, each $\mathcal{F}_i$ is a covering of $\mathbb{R}^3$, completing the proof of the theorem. ■

It is worth noting that Theorem 5.4.1 can easily be generalized to higher dimensions. We omit the detailed statement of this generalization.

5.5 THE LINEAR ARBORICITY OF GRAPHS

A *linear forest* is a forest (i.e., an acyclic simple graph) in which every connected component is a path. The *linear arboricity* $\mathrm{la}(G)$ of a graph G is the minimum number of linear forests in G, whose union is the set of all edges of G. This notion was introduced by Harary as one of the covering invariants of graphs. The following conjecture, known as the *Linear Arboricity Conjecture*, was raised in Akiyama, Exoo and Harary (1981).

Conjecture 5.5.1 [The Linear Arboricity Conjecture] *The linear arboricity of every d-regular graph is $\lceil (d+1)/2 \rceil$.*

Note that since every d-regular graph G on n vertices has $nd/2$ edges, and every linear forest in it has at most $n-1$ edges, the inequality

$$\mathrm{la}(G) \geq \frac{nd}{2(n-1)} > \frac{d}{2}$$

is immediate. Since $\mathrm{la}(G)$ is an integer, this gives $\mathrm{la}(G) \geq \lceil (d+1)/2 \rceil$. The difficulty in Conjecture 5.5.1 lies in proving the converse inequality: $\mathrm{la}(G) \leq \lceil (d+1)/2 \rceil$. Note also that, since every graph G with maximum degree Δ is a subgraph of a Δ-regular graph (which may have more vertices, as well as more edges than G), the Linear Arboricity Conjecture is equivalent to the statement that the linear arboricity of every graph G with maximum degree Δ is at most $\lceil (\Delta+1)/2 \rceil$.

Although this conjecture received a considerable amount of attention, the best general result concerning it, proved without any probabilistic arguments, is that $\mathrm{la}(G) \leq \lceil 3\Delta/5 \rceil$ for even Δ and that $\mathrm{la}(G) \leq \lceil (3\Delta+2)/5 \rceil$ for odd Δ. In this section we prove that, for every $\epsilon > 0$, there is a $\Delta_0 = \Delta_0(\epsilon)$ such that, for every $\Delta \geq \Delta_0$, the linear arboricity of every graph with maximum degree Δ is less than $\left(\frac{1}{2} + \epsilon\right)\Delta$. This result (with a somewhat more complicated proof) appears in Alon (1988) and its proof relies heavily on the Local Lemma. We note that this proof is more complicated than the other proofs given in this chapter and requires certain preparations, some of which are of independent interest.

It is convenient to deduce the result for undirected graphs from its directed version. A *d-regular digraph* is a directed graph in which the indegree and the outdegree of every vertex is precisely d. A linear directed forest is a directed graph in which every connected component is a directed path. The *dilinear arboricity* $\mathrm{dla}(G)$ of a directed graph G is the minimum number of linear directed forests in G whose union covers all edges of G. The directed version of the Linear Arboricity Conjecture, first stated in Nakayama and Peroche (1987), is the following:

Conjecture 5.5.2 *For every d-regular digraph D*

$$\mathrm{dla}(D) = d + 1.$$

Note that since the edges of any (connected) undirected $2d$-regular graph G can be oriented along a Euler cycle, so that the resulting oriented digraph is d-regular, the validity of Conjecture 5.5.2 for d implies that of Conjecture 5.5.1 for $2d$.

It is easy to prove that any graph with n vertices and maximum degree d contains an independent set of size at least $n/(d+1)$. The following proposition shows that, at the price of decreasing the size of such a set by a constant factor, we can guarantee that it has a certain structure.

Proposition 5.5.3 *Let $H = (V, E)$ be a graph with maximum degree d, and let $V = V_1 \cup V_2 \cup \cdots \cup V_r$ be a partition of V into r pairwise disjoint sets. Suppose each set V_i is of cardinality $|V_i| \geq 2ed$, where e is the basis of the natural logarithm. Then there is an independent set of vertices $W \subseteq V$ that contains a vertex from each V_i.*

Proof. Clearly, we may assume that each set V_i is of cardinality precisely $g = \lceil 2ed \rceil$ (otherwise, simply replace each V_i by a subset of cardinality g of it, and replace H by its induced subgraph on the union of these r new sets). Let us pick from each set V_i randomly and independently a single vertex according to a uniform distribution. Let W be the random set of the vertices picked. To complete the proof, we show that with positive probability W is an independent set of vertices in H.

For each edge f of H, let A_f be the event where W contains both ends of f. Clearly, $\Pr[A_f] \leq 1/g^2$. Moreover, if the endpoints of f are in V_i and in V_j, then the event A_f is mutually independent of all the events corresponding to edges whose endpoints do not lie in $V_i \cup V_j$. Thus there is a dependency digraph for the events in which the maximum degree is less than $2gd$, and since $e \cdot 2gd \cdot 1/g^2 = 2ed/g < 1$, we conclude, by Corollary 5.1.2, that with positive probability none of the events A_f holds. But this means that W is an independent set containing a vertex from each V_i, completing the proof. ∎

Proposition 5.5.3 suffices to prove Conjecture 5.5.2 for digraphs with no short directed cycle. Recall that the directed girth of a digraph is the minimum length of a directed cycle in it.

Theorem 5.5.4 *Let $G = (U, F)$ be a d-regular digraph with directed girth $g \geq 8ed$. Then*

$$\mathrm{dla}(G) = d + 1.$$

Proof. As is well known, F can be partitioned into d pairwise disjoint 1-regular spanning subgraphs $F_1, \ldots, F_d$ of G. [This is an easy consequence of the Hall–König theorem; let H be the bipartite graph whose two classes of vertices A and B are copies of U, in which $u \in A$ is joined to $v \in B$ if $(u, v) \in F$. Since H is d-regular its edges can be decomposed into d perfect matchings, which correspond to d 1-regular spanning subgraphs of G.] Each F_i is a union of vertex disjoint directed cycles $C_{i1}, C_{i2}, \ldots, C_{ir_i}$. Let $V_1, V_2, \ldots, V_r$ be the sets of *edges* of all the cycles $\{C_{ij} : 1 \leq i \leq d, 1 \leq j \leq r_i\}$. Clearly, $V_1, V_2, \ldots, V_r$ is a partition of the set F of all edges of G, and by the girth

condition, $|V_i| \geq g \geq 8ed$ for all $1 \leq i \leq r$. Let H be the line graph of G, that is, the graph whose set of vertices is the set F of edges of G in which two edges are adjacent iff they share a common vertex in G. Clearly, H is $4d-2$ regular. As the cardinality of each V_i is at least $8ed \geq 2e(4d-2)$, there is, by Proposition 5.5.3, an independent set of H containing a member from each V_i. But this means that there is a matching M in G, containing at least one edge from each cycle C_{ij} of the 1-factors $F_1, \ldots, F_d$. Therefore $M, F_1 \setminus M, F_2 \setminus M, \ldots, F_d \setminus M$ are $d+1$ directed forests in G (one of which is a matching) that cover all its edges. Hence

$$\mathrm{dla}(G) \leq d+1.$$

As G has $|U| \cdot d$ edges, and each directed linear forest can have at most $|U|-1$ edges,

$$\mathrm{dla}(G) \geq |U|d/(|U|-1) > d.$$

Thus $\mathrm{dla}(G) = d+1$, completing the proof. ■

The last theorem shows that the assertion of Conjecture 5.5.2 holds for digraphs with sufficiently large (directed) girth. In order to deal with digraphs with small girth, we show that most of the edges of each regular digraph can be decomposed into a relatively small number of almost regular digraphs with high girth. To do this, we need the following statement, which is proved using the Local Lemma:

Lemma 5.5.5 *Let $G = (V, E)$ be a d-regular directed graph, where d is sufficiently large, and let p be an integer satisfying $10\sqrt{d} \leq p \leq 20\sqrt{d}$. Then, there is a p-coloring of the vertices of G by the colors $0, 1, 2, \ldots, p-1$ with the following property: for each vertex $v \in V$ and each color i, the numbers*

$$N^+(v,i) = |\{u \in V : (v,u) \in E \text{ and } u \text{ is colored } i\}|$$

and

$$N^-(v,i) = |\{u \in V : (u,v) \in E \text{ and } u \text{ is colored } i\}|$$

satisfy

$$|N^-(v,i) - d/p|, |N^+(v,i) - d/p| \leq 3\sqrt{d/p}\sqrt{\log d}. \tag{5.7}$$

Proof. Let $f : V \to \{0, 1, \ldots, p-1\}$ be a random vertex coloring of V by p colors, where for each $v \in V$, $f(v) \in \{0, 1, \ldots, p-1\}$ is chosen according to a uniform distribution. For every vertex $v \in V$ and every color $i, 0 \leq i < p$, let $A^+_{v,i}$ be the event where the number $N^+(v,i)$ of neighbors of v in G whose color is i does not satisfy inequality (5.7). Clearly, $N^+(v,i)$ is a binomial random variable with expectation d/p and standard deviation $\sqrt{(d/p)(1-1/p)} < \sqrt{d/p}$. Hence, by the standard estimates for binomial distribution given in Appendix A, for every $v \in V$ and $0 \leq i < p$,

$$\Pr[A^+_{v,i}] < 1/d^4.$$

Similarly, if $A^-_{v,i}$ is the event where the number $N^-(v,i)$ violates (5.7), then

$$\Pr[A^-_{v,i}] < 1/d^4.$$

Clearly, each of the events $A^+_{v,i}$ or $A^-_{v,i}$ is mutually independent of all the events $A^+_{u,j}$ or $A^-_{u,j}$ for all vertices $u \in V$ that do not have a common neighbor with v in G. Thus there is a dependency digraph for all our events with maximum degree $\leq (2d)^2 \cdot p$. Since $e \cdot (1/d^4)((2d)^2p + 1) < 1$, Corollary 5.1.2 (i.e., the symmetric form of the Local Lemma) implies that, with positive probability, no event $A^+_{v,i}$ or $A^-_{v,i}$ occurs. Hence there is a coloring f that satisfies (5.7) for all $v \in V$ and $0 \leq i < p$, completing the proof. ■

We are now ready to deal with general regular digraphs. Let $G = (V, E)$ be an arbitrary d-regular digraph. Throughout the argument, we assume, whenever needed, that d is sufficiently large. Let p be a prime satisfying $10d^{1/2} \leq p \leq 20d^{1/2}$ (it is well known that for every n there is a prime between n and $2n$). By Lemma 5.5.5, there is a vertex coloring $f : V \to \{0, 1, \ldots, p-1\}$ satisfying (5.7). For each i, $0 \leq i < p$, let $G_i = (V, E_i)$ be the spanning subdigraph of G defined by $E_i = \{(u,v) \in E : f(v) \equiv f(u) + i (\mathrm{mod}\ p)\}$. By inequality (5.7), the maximum indegree Δ^-_i and the maximum outdegree Δ^+_i in each G_i are at most $(d/p) + 3\sqrt{d/p}\sqrt{\log d}$. Moreover, for each $i > 0$, the length of every directed cycle in G_i is divisible by p. Thus the directed girth g_i of G_i is at least p. Since each G_i can be completed, by adding vertices and edges to a Δ_i-regular digraph with the same girth g_i and with $\Delta_i = \max(\Delta^+_i, \Delta^-_i)$, and since $g_i > 8e\Delta_i$ (for all sufficiently large d), we conclude, by Theorem 5.5.4, that $\mathrm{dla}(G_i) \leq \Delta_i + 1 \leq (d/p) + 3\sqrt{d/p}\sqrt{\log d} + 1$ for all $1 \leq i < p$. For G_0, we only apply the trivial inequality

$$\mathrm{dla}(G_0) \leq 2\Delta_0 \leq 2\frac{d}{p} + 6\sqrt{\frac{d}{p}}\sqrt{\log d}$$

obtained, for example, by embedding G_0 as a subgraph of a Δ_0-regular graph, splitting the edges of this graph into Δ_0 1-regular spanning subgraphs, and breaking each of these 1-regular spanning subgraphs into two linear directed forests. The last two inequalities, together with the fact that $10\sqrt{d} \leq p \leq 20\sqrt{d}$, imply

$$\mathrm{dla}(G) \leq d + 2\frac{d}{p} + 3\sqrt{pd}\sqrt{\log d} + 3\sqrt{\frac{d}{p}}\sqrt{\log d} + p - 1 \leq d + c \cdot d^{3/4}(\log d)^{1/2}.$$

We have thus proved the following:

Theorem 5.5.6 *There is an absolute constant $c > 0$ such that for every d-regular digraph G*

$$\mathrm{dla}(G) \leq d + cd^{3/4}(\log d)^{1/2}.$$

We note that by being a little more careful, we can improve the error term to $c'd^{2/3}(\log d)^{1/3}$. Since the edges of any undirected $d = 2f$-regular graph can be oriented so that the resulting digraph is f-regular, and since any $(2f-1)$-regular undirected graph is a subgraph of a $2f$-regular graph, the last theorem implies the following:

Theorem 5.5.7 *There is an absolute constant $c > 0$ such that, for every undirected d-regular graph G,*

$$la(G) \leq \frac{d}{2} + cd^{3/4}(\log d)^{1/2}.$$

5.6 LATIN TRANSVERSALS

Following the proof of the Local Lemma, we noted that the mutual independency assumption in this lemma can be replaced by the weaker assumption that the conditional probability of each event, given the mutual nonoccurrence of an arbitrary set of events, each nonadjacent to it in the dependency digraph, is sufficiently small. In this section we describe an application, from Erdős and Spencer (1991), of this modified version of the lemma. Let $A = (a_{ij})$ be an $n \times n$ matrix with, say, integer entries. A permutation π is called a *Latin transversal* (of A) if the entries $a_{i\pi(i)}$ $(1 \leq i \leq n)$ are all distinct.

Theorem 5.6.1 *Suppose $k \leq (n-1)/(4e)$, and suppose no integer appears in more than k entries of A. Then A has a Latin transversal.*

Proof. Let π be a random permutation of $\{1, 2, \ldots, n\}$, chosen according to a uniform distribution among all possible $n!$ permutations. Denote by T the set of all ordered four-tuples (i, j, i', j') satisfying $i < i', j \neq j'$ and $a_{ij} = a_{i'j'}$. For each $(i, j, i', j') \in T$, let $A_{iji'j'}$ denote the event where $\pi(i) = j$ and $\pi(i') = j'$. The existence of a Latin transversal is equivalent to the statement that, with positive probability, none of these events hold. Let us define a symmetric digraph (i.e., a graph) G on the vertex set T by making (i, j, i', j') adjacent to (p, q, p', q') if and only if $\{i, i'\} \cap \{p, p'\} \neq \emptyset$ or $\{j, j'\} \cap \{q, q'\} \neq \emptyset$. Thus these two four-tuples are not adjacent if the four cells (i, j), (i', j'), (p, q), and (p', q') occupy four distinct rows and columns of A. The maximum degree of G is less than $4nk$; indeed, for a given $(i, j, i', j') \in T$, there are at most $4n$ choices of (s, t) with either $s \in \{i, i'\}$ or $t \in \{j, j'\}$, and for each of these choices of (s, t) there are less than k choices for $(s', t') \neq (s, t)$ with $a_{st} = a_{s't'}$. Each such four-tuple (s, t, s', t') can be uniquely represented as (p, q, p', q') with $p < p'$. Since $e \cdot 4nk \cdot [1/n(n-1)] \leq 1$, the desired result follows from the above-mentioned strengthening of the symmetric version of the Local Lemma, if we can show that

$$\Pr\left[A_{iji'j'} \mid \bigwedge_{S} \overline{A_{pqp'q'}}\right] \leq \frac{1}{n(n-1)} \tag{5.8}$$

for any $(i,j,i',j') \in T$ and any set S of members of T that are nonadjacent in G to (i,j,i',j'). By symmetry, we may assume that $i = j = 1, i' = j' = 2$, and that hence none of the p's or q's are either 1 or 2. Let us call a permutation π *good* if it satisfies $\bigwedge_S \overline{A_{pqp'q'}}$, and let S_{ij} denote the set of all good permutations π satisfying $\pi(1) = i$ and $\pi(2) = j$. We claim that $|S_{12}| \leq |S_{ij}|$ for all $i \neq j$. Indeed, suppose, first, that $i, j > 2$. For each good $\pi \in S_{12}$, define a permutation π^* as follows: Suppose $\pi(x) = i$, $\pi(y) = j$. Then define $\pi^*(1) = i, \pi^*(2) = j, \pi^*(x) = 1, \pi^*(y) = 2$, and $\pi^*(t) = \pi(t)$ for all $t \neq 1, 2, x, y$. One can easily check that π^* is good, since the cells $(1,i), (2,j), (x,1), (y,2)$ are not part of any $(p,q,p',q') \in S$. Thus $\pi^* \in S_{ij}$, and since the mapping $\pi \to \pi^*$ is injective, $|S_{12}| \leq |S_{ij}|$, as claimed. Similarly, one can define injective mappings showing that $|S_{12}| \leq |S_{ij}|$ even when $\{i,j\} \cap \{1,2\} \neq \emptyset$. It follows that

$$\Pr\left[A_{1122} \wedge \bigwedge_S \overline{A_{pqp'q'}}\right] \leq \Pr\left[A_{1i2j} \wedge \bigwedge_S \overline{A_{pqp'q'}}\right]$$

for all $i \neq j$ and hence that

$$\Pr\left[A_{1122} \mid \bigwedge_S \overline{A_{pqp'q'}}\right] \leq \frac{1}{n(n-1)}.$$

By symmetry, this implies (5.8) and completes the proof. ■

5.7 MOSER'S FIX-IT ALGORITHM

When the probabilistic method is applied to prove that a certain event holds with high probability, it often supplies an efficient deterministic, or at least randomized, algorithm for the corresponding problem.

By applying the Local Lemma, we often manage to prove that a given event holds with positive probability, although this probability may be exponentially small in the dimensions of the problem. Consequently, it is not clear if any of these proofs can provide polynomial algorithms for the corresponding algorithmic problems. For many years, there was no known method of converting the proofs of any of the examples discussed in this chapter into an efficient algorithm. In 1991, József Beck found such a method that works for some of these examples, with a little loss in the constants. This has been extended and modified by several researchers. In 2009, Robin Moser found a remarkably simple algorithm, combined with a subtle analysis. This approach has been extended in his joint work with Tardos, which provides an efficient algorithm for essentially all known applications of the Local Lemma, with no loss in the constants. In fact, in some examples the constants obtained are even better than those that follow from the earlier existence proofs. We proceed with the details, following Moser and Tardos (2010).

We first give the context. Let Ω be a finite set. For each $v \in \Omega$ let $C[v]$ denote a random variable. The variables $C[v]$ may have different distributions but, critically,

they are mutually independent. Ω and the $C[v]$ then define a probability space. Let I denote an index set. For each $\alpha \in I$, there is an associated set $A[\alpha] \subset \Omega$ and an event BAD$[\alpha]$. Set $p[\alpha] = \Pr[\text{BAD}[\alpha]]$. Critically, the event BAD$[\alpha]$ depends only on the values $C[v]$ for $v \in A[\alpha]$. Our goal shall be to find (under suitable side conditions) specific values for $C[v]$ such that none of the BAD$[\alpha]$, $\alpha \in I$, holds. We define $\sim$ on I by setting $\alpha \sim \beta$ if $A[\alpha] \cap A[\beta] \neq \emptyset$. (Note that $\alpha \sim \alpha$.) The relation $\sim$ does yield a dependency graph D on the events BAD$[\alpha]$ with (α, β) an edge if $\alpha \sim \beta$. The event BAD$[\alpha]$ is mutually independent of all BAD$[\beta]$ with (α, β) not an edge since those events are determined by $C[v]$ for $v \notin A[\alpha]$.

As an instructive example, let $A[\alpha] \subset \Omega$ be a family of k element sets. For $v \in \Omega$, let $\Pr[C[v] = \text{Red}] = \frac{1}{2}$, and $\Pr[C[v] = \text{Blue}] = \frac{1}{2}$. That is, C is a random 2-coloring of Ω. For each α let BAD$[\alpha]$ denote the event where $A[\alpha]$ is monochromatic. Thus all $p[\alpha] = 2^{1-k}$. Our goal is then to find (under suitable side conditions) a coloring of Ω for which no $A[\alpha]$ is monochromatic. In other cases, Ω consists of Boolean variables $x_1, \ldots, x_n$ and $C[x_i]$ is true or false with independent probability $1/2$. Another type $S \subset \Omega$ is a random set, $C[v]$ is that $v \in S$, and the $C[v]$ are determined by independent, though not necessarily identical, coin flips. When Ω is the set of pairs $\{i,j\} \subset \{1, \ldots, n\}$ and $\Pr[C(\{i,j\}) = 1] = p_{i,j}$, $\Pr[C(\{i,j\}) = 0] = 1 - p_{i,j}$, one gets a broad generalization of $G(n,p)$. It is even possible for $C[v]$ to be multivalued. Indeed, essentially all applications of the Local Lemma can be placed in this context.

MOSER's FIX-IT ALGORITHM
FIX-IT I: For each $v \in \Omega$, choose $C[v]$ according to its distribution.
FIX-IT II: WHILE at least one BAD$[\alpha]$ holds
    FIX-IT III: Select one $\alpha \in I$ for which BAD$[\alpha]$.
    FIX-IT IV: Reset $C[v]$ for each $v \in A[\alpha]$.
    END WHILE

Tautologically, when and if the FIX-IT Algorithm terminates, the desired values of $C[v]$ have been found. FIX-IT III allows an arbitrary deterministic selection process; for definiteness we may imagine I to be linearly ordered and select the first $\alpha \in I$ for which BAD$[\alpha]$. We define the *LOG* to be the sequence $\alpha_1 \alpha_2 \cdots \alpha_u$ where α_t is the α selected (*not* all the α with BAD$[\alpha]$!) the t-th time FIX-IT III is applied. We define TLOG $= u$, the number of times FIX-IT III is applied. A priori, TLOG $= \infty$ is possible, but we shall give conditions that imply $E[\text{TLOG}] < \infty$, which in turn implies that the FIX-IT Algorithm will terminate with probability 1.

We define a *Moser Tree* to be a finite rooted tree T whose vertices are labeled by $\alpha \in I$. The depth of a node is its distance from the root, and the depth of a tree is the maximal depth of its nodes. We require the following:

1. If a node labeled β is a child of a node labeled α, then $\beta \sim \alpha$.
2. If two nodes at the same depth have labels β, γ, then it *cannot* be that $\beta \sim \gamma$. (In particular, the labels at a given depth are distinct.)

Note, however, that many nodes may, and often will, have the same label. We define $p[T]$ as the product of the $p[\alpha]$, where α ranges over the labels of the nodes. To clarify, when α appears u times, the factor $p[\alpha]$ appears u times.

Theorem 5.7.1 *Suppose the sum of $p[T]$ over all Moser Trees T is convergent, and let s denote its sum. Then $E[TLOG] \leq s$.*

Let $\alpha_1, \ldots, \alpha_u$ be a prefix of *LOG*. We associate to it a Moser Tree T_u. The root is labeled α_u. Now let t run from $u-1$ down to 1. If we do not have $\alpha_t \sim \alpha_{t'}$ for some $t < t' \leq u$ for which $\alpha_{t'}$ has been already placed in T_u, then we ignore α_t. Otherwise, among all such t', select one such that the node labeled $\alpha_{t'}$ is at the greatest depth. (In case of ties, select any one such $\alpha_{t'}$.) Add a node with label α_t and make it the child of the node labeled $\alpha_{t'}$.

We claim $T = T_u$ will be a Moser Tree. When a node with label β is created as a child of a node with label α, we must have $\alpha \sim \beta$. Now suppose at depth D there are two nodes with labels β, γ and $\beta \sim \gamma$. One of them would have been created first, say the one with label $\beta = \alpha_{t'}$. Later in the process, we reach $\gamma = \alpha_t$. As $\gamma \sim \beta$, the node with label γ will be placed at depth at least $D+1$, contradicting our assumption.

Remark. The prefix $\alpha_1 \cdots, \alpha_u$ depends on the selection process used in FIX-IT III. In some sense, the Moser Tree T_u encapsulates the critical information leading to the choice of α_u. If, say, FIX-IT III had given priority to the nodes in the tree, then it would have begun with precisely the nodes in the tree in any order for which children come before their parent.

Example. Suppose Ω is the English alphabet and $\alpha \sim \beta$ if they are equal or one or two apart in alphabetical order. Consider the prefix *RFSPTR*. The Moser Tree begins (at the end) with root *R*. *T* is a child of *R*. *P* is a child of *R*. *S* is a child of *P*, *F* is ignored, and *R* is a child of *S*. Had *F* been given low priority in FIX-IT III, the prefix would have been *RSPTR*. Had, further, *P* been given high priority, the prefix would have been *RPSTR*.

Let $u < v$, and assume LOG has prefix $\alpha_1 \cdots \alpha_v$. We claim the Moser Trees T_u, T_v are not equal. Indeed, suppose they are. As they have roots α_u, α_v, it must be that $\alpha_u = \alpha_v$. All $1 \leq j \leq u$ with α_j in T_u would also have α_j in T_v, and T_v would have the additional node α_v, showing the two trees differ. A *LOG* of length u will then generate u distinct Moser Trees. Hence

$$\mathrm{E[TLOG]} = \sum_T \Pr[T = T_n \text{ for some } n]$$

Theorem 5.7.2 *For any Moser Tree T*

$$\Pr[T = T_n \textit{ for some } n] \leq p(T)$$

It is helpful to *preprocess* the random choices of $C[v]$. For all integers $t \geq 0$, let $C[v,t]$ have the distribution of $C[v]$, and let all $C[v,t]$ be mutually independent. At step FIX-IT I, v is given $C[v,0]$. At step FIX-IT IV, if $C[v,t]$ has already been used, then v is given $C[v,t+1]$. Given the Moser Tree T, say vertex v appears in $A[\alpha_{i_0}], \ldots, A[\alpha_{i_u}]$, where the nodes are listed in order of depth, the highest depth first. Any $s < i_u$ with $v \in A[\alpha_s]$ must appear in the Moser Tree. Hence, when FIX-IT IV is applied to $A[\alpha_{i_t}]$, it will use $C[v,t]$. For some $T_n = T$ (regardless of n), it is necessary that BAD$[\alpha_t]$ for all t. The BAD$[\alpha_t]$ have probability $p[\alpha_t]$. Critically, as the "coin flips" $C[v,t]$ are mutually independent and none is used twice, the events BAD$[\alpha_t]$ are mutually independent and so the probability that they all hold is the product of the probabilities, namely $p(T)$.

Example. Let $A[\alpha] = \{1,2,3\}$, $A[\beta] = \{2,3,4\}$, and let T consist of root β with single child α. Let BAD$[\alpha]$, BAD$[\beta]$ be the events that $A[\alpha], A[\beta]$ are monochromatic. In order for $T = T_n$ for some n, it is necessary that $C[1,0]$, $C[2,0]$, $C[3,0]$ be the same and that $C[2,1]$, $C[3,1]$, $C[4,0]$ be the same. These are mutually independent events, and the probability they both hold is $(1/4)^2$. If $T = T_n$, there cannot be any other γ with 2 or 3 or 4 in $A[\gamma]$ that appears before β, as that γ would have been a node in the Moser Tree T_n. There cannot be any other γ with $1 \in A[\gamma]$ that appears before α as that α would also have been a node in the Moser Tree T_n.

Calculating the sum of $p[T]$ over all Moser Trees can be a daunting task. Instead, we find a larger sum. We call a labeled (with I) rooted tree a *weak* Moser Tree if the following conditions hold:

1. If a node labeled β is a child of a node labeled α, then $\beta \sim \alpha$.
2. The labels of the children of a node are distinct.

Weak Moser Trees have a nice recursive structure. For $\alpha \in I$, let $w(\alpha)$ denote the (possibly infinite) sum of $p[T]$ over all weak Moser Trees with root labeled α. For $\alpha \in I$, let $w(D,\alpha)$ denote the sum of $p[T]$ over all weak Moser Trees with root labeled α and depth at most D. Weak Moser Trees with root α and depth at most D decompose into the root α and some (maybe none) Moser Trees with roots $\beta \sim \alpha$, all of whom have depth at most $D-1$. Thus w is given by the recursive system

$$w(D,\alpha) = p(\alpha) \prod_{\beta \sim \alpha} (1 + w(D-1,\beta)). \tag{5.9}$$

The only tree with root α and depth 0 consists solely of the root and has $p(T) = p(\alpha)$. This yields the initial condition $w(0,\alpha) = p(\alpha)$.

Theorem 5.7.3 *Suppose there exist $x[\alpha] \geq p[\alpha]$ for $\alpha \in I$ such that*

$$x(\alpha) \geq p(\alpha) \prod_{\beta \sim \alpha} (1 + x(\beta)). \tag{5.10}$$

Then $w(\alpha) \leq x(\alpha)$. Further, $E[TLOG] \leq \sum_{\alpha \in I} x(\alpha)$.

Proof. We show $w(D, \alpha) \leq x(\alpha)$ for all $\alpha \in I$ by induction on D. For $D = 0$, $w(0, \alpha) = p(\alpha) \leq x(\alpha)$. Suppose, by induction on D, that $w(D-1, \beta) \leq x(\beta)$ for all $\beta \in I$. From (5.9), (5.10), $w(D, \alpha) \leq x(\alpha)$ for all $\alpha \in I$. Thus $w(\alpha) = \lim_{D \to \infty} w(D, \alpha) \leq x(\alpha)$. As all Moser Trees are weak Moser Trees, E[TLOG] is at most the sum of $p(T)$ over all weak Moser Trees, which is at most $\sum_{\alpha \in I} x(\alpha)$.

The symmetric case is particularly nice, and often occurs in applications. Suppose all $p[\alpha] \leq p$ and for all α, $|\{\beta : \beta \sim \alpha\}| \leq d$. We apply Theorem 5.7.3 with all $x(\alpha) = x$. If there exists $x \geq p$ such that $x \geq p(1+x)^d$, then E[TLOG] $\leq x|I|$. $x(1+x)^{-d}$ has maximal value $(d-1)^{d-1}d^{-d}$, given at $x = (d-1)^{-1}$. ■

Theorem 5.7.4 *If $p \leq (d-1)^{d-1}d^{-d}$, then $E[TLOG] \leq |I|/(d-1)$.*

Theorem 5.7.5 *If $epd \leq 1$, then $E[TLOG] \leq |I|/(d-1)$.*

Moser has also given an alternative analysis of the algorithm based on what he calls an entropy compression argument. The basic idea here is to show that, if the algorithm runs without terminating for a long time, then the LOG constructed enables us to compress the random string to a shorter one, and this is impossible. Rather than describing the argument for the general case, we give here only one simple and elegant illustration, given in Grytczuk, Kozik and Micek (2013).

A *repetition of length h* in a sequence is two identical adjacent blocks, each consisting of h consecutive elements. A sequence is *nonrepetitive* if it contains no repetitions. Thus, for example, the sequence 1231241 is nonrepetitive, while 1213413451 is not, as it contains the repetition 134134. Thue proved in 1906 that there is an infinite nonrepetitive sequence over an alphabet of three symbols. An extension is proved in Alon et al. (2002) using the Local Lemma: the vertices of any graph with maximum degree d can be colored by $O(d^2)$ colors so that every path in the graph is nonrepetitive. A variant in which the allowed colors for each vertex must lie in a list associated with the vertex has been considered as well. Here we prove the following:

Theorem 5.7.6 [Grytczuk et al. (2013)] *For every $n \geq 1$ and every sequence of lists of symbols $L_1, L_2, \ldots, L_n$, each of size 5, there is a nonrepetitive sequence $s_1, s_2, \ldots, s_n$, where $s_i \in L_i$. Moreover, there is a randomized algorithm that finds such a sequence in expected time polynomial in n, given the lists.*

Note that, by König's Lemma, the above implies that any infinite sequence of lists of size 5 admits a nonrepetitive sequence of symbols chosen from the lists. We note also that it has been conjectured that lists of size 3 suffice (see the paragraph following the proof for more about lists of size 3 and 4).

The proof of the theorem is by a simple algorithm: the sequence is generated by choosing symbols randomly, independently, and uniformly from the lists, where every time a repetition occurs, the repeated block is erased and the process continues. More formally, consider the following algorithm for generating a nonrepetitive sequence $s_1, s_2, \ldots, s_n$ from the lists $L_1, L_2, \ldots, L_s$, each of size 5:

Starting with $i = 1$, as long as $i \leq n$, perform an iteration as follows: let s_i be a random element of L_i. If the sequence $s_1, s_2, \ldots, s_i$ is nonrepetitive, increase i to $i + 1$ and go to the next iteration. Otherwise, note that, crucially, there is a unique repetition ending at s_i, and let it be $s_{i-2h+1}, \ldots, s_{i-h}, s_{i-h+1}, \ldots, s_i$. In this case, replace i by $i - h + 1$ and proceed to the next iteration. (This includes the case $h = 1$ where $s_i = s_{i-1}$ and s_i is deleted.)

Note that, if the algorithm terminates, it generates a nonrepetitive sequence of length n, as required. To complete the proof, we show that with high probability it terminates after $O(n)$ iterations. To estimate the probability that it does not terminate in M iterations, fix an arbitrary order of the symbols in each list and let $r_j \in [5]$ be the position of the chosen element in iteration j ($1 \leq j \leq M$). Define a sequence $d_1, \ldots, d_m$, where $d_1 = 1$ and d_j is the difference between the value of i at the end of iteration j and its value at the end of iteration $j - 1$. Thus $d_j = 1$ if in iteration number j no repetition is obtained; otherwise it is $-h + 1$ where h is the length of the repeated block obtained. The LOG corresponding to the run of the algorithm is (D_M, S_M), where $D_M = (d_1, \ldots, d_M)$ and $S_M = (s_1, s_2 \cdots, s_\ell)$ is the sequence obtained after M iterations. The crucial fact is the following:

Fact: Every LOG which can be obtained corresponds to exactly one sequence $r_1, r_2, \ldots, r_M$. Therefore, the probability to get it is 5^{-M}.

To prove the fact, it suffices to show that by knowing the LOG (D_M, S_M) we can reconstruct r_M and (D_{M-1}, S_{M-1}), where $D_{M-1} = (d_1, \ldots, d_{M-1})$ and S_{M-1} is the sequence the algorithm generates after $M - 1$ iterations. The same process will then enable us to reconstruct, by induction, $r_{M-1}, \ldots, r_1$.

Knowing D_M clearly reveals D_{M-1}. If $d_M = 1$, then r_M is just the position of s_ℓ in the list L_ℓ, and S_{M-1} is S_M without its last symbol. Otherwise, $d_M = -h + 1$, where h is the length of the last repeated block. In this case, the h symbols that have been erased from S_{M-1} together with the last chosen symbol to get S_M are equal to the last h symbols of S_M, in order, and thus in this case $S_{M-1} = s_1, \ldots, s_\ell, s_{\ell-h+1}, \ldots, s_{\ell-1}$ and r_M is the location of s_ℓ in the list $L_{\ell+h}$. This proves the fact.

Each possible sequence $D_M = (d_1, \ldots, d_m)$ can be encoded by a sequence over $+, -$ as follows: for each j, $1 \leq j \leq M$, in order, write $+$ once, followed by $|d_j - 1|$ times $-$ (note that $d_j - 1$ is zero when $d_j = 1$). Thus the total number of $+$ in the sequence is M, and the total number of $-$ is exactly the number of symbols discarded during the process, which is at most $M - 1$ (and at least $M - n$, but we will not use this last fact here). This shows that the number of possibilities for D_M in the LOG is smaller than 2^{2M}, and hence the number of possibilities for the LOG (D_M, S_M) is at most $2^{2M}5^n$. By the fact, this implies that the probability the algorithm does not terminate after M iterations is at most $4^M 5^{n-M}$, which is tiny for, say, $M = 8n$. Therefore, the algorithm terminates, with high probability, after less than $8n$ iterations and the expected number of iterations until termination is $O(n)$. This completes the proof of the theorem.

The randomized algorithm described above is well defined for lists of any size. It has been conjectured that, for some $M = M(n)$, the algorithm terminates with high probability when all lists have size 3. We give this as an Exercise when the lists all

have size 4. Extensive simulations (at least, in the case where all L_i are the same) indicate that one can take $M = O(n)$ even if each list is of size 3, but this remains a conjecture.

5.8 EXERCISES

1. (*) Prove that, for every integer $d > 1$, there is a finite $c(d)$ such that the edges of any bipartite graph with maximum degree d in which every cycle has at least $c(d)$ edges can be colored by $d + 1$ colors so that there are no two adjacent edges with the same color and there is no two-colored cycle.
2. (*) Prove that, for every $\epsilon > 0$, there is a finite $l_0 = l_0(\epsilon)$ and an infinite sequence of bits $a_1, a_2, a_3, \cdots$, where $a_i \in \{0, 1\}$, such that for every $l > l_0$ and every $i \geq 1$, the two binary vectors $u = (a_i, a_{i+1}, \ldots, a_{i+l-1})$ and $v = (a_{i+l}, a_{i+l+1}, \ldots, a_{i+2l-1})$ differ in at least $(\frac{1}{2} - \epsilon)l$ coordinates.
3. Let $G = (V, E)$ be a simple graph, and suppose each $v \in V$ is associated with a set $S(v)$ of colors of size at least $10d$, where $d \geq 1$. Suppose, in addition, that for each $v \in V$ and $c \in S(v)$ there are at most d neighbors u of v such that c lies in $S(u)$. Prove that there is a proper coloring of G assigning to each vertex v a color from its class $S(v)$.
4. Let $G = (V, E)$ be a cycle of length $4n$, and let $V = V_1 \cup V_2 \cup \cdots \cup V_n$ be a partition of its $4n$ vertices into n pairwise disjoint subsets, each of cardinality 4. Is it true that there must be an independent set of G containing precisely one vertex from each V_i? (Prove or supply a counterexample.)
5. (*) Prove that there is an absolute constant $c > 0$ such that for every k there is a set S_k of at least $ck \ln k$ integers, such that for every coloring of the integers by k colors there is an integer x for which the set $x + S$ does not intersect all color classes.
6. Suppose $|I| = m$ and $|\{\beta : \alpha \sim \beta\}| = d$ for all $\alpha \in I$. Suppose I and, for all $\alpha \in I$, $\{\beta : \alpha \sim \beta\}$ are given in linked lists. Suppose BAD[α] can be checked in unit time. Give a data structure so that FIX-IT runs in $O(m + (d + 1)\text{TLOG})$ time. (Note that simply checking all BAD[α] at step FIX-IT III could take time $\Theta(m \cdot TLOG)$.)
7. Set $m = 10^5$ (this can be adjusted for computer speed) and $n = 10m$. Let $x[1], \ldots, x[n]$ be Boolean variables. Let $\sigma_1, \ldots, \sigma_{30}$ be permutations of $1, \ldots, n$. For each $1 \leq i \leq 30$ and $0 \leq t < m$, create a clause $C = y[1] \vee \cdots \vee y[10]$ where y_l is either $x[\sigma_i(10t + l)]$ or its negation. Show, using the Local Lemma, that the conjunction of all $30m$ clauses is satisfiable. Selecting the permutations at random and the $y[l]$ at random, implement the FIX-IT algorithm to find a satisfying assignment.
8. Consider the proof of Theorem 5.7.6. Prove that in the encoding of D_M as a sequence of $+1$ and -1 (regardless of the sizes of the lists), the string $++--$ $-++---+$ cannot occur. Now suppose the lists $L_1, \ldots, L_s$ all have size four. Show that, with high probability, the algorithm terminates after $O(n)$ iterations.

THE PROBABILISTIC LENS:
Directed Cycles

Let $D = (V, E)$ be a simple directed graph with minimum outdegree δ and maximum indegree Δ.

Theorem 1 [Alon and Linial (1989)] *If* $e(\Delta\delta + 1)(1 - 1/k)^{\delta} < 1$, *then* D *contains a (directed, simple) cycle of length* $0 \pmod k$.

Proof. Clearly, we may assume that every outdegree is precisely δ, since otherwise we can consider a subgraph of D with this property.

Let $f : V \to \{0, 1, \ldots, k-1\}$ be a random coloring of V, obtained by choosing, for each $v \in V$, $f(v) \in \{0, \ldots, k-1\}$ independently, according to a uniform distribution. For each $v \in V$, let A_v denote the event where there is no $u \in V$, with $(v, u) \in E$ and $f(u) \equiv f(v) + 1 \pmod k$. Clearly, $\Pr[A_v] = (1 - 1/k)^{\delta}$. One can easily check that each event A_v is mutually independent of all the events A_u but those satisfying

$$N^+(v) \cap (\{u\} \cup N^+(u)) \neq \emptyset \,,$$

where here $N^+(v) = \{w \in V : (v, w) \in E\}$. The number of such u's is at most $\Delta\delta$ and hence, by our assumption and by the Local Lemma (Corollary 5.1.2), $\Pr\left[\bigwedge_{v \in V} \overline{A_v}\right] > 0$. Thus there is an $f : V \to \{0, 1, \ldots, k-1\}$ such that for every $v \in V$ there is a $u \in V$ with

$$(v, u) \in E \text{ and } f(u) \equiv f(v) + 1 \pmod k. \tag{1}$$

Starting at an arbitrary $v = v_0 \in V$ and applying (1) repeatedly, we obtain a sequence $v_0, v_1, v_2, \cdots$ of vertices of D so that $(v_i, v_{i+1}) \in E$ and $f(v_{i+1}) \equiv f(v_i) + 1 \pmod k$ for all $i \geq 0$. Let j be the minimum integer so that there is an $\ell < j$ with $v_\ell = v_j$. The cycle $v_\ell v_{\ell+1} v_{\ell+2} \cdots v_j = v_\ell$ is a directed simple cycle of D whose length is divisible by k. ■

6

Correlation Inequalities

> Sandwiched as we are between the "everything" that is behind us and the "zero" beyond us, ours is an ephemeral existence in which there is neither coincidence nor possibility.
> –from *A Wild Sheep Chase*, by Haruki Murakami

Let $G = (V, E)$ be a random graph on the set of vertices $V = \{1, 2, \ldots, n\}$ generated by choosing, for each $i, j \in V, i \neq j$ independently, the pair $\{i, j\}$ to be an edge with probability p, where $0 < p < 1$. Let H be the event where G is Hamiltonian, and let P be the event where G is planar. Suppose one wants to compare the two quantities $\Pr[P \wedge H]$ and $\Pr[P] \cdot \Pr[H]$. Intuitively, knowing that G is Hamiltonian suggests that it has many edges and hence seems to indicate that G is less likely to be planar. Therefore, it seems natural to expect that $\Pr[P \mid H] \leq \Pr[P]$, implying

$$\Pr[P \wedge H] \leq \Pr[H] \cdot \Pr[P].$$

This inequality, which is, indeed, correct, is a special case of the FKG Inequality of Fortuin, Kasteleyn and Ginibre (1971). In this chapter, we present the proof of this inequality and several related results, which deal with the correlation between certain events in probability spaces. The proofs of all these results are rather simple, and still they supply many interesting consequences. The first inequality of this type is due to Harris (1960). A result closer to the ones considered here is a lemma of Kleitman (1966), stating that, if $\mathcal{A}$ and $\mathcal{B}$ are two *monotonically decreasing* families

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

of subsets of $\{1, 2, \ldots, n\}$ (i.e., $A \in \mathcal{A}$ and $A' \subseteq A \Rightarrow A' \in \mathcal{A}$ and, similarly, $B \in \mathcal{B}$ and $B' \subseteq B \Rightarrow B' \in \mathcal{B}$), then

$$|\mathcal{A} \cap \mathcal{B}| \cdot 2^n \geq |\mathcal{A}| \cdot |\mathcal{B}|.$$

This lemma was followed by many extensions and generalizations until Ahlswede and Daykin (1978) obtained a very general result, which implies all these extensions. In the next section we present this result and its proof. Some of its many applications are discussed in the rest of the chapter.

6.1 THE FOUR FUNCTIONS THEOREM OF AHLSWEDE AND DAYKIN

Suppose $n \geq 1$ and put $N = \{1, 2, \ldots, n\}$. Let $P(N)$ denote the set of all subsets of N, and let $\mathbb{R}^+$ denote the set of nonnegative real numbers. For a function $\varphi : P(N) \to \mathbb{R}^+$ and for a family $\mathcal{A}$ of subsets of N, denote $\varphi(\mathcal{A}) = \sum_{A \in \mathcal{A}} \varphi(A)$. For two families $\mathcal{A}$ and $\mathcal{B}$ of subsets of N, define $\mathcal{A} \cup \mathcal{B} = \{A \cup B : A \in \mathcal{A}, B \in \mathcal{B}\}$ and $\mathcal{A} \cap \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$.

Theorem 6.1.1 [The Four Functions Theorem] *Let $\alpha, \beta, \gamma, \delta : P(N) \to \mathbb{R}^+$ be four functions from the set of all subsets of N to the nonnegative reals. If, for every two subsets $A, B \subseteq N$ the inequality*

$$\alpha(A)\beta(B) \leq \gamma(A \cup B)\delta(A \cap B) \tag{6.1}$$

holds, then, for every two families of subsets $\mathcal{A}, \mathcal{B} \subseteq P(N)$,

$$\alpha(\mathcal{A})\beta(\mathcal{B}) \leq \gamma(\mathcal{A} \cup \mathcal{B})\delta(\mathcal{A} \cap \mathcal{B}). \tag{6.2}$$

Proof. Observe, first, that we may modify the four functions $\alpha, \beta, \gamma, \delta$ by defining $\alpha(A) = 0$ for all $A \notin \mathcal{A}$, $\beta(B) = 0$ for all $B \notin \mathcal{B}$, $\gamma(C) = 0$ for all $C \notin \mathcal{A} \cup \mathcal{B}$, and $\delta(D) = 0$ for all $D \notin \mathcal{A} \cap \mathcal{B}$. Clearly, (6.1) still holds for the modified functions, and in inequality (6.2) we may assume now that $\mathcal{A} = \mathcal{B} = \mathcal{A} \cup \mathcal{B} = \mathcal{A} \cap \mathcal{B} = P(N)$.

To prove this inequality, we apply induction on n. The only step that requires some computation is $n = 1$. In this case, $P(N) = \{\phi, N\}$. For each function $\varphi \in \{\alpha, \beta, \gamma, \delta\}$, define $\varphi_0 = \varphi(\phi)$ and $\varphi_1 = \varphi(N)$. By (6.1), we have

$$\begin{aligned} \alpha_0\beta_0 &\leq \gamma_0\delta_0, \\ \alpha_0\beta_1 &\leq \gamma_1\delta_0, \\ \alpha_1\beta_0 &\leq \gamma_1\delta_0, \\ \alpha_1\beta_1 &\leq \gamma_1\delta_1. \end{aligned} \tag{6.3}$$

By the above paragraph, we only have to prove inequality (6.2), where $\mathcal{A} = \mathcal{B} = P(N)$, that is, to prove that

$$(\alpha_0 + \alpha_1)(\beta_0 + \beta_1) \leq (\gamma_0 + \gamma_1)(\delta_0 + \delta_1). \tag{6.4}$$

If either $\gamma_1 = 0$ or $\delta_0 = 0$, this follows immediately from (6.3). Otherwise, by (6.3), $\gamma_0 \geq \alpha_0\beta_0/\delta_0$ and $\delta_1 \geq \alpha_1\beta_1/\gamma_1$. It thus suffices to show that

$$\left(\frac{\alpha_0\beta_0}{\delta_0} + \gamma_1\right)\left(\delta_0 + \frac{\alpha_1\beta_1}{\gamma_1}\right) \geq (\alpha_0 + \alpha_1)(\beta_0 + \beta_1)$$

or, equivalently, that

$$(\alpha_0\beta_0 + \gamma_1\delta_0)(\delta_0\gamma_1 + \alpha_1\beta_1) \geq (\alpha_0 + \alpha_1)(\beta_0 + \beta_1)\delta_0\gamma_1\,.$$

The last inequality is equivalent to

$$(\gamma_1\delta_0 - \alpha_0\beta_1)(\gamma_1\delta_0 - \alpha_1\beta_0) \geq 0\,,$$

which follows from (6.3), as both factors on the left-hand side are nonnegative. This completes the proof for $n = 1$.

Suppose, now, that the theorem holds for $n-1$, and let us prove it for $n\,(\geq 2)$. Put $N' = N \setminus \{n\}$, and define for each $\varphi \in \{\alpha, \beta, \gamma, \delta\}$ and each $A \subseteq N'$, $\varphi'(A) = \varphi(A) + \varphi(A \cup \{n\})$. Clearly, for each function $\varphi \in \{\alpha, \beta, \gamma, \delta\}$, $\varphi'(P(N')) = \varphi(P(N))$. Therefore, the desired inequality (6.3) would follow from applying the induction hypothesis to the functions $\alpha', \beta', \gamma', \delta' : P(N') \to \mathbb{R}^+$. However, in order to apply this hypothesis we have to check that these new functions satisfy the assumption of Theorem 6.1.1 on N'; that is, that for every $A', B' \subseteq N'$,

$$\alpha'(A')\beta'(B') \leq \gamma'(A' \cup B')\delta'(A' \cap B')\,. \tag{6.5}$$

Not surprisingly, this last inequality follows easily from the case $n = 1$, which we have already proved. Indeed, let T be a 1-element set and define

$$\begin{aligned}
\overline{\alpha}(\phi) &= \alpha(A'), & \overline{\alpha}(T) &= \alpha(A' \cup \{n\}),\\
\overline{\beta}(\phi) &= \beta(B'), & \overline{\beta}(T) &= \beta(B' \cup \{n\}),\\
\overline{\gamma}(\phi) &= \gamma(A' \cup B'), & \overline{\gamma}(T) &= \gamma(A' \cup B' \cup \{n\}),\\
\overline{\delta}(\phi) &= \delta(A' \cap B'), & \overline{\delta}(T) &= \delta((A' \cap B') \cup \{n\}).
\end{aligned}$$

By the assumption (6.1), $\overline{\alpha}(S)\overline{\beta}(R) \leq \overline{\gamma}(S \cup R)\overline{\delta}(S \cap R)$ for all $S, R \subseteq T$ and, hence, by the case $n = 1$ already proved,

$$\alpha'(A')\beta'(B') = \overline{\alpha}(P(T))\overline{\beta}(P(T)) \leq \overline{\gamma}(P(T))\overline{\delta}(P(T)) = \gamma'(A' \cup B')\delta'(A' \cap B')\,,$$

which is the desired inequality (6.5). Therefore, inequality (6.2) holds, completing the proof. ∎

The Ahlswede–Daykin theorem can be extended to any arbitrary finite distributive lattices. A *lattice* is a partially ordered set in which every two elements, x and y,

have a unique minimal upper bound, denoted by $x \vee y$ and called the *join* of x and y, and a unique maximal lower bound, denoted by $x \wedge y$ and called the *meet* of x and y. A lattice L is *distributive* if, for all $x, y, z \in L$,

$$x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$$

or, equivalently, if for all $x, y, z \in L$,

$$x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

For two sets $X, Y \subseteq L$, define

$$X \vee Y = \{x \vee y : x \in X, y \in Y\}$$

and

$$X \wedge Y = \{x \wedge y : x \in X, y \in Y\}.$$

Any subset L of $P(N)$, where $N = \{1, 2, \ldots, n\}$, ordered by inclusion, which is closed under the union and intersection operations, is a distributive lattice. Here, the join of two members $A, B \in L$ is simply their union $A \cup B$, and their meet is the intersection $A \cap B$. It is somewhat more surprising (but easy to check) that every finite distributive lattice L is isomorphic to a sublattice of $P(\{1, 2, \ldots, n\})$ for some n. [To see this, call an element $x \in L$ *join-irreducible* if, whenever $x = y \vee z$, then either $x = y$ or $x = z$. Let $x_1, x_2, \ldots, x_n$ be the set of all join-irreducible elements in L, and associate each element $x \in L$ with the set $A = A(x) \subseteq N$, where $x = \vee_{i \in A} x_i$, and $\{x_i : i \in A\}$ are all the join-irreducibles y satisfying $y \leq x$. The mapping $x \to A(x)$ is the desired isomorphism.] This fact enables us to generalize Theorem 6.1.1 to arbitrary finite distributive lattices as follows:

Corollary 6.1.2 *Let L be a finite distributive lattice, and let α, β, γ and δ be four functions from L to $\mathbb{R}^+$. If*

$$\alpha(x)\beta(y) \leq \gamma(x \vee y)\delta(x \wedge y)$$

for all $x, y \in L$, then for every $X, Y \subseteq L$

$$\alpha(X)\beta(Y) \leq \gamma(X \vee Y)\delta(X \wedge Y).$$

The simplest case in the last corollary is the case where all the four functions α, β, γ, and δ are identically 1, stated below.

Corollary 6.1.3 *Let L be a finite distributive lattice, and suppose $X, Y \subseteq L$. Then*

$$|X| \cdot |Y| \leq |X \vee Y| \cdot |X \wedge Y|.$$

We close this section by presenting a very simple consequence of the last corollary, first proved by Marica and Schonheim (1969).

Corollary 6.1.4 *Let $\mathcal{A}$ be a family of subsets of a finite set N, and define*

$$\mathcal{A} \setminus \mathcal{A} = \{F \setminus F' : F, \quad F' \in \mathcal{A}\}.$$

Then $|\mathcal{A} \setminus \mathcal{A}| \geq |\mathcal{A}|$.

Proof. Let L be the distributive lattice of all subsets of N. By applying Corollary 6.1.3 to $\mathcal{A}$ and $\mathcal{B} = \{N \setminus F : F \in \mathcal{A}\}$, we obtain

$$|\mathcal{A}|^2 = |\mathcal{A}| \cdot |\mathcal{B}| \leq |\mathcal{A} \cup \mathcal{B}| \cdot |\mathcal{A} \cap \mathcal{B}| = |\mathcal{A} \setminus \mathcal{A}|^2.$$

The desired result follows. ∎

6.2 THE FKG INEQUALITY

A function $\mu : L \to \mathbb{R}^+$, where L is a finite distributive lattice, is called *log-supermodular* if

$$\mu(x)\mu(y) \leq \mu(x \vee y)\mu(x \wedge y)$$

for all $x, y \in L$. A function $f : L \to \mathbb{R}^+$ is *increasing* if $f(x) \leq f(y)$ whenever $x \leq y$ and is *decreasing* if $f(x) \geq f(y)$ whenever $x \leq y$.

Motivated by a problem from statistical mechanics, Fortuin et al. (1971) proved the following useful inequality, which has become known as the FKG inequality:

Theorem 6.2.1 [The FKG inequality] *Let L be a finite distributive lattice, and let $\mu : L \to \mathbb{R}^+$ be a log-supermodular function. Then, for any two increasing functions $f, g : L \to \mathbb{R}^+$, we have*

$$\left(\sum_{x \in L} \mu(x) f(x)\right)\left(\sum_{x \in L} \mu(x) g(x)\right) \leq \left(\sum_{x \in L} \mu(x) f(x) g(x)\right)\left(\sum_{x \in L} \mu(x)\right). \tag{6.6}$$

Proof. Define four functions $\alpha, \beta, \gamma, \delta : L \to \mathbb{R}^+$ as Follows: For each $x \in L$,

$$\alpha(x) = \mu(x) f(x), \qquad \beta(x) = \mu(x) g(x),$$
$$\gamma(x) = \mu(x) f(x) g(x), \qquad \delta(x) = \mu(x).$$

We claim that these functions satisfy the hypothesis of the Ahlswede–Daykin theorem, stated in Corollary 6.1.2. Indeed, if $x, y \in L$ then, by the supermodularity of μ and since f and g are increasing,

$$\alpha(x)\beta(y) = \mu(x) f(x) \mu(y) g(y) \leq \mu(x \vee y) f(x) g(y) \mu(x \wedge y)$$
$$\leq \mu(x \vee y) f(x \vee y) g(x \vee y) \mu(x \wedge y) = \gamma(x \vee y)\delta(x \wedge y).$$

Therefore, by Corollary 6.1.2 (with $X = Y = L$)

$$\alpha(L)\beta(L) \leq \gamma(L)\delta(L),$$

which is the desired result. ■

Note that the conclusion of Theorem 6.2.1 holds also if both f and g are decreasing (simply interchange γ and δ in the proof). In case f is increasing and g is decreasing (or vice versa), the opposite inequality holds:

$$\left(\sum_{x\in L}\mu(x)f(x)\right)\left(\sum_{x\in L}\mu(x)g(x)\right) \geq \left(\sum_{x\in L}\mu(x)f(x)g(x)\right)\left(\sum_{x\in L}\mu(x)\right).$$

To prove it, simply apply Theorem 6.2.1 to the two increasing functions $f(x)$ and $k - g(x)$, where k is the constant $\max_{x\in L} g(x)$. [This constant is needed to guarantee that $k - g(x) \geq 0$ for all $x \in L$.]

It is helpful to view μ as a measure on L. Assuming μ is not identically zero, we can define, for any function $f : L \to \mathbb{R}^+$, its expectation

$$\langle f\rangle = \frac{\sum_{x\in L} f(x)\mu(x)}{\sum_{x\in L}\mu(x)}.$$

With this notation, the FKG inequality asserts that, if μ is log-supermodular and $f, g : L \to \mathbb{R}^+$ are both increasing or both decreasing, then

$$\langle fg\rangle \geq \langle f\rangle\langle g\rangle.$$

Similarly, if f is increasing and g is decreasing (or vice versa), then

$$\langle fg\rangle \leq \langle f\rangle\langle g\rangle.$$

This formulation demonstrates clearly the probabilistic nature of the inequality, some of whose many interesting consequences are presented in the rest of this chapter.

6.3 MONOTONE PROPERTIES

Recall that a family $\mathcal{A}$ of subsets of $N = \{1, 2, \ldots, n\}$ is *monotonically decreasing* if $A \in \mathcal{A}$ and $A' \subseteq A \Rightarrow A' \in \mathcal{A}$. Similarly, it is *monotonically increasing* if $A \in \mathcal{A}$ and $A \subseteq A' \Rightarrow A' \in \mathcal{A}$. By considering the power set $P(N)$ as a symmetric probability space, one naturally defines the *probability* of $\mathcal{A}$ by

$$\Pr[\mathcal{A}] = \frac{|\mathcal{A}|}{2^n}.$$

Thus $\Pr[\mathcal{A}]$ is simply the probability that a randomly chosen subset of N lies in $\mathcal{A}$.

Kleitman's Lemma, which was the starting point of all the correlation inequalities considered in this chapter, is the following:

Proposition 6.3.1 *Let $\mathcal{A}$ and $\mathcal{B}$ be two monotonically increasing families of subsets of $N = \{1, 2, \ldots, n\}$, and let $\mathcal{C}$ and $\mathcal{D}$ be two monotonically decreasing families of subsets of N. Then*

$$\Pr[\mathcal{A} \cap \mathcal{B}] \geq \Pr[\mathcal{A}] \cdot \Pr[\mathcal{B}],$$
$$\Pr[\mathcal{C} \cap \mathcal{D}] \geq \Pr[\mathcal{C}] \cdot \Pr[\mathcal{D}],$$
$$\Pr[\mathcal{A} \cap \mathcal{C}] \leq \Pr[\mathcal{A}] \cdot \Pr[\mathcal{C}].$$

In terms of cardinalities, this can be read as follows:

$$2^n |\mathcal{A} \cap \mathcal{B}| \geq |\mathcal{A}| \cdot |\mathcal{B}|,$$
$$2^n |\mathcal{C} \cap \mathcal{D}| \geq |\mathcal{C}| \cdot |\mathcal{D}|,$$
$$2^n |\mathcal{A} \cap \mathcal{C}| \leq |\mathcal{A}| \cdot |\mathcal{C}|,$$

where here and in what follows, $\mathcal{A} \cap \mathcal{B}$, $\mathcal{C} \cap \mathcal{D}$, and $\mathcal{A} \cap \mathcal{C}$ denote usual intersections of families.

Proof. Let $f : P(N) \to \mathbb{R}^+$ be the characteristic function of $\mathcal{A}$; that is, $f(A) = 0$ if $A \notin \mathcal{A}$ and $f(A) = 1$ if $A \in \mathcal{A}$. Similarly, let g be the characteristic function of B. By the assumptions, f and g are both increasing. Applying the FKG inequality with the trivial measure $\mu \equiv 1$, we get

$$\Pr[\mathcal{A} \cap \mathcal{B}] = \langle fg \rangle \geq \langle f \rangle \langle g \rangle = \Pr[\mathcal{A}] \cdot \Pr[\mathcal{B}].$$

The other two inequalities follow similarly from Theorem 6.2.1 and the paragraph following it.

It is worth noting that the proposition can be also derived easily from the Ahlswede–Daykin theorem or from Corollary 6.1.3. ■

The last proposition has several interesting combinatorial consequences, some of which appear already in Kleitman's original paper. Since those are direct combinatorial consequences and do not contain any additional probabilistic ideas, we omit their exact statement and turn to a version of Proposition 6.3.1 in a more general probability space.

For a real vector $p = (p_1, \ldots, p_n)$, where $0 \leq p_i \leq 1$, consider the probability space whose elements are all members of the power set $P(N)$, where, for each $A \subseteq N$, $\Pr[A] = \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j)$. Clearly, this probability distribution is obtained if we choose a random $A \subseteq N$ by choosing each element $i \in N$, independently, with probability p_i. Let us denote, for each $\mathcal{A} \subseteq P(N)$, its probability in this space by $\Pr_p[\mathcal{A}]$. In particular, if all the probabilities p_i are $1/2$, then $\Pr_p[\mathcal{A}]$ is the quantity

denoted as $\Pr[A]$ in Proposition 6.3.1. Define $\mu = \mu_p : P(N) \to \mathbb{R}^+$ by $\mu(A) = \prod_{i \in A} p_i \prod_{j \notin A} (1 - p_j)$.

It is easy to check that μ is log-supermodular. This is because, for $A, B \subseteq N$, $\mu(A)\mu(B) = \mu(A \cup B)\mu(A \cap B)$, as can be checked by comparing the contribution arising from each $i \in N$ to the left-hand side and to the right-hand side of the last equality. Hence, one can apply the FKG inequality and obtain the following generalization of Proposition 6.3.1:

Theorem 6.3.2 *Let $\mathcal{A}$ and $\mathcal{B}$ be two monotonically increasing families of subsets of N, and let $\mathcal{C}$ and $\mathcal{D}$ be two monotonically decreasing families of subsets of N. Then, for any real vector $p = (p_1, \ldots, p_n)$, $0 \leq p_i \leq 1$,*

$$Pr_p[\mathcal{A} \cap \mathcal{B}] \geq Pr_p[\mathcal{A}] \cdot Pr_p[\mathcal{B}],$$
$$Pr_p[\mathcal{C} \cap \mathcal{D}] \geq Pr_p[\mathcal{C}] \cdot Pr_p[\mathcal{D}],$$
$$Pr_p[\mathcal{A} \cap \mathcal{C}] \leq Pr_p[\mathcal{A}] \cdot Pr_p[\mathcal{C}].$$

This theorem can be applied in many cases and will be used in Chapter 8 to derive the Janson inequalities. As a simple illustration, suppose that $A_1, A_2, \ldots, A_k$ are arbitrary subsets of N, and one chooses a random subset A of N by choosing each $i \in N$, independently, with probability p. Then, Theorem 6.3.2 easily implies that

$$\Pr[A \text{ intersects each } A_i] \geq \prod_{i=1}^{k} \Pr[A \text{ intersects } A_i].$$

Note that this is false, in general, for other similar probabilistic models. For example, if A is a randomly chosen ℓ-element subset of N, then the last inequality may fail.

By viewing the members of N as the $n = \binom{m}{2}$ edges of the complete graph on the set of vertices $V = \{1, 2, \ldots, m\}$, we can derive a correlation inequality for random graphs. Let $G = (V, E)$ be a random graph on the set of vertices V generated by choosing, for each $i, j \in V$, $i \neq j$, independently, the pair $\{i, j\}$ to be an edge with probability p. (This model of random graphs is discussed in detail in Chapter 10.) A *property of graphs* is a subset of the set of all graphs on V, closed under isomorphism. Thus, for example, connectivity is a property (corresponding to all connected graphs on V), and planarity is another property. A property Q is *monotonically increasing* if whenever G has Q and H is obtained from G by adding edges, then H has Q too. A *monotonically decreasing* property is defined in a similar manner. By interpreting the members of N in Theorem 6.3.2 as the $\binom{m}{2}$ pairs $\{i, j\}$ with $i, j \in V$, $i \neq j$, we obtain the following:

Theorem 6.3.3 *Let Q_1, Q_2, Q_3, and Q_4 be graph properties, where Q_1, Q_2 are monotonically increasing and Q_3, Q_4 are monotonically decreasing. Let $G = (V, E)$ be a random graph on V obtained by picking every edge, independently, with probability p.*

Then

$$\Pr[G \in Q_1 \cap Q_2] \geq \Pr[G \in Q_1] \cdot \Pr[G \in Q_2],$$
$$\Pr[G \in Q_3 \cap Q_4] \geq \Pr[G \in Q_3] \cdot \Pr[G \in Q_4],$$
$$\Pr[G \in Q_1 \cap Q_3] \leq \Pr[G \in Q_1] \cdot \Pr[G \in Q_3].$$

Thus, for example, the probability that G is both Hamiltonian and planar does not exceed the product of the probability that it is Hamiltonian by that it is planar. It seems hopeless to try and prove such a statement directly, without using one of the correlation inequalities.

6.4 LINEAR EXTENSIONS OF PARTIALLY ORDERED SETS

Let $(P, \leq)$ be a partially ordered set with n elements. A *linear extension* of P is a one-to-one mapping $\sigma : P \to \{1, 2, \ldots, n\}$, which is order-preserving; that is, if $x, y \in P$ and $x \leq y$, then $\sigma(x) \leq \sigma(y)$. Intuitively, σ is a ranking of the elements of P that preserves the partial order of P. Consider the probability space of all linear extensions of P, where each possible extension is equally likely. In this space, we can consider events of the form, for example, $x \leq y$ or $(x \leq y) \wedge (x \leq z)$ (for $x, y, z \in P$) and compute their probabilities. It turns out that the FKG inequality is a very useful tool for studying the correlation between such events. The best known result of this form was conjectured by Rival and Sands and proved by Shepp (1982). (See also Fishburn (1992) for a strengthening.) It asserts that for any partially ordered set P and any three elements $x, y, z \in P$: $\Pr[x \leq y \wedge x \leq z] \geq \Pr[x \leq y] \Pr[x \leq z]$.

This result became known as the *XYZ* theorem. Although it looks intuitively obvious, its proof is nontrivial and contains a clever application of the FKG inequality. In this section we present this result and its elegant proof.

Theorem 6.4.1 *Let P be a partially ordered set with n elements $a_1, a_2, \ldots, a_n$. Then*

$$\Pr[a_1 \leq a_2 \wedge a_1 \leq a_3] \geq \Pr[a_1 \leq a_2] \Pr[a_1 \leq a_3].$$

Proof. Let m be a large integer (which will later tend to infinity), and let L be the set of all ordered n-tuples $\mathbf{x} = (x_1, \ldots, x_n)$, where $x_i \in M = \{1, 2, \ldots, m\}$. (Note that we do *not* assume that the numbers x_i are distinct.) Define an order relation $\leq$ on L as follows. For $\mathbf{y} = (y_1, \ldots, y_n) \in L$ and $\mathbf{x}$ as above, $\mathbf{x} \leq \mathbf{y}$ iff $x_1 \geq y_1$ and $x_i - x_1 \leq y_i - y_1$ for all $2 \leq i \leq n$. It is not too difficult to check that $(L, \leq)$ is a lattice in which the ith component of the meet $\mathbf{x} \wedge \mathbf{y}$ is $(\mathbf{x} \wedge \mathbf{y})_i = \min(x_i - x_1, y_i - y_1) + \max(x_1, y_1)$ and the ith component of the join $\mathbf{x} \vee \mathbf{y}$ is $(\mathbf{x} \vee \mathbf{y})_i = \max(x_i - x_1, y_i - y_1) + \min(x_1, y_1)$.

Moreover, the lattice L is distributive. This follows by an easy computation from the fact that the trivial lattice of integers (with respect to the usual order) is distributive and hence, for any three integers a, b, and c,

$$\min(a, \max(b, c)) = \max(\min(a, b), \min(a, c)), \tag{6.7}$$

and

$$\max(a, \min(b, c)) = \min(\max(a, b), \max(a, c)). \tag{6.8}$$

Let us show how this implies that L is distributive. Let $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$, and $\mathbf{z} = (z_1, \dots, z_n)$ be three elements of L. We must show that

$$\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z}) = (\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z}).$$

The ith component of $\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z})$ is

$$\begin{aligned}(\mathbf{x} \wedge (\mathbf{y} \vee \mathbf{z}))_i &= \min(x_i - x_1, (\mathbf{y} \vee \mathbf{z})_i - (\mathbf{y} \vee \mathbf{z})_1) \\ &\quad + \max(x_1, (\mathbf{y} \vee \mathbf{z})_1) \\ &= \min(x_i - x_1, \max(y_i - y_1, z_i - z_1)) \\ &\quad + \max(x_1, \min(y_1, z_1)).\end{aligned}$$

Similarly, the ith component of $(\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z})$ is

$$\begin{aligned}((\mathbf{x} \wedge \mathbf{y}) \vee (\mathbf{x} \wedge \mathbf{z}))_i &= \max((\mathbf{x} \wedge \mathbf{y})_i - (\mathbf{x} \wedge \mathbf{y})_1, (\mathbf{x} \wedge \mathbf{z})_i - (\mathbf{x} \wedge \mathbf{z})_1) \\ &\quad + \min((\mathbf{x} \wedge \mathbf{y})_1, (\mathbf{x} \wedge \mathbf{z})_1) \\ &= \max(\min(x_i - x_1, y_i - y_1), \min(x_i - x_1, z_i - z_1)) \\ &\quad + \min(\max(x_1, y_1), \max(x_1, z_1)).\end{aligned}$$

These two quantities are equal, as follows by applying (6.7) with $a = x_i - x_1$, $b = y_i - y_1$, $c = z_i - z_1$, and (6.8) with $a = x_1$, $b = y_1$, $c = z_1$.

Thus L is distributive. To apply the FKG inequality we need the measure function μ and the two functions f and g. Let μ be the characteristic function of P; that is, for $\mathbf{x} = (x_1, \dots, x_n) \in L$, $\mu(\mathbf{x}) = 1$ if $x_i \le x_j$ whenever $a_i \le a_j$ in P and $\mu(\mathbf{x}) = 0$ otherwise. To show that μ is log-supermodular, it suffices to check that if $\mu(\mathbf{x}) = \mu(\mathbf{y}) = 1$, then $\mu(\mathbf{x} \vee \mathbf{y}) = \mu(\mathbf{x} \wedge \mathbf{y}) = 1$. However, if $\mu(\mathbf{x}) = \mu(\mathbf{y}) = 1$ and $a_i \le a_j$ in P, then $x_i \le x_j$ and $y_i \le y_j$ and hence

$$\begin{aligned}(x \vee y)_i &= \max(x_i - x_1, y_i - y_1) + \min(x_1, y_1) \\ &\le \max(x_j - x_1, y_j - y_1) + \min(x_1, y_1) = (x \vee y)_j,\end{aligned}$$

that is, $\mu(\mathbf{x} \vee \mathbf{y}) = 1$. Similarly, $\mu(\mathbf{x}) = \mu(\mathbf{y}) = 1$ implies $\mu(\mathbf{x} \wedge \mathbf{y}) = 1$ too.

Not surprisingly, we define the functions f and g as the characteristic functions of the two events $x_1 \le x_2$ and $x_1 \le x_3$, respectively; that is, $f(\mathbf{x}) = 1$ if $x_1 \le x_2$ and $f(\mathbf{x}) = 0$ otherwise, and $g(\mathbf{x}) = 1$ if $x_1 \le x_3$ and $g(\mathbf{x}) = 0$ otherwise. Trivially, both f and g are increasing. Indeed, if $\mathbf{x} \le \mathbf{y}$ and $f(\mathbf{x}) = 1$, then $0 \le x_2 - x_1 \le y_2 - y_1$ and hence $f(\mathbf{y}) = 1$, and similarly for g.

We, therefore, have all the necessary ingredients for applying the FKG inequality (Theorem 6.2.1). This gives that in L the probability that an n-tuple $(x_1, \dots, x_n)$ that

satisfies the inequalities in P satisfies both $x_1 \le x_2$ and $x_1 \le x_3$ is at least as big as the product of the probability that it satisfies $x_1 \le x_2$ by that it satisfies $x_1 \le x_3$. Note that this is not yet what we wanted to prove; the n-tuples in L are not n-tuples of distinct integers and thus do not correspond to linear extensions of P. However, as $m \to \infty$, the probability that $x_i = x_j$ for some $i \neq j$ in a member $\mathbf{x} = (x_1, \ldots, x_n)$ of L tends to 0, and the assertion of the theorem follows. ■

6.5 EXERCISES

1. Let G be a graph and let P denote the probability that a random subgraph of G obtained by picking each edge of G with probability $1/2$, independently, is connected (and spanning). Let Q denote the probability that in a random two-coloring of G, where each edge is chosen randomly and independently to be either red or blue, the red graph and the blue graph are both connected (and spanning). Is $Q \le P^2$?
2. A family of subsets $\mathcal{G}$ is called *intersecting* if $G_1 \cap G_2 \neq \emptyset$ for all $G_1, G_2 \in \mathcal{G}$. Let $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_k$ be k intersecting families of subsets of $\{1, 2, \ldots, n\}$. Prove that

$$\left| \bigcup_{i=1}^{k} \mathcal{F}_i \right| \le 2^n - 2^{n-k}.$$

3. Show that the probability that in the random graph $G(2k, 1/2)$ the maximum degree is at most $k - 1$ is at least $1/4^k$.

THE PROBABILISTIC LENS:
Turán's Theorem

In a graph $G = (V, E)$, let d_v denote the degree of a vertex v and let $\alpha(G)$ be the maximal size of an independent set of vertices. The following result was proved by Caro and Wei.

Theorem 1 $\alpha(G) \geq \sum_{v \in V} \frac{1}{d_v + 1}$.

Proof. Let $<$ be a uniformly chosen total ordering of V. Define

$$I = \{v \in V : \{v, w\} \in E \Rightarrow v < w\}.$$

Let X_v be the indicator random variable for $v \in I$ and $X = \sum_{v \in V} X_v = |I|$. For each v,

$$\mathrm{E}[X_v] = \Pr[v \in I] = \frac{1}{d_v + 1},$$

since $v \in I$ if and only if v is the least element among v and its neighbors. Hence

$$\mathrm{E}[X] = \sum_{v \in V} \frac{1}{d_v + 1}$$

and so there exists a specific ordering $<$ with

$$|I| \geq \sum_{v \in V} \frac{1}{d_v + 1}.$$

But if $x, y \in I$ and $\{x, y\} \in E$, then $x < y$ *and* $y < x$, a contradiction. Thus I is independent and $\alpha(G) \geq |I|$. ■

For any $m \le n$, let q, r satisfy $n = mq + r, 0 \le r < m$, and let $e = r\binom{q+1}{2} + (m-r)\binom{q}{2}$. Define a graph $G = G_{n,e}$ on n vertices and e edges by splitting the vertex set into m classes as evenly as possible and joining two vertices if and only if they lie in the same class. Clearly, $\alpha(G_{n,e}) = m$.

Theorem 2 Turán (1941) *Let H have n vertices and e edges. Then $\alpha(H) \ge m$ and $\alpha(H) = m \Longleftrightarrow H \cong G_{n,e}$.*

Proof. $G_{n,e}$ has $\sum_{v\in V}(d_v + 1)^{-1} = m$ since each clique contributes 1 to the sum. Fixing $e = \sum_{v\in V} d_v/2$, $\sum_{v\in V}(d_v + 1)^{-1}$ is minimized with the d_v as close together as possible. Thus for any H,

$$\alpha(H) \ge \sum_{v\in V} \frac{1}{d_v + 1} \ge m.$$

For $\alpha(H) = m$, we must have equality on both sides above. The second equality implies that d_v must be as close together as possible. Letting $X = |I|$ as in the previous theorem, assume $\alpha(H) = \mathrm{E}[X]$. But $\alpha(H) \ge X$ for all values of $<$, so X must be a constant. Suppose H is not a union of cliques. Then, there exist $x, y, z \in V$ with $\{x, y\}, \{x, z\} \in E$, $\{y, z\} \notin E$. Let $<$ be an ordering that begins x, y, z and $<'$ the same ordering except that it begins y, z, x, and let I, I' be the corresponding sets of vertices all of whose neighbors are "greater." Then I, I' are identical except that $x \in I, y, z \notin I$, whereas $x \notin I', y, z \in I'$. Thus X is not constant. That is, $\alpha(H) = \mathrm{E}[X]$ implies that H is the union of cliques and so $H \cong G_{n,e}$. ■

7

Martingales and Tight Concentration

Mathematics seems much more real to me than business—in the sense that, well, what's the reality in a McDonald's stand? It's here today and gone tomorrow. Now, the integers—that's reality. When you prove a theorem, you've really done something that has substance to it, to which no business venture can compare for reality.
–Jim Simons

7.1 DEFINITIONS

A *martingale* is a sequence $X_0, \ldots, X_m$ of random variables so that for $0 \leq i < m$,

$$\mathrm{E}[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = X_i.$$

Imagine a gambler walking into a casino with X_0 dollars. The casino contains a variety of games of chance. All games are "fair" in that their expectations are zero. The gambler may allow previous history to determine his choice of game and bet. He might employ the gambler's definition of martingale—double the bet until you win. He might play roulette until he wins three times, and then switch to keno. Let X_i be the gambler's fortune at time i. Given that $X_i = a$, the conditional expectation of X_{i+1} must be a and so this is a martingale.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

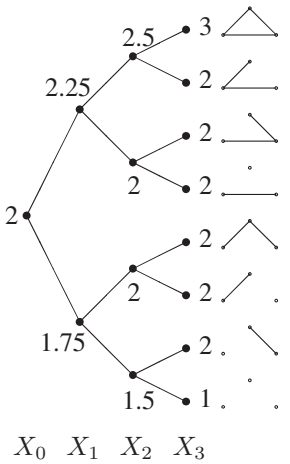
A simple but instructive martingale occurs when the gambler plays "flip a coin" for stakes of one dollar each time. Let $Y_1, \dots, Y_m$ be independent coin flips, each $+1$ or -1 with probability $1/2$. Normalize so that $X_0 = 0$ is the gambler's initial stake, though he has unlimited credit. Then $X_i = Y_1 + \cdots + Y_i$ has distribution S_i.

Our martingales will look quite different, at least from the outside.

The Edge Exposure Martingale. Let the random graph $G(n,p)$ be the underlying probability space. Label the potential edges $\{i,j\} \subseteq [n]$ by $e_1, \dots, e_m$, setting $m = \binom{n}{2}$ for convenience, in any specific manner. Let f be any graph theoretic function. We define a martingale $X_0, \dots, X_m$ by giving the values $X_i(H)$, where H is sampled from $G(n,p)$. $X_m(H)$ is simply $f(H)$. $X_0(H)$ is the expected value of $f(G)$ with $G \sim G(n,p)$. Note that X_0 is a constant. In general (including the cases $i = 0$ and $i = m$)

$$X_i(H) = \mathrm{E}[f(G) \mid e_j \in G \Longleftrightarrow e_j \in H, 1 \le j \le i].$$

In words, to find $X_i(H)$ we first expose the first i pairs $e_1, \dots, e_i$ and see if they are in H. The remaining edges are not seen and considered to be random. $X_i(H)$ is then the conditional expectation of $f(G)$ with this partial information. When $i = 0$, nothing is exposed and X_0 is a constant. When $i = m$, all is exposed and X_m is the function f. The martingale moves from no information to full information in small steps.



The edge exposure martingale with $n = m = 3$, f is the chromatic number and the edges exposed in the order "bottom, left, right." The values $X_i(H)$ are given by tracing from the central node to the leaf labeled H.

The figure shows why this is a martingale. The conditional expectation of $f(H)$ knowing the first $i-1$ edges is the weighted average of the conditional expectations of $f(H)$, where the ith edge has been exposed. More generally—in what is sometimes referred to as a Doob martingale process—X_i may be the conditional expectation of $f(H)$ after certain information is revealed as long as the information known at time i includes the information known at time $i-1$.

The Vertex Exposure Martingale. Again, let $G(n,p)$ be the underlying probability space and f any graph theoretic function. Define $X_1, \ldots, X_n$ by

$$X_i(H) = \mathrm{E}[f(G) \mid \text{ for } x,y \le i, \{x,y\} \in G \Longleftrightarrow \{x,y\} \in H].$$

In words, to find $X_i(H)$ we expose the first i vertices and all their internal edges, and take the conditional expectation of $f(G)$ with that partial information. By ordering the edges appropriately, the vertex exposure martingale may be considered a subsequence of the edge exposure martingale. Note that $X_1(H) = \mathrm{E}[f(G)]$ is constant as no edges have been exposed, and $X_n(H) = f(H)$ as all edges have been exposed.

7.2 LARGE DEVIATIONS

Maurey (1979) applied a large deviation inequality for martingales to prove an isoperimetric inequality for the symmetric group S_n. This inequality was useful in the study of normed spaces; see Milman and Schechtman (1986) for many related results. The applications of martingales in graph theory also all involve the same underlying martingale result used by Maurey, which is the following:

Theorem 7.2.1 [Azuma's Inequality] *Let $0 = X_0, \ldots, X_m$ be a martingale with*

$$|X_{i+1} - X_i| \le 1$$

for all $0 \le i < m$. Let $\lambda > 0$ be arbitrary. Then

$$\Pr\left[X_m > \lambda\sqrt{m}\right] < e^{-\lambda^2/2}.$$

In the "flip a coin" martingale, X_m has distribution S_m and this result is Theorem A.1.1. Indeed, the general proof is quite similar.

Proof. Set, with foresight, $\alpha = \lambda/\sqrt{m}$. Set $Y_i = X_i - X_{i-1}$ so that $|Y_i| \le 1$ and $\mathrm{E}[Y_i \mid X_{i-1}, X_{i-2}, \ldots, X_0] = 0$. Then, as in Theorem A.1.16,

$$\mathrm{E}[e^{\alpha Y_i} \mid X_{i-1}, X_{i-2}, \ldots, X_0] \le \cosh(\alpha) \le e^{\alpha^2/2}.$$

Hence

$$\begin{aligned}
\mathrm{E}[e^{\alpha X_m}] &= \mathrm{E}\left[\prod_{i=1}^{m} e^{\alpha Y_i}\right] \\
&= \mathrm{E}\left[\left(\prod_{i=1}^{m-1} e^{\alpha Y_i}\right)\mathrm{E}[e^{\alpha Y_m} \mid X_{m-1}, X_{m-2}, \ldots, X_0]\right] \\
&\le \mathrm{E}\left[\prod_{i=1}^{m-1} e^{\alpha Y_i}\right] e^{\alpha^2/2} \le e^{\alpha^2 m/2}.
\end{aligned}$$

Therefore

$$\begin{aligned}\Pr\left[X_m > \lambda\sqrt{m}\right] &= \Pr\left[e^{\alpha X_m} > e^{\alpha\lambda\sqrt{m}}\right] \\ &< \mathrm{E}\left[e^{\alpha X_m}\right] e^{-\alpha\lambda\sqrt{m}} \\ &\le e^{\alpha^2 m/2 - \alpha\lambda\sqrt{m}} = e^{-\lambda^2/2},\end{aligned}$$

as needed. ■

Corollary 7.2.2 *Let $c = X_0, \ldots, X_m$ be a martingale with*

$$|X_{i+1} - X_i| \le 1$$

for all $0 \le i < m$. Then

$$\Pr\left[|X_m - c| > \lambda\sqrt{m}\right] < 2e^{-\lambda^2/2}.$$

A graph theoretic function f is said to satisfy the edge Lipschitz condition if, whenever H and H' differ in only one edge, $|f(H) - f(H')| \le 1$. It satisfies the vertex Lipschitz condition if, whenever H and H' differ at only one vertex, $|f(H) - f(H')| \le 1$.

Theorem 7.2.3 *When f satisfies the edge Lipschitz condition, the corresponding edge exposure martingale satisfies $|X_{i+1} - X_i| \le 1$. When f satisfies the vertex Lipschitz condition, the corresponding vertex exposure martingale satisfies $|X_{i+1} - X_i| \le 1$.*

We prove these results in a more general context later. They have the intuitive sense that, if knowledge of a particular vertex or edge cannot change f by more than 1, then exposing a vertex or edge should not change the expectation of f by more than 1. Now we give a simple application of these results.

Theorem 7.2.4 Shamir and Spencer (1987) *Let n, p be arbitrary, and let $c = \mathrm{E}[\chi(G)]$, where $G \sim G(n,p)$. Then*

$$\Pr\left[|\chi(G) - c| > \lambda\sqrt{n-1}\right] < 2e^{-\lambda^2/2}.$$

Proof. Consider the vertex exposure martingale $X_1, \ldots, X_n$ on $G(n,p)$ with $f(G) = \chi(G)$. A single vertex can always be given a new color, so the vertex Lipschitz condition applies. Now apply Azuma's Inequality in the form of Corollary 7.2.2. ■

Letting $\lambda \to \infty$ arbitrarily slowly, this result shows that the distribution of $\chi(G)$ is "tightly concentrated" around its mean. The proof gives no clue as to where the mean is.

7.3 CHROMATIC NUMBER

In Theorem 10.3.1, we prove that $\chi(G) \sim n/2\log_2 n$ almost surely, where $G \sim G(n, \frac{1}{2})$. Here we give the original proof of Béla Bollobás using martingales. We follow the notations of Section 10.3, setting $f(k) = \binom{n}{k} 2^{-\binom{k}{2}}$, k_0 so that $f(k_0 - 1) > 1 > f(k_0)$, $k = k_0 - 4$ so that $k \sim 2\log_2 n$, and $f(k) > n^{3+o(1)}$. Our goal is to show

$$\Pr[\omega(G) < k] = e^{-n^{2+o(1)}},$$

where $\omega(G)$ is the size of the maximum clique of G. We shall actually show in Theorem 7.3.2 a more precise bound. The remainder of the argument is given in Section 10.3.

Let $Y = Y(H)$ be the maximal size of a family of edge disjoint cliques of size k in H. This ingenious and unusual choice of function is key to the martingale proof.

Lemma 7.3.1 $\mathrm{E}[Y] \geq (1 + o(1))(n^2/2k^4)$.

Proof. Let $\mathcal{K}$ denote the family of k-cliques of G so that $f(k) = \mu = \mathrm{E}[|\mathcal{K}|]$. Let W denote the number of unordered pairs $\{A, B\}$ of k-cliques of G with $2 \leq |A \cap B| < k$. Then, $\mathrm{E}[W] = \Delta/2$, with Δ as described in Section 10.3 (see also Section 4.5), $\Delta \sim \mu^2 k^4 n^{-2}$. Let $\mathcal{C}$ be a random subfamily of $\mathcal{K}$ defined by setting, for each $A \in \mathcal{K}$, $\Pr[A \in \mathcal{C}] = q$, q to be determined. Let W' be the number of unordered pairs $\{A, B\}$, $A, B \in \mathcal{C}$ with $2 \leq |A \cap B| < k$. Then

$$\mathrm{E}[W'] = \mathrm{E}[W]q^2 = \Delta q^2/2.$$

Delete from $\mathcal{C}$ one set from each such pair $\{A, B\}$. This yields a set $\mathcal{C}^*$ of edge disjoint k-cliques of G and

$$\mathrm{E}[Y] \geq \mathrm{E}[|\mathcal{C}^*|] \geq \mathrm{E}[|\mathcal{C}|] - \mathrm{E}[W'] = \mu q - \Delta q^2/2 = \mu^2/2\Delta \sim n^2/2k^4,$$

where we choose $q = \mu/\Delta$ (< 1) to minimize the quadratic. ∎

We conjecture that Lemma 7.3.1 may be improved to $\mathrm{E}[Y] > cn^2/k^2$. That is, with positive probability there is a family of k-cliques that are edge disjoint and cover a positive proportion of the edges.

Theorem 7.3.2

$$\Pr[\omega(G) < k] < e^{-(c+o(1))(n^2/\ln^8 n)}$$

with c a positive constant.

Proof. Let $Y_0, \ldots, Y_m$, $m = \binom{n}{2}$, be the edge exposure martingale on $G(n, \frac{1}{2})$ with the function Y just defined. The function Y satisfies the edge Lipschitz condition, as

adding a single edge can only add at most one clique to a family of edge disjoint cliques. (Note that the Lipschitz condition would not be satisfied for the number of k-cliques, as a single edge might yield many new cliques.) G has no k-clique if and only if $Y = 0$. Apply Azuma's Inequality with $m = \binom{n}{2} \sim n^2/2$ and $\mathrm{E}[Y] \geq (1 + o(1))(n^2/2k^4)$. Then

$$\begin{aligned}
\Pr[\omega(G) < k] &= \Pr[Y = 0] \\
&\leq \Pr[Y - \mathrm{E}[Y] \leq -\mathrm{E}[Y]] \\
&\leq e^{-\mathrm{E}[Y]^2/2\binom{n}{2}} \\
&\leq e^{-(c'+o(1))n^2/k^8} \\
&= e^{-(c+o(1))n^2/\ln^8 n},
\end{aligned}$$

as desired. ∎

Here is another example where the martingale approach requires an inventive choice of graph theoretic function.

Theorem 7.3.3 *Let $p = n^{-\alpha}$, where $\alpha > \frac{5}{6}$ is fixed. Let $G = G(n,p)$. Then there exists $u = u(n,p)$ so that almost always*

$$u \leq \chi(G) \leq u + 3.$$

That is, $\chi(G)$ is concentrated in four values.

We first require a technical lemma that has been well known.

Lemma 7.3.4 *Let α, c be fixed, $\alpha > \frac{5}{6}$. Let $p = n^{-\alpha}$. Then almost always every $c\sqrt{n}$ vertices of $G = G(n,p)$ may be three-colored.*

Proof. If not, let T be a minimal set that is not three-colorable. As $T - \{x\}$ is three-colorable, x must have internal degree at least 3 in T for all $x \in T$. Thus, if T has t vertices, it must have at least $3t/2$ edges. The probability of this occurring for some T with at most $c\sqrt{n}$ vertices is bounded from above by

$$\sum_{t=4}^{c\sqrt{n}} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} p^{3t/2}.$$

We bound

$$\binom{n}{t} \leq \left(\frac{ne}{t}\right)^t \quad \text{and} \quad \binom{\binom{t}{2}}{3t/2} \leq \left(\frac{te}{3}\right)^{3t/2},$$

so each term is at most

$$\left(\frac{ne}{t}\frac{t^{3/2}e^{3/2}}{3^{3/2}}n^{-3\alpha/2}\right)^t \le \left(c_1 n^{1-3\alpha/2}t^{1/2}\right)^t \le \left(c_2 n^{1-3\alpha/2}n^{1/4}\right)^t = (c_2 n^{-\epsilon})^t$$

with $\epsilon = \frac{3}{2}\alpha - \frac{5}{4} > 0$ and the sum is therefore $o(1)$. ■

Proof. [Theorem 7.3.3] Let $\epsilon > 0$ be arbitrarily small and let $u = u(n, p, \epsilon)$ be the least integer so that

$$\Pr[\chi(G) \le u] > \epsilon\,.$$

Now define $Y(G)$ to be the minimal size of a set of vertices S for which $G - S$ may be u-colored. This Y satisfies the vertex Lipschitz condition since at worst one could add a vertex to S. Apply the vertex exposure martingale on $G(n, p)$ to Y. Letting $\mu = \mathrm{E}[Y]$

$$\Pr[Y \le \mu - \lambda\sqrt{n-1}] < e^{-\lambda^2/2},$$
$$\Pr[Y \ge \mu + \lambda\sqrt{n-1}] < e^{-\lambda^2/2}.$$

Let λ satisfy $e^{-\lambda^2/2} = \epsilon$ so that these tail events each have probability less than ϵ. We defined u so that, with probability at least ϵ, G would be u-colorable and hence $Y = 0$. That is, $\Pr[Y = 0] > \epsilon$. The first inequality therefore forces $\mu \le \lambda\sqrt{n-1}$. Now employing the second inequality,

$$\Pr\left[Y \ge 2\lambda\sqrt{n-1}\right] \le \Pr\left[Y \ge \mu + \lambda\sqrt{n-1}\right] \le \epsilon\,.$$

With probability at least $1 - \epsilon$ there is a u-coloring of all but at most $c'\sqrt{n}$ vertices. By the lemma, almost always, and so with probability at least $1 - \epsilon$, these points may be colored with three further colors, giving a $(u + 3)$-coloring of G. The minimality of u guarantees that with probability at least $1 - \epsilon$ at least u colors are needed for G. Altogether

$$\Pr[u \le \chi(G) \le u + 3] \ge 1 - 3\epsilon$$

and ϵ was arbitrarily small. ■

Using the same technique, similar results can be achieved for other values of α. Together with some related ideas, it can be shown that, for any fixed $\alpha > \frac{1}{2}$, $\chi(G)$ is concentrated on at most two values. See Łuczak (1991) and Alon and Krivelevich (1997) for the detailed proofs.

7.4 TWO GENERAL SETTINGS

The martingales, useful in studying random graphs generally, can be placed in the following general setting, which is essentially the one considered in Maurey (1979) and in Milman and Schechtman (1986): Let $\Omega = A^B$ denote the set of functions

$g : B \to A$. (With B the set of pairs of vertices on n vertices and $A = \{0, 1\}$, we may identify $g \in A^B$ with a graph on n vertices.) We define a measure by giving values p_{ab} and setting

$$\Pr[g(b) = a] = p_{ab}$$

with the values $g(b)$ assumed mutually independent. [In $G(n,p)$ all $p_{1b} = p, p_{0b} = 1 - p$.] Now fix a gradation

$$\emptyset = B_0 \subset B_1 \subset \cdots \subset B_m = B\,.$$

Let $L : A^B \to R$ be a functional (e.g., clique number). We define a martingale $X_0, X_1, \ldots, X_m$ by setting

$$X_i(h) = \mathrm{E}[L(g) \mid g(b) = h(b) \text{ for all } b \in B_i]\,.$$

X_0 is a constant, the expected value of L of the random g. X_m is L itself. The values $X_i(g)$ approach $L(g)$ as the values of $g(b)$ are "exposed." We say the functional L satisfies the Lipschitz condition relative to the gradation, if for all $0 \le i < m$,

$$h, h' \text{ differ only on } B_{i+1} - B_i \qquad \Rightarrow \qquad |L(h') - L(h)| \le 1\,.$$

Theorem 7.4.1 *Let L satisfy the Lipschitz condition. Then the corresponding martingale satisfies*

$$|X_{i+1}(h) - X_i(h)| \le 1$$

for all $0 \le i < m$, $h \in A^B$.

Proof. Let H be the family of h' that agree with h on B_{i+1}. Then

$$X_{i+1}(h) = \sum_{h' \in H} L(h') w_{h'}\,,$$

where $w_{h'}$ is the conditional probability that $g = h'$, given that $g = h$ on B_{i+1}. For each $h' \in H$, let $H[h']$ denote the family of h^* that agree with h' on all points except (possibly) $B_{i+1} - B_i$. The $H[h']$ partition the family of h^*, agreeing with h on B_i. Thus we may express

$$X_i(h) = \sum_{h' \in H} \sum_{h^* \in H[h']} [L(h^*) q_{h^*}] w_{h'}\,,$$

where q_{h^*} is the conditional probability that g agrees with h^* on B_{i+1} given that it agrees with h on B_i. (This is because for $h^* \in H[h']$, $w_{h'}$ is also the conditional probability that $g = h^*$, given that $g = h^*$ on B_{i+1}.) Thus

$$\begin{aligned} |X_{i+1}(h) - X_i(h)| &= \left| \sum_{h' \in H} w_{h'} \left[L(h') - \sum_{h^* \in H[h']} L(h^*) q_{h^*} \right] \right| \\ &\le \sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} |q_{h^*} [L(h') - L(h^*)]|\,. \end{aligned}$$

The Lipschitz condition gives $|L(h') - L(h^*)| \leq 1$, so

$$|X_{i+1}(h) - X_i(h)| \leq \sum_{h' \in H} w_{h'} \sum_{h^* \in H[h']} q_{h^*} = \sum_{h' \in H} w_{h'} = 1 . \qquad \blacksquare$$

Now we can express Azuma's Inequality in a general form.

Theorem 7.4.2 *Let L satisfy the Lipschitz condition relative to a gradation of length m and let $\mu = \mathrm{E}[L(g)]$. Then for all $\lambda > 0$,*

$$\Pr\left[L(g) \geq \mu + \lambda\sqrt{m}\right] < e^{-\lambda^2/2},$$
$$\Pr\left[L(g) \leq \mu - \lambda\sqrt{m}\right] < e^{-\lambda^2/2}.$$

The second general setting is taken from Alon, Kim, and Spencer (1997). We assume our underlying probability space is generated by a finite set of mutually independent Yes/No choices, indexed by $i \in I$. We are given a random variable Y on this space. Let p_i denote the probability that choice i is Yes. Let c_i be such that changing choice i (keeping all else the same) can change Y by at most c_i. We call c_i the *effect* of i. Let C be an upper bound on all c_i. We call $p_i(1 - p_i)c_i^2$ the *variance* of choice i.

Now consider a solitaire game in which Paul finds the value of Y by making queries of an always truthful oracle Carole. The queries are always of a choice $i \in I$. Paul's choice of query can depend on Carole's previous responses. A strategy for Paul can then naturally be represented in a decision tree form. A "line of questioning" is a path from the root to a leaf of this tree, a sequence of questions and responses that determine Y. The total variance of a line of questioning is the sum of the variances of the queries in it.

Theorem 7.4.3 *For all $\epsilon > 0$, there exists $\delta > 0$ so that the following holds: Suppose Paul has a strategy for finding Y such that every line of questioning has total variance at most σ^2. Then*

$$\Pr[|Y - \mathrm{E}[Y]| > \alpha\sigma] \leq 2e^{-\alpha^2/2(1+\epsilon)} \tag{7.1}$$

for all positive α with $\alpha C < \sigma(1 + \epsilon)\delta$.

Applications. For a specific suboptimal bound we may take $\epsilon = \delta = 1$. If $C = O(1)$, $\alpha \to \infty$, and $\alpha = o(\sigma)$, the upper bound of (7.1) is $\exp[-\Omega(\alpha^2)]$. In many cases, Paul queries *all* $i \in I$. Then we may take σ with $\sigma^2 = \sum_{i \in I} p_i(1 - p_i)c_i^2$. For example, consider an edge Lipschitz Y on $G(n, p)$ with $p = p(n) \to 0$. I is the set of $m = \binom{n}{2}$ potential edges, all $p_i = p$, $C = 1$, so that $\sigma = \Theta(\sqrt{n^2 p})$. If $\alpha \to \infty$ with $\alpha = o(\sqrt{n^2 p})$, the upper bound of (7.1) is again $\exp[-\Omega(\alpha^2)]$.

Proof. For simplicity we replace Y by $Y - \mathrm{E}[Y]$ so that we shall henceforth assume $\mathrm{E}[Y] = 0$. By symmetry, we shall bound only the upper tail of Y. We set, with foresight, $\lambda = \alpha/[\sigma(1+\epsilon)]$. Our side assumption gives that $C\lambda < \delta$. We will show

$$\mathrm{E}[e^{\lambda Y}] \le e^{(1+\epsilon)\lambda^2\sigma^2/2}. \tag{7.2}$$

The martingale inequality then follows by the Markov bound

$$\Pr[Y > \alpha\sigma] < e^{-\lambda\alpha\sigma}\mathrm{E}[e^{\lambda Y}] \le e^{-\alpha^2/2(1+\epsilon)}.$$

We first claim that, for all $\epsilon > 0$, there exists $\delta > 0$ so that for $0 \le p \le 1$ and $|a| \le \delta$

$$pe^{(1-p)a} + (1-p)e^{-pa} \le e^{(1+\epsilon)p(1-p)a^2/2}. \tag{7.3}$$

Take the Taylor series in a of the left-hand side. The constant term is 1, the linear term 0, the coefficient of a^2 is $\frac{1}{2}p(1-p)$, and, for $j \ge 3$, the coefficient of a^j is at most

$$\frac{1}{j!}p(1-p)(p^{j-1} + (1-p)^{j-1}) \le \frac{1}{j!}p(1-p).$$

Pick δ so that $|a| \le \delta$ implies

$$\sum_{j=3}^{\infty} \frac{a^j}{j!} < \epsilon a^2/2.$$

(In particular, this holds for $\epsilon = \delta = 1$.) Then

$$pe^{(1-p)a} + (1-p)e^{-pa} \le 1 + p(1-p)\frac{a^2}{2}(1+\epsilon)$$

and (7.3) follows from the inequality $1 + x \le e^x$.

Using this δ, we show (7.2) by induction on the depth M of the decision tree. For $M = 0$, Y is constant and (7.2) is immediate. Otherwise, let $p, c, v = p(1-p)c^2$ denote the probability, effect, and variance, respectively, of Paul's first query. Let μ_y, μ_n denote the conditional expectations of Y if Carole's response is Yes or No, respectively. Then $0 = \mathrm{E}[Y]$ can be split into

$$0 = p\mu_y + (1-p)\mu_n.$$

The difference $\mu_y - \mu_n$ is the expected *change* in Y when all other choices are made independent with their respective probabilities and the root choice is changed from Yes to No. As this always changes Y by at most c,

$$|\mu_y - \mu_n| \le c.$$

Thus we may parameterize

$$\mu_y = (1-p)b \quad \text{and} \quad \mu_n = -pb$$

with $|b| \le c$. From (7.3)

$$pe^{\lambda\mu_y} + (1-p)e^{\lambda\mu_n} \le e^{(1+\epsilon)p(1-p)b^2\lambda^2/2} \le e^{(1+\epsilon)v\lambda^2/2}.$$

Let A_y denote the expectation of $e^{\lambda(Y-\mu_y)}$, conditional on Carole's first response being Yes, and let A_n denote the analogous quantity for No. Given Carole's first response, Paul has a decision tree (one of the two main subtrees) that determines Y with total variation at most $\sigma^2 - v$ and the tree has depth at most $M-1$. So, by induction, $A_y, A_n \le A^-$, where we set

$$A^- = e^{(1+\epsilon)\lambda^2(\sigma^2-v)/2}.$$

Now we split

$$\begin{aligned} \mathrm{E}[e^{\lambda Y}] &= pe^{\lambda\mu_y}A_y + (1-p)e^{\lambda\mu_n}A_n \\ &\le [pe^{\lambda\mu_y} + (1-p)e^{\lambda\mu_n}]A^- \\ &\le e^{(1+\epsilon)\lambda^2(v+(\sigma^2-v))/2}, \end{aligned}$$

completing the proof of (7.2) and hence of Theorem 7.4.3. ■

We remark that this formal inductive proof somewhat masks the martingale. A martingale $\mathrm{E}[Y] = Y_0, \dots, Y_M = Y$ can be defined with Y_t, the conditional expectation of Y after the first t queries and responses. Theorem 7.4.3 can be thought of as bounding the tail of Y by that of a normal distribution of greater or equal variance. For very large distances from the mean, large α, this bound fails.

7.5 FOUR ILLUSTRATIONS

Let g be the random function from $\{1, \dots, n\}$ to itself, all n^n possible functions equally likely. Let $L(g)$ be the number of values not hit, that is, the number of y for which $g(x) = y$ has no solution. By Linearity of Expectation,

$$\mathrm{E}[L(g)] = n\left(1 - \frac{1}{n}\right)^n,$$

and this quantity is at most n/e and at least $n(1-1/n)^{n-1} \cdot (1-1/n) > (n-1)/e$.

Set $B_i = \{1, \dots, i\}$. L satisfies the Lipschitz condition relative to this gradation since changing the value of $g(i)$ can change $L(g)$ by at most 1. Thus we have the following:

Theorem 7.5.1 $\Pr\left[\left|L(g) - \frac{n}{e}\right| > \lambda\sqrt{n} + 1\right] < 2e^{-\lambda^2/2}$.

Deriving these asymptotic bounds from first principles is quite cumbersome.

As a second illustration, let B be any normed space, and let $v_1, \ldots, v_n \in B$ with all $|v_i| \leq 1$. Let $\epsilon_1, \ldots, \epsilon_n$ be independent with $\Pr[\epsilon_i = +1] = \Pr[\epsilon_i = -1] = 1/2$, and set

$$X = |\epsilon_1 v_1 + \cdots + \epsilon_n v_n| .$$

Theorem 7.5.2

$$\Pr\left[X - \mathrm{E}[X] > \lambda\sqrt{n}\right] < e^{-\lambda^2/2},$$
$$\Pr\left[X - \mathrm{E}[X] < -\lambda\sqrt{n}\right] < e^{-\lambda^2/2}.$$

Proof. Consider $\{-1, +1\}^n$ as the underlying probability space with all $(\epsilon_1, \ldots, \epsilon_n)$ equally likely. Then X is a random variable, and we define a martingale $X_0, \ldots, X_n = X$ by exposing one ϵ_i at a time. The value of ϵ_i can only change X by 2, so direct application of Theorem 7.4.1 gives $|X_{i+1} - X_i| \leq 2$. But let ϵ, ϵ' be two n-tuples differing only in the ith coordinate:

$$X_i(\epsilon) = (X_{i+1}(\epsilon) + X_{i+1}(\epsilon'))/2$$

so that

$$|X_i(\epsilon) - X_{i+1}(\epsilon)| = |X_{i+1}(\epsilon') - X_{i+1}(\epsilon)|/2 \leq 1 .$$

Now apply Azuma's Inequality. ∎

For a third illustration, let ρ be the Hamming metric on $\{0, 1\}^n$. For $A \subseteq \{0, 1\}^n$, let $B(A, s)$ denote the set of $y \in \{0, 1\}^n$ so that $\rho(x, y) \leq s$ for some $x \in A$. [$A \subseteq B(A, s)$ as we may take $x = y$.]

Theorem 7.5.3 *Let $\epsilon, \lambda > 0$ satisfy $e^{-\lambda^2/2} = \epsilon$. Then*

$$|A| \geq \epsilon 2^n \qquad \Rightarrow \qquad |B(A, 2\lambda\sqrt{n})| \geq (1 - \epsilon)2^n.$$

Proof. Consider $\{0, 1\}^n$ as the underlying probability space, all points equally likely. For $y \in \{0, 1\}^n$, set

$$X(y) = \min_{x \in A} \rho(x, y) .$$

Let $X_0, X_1, \ldots, X_n = X$ be the martingale given by exposing one coordinate of $\{0, 1\}^n$ at a time. The Lipschitz condition holds for X: If y, y' differ in just one coordinate, then $|X(y) - X(y')| \leq 1$. Thus, with $\mu = \mathrm{E}[X]$

$$\Pr\left[X < \mu - \lambda\sqrt{n}\right] < e^{-\lambda^2/2} = \epsilon ,$$
$$\Pr\left[X > \mu + \lambda\sqrt{n}\right] < e^{-\lambda^2/2} = \epsilon .$$

But

$$\Pr[X = 0] = |A|2^{-n} \geq \epsilon,$$

so $\mu \leq \lambda\sqrt{n}$. Thus

$$\Pr\left[X > 2\lambda\sqrt{n}\right] < \epsilon$$

and

$$|B(A, 2\lambda\sqrt{n})| = 2^n \Pr\left[X \leq 2\lambda\sqrt{n}\right] \geq 2^n(1-\epsilon). \qquad \blacksquare$$

Actually, a much stronger result is known. Let $B(s)$ denote a ball of radius s about $(0, \ldots, 0)$. The isoperimetric inequality proved by Harper (1966) states that

$$|A| \geq |B(r)| \qquad \Rightarrow \qquad |B(A, s)| \geq |B(r+s)|.$$

One may actually use this inequality as a beginning to give an alternate proof that $\chi(G) \sim n/2\log_2 n$ and to prove a number of the other results we have shown using martingales.

We illustrate Theorem 7.4.3 with a key technical lemma (in simplified form) from Alon, Kim, and Spencer (1997). Let $G = (V, E)$ be a graph on N vertices, each vertex having degree D. Asymptotics will be for $N, D \to \infty$. Set $p = 1/D$. Define a random subgraph $H \subseteq G$ by placing each edge $e \in E$ in H with independent probability p. Let M (for matching) be the set of isolated edges of H. Let V^* be those $v \in V$ *not* in any $\{v, w\} \in M$. For $v \in V$, set $\deg^*(v)$ equal to the number of $w \in V^*$ with $\{v, w\} \in E$. As

$$\Pr[w \notin V^*] = \sum_{\{v,w\} \in E} p(1-p)^{2D-1} = e^{-2} + O(D^{-1}),$$

Linearity of Expectation gives

$$\mathrm{E}[\deg^*(v)] = D(1 - e^{-2}) + O(1).$$

We want $\deg^*(v)$ tightly concentrated about its mean.

In the notation of Theorem 7.4.3, the probability space is determined by the choices $e \in H$ for all $e \in E$. All $p_i = p$. Changing $e \in H$ to $e \notin H$ can change $\deg^*(v)$ by at most $C = 4$.

Paul needs to find $\deg^*(v)$ by queries of the form "Is $e \in H$?" For each w with $\{v, w\} \in E$, he determines if $w \in V^*$ by the following line of inquiry: First, for all u with $\{w, u\} \in E$, he queries if $\{w, u\} \in H$. If no $\{w, u\} \in H$, then $w \in V^*$. If two (or more) $\{w, u_1\}, \{w, u_2\} \in H$, then w cannot be in an *isolated* edge of H, so $w \in V^*$. Now suppose $\{w, u\} \in H$ for precisely one u. Paul then asks (using his acquired knowledge!) for each $z \neq w$ with $\{u, z\} \in E$ if $\{u, z\} \in H$. The replies determine if $\{w, u\}$ is an isolated edge of H and hence if $w \in V^*$. Paul has made at most $D + (D-1)$ queries for each w for a total of at most $D(2D-1) = O(D^2)$ queries. We deduce

$$\Pr\left[|\deg^*(v) - D(1 - e^{-2})| > \lambda D^{1/2}\right] = \exp[-\Omega(\lambda^2)]$$

when $\lambda \to \infty$ and $\lambda = o(D^{1/2})$.

In application, one wishes to iterate this procedure (now applying it to the restriction of G to V^*) in order to find a large matching. This is somewhat akin to the Rödl nibble of Section 4.7. There are numerous further complications, but the tight concentration of $\deg^*(v)$ about its mean plays an indispensable role.

7.6 TALAGRAND'S INEQUALITY

Let $\Omega = \prod_{i=1}^{n} \Omega_i$, where each Ω_i is a probability space and Ω has the product measure. Let $A \subseteq \Omega$, and let $\vec{x} = (x_1, \ldots, x_n) \in \Omega$. Talagrand (1996) gives an unusual, subtle, and ultimately powerful notion of the distance—denoted $\rho(A, \vec{x})$—from $\vec{x}$ to A. We imagine moving from $\vec{x}$ to some $\vec{y} = (y_1, \ldots, y_n) \in A$ by changing coordinates. $\rho(A, \vec{x})$ will measure the minimal cost of such a move when a suitably restricted adversary sets the cost of each change.

Definition 1 *$\rho(A, \vec{x})$ is the least value such that, for any $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n) \in R^n$ with $|\vec{\alpha}| = 1$, there exists $\vec{y} = (y_1, \ldots, y_n) \in A$ with*

$$\sum_{x_i \neq y_i} \alpha_i \leq \rho(A, \vec{x}).$$

Note that $\vec{y}$ can, and generally will, depend on $\vec{\alpha}$.

We define for any real $t \geq 0$

$$A_t = \{\vec{x} \in \Omega : \rho(A, \vec{x}) \leq t\}.$$

Note that $A_0 = A$ as, when $\vec{x} \in A$, one can select $\vec{y} = \vec{x}$.

Talagrand's Inequality

$$\Pr[A](1 - \Pr[A_t]) \leq e^{-t^2/4}.$$

In particular, if $\Pr[A] \geq \frac{1}{2}$ (or any fixed constant) and t is "very large," then all but a very small proportion of Ω is within "distance" t of A.

Example. Take $\Omega = \{0, 1\}^n$ with uniform distribution, and let τ be the Hamming (L^1) metric. Then, $\rho(A, \vec{x}) \geq \min_{\vec{y} \in A} \tau(\vec{x}, \vec{y}) n^{-1/2}$, as the adversary can choose all $\alpha_i = n^{-1/2}$. Suppose to move from $\vec{x}$ to A, the values $x_1, \ldots, x_l$ (or any particular l coordinates) must be changed. Then $\rho(A, \vec{x}) \geq l^{1/2}$, as the adversary could choose $\alpha_i = l^{-1/2}$ for $1 \leq i \leq l$ and zero elsewhere.

Define $U(A, \vec{x})$ to be the set of $\vec{s} = (s_1, \ldots, s_n) \in \{0, 1\}^n$ with the property that there exists $\vec{y} \in A$ such that

$$x_i \neq y_i \Rightarrow s_i = 1.$$

We may think of $U(A, \vec{x})$ as representing the possible paths from $\vec{x}$ to A. Note that, when $s_i = 1$, we, for somewhat technical reasons, do not require $x_i \neq y_i$. With this

notation, $\rho(A,\vec{x})$ is the least real so that for all $\vec{\alpha}$ with $|\vec{\alpha}| = 1$ there exists $\vec{s} \in U(A,\vec{x})$ with $\vec{\alpha} \cdot \vec{s} \le \rho(A,\vec{x})$.

Now define $V(A,\vec{x})$ to be the convex hull of $U(A,\vec{x})$. The following result gives an alternate characterization of ρ that supplies the concept with much of its richness:

Theorem 7.6.1

$$\rho(A,\vec{x}) = \min_{\vec{v} \in V(A,\vec{x})} |\vec{v}|.$$

Proof. Let $\vec{v} \in V(A,\vec{x})$ achieve this minimum. The hyperplane through $\vec{v}$ perpendicular to the line from the origin to $\vec{v}$ then separates $V(A,\vec{x})$ from the origin so that all $\vec{s} \in V(A,\vec{x})$ have $\vec{s} \cdot \vec{v} \ge \vec{v} \cdot \vec{v}$. Set $\vec{\alpha} = \vec{v}/|\vec{v}|$. Then all $\vec{s} \in U(A,\vec{x}) \subseteq V(A,\vec{x})$ have $\vec{s} \cdot \vec{\alpha} \ge \vec{v} \cdot \vec{v}/|\vec{v}| = |\vec{v}|$. Conversely, take any $\vec{\alpha}$ with $|\vec{\alpha}| = 1$. Then $\vec{\alpha} \cdot \vec{v} \le |\vec{v}|$. As $\vec{v} \in V(A,\vec{x})$, we may write $\vec{v} = \sum \lambda_i \vec{s}_i$ for some $\vec{s}_i \in U(A,\vec{x})$, with all $\lambda_i \ge 0$ and $\sum \lambda_i = 1$. Then

$$|\vec{v}| \ge \sum \lambda_i (\vec{\alpha} \cdot \vec{s}_i)$$

and hence some $\vec{\alpha} \cdot \vec{s}_i \le |\vec{v}|$. ∎

The case $\Omega = \{0,1\}^n$ is particularly important and instructive. There, $\rho(A,\vec{x})$ is simply the Euclidean distance from $\vec{x}$ to the convex hull of A.

Theorem 7.6.2

$$\int_\Omega \exp\left[\frac{1}{4}\rho^2(A,\vec{x})\right] d\vec{x} \le \frac{1}{\Pr[A]}.$$

Talagrand's theorem is an immediate corollary of the above result. Indeed, fix A and consider the random variable $X = \rho(A,\vec{x})$. Then

$$\Pr[\overline{A_t}] = \Pr[X \ge t] = \Pr\left[e^{X^2/4} \ge e^{t^2/4}\right] \le \mathrm{E}[e^{X^2/4}]e^{-t^2/4}$$

and the theorem states $\mathrm{E}[e^{X^2/4}] \le 1/\Pr[A]$.

Proof. [Theorem 7.6.2]. We use induction on the dimension n. For $n = 1$, $\rho(A,\vec{x}) = 1$ if $\vec{x} \notin A$, and zero otherwise, so that

$$\int \exp\left[\frac{1}{4}\rho^2(A,\vec{x})\right] = \Pr[A] + (1 - \Pr[A])e^{1/4} \le \frac{1}{\Pr[A]},$$

as the inequality $u + (1-u)e^{1/4} \le u^{-1}$ for $0 < u \le 1$ is a simple calculus exercise.

Assume the result for n. Write $\mathrm{OLD} = \prod_{i=1}^n \Omega_i$, $\mathrm{NEW} = \Omega_{n+1}$, so that $\Omega = \mathrm{OLD} \times \mathrm{NEW}$ and any $z \in \Omega$ can be uniquely written $z = (x,\omega)$ with $x \in \mathrm{OLD}$, $\omega \in \mathrm{NEW}$. Set

$$B = \{x \in \mathrm{OLD} : (x,\omega) \in A \text{ for some } \omega \in \mathrm{NEW}\}$$

and for any $\omega \in$ NEW set

$$A_\omega = \{x \in \text{OLD} : (x, \omega) \in A\}.$$

Given $z = (x, \omega) \in \Omega$, we can move to A in two basic ways—either by changing ω, which reduces the problem to moving from x to B, or by not changing ω, which reduces the problem to moving from x to A_ω. Thus

$$\vec{s} \in U(B, x) \qquad \Rightarrow \qquad (\vec{s}, 1) \in U(A, (x, \omega))$$

and

$$\vec{t} \in U(A_\omega, x) \qquad \Rightarrow \qquad (\vec{t}, 0) \in U(A, (x, \omega)).$$

Taking the convex hulls, if $\vec{s} \in V(B, x)$ and $\vec{t} \in V(A_\omega, x)$, then $(\vec{s}, 1)$ and $(\vec{t}, 0)$ are in $V(A, (x, \omega))$, and hence for any $\lambda \in [0, 1]$

$$((1-\lambda)\vec{s} + \lambda\vec{t}, 1-\lambda) \in V(A, (x, \omega)).$$

Then, by convexity

$$\rho^2(A, (x, \omega)) \le (1-\lambda)^2 + |(1-\lambda)\vec{s} + \lambda\vec{t}|^2 \le (1-\lambda)^2 + (1-\lambda)|\vec{s}|^2 + \lambda|\vec{t}|^2.$$

Selecting $\vec{s}, \vec{t}$ with minimal norms yields the critical inequality

$$\rho^2(A, (x, \omega)) \le (1-\lambda)^2 + \lambda\rho^2(A_\omega, x) + (1-\lambda)\rho^2(B, x).$$

Quoting from Talagrand, "The main trick of the proof is to resist the temptation to optimize now over λ." Rather, we first fix ω and bound

$$\int_x \exp\left[\frac{1}{4}\rho^2(A, (x, \omega))\right] \le e^{(1-\lambda)^2/4} \int_x \left(\exp\left[\frac{1}{4}\rho^2(A_\omega, x)\right]\right)^\lambda \left(\exp\left[\frac{1}{4}\rho^2(B, x)\right]\right)^{1-\lambda}.$$

By Hölder's Inequality, this is at most

$$e^{(1-\lambda)^2/4}\left[\int_x \exp\left[\frac{1}{4}\rho^2(A_\omega, x)\right]\right]^\lambda \left[\int_x \exp\left[\frac{1}{4}\rho^2(B, x)\right]\right]^{1-\lambda},$$

which by induction is at most

$$e^{(1-\lambda)^2/4}\left(\frac{1}{\Pr\left[A_\omega\right]}\right)^\lambda \left(\frac{1}{\Pr\left[B\right]}\right)^{1-\lambda} = \frac{1}{\Pr[B]} e^{(1-\lambda)^2/4} r^{-\lambda},$$

where $r = \Pr[A_\omega]/\Pr[B] \le 1$. *Now* we use calculus and minimize $e^{(1-\lambda)^2/4} r^{-\lambda}$ by choosing $\lambda = 1 + 2\ln\ r$ for $e^{-1/2} \le r \le 1$, and $\lambda = 0$ otherwise. Further (somewhat tedious but simple) calculation shows $e^{(1-\lambda)^2/4} r^{-\lambda} \le 2 - r$ for this $\lambda = \lambda(r)$. Thus

$$\int_x \exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] \le \frac{1}{\Pr[B]}\left(2 - \frac{\Pr\left[A_\omega\right]}{\Pr[B]}\right).$$

We integrate over ω, giving

$$\int_\omega \int_x \exp\left[\frac{1}{4}\rho^2(A,(x,\omega))\right] \le \frac{1}{\Pr[B]}\left(2 - \frac{\Pr\left[A\right]}{\Pr[B]}\right) = \frac{1}{\Pr[A]} x(2-x),$$

where $x = \Pr[A]/\Pr[B] \in [0,1]$. But $x(2-x) \le 1$, completing the induction and hence the theorem.

7.7 APPLICATIONS OF TALAGRAND'S INEQUALITY

Let $\Omega = \prod_{i=1}^n \Omega_i$, where each Ω_i is a probability space and Ω has the product measure. Let $h : \Omega \to R$. Talagrand's Inequality enables us, under certain conditions, to show that the random variable $X = h(\cdot)$ is tightly concentrated. In this sense, it can serve the same function Azuma's Inequality does for martingales and there are many cases in which it gives far stronger results.

We call $h : \Omega \to R$ *Lipschitz* if $|h(x) - h(y)| \le 1$ whenever x, y differ in at most one coordinate. Talagrand's Inequality is most effective on those Lipschitz functions with the property that, when $h(x) \ge s$, there are a relatively small number of coordinates that will certify that $h(x) \ge s$. We formalize this notion as follows:

Definition 2 *Let $f : N \to N$. h is f-*certifiable *if, whenever $h(x) \ge s$, there exists $I \subseteq \{1, \ldots, n\}$ with $|I| \le f(s)$ so that all $y \in \Omega$ that agree with x on the coordinates I have $h(y) \ge s$.*

Example. Consider $G(n,p)$ as the product of $\binom{n}{2}$ coin flips, and let $h(G)$ be the number of triangles in G. Then h is f-certifiable with $f(s) = 3s$. For, if $h(G) \ge s$, there exist s triangles that together have at most $3s$ edges and any other G' with those $3s$ edges has $h(G') \ge s$. Note that I, here the indices for those $3s$ edges, very much depends on G. Also note that we need certify only lower bounds for h.

Theorem 7.7.1 *Under the above assumptions and for all b, t,*

$$\Pr\left[X \le b - t\sqrt{f(b)}\right] \Pr[X \ge b] \le e^{-t^2/4}.$$

Proof. Set $A = \{x : h(x) < b - t\sqrt{f(b)}\}$. Now suppose $h(y) \ge b$. We claim $y \notin A_t$. Let I be a set of indices of size at most $f(b)$ that certifies $h(y) \ge b$ as given above.

Define $\alpha_i = 0$ when $i \notin I$, and $\alpha_i = |I|^{-1/2}$ when $i \in I$. If $y \in A_t$, there exists a $z \in A$ that differs from y in at most $t|I|^{1/2} \leq t\sqrt{f(b)}$ coordinates of I though at arbitrary coordinates outside I. Let y' agree with y on I and agree with z outside I. By the certification, $h(y') \geq b$. Now y', z differ in at most $t\sqrt{f(b)}$ coordinates and so, by Lipschitz,

$$h(z) \geq h(y') - t\sqrt{f(b)} \geq b - t\sqrt{f(b)},$$

but then $z \notin A$, a contradiction. So $\Pr[X \geq b] \leq \Pr[\overline{A_t}]$, and from Talagrand's theorem,

$$\Pr[X < b - t\sqrt{f(b)}] \Pr[X \geq b] \leq e^{-t^2/4}.$$

As the right-hand side is continuous in t, we may replace "$<$" by "$\leq$", giving the theorem. ∎

A small generalization is sometimes useful. Call $h : \Omega \to R$ *K-Lipschitz* if $|h(x) - h(y)| \leq K$ whenever x, y differ in only one coordinate. Applying the above theorem to h/K, which is Lipschitz, we find

$$\Pr\left[X \leq b - tK\sqrt{f(b)}\right] \Pr[X \geq b] \leq e^{-t^2/4}.$$

In applications, one often takes b to be the median so that, for t large, the probability of being $t\sqrt{f(b)}$ under the median goes sharply to zero. But it works both ways: by parameterizing so that $m = b - t\sqrt{f(b)}$ is the median, one usually gets $b \sim m + t\sqrt{f(m)}$ and that the probability of being $t\sqrt{f(b)}$ above the median goes sharply to zero. Martingales, via Azuma's Inequality, generally produce a concentration result around the mean μ of X, while Talagrand's Inequality yields a concentration result about the median m. Means tend to be easy to compute, and medians notoriously difficult, but a tight concentration result will generally allow us to show that the mean and median are not far away.

Let $x = (x_1, \ldots, x_n)$, where the x_i are independently and uniformly chosen from $[0, 1]$. Set $X = h(x)$ to be the length of the longest increasing subsequence of x. Elementary methods give that $c_1 n^{1/2} < X < c_2 n^{1/2}$ almost surely for some positive constants c_1, c_2 and that the mean μ and median m of X are both in that range. Also, X is Lipschitz, as changing one x_i can only change X by at most one. How concentrated is X? We can apply Azuma's Inequality to deduce that, if $s \gg n^{1/2}$, then $|X - \mu| \leq s$ almost surely. This is not particularly good since X itself is only of order $n^{1/2}$. Now consider Talagrand's Inequality. X is f-certifiable with $f(s) = s$ since, if x has an increasing subsequence of length s, those s coordinates certify that $X \geq s$. Then $\Pr[X < m - tm^{1/2}] \leq e^{-t^2/4} / \Pr[X \geq m] \leq 2e^{-t^2/4}$, as m is the median value. But $m = \Theta(n^{1/2})$. Thus, when $s \gg n^{1/4}$, we have $X > m - s$ almost surely. For the other side, suppose $t \to \infty$ slowly and let b be such that $b - tb^{1/2} = m$. Then, $\Pr[X \geq b] \leq e^{-t^2/4} / \Pr[X \leq m] \leq 2e^{-t^2/4}$. Then $X \leq b$ almost surely. But $b = m + (1 + o(1))tm^{1/2}$ so that $X \leq m + tm^{1/2}$ almost surely. Combining, if $s \gg n^{1/4}$, then $|X - m| < s$ almost surely. A much stronger result, determining the

precise asymptotic distribution of X, has been obtained by Baik, Deift and Johansson (1999), using deep analytic tools.

Let us re-examine the bound (Theorem 7.3.2) that $G(n, \frac{1}{2})$ has no clique of size k with k as defined there. We let, as there, Y be the maximal number of edge disjoint k-cliques. From the work there, $\mathrm{E}[Y] = \Omega(n^2k^{-4})$ and Y is tightly concentrated about $\mathrm{E}[Y]$ so that the median m of Y must also have $m = \Omega(n^2k^{-4})$. As before, Y is Lipschitz. Further, Y is f-certifiable with $f(s) = \binom{k}{2} s$, as the edges of the s-cliques certify that $Y \geq s$. Hence

$$\Pr\left[Y \leq m - tm^{1/2}\binom{k}{2}^{1/2}\right] \Pr[Y \geq m] < e^{-t^2/4}.$$

Set $t = \Theta(m^{1/2}/k)$ so that $m = tm^{1/2}\binom{k}{2}^{1/2}$. Then

$$\Pr[\omega(G) < k] = \Pr[Y \leq 0] < 2e^{-t^2/4} < \exp\left[-\Omega\left(\frac{n^2}{\ln^{6} n}\right)\right],$$

which improves the bound of Theorem 7.3.2. Still, we should note that application of the Extended Janson Inequality in Section 10.3 does even better.

7.8 KIM–VU POLYNOMIAL CONCENTRATION

The approach of Kim and Vu (2000) is often useful. Let $H = (V(H), E(H))$ be a hypergraph, and let each edge $e \in E(H)$ have a nonnegative weight $w(e)$. Let t_i, $i \in V(H)$ be mutually independent indicator random variables with $\mathrm{E}[t_i] = p_i$. Consider the random variable polynomial

$$Y = \sum_{e \in E(H)} w_e \prod_{i \in e} t_i.$$

We allow $e = \emptyset$, in which case $\prod_{i \in e} t_i$ is by convention 1. We want to show that Y is concentrated about its mean.

Let $S \subseteq V(H)$ be a random set given by $\Pr[i \in S] = p_i$, these events being mutually independent over $i \in V(H)$. Then Y is the weighted number of hyperedges e in the restriction of H to S. In applications, we generally have all weights equal, so that Y simply counts the hyperedges in the random S. But we may also think abstractly of Y as simply any polynomial over the indicators t_i having all nonnegative coefficients.

We set $n = |V(H)|$, the number of vertices of H (number of variables t_i). Let k be an upper bound on the size of all hyperedges (upper bound on the degree of the polynomial Y).

Let $A \subseteq V(H)$ with $|A| \leq k$. We truncate Y to Y_A as follows: for those terms $\prod_{i \in e} t_i$ with $A \subseteq e$, we set $t_i = 1$ for all $i \in A$, replacing the term by $\prod_{i \in e - A} t_i$. All other terms (where e does not contain A) are deleted. For example, with $A = \{1\}$, $2t_1t_2 + 5t_1t_3t_4 + 7t_2t_4$ becomes $2t_2 + 5t_3t_4$. Intriguingly, as polynomials in the t_i, Y_A is the

partial derivative of Y with respect to the t_i, $i \in A$. Set $E_A = \mathrm{E}[Y_A]$. That is, E_A is the expected number of hyperedges in S that contain A, conditional on all vertices of A being in S. Set E_i equal to the maximal E_A over all $A \subseteq V(H)$ of size i. Set $\mu = \mathrm{E}[Y]$ for convenience, and set

$$E' = \max_{1 \le i \le k} E_i \quad \text{and} \quad E = \max[\mu, E'].$$

Theorem 7.8.1 [Kim–Vu Polynomial Concentration] *With the above hypotheses*

$$\Pr\left[|Y - \mu| > a_k (EE')^{1/2} \lambda^k\right] < d_k e^{-\lambda} n^{k-1}$$

for any $\lambda > 1$.

Here, for definiteness, we may take $a_k = 8^k k!^{1/2}$ and $d_k = 2e^2$.

We omit the proof, which combines martingale inequalities similar to those of Theorem 7.4.3 with a subtle induction on the degree k. There may well be room for improvement in the a_k, d_k, and n^{k-1} terms. In applications, one generally has k fixed and $\lambda \gg \ln\ n$ so that the $e^{-\lambda}$ term dominates the probability bound.

Applications of Kim–Vu polynomial concentration tend to be straightforward. Let $G \sim G(n,p)$ with $p = n^{-\alpha}$, and assume $0 < \alpha < 2/3$. Fix a vertex x of G, and let $Y = Y(x)$ be the number of triangles containing x. Set $\mu = \mathrm{E}[Y] = \binom{n-1}{2} p^3 \sim \frac{1}{2} n^{2-3\alpha}$. Let $\delta > 0$ be fixed. We want to bound $\Pr[|Y - \mu| > \delta\mu]$.

The random graph G is defined by the random variables t_{ij}, one for each unordered pair of vertices, which are indicators of the adjacency of the two vertices. In that context

$$Y = \sum_{i,j \ne x} t_{xi} t_{xj} t_{ij}.$$

This is a polynomial of degree $k = 3$. When A consists of a single edge $\{x, i\}$, we find $E_A = (n-2)p^2$; when it consists of three edges forming a triangle containing x, we find $E_A = 1$. When $A = \emptyset$, $E_A = \mu$. Other cases give smaller E_A. Basically, $E' \sim \max[np^2, 1]$. Calculation gives $E' \sim c\mu n^{-\epsilon}$ for some positive ϵ (dependent on α) throughout our range. We apply Kim–Vu polynomial concentration with $\lambda = c' n^{\epsilon/6}$, with c' a small positive constant, to bound $\Pr[|Y - \mu| > \delta\mu]$ by $\exp[-\Omega(n^{\epsilon/6})]$. Note that the n^{k-1} factor is absorbed by the exponential.

In particular, as this probability is $o(n^{-1})$, we have that almost surely every vertex x is in $\sim \mu$ triangles. This result generalizes. Fix $\alpha \in (0,1)$, and suppose (R,H) is a rooted graph, safe, in the sense of Section 10.4, with respect to α. Let $G \sim G(n,p)$ with $p = n^{-\alpha}$. For distinct vertices $x_1, \ldots, x_r$, let $Y = Y(x_1, \ldots, x_r)$ denote the number of extensions in G to H. Set $\mu = \mathrm{E}[Y]$. Kim–Vu polynomial concentration gives an exponentially small upper bound on the probability that Y is not near μ. In particular, this probability is $o(n^{-r})$. Hence, almost surely, every r vertices have $\sim \mu$ extensions to H.

7.9 EXERCISES

1. Let $G = (V, E)$ be the graph whose vertices are all 7^n vectors of length n over Z_7, in which two vertices are adjacent iff they differ in precisely one coordinate. Let $U \subset V$ be a set of 7^{n-1} vertices of G, and let W be the set of all vertices of G whose distance from U exceeds $(c+2)\sqrt{n}$, where $c > 0$ is a constant. Prove that $|W| \leq 7^n \cdot e^{-c^2/2}$.
2. (*) Let $G = (V, E)$ be a graph with chromatic number $\chi(G) = 1000$. Let $U \subset V$ be a random subset of V chosen uniformly from among all $2^{|V|}$ subsets of V. Let $H = G[U]$ be the induced subgraph of G on U. Prove that

$$\Pr[\chi(H) \leq 400] < 1/100.$$

3. Prove that there is an absolute constant c such that, for every $n > 1$, there is an interval I_n of at most $c\sqrt{n}/\log\ n$ consecutive integers such that the probability that the chromatic number of $G(n, 0.5)$ lies in I_n is at least 0.99.

THE PROBABILISTIC LENS:
Weierstrass Approximation Theorem

The well-known Weierstrass approximation theorem asserts that the set of real polynomials over $[0, 1]$ is dense in the space of all continuous real functions over $[0, 1]$. This is stated in the following theorem:

Theorem 1 [Weierstrass approximation theorem] *For every continuous real function $f : [0, 1] \to \mathbb{R}$ and every $\epsilon > 0$, there is a polynomial $p(x)$ such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in [0, 1]$.*

Bernstein (1912) gave a charming probabilistic proof of this theorem, based on the properties of the binomial distribution. His proof is as follows:

Proof. Since a continuous $f : [0, 1] \to \mathbb{R}$ is uniformly continuous, there is a $\delta > 0$ such that, if $x, x' \in [0, 1]$ and $|x - x'| \leq \delta$, $|f(x) - f(x')| \leq \epsilon/2$. In addition, since f must be bounded, there is an $M > 0$ such that $|f(x)| \leq M$ in $[0, 1]$.

Let $B(n, x)$ denote the binomial random variable with n independent trials and probability of success x for each of them. Thus the probability that $B(n, x) = j$ is precisely $\binom{n}{j} x^j (1 - x)^{n-j}$. The expectation of $B(n, x)$ is nx, and its standard deviation is $\sqrt{nx(1 - x)} \leq \sqrt{n}$. Therefore, by Chebyshev's Inequality discussed in Chapter 4, for every integer n, $\Pr[|B(n, x) - nx| > n^{2/3}] \leq 1/n^{1/3}$. It follows that there is an integer n such that

$$\Pr\left[|B(n, x) - nx| > n^{2/3}\right] < \frac{\epsilon}{4M}$$

and

$$\frac{1}{n^{1/3}} < \delta .$$

Define

$$P_n(x) = \sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} f(i/n).$$

We claim that, for every $x \in [0,1]$, $|P_n(x) - f(x)| \le \epsilon$. Indeed, since

$$\sum_{i=0}^{n} \binom{n}{i} x^i (1-x)^{n-i} = 1,$$

we have

$$\begin{aligned}
|P_n(x) - f(x)| \le & \sum_{i:|i-nx|\le n^{2/3}} \binom{n}{i} x^i (1-x)^{n-i} |f(i/n) - f(x)| \\
& + \sum_{i:|i-nx|>n^{2/3}} \binom{n}{i} x^i (1-x)^{n-i} [|f(i/n)| + |f(x)|] \\
\le & \sum_{i:|i/n-x|\le n^{-1/3}<\delta} \binom{n}{i} x^i (1-x)^{n-i} |f(i/n) - f(x)| \\
& + 2M \Pr\left[|B(n,x) - nx| > n^{2/3}\right] \\
\le & \frac{\epsilon}{2} + 2M \frac{\epsilon}{4M} = \epsilon.
\end{aligned}$$

This completes the proof. ■

8

The Poisson Paradigm

> One of the things that attracts us most when we apply ourselves to a mathematical problem is precisely that within us we always hear the call: here is the problem, search for the solution, you can find it by pure thought, for in mathematics there is no *ignorabimus*.
> –David Hilbert

When X is the sum of many rare indicator "mostly independent" random variables, and $\mu = \mathrm{E}[X]$, we would like to say that X is close to a Poisson distribution with mean μ and, in particular, that $\Pr[X = 0]$ is nearly $e^{-\mu}$. We call this rough statement the Poisson Paradigm. In this chapter, we give a number of situations in which this paradigm may be rigorously proved.

8.1 THE JANSON INEQUALITIES

In many instances, we would like to bound the probability that none of a set of bad events $\{B_i\}_{i \in I}$ occurs. If the events are mutually independent, then

$$\Pr\left[\bigwedge_{i \in I} \overline{B_i}\right] = \prod_{i \in I} \Pr[\overline{B_i}].$$

When the B_i are "mostly" independent, the Janson inequalities allow us, sometimes, to say that these two quantities are "nearly" equal.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

Let Ω be a finite universal set, and let R be a random subset of Ω given by

$$\Pr[r \in R] = p_r,$$

these events being mutually independent over $r \in \Omega$. Let $\{A_i\}_{i\in I}$ be subsets of Ω, and I a finite index set. Let B_i be the *event* $A_i \subseteq R$. (That is, each point $r \in \Omega$ "flips a coin" to determine if it is in R. B_i is the event where the coins for all $r \in A_i$ came up "heads.") Let X_i be the indicator random variable for B_i, and $X = \sum_{i\in I} X_i$ the number of $A_i \subseteq R$. The events $\bigwedge_{i\in I} \overline{B_i}$ and $X = 0$ are then identical. For $i, j \in I$, we write $i \sim j$ if $i \neq j$ and $A_i \cap A_j \neq \emptyset$. Note that, when $i \neq j$ and not $i \sim j$, then B_i, B_j are independent events since they involve separate coin flips. Furthermore, and this plays a crucial role in the proofs, if $i \notin J \subset I$ and not $i \sim j$ for all $j \in J$, then B_i is mutually independent of $\{B_j\}_{j\in J}$, that is, independent of any Boolean function of those B_j. This is because the coin flips on A_i and on $\cup_{j\in J} A_j$ are independent. We define

$$\Delta = \sum_{i\sim j} \Pr[B_i \wedge B_j].$$

Here, the sum is over ordered pairs so that $\Delta/2$ gives the same sum over unordered pairs. We set

$$M = \prod_{i\in I} \Pr[\overline{B_i}],$$

which is the value of $\Pr\left[\bigwedge_{i\in I} \overline{B_i}\right]$ if the B_i were independent. Finally, we set

$$\mu = \mathrm{E}[X] = \sum_{i\in I} \Pr[B_i].$$

The following results were given in Janson, Łuczak and Ruciński (1990).

Theorem 8.1.1 [The Janson Inequality] *Let $\{B_i\}_{i\in I}$, Δ, M, μ be as above, and assume all $\Pr[B_i] \leq \epsilon$. Then*

$$M \leq \Pr\left[\bigwedge_{i\in I} \overline{B_i}\right] \leq M e^{[1/(1-\epsilon)]\Delta/2}$$

and, further

$$\Pr\left[\bigwedge_{i\in I} \overline{B_i}\right] \leq e^{-\mu+\Delta/2}.$$

For each $i \in I$

$$\Pr[\overline{B_i}] = 1 - \Pr[B_i] \leq e^{-\Pr[B_i]}$$

so, multiplying over $i \in I$

$$M \leq e^{-\mu}.$$

The two upper bounds for Theorem 8.1.1 are generally quite similar; we tend to use the second for convenience. In many asymptotic instances, a simple calculation gives $M \sim e^{-\mu}$. In particular, this is always the case when $\epsilon = o(1)$ and $\epsilon\mu = o(1)$.

Perhaps the simplest example of Theorem 8.1.1 is the asymptotic probability that $G(n, c/n)$ is triangle-free, given in Section 10.1. There, as is often the case, $\epsilon = o(1)$, $\Delta = o(1)$, and μ approaches a constant k. In those instances, $\Pr[\bigwedge_{i\in I}\overline{B_i}] \to e^{-k}$. This is no longer the case when Δ becomes large. Indeed, when $\Delta \geq 2\mu$, the upper bound of Theorem 8.1.1 becomes useless. Even for Δ slightly less, it is improved by the following result:

Theorem 8.1.2 [The Extended Janson Inequality] *Under the assumptions of Theorem 8.1.1 and the further assumption that* $\Delta \geq \mu$

$$\Pr\left[\bigwedge_{i\in I}\overline{B_i}\right] \leq e^{-\mu^2/2\Delta}.$$

Theorem 8.1.2 (when it applies) often gives a much stronger result than Chebyshev's Inequality as used in Chapter 4. In Section 4.3 we saw $\mathrm{Var}[X] \leq \mu + \Delta$, so that

$$\Pr\left[\bigwedge_{i\in I}\overline{B_i}\right] = \Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathrm{E}[X]^2} \leq \frac{\mu + \Delta}{\mu^2}.$$

Suppose $\mu \to \infty$, $\mu \ll \Delta$, and $\gamma = \mu^2/\Delta \to \infty$. Chebyshev's upper bound on $\Pr[X = 0]$ is then roughly γ^{-1}, while Janson's upper bound is roughly $e^{-\gamma}$.

8.2 THE PROOFS

The original proofs of Janson are based on estimates of the Laplace transform of an appropriate random variable. The proof we present here follows that of Boppana and Spencer (1989). We shall use the inequalities

$$\Pr\left[B_i \,|\, \bigwedge_{j\in J}\overline{B_j}\right] \leq \Pr[B_i],$$

valid for all index sets $J \subset I, i \notin J$, and

$$\Pr\left[B_i \,|\, B_k \wedge \bigwedge_{j\in J}\overline{B_j}\right] \leq \Pr[B_i \,|\, B_k],$$

valid for all index sets $J \subset I, i, k \notin J$. The first follows from Theorem 6.3.2. The second is equivalent to the first since conditioning on B_k is the same as assuming $p_r = \Pr[r \in R] = 1$ for all $r \in A_k$.

Proof [Theorem 8.1.1.] The lower bound follows immediately. Order the index set $I = \{1, \ldots, m\}$ for convenience. For $1 \le i \le m$

$$\Pr\left[B_i \,|\, \bigwedge_{1 \le j < i} \overline{B_j}\right] \le \Pr[B_i]$$

so

$$\Pr\left[\overline{B_i} \,|\, \bigwedge_{1 \le j < i} \overline{B_j}\right] \ge \Pr[\overline{B_i}]$$

and

$$\Pr\left[\bigwedge_{i \in I} \overline{B_i}\right] = \prod_{i=1}^{m} \Pr\left[\overline{B_i} \,|\, \bigwedge_{1 \le j < i} \overline{B_j}\right] \ge \prod_{i=1}^{m} \Pr[\overline{B_i}].$$

Now the first upper bound. For a given i, renumber, for convenience, so that $i \sim j$ for $1 \le j \le d$ and not for $d+1 \le j < i$. We use the inequality $\Pr[A \,|\, B \wedge C] \ge \Pr[A \wedge B \,|\, C]$, valid for any A, B, C. With $A = B_i$, $B = \overline{B_1} \wedge \ldots \wedge \overline{B_d}$, and $C = \overline{B_{d+1}} \wedge \ldots \wedge \overline{B_{i-1}}$,

$$\begin{aligned}\Pr\left[B_i \,|\, \bigwedge_{1 \le j < i} \overline{B_j}\right] &= \Pr[A \,|\, B \wedge C] \ge \Pr[A \wedge B \,|\, C] \\ &= \Pr[A \,|\, C] \Pr[B \,|\, A \wedge C]\ .\end{aligned}$$

From the mutual independence, $\Pr[A \,|\, C] = \Pr[A]$. We bound

$$\Pr[B \,|\, A \wedge C] \ge 1 - \sum_{j=1}^{d} \Pr[B_j \,|\, B_i \wedge C] \ge 1 - \sum_{j=1}^{d} \Pr[B_j \,|\, B_i]$$

from the Correlation Inequality. Thus

$$\Pr\left[B_i \,|\, \bigwedge_{1 \le j < i} \overline{B_j}\right] \ge \Pr[B_i] - \sum_{j=1}^{d} \Pr[B_j \wedge B_i].$$

Reversing

$$\begin{aligned}\Pr\left[\overline{B_i} \,|\, \bigwedge_{1 \le j < i} \overline{B_j}\right] &\le \Pr[\overline{B_i}] + \sum_{j=1}^{d} \Pr[B_j \wedge B_i] \\ &\le \Pr[\overline{B_i}] \left(1 + \frac{1}{1-\epsilon} \sum_{j=1}^{d} \Pr[B_j \wedge B_i]\right)\end{aligned}$$

since $\Pr[\overline{B_i}] \geq 1 - \epsilon$. Employing the inequality $1 + x \leq e^x$

$$\Pr\left[\overline{B_i} \mid \bigwedge_{1 \leq j < i} \overline{B_j}\right] \leq \Pr[\overline{B_i}] \exp\left(\frac{1}{1-\epsilon} \sum_{j=1}^{d} \Pr[B_j \wedge B_i]\right).$$

For each $1 \leq i \leq m$, we plug this inequality into

$$\Pr\left[\bigwedge_{i \in I} \overline{B_i}\right] = \prod_{i=1}^{m} \Pr\left[\overline{B_i} \mid \bigwedge_{1 \leq j < i} \overline{B_j}\right].$$

The terms $\Pr[\overline{B_i}]$ multiply to M. The exponents add: for each $i, j \in I$ with $j < i$ and $j \sim i$, the term $\Pr[B_j \wedge B_i]$ appears once so they add to $\Delta/2$.

For the second upper bound, we instead bound

$$\begin{aligned}\Pr\left[\overline{B_i} \mid \bigwedge_{1 \leq j < i} \overline{B_j}\right] &\leq 1 - \Pr[B_i] + \sum_{j=1}^{d} \Pr[B_j \wedge B_i] \\ &\leq \exp\left(-\Pr[B_i] + \sum_{j=1}^{d} \Pr[B_j \wedge B_i]\right).\end{aligned}$$

Now the $-\Pr[B_i]$ terms add to $-\mu$, while the $\Pr[B_j \wedge B_i]$ terms again add to $\Delta/2$. ■

Proof [Theorem 8.1.2.] The second upper bound of Theorem 8.1.1 may be rewritten as

$$-\ln\left(\Pr\left[\bigwedge_{i \in I} \overline{B_i}\right]\right) \geq \sum_{i \in I} \Pr[B_i] - \frac{1}{2} \sum_{i \sim j} \Pr[B_i \wedge B_j].$$

For any set of indices $S \subset I$, the same inequality applied only to $\{B_i\}_{i \in S}$ gives

$$-\ln\left(\Pr\left[\bigwedge_{i \in S} \overline{B_i}\right]\right) \geq \sum_{i \in S} \Pr[B_i] - \frac{1}{2} \sum_{i,j \in S, i \sim j} \Pr[B_i \wedge B_j].$$

Let now S be a random subset of I given by

$$\Pr[i \in S] = p$$

with p a constant to be determined, and the events are mutually independent. (Here we are using probabilistic methods to prove a probability theorem!) Each term $\Pr[B_i]$ then appears with probability p, and each term $\Pr[B_i \wedge B_j]$ with probability p^2, so that

$$\begin{aligned}\mathrm{E}\left[-\ln\left(\Pr\left[\bigwedge_{i \in S} \overline{B_i}\right]\right)\right] &\geq \mathrm{E}\left[\sum_{i \in S} \Pr[B_i]\right] - \frac{1}{2}\mathrm{E}\left[\sum_{i,j \in S, i \sim j} \Pr[B_i \wedge B_j]\right] \\ &= p\mu - p^2 \frac{\Delta}{2}.\end{aligned}$$

We set

$$p = \frac{\mu}{\Delta}$$

so as to maximize this quantity. The added assumption of Theorem 8.1.2 assures us that the probability p is at most 1. Then

$$\mathrm{E}\left[-\ln\left(\Pr\left[\bigwedge_{i\in S}\overline{B_i}\right]\right)\right] \geq \frac{\mu^2}{2\Delta}.$$

Thus there is a specific $S \subset I$ for which

$$-\ln\left(\Pr\left[\bigwedge_{i\in S}\overline{B_i}\right]\right) \geq \frac{\mu^2}{2\Delta}.$$

That is

$$\Pr\left[\bigwedge_{i\in S}\overline{B_i}\right] \leq e^{-\mu^2/2\Delta}.$$

But

$$\Pr\left[\bigwedge_{i\in I}\overline{B_i}\right] \leq \Pr\left[\bigwedge_{i\in S}\overline{B_i}\right],$$

completing the proof. ■

8.3 BRUN'S SIEVE

The more traditional approach to the Poisson Paradigm is called Brun's sieve, after its use by the number theorist T. Brun. Let $B_1, \ldots, B_m$ be events, X_i the indicator random variable for B_i, and $X = X_1 + \ldots + X_m$ the number of B_i that hold. Let there be a hidden parameter n (so that actually $m = m(n), B_i = B_i(n), X = X(n)$), which will define our o, O notation. Define

$$S^{(r)} = \sum \Pr[B_{i_1} \wedge \ldots \wedge B_{i_r}],$$

the sum over all sets $\{i_1, \ldots, i_r\} \subseteq \{1, \ldots, m\}$, and put

$$X^{(r)} = X(X-1)\ldots(X-r+1).$$

The inclusion–exclusion principle gives

$$\Pr[X = 0] = \Pr[\overline{B_1} \wedge \ldots \wedge \overline{B_m}] = 1 - S^{(1)} + S^{(2)} - \ldots + (-1)^r S^{(r)} \ldots .$$

Theorem 8.3.1 *Suppose there is a constant μ so that*

$$\mathrm{E}[X] = S^{(1)} \to \mu$$

and such that, for every fixed r

$$\mathrm{E}\left[\binom{X}{r}\right] = S^{(r)} \to \frac{\mu^r}{r!}.$$

Then

$$\Pr[X = 0] \to e^{-\mu}$$

and indeed for every t

$$\Pr[X = t] \to \frac{\mu^t}{t!}e^{-\mu}.$$

Proof. We do only the case $t = 0$. Fix $\epsilon > 0$. Choose s so that

$$\left|\sum_{r=0}^{2s}(-1)^r\frac{\mu^r}{r!} - e^{-\mu}\right| \le \frac{\epsilon}{2}.$$

The Bonferroni Inequalities state that, in general, the inclusion–exclusion formula alternately over- and underestimates $\Pr[X = 0]$. In particular,

$$\Pr[X = 0] \le \sum_{r=0}^{2s}(-1)^r S^{(r)}.$$

Select n_0 (the hidden variable) so that for $n \ge n_0$

$$\left|S^{(r)} - \frac{\mu^r}{r!}\right| \le \frac{\epsilon}{2(2s+1)}$$

for $0 \le r \le 2s$. For such n

$$\Pr[X = 0] \le e^{-\mu} + \epsilon.$$

Similarly, taking the sum to $2s + 1$ we find n_0 so that for $n \ge n_0$

$$\Pr[X = 0] \ge e^{-\mu} - \epsilon.$$

As ϵ was arbitrary, $\Pr[X = 0] \to e^{-\mu}$. ∎

The threshold functions for $G \sim G(n, p)$ to contain a copy of a given graph H, derived in Section 10.1 via the Janson Inequality, were originally found using Brun's sieve. Here is an example where both methods are used. Let $G \sim G(n, p)$, the random

graph of Chapter 10. Let EPIT represent the statement that *every* vertex lies in a triangle.

Theorem 8.3.2 *Let $c > 0$ be fixed and let $p = p(n)$, $\mu = \mu(n)$ satisfy*

$$\binom{n-1}{2} p^3 = \mu ,$$
$$e^{-\mu} = \frac{c}{n}.$$

Then

$$\lim_{n\to\infty} \Pr[G(n,p) \text{ satisfies EPIT}] = e^{-c}.$$

In Spencer (1990a) threshold functions are found for a very wide class of "extension statements" that every r vertices lie in a copy of some fixed H.

Proof. First, fix $x \in V(G)$. For each unordered $y, z \in V(G) - \{x\}$, let B_{xyz} be the event where $\{x, y, z\}$ is a triangle of G. Let C_x be the event $\bigwedge_{y,z} \overline{B_{xyz}}$ and X_x the corresponding indicator random variable. We use Janson's Inequality to bound $\mathrm{E}[X_x] = \Pr[C_x]$. Here $p = o(1)$ so $\epsilon = o(1)$. $\sum \Pr[B_{xyz}] = \mu$ as defined above. Dependency $xyz \sim xuv$ occurs if and only if the sets overlap (other than in x). Hence

$$\Delta = \sum_{y,z,z'} \Pr[B_{xyz} \wedge B_{xyz'}] = O(n^3 p^5) = o(1)$$

since $p = n^{-2/3+o(1)}$. Thus

$$\mathrm{E}[X_x] \sim e^{-\mu} = \frac{c}{n}.$$

Now define

$$X = \sum_{x\in V(G)} X_x,$$

the number of vertices x not lying in a triangle. Then from Linearity of Expectation,

$$\mathrm{E}[X] = \sum_{x\in V(G)} \mathrm{E}[X_x] \to c.$$

We need to show that the Poisson Paradigm applies to X. Fix r. Then

$$\mathrm{E}\left[\binom{X}{r}\right] = S^{(r)} = \sum \Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}],$$

the sum over all sets of vertices $\{x_1, \ldots, x_r\}$. All r-sets look alike, so

$$\mathrm{E}\left[\binom{X}{r}\right] = \binom{n}{r} \Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}] \sim \frac{n^r}{r!} \Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}],$$

where $x_1, \ldots, x_r$ are some particular vertices. But

$$C_{x_1} \wedge \ldots \wedge C_{x_r} = \bigwedge \overline{B_{x_i yz}},$$

the conjunction over $1 \le i \le r$ and all y, z. We apply Janson's Inequality to this conjunction. Again, $\epsilon = p^3 = o(1)$. The number of $\{x_i, y, z\}$ is $r\binom{n-1}{2} - O(n)$, the overcount coming from those triangles containing two (or three) of the x_i. (Here it is crucial that r is fixed.) Thus

$$\sum \Pr[B_{x_i yz}] = p^3 \left(r \binom{n-1}{2} - O(n) \right) = r\mu + O(n^{-1+o(1)}).$$

As before, Δ is p^5 times the number of pairs $x_i yz \sim x_j y' z'$. There are $O(rn^3) = O(n^3)$ terms with $i = j$ and $O(r^2 n^2) = O(n^2)$ terms with $i \ne j$, so again $\Delta = o(1)$. Therefore

$$\Pr[C_{x_1} \wedge \ldots \wedge C_{x_r}] \sim e^{-r\mu}$$

and

$$\mathrm{E}\left[\binom{X}{r}\right] \sim \frac{(ne^{-\mu})^r}{r!} = \frac{c^r}{r!}.$$

Hence the conditions of Theorem 8.3.1 are met for X. ■

8.4 LARGE DEVIATIONS

We return to the formulation of Section 8.1. Our object is to derive large deviation results on X similar to those in Appendix A. Given a point in the probability space (i.e., a selection of R), we call an index set $J \subseteq I$ a disjoint family (abbreviated disfam) if

- B_j for every $j \in J$.
- For no $j, j' \in J$ is $j \sim j'$.

If, in addition

- If $j' \notin J$ and $B_{j'}$, then $j \sim j'$ for some $j \in J$,

then we call J a maximal disjoint family (maxdisfam). We give some general results on the possible sizes of maxdisfams. The connection to X must then be done on an *ad hoc* basis.

Lemma 8.4.1 *With the above notation and for any integer s,*

$$\Pr[\textit{there exists a disfam } J, |J| = s] \le \frac{\mu^s}{s!}.$$

Proof. Let $\sum^*$ denote the sum over all s-sets $J \subseteq I$ with no $j \sim j'$. Let $\sum^o$ denote the sum over ordered s-tuples $(j_1, \dots, j_s)$, with $\{j_1, \dots, j_s\}$ forming such a J. Let $\sum^a$ denote the sum over *all* ordered s-tuples $(j_1, \dots, j_s)$. Then

$$\begin{aligned}\Pr[\text{there exists a disfam } J, |J| = s] &\le \sum^* \Pr\left[\bigwedge_{j\in J} B_j\right] \\ &= \sum^* \prod_{j\in J} \Pr[B_j] = \frac{1}{s!}\sum^o \Pr[B_{j_1}] \dots \Pr[B_{j_s}] \\ &\le \frac{1}{s!}\sum^a \Pr[B_{j_1}] \dots \Pr[B_{j_s}] \le \frac{1}{s!}\left(\sum_{i\in I} \Pr[B_i]\right)^s = \frac{\mu^s}{s!}\ .\end{aligned}$$

■

Lemma 8.4.1 gives an effective upper bound when $\mu^s \ll s!$ – basically if $s > \mu\alpha$ for $\alpha > e$. For smaller s, we look at the further condition of J being a *max*disfam. To that end, we let μ_s denote the minimum, over all $j_1, \dots, j_s \in I$ of $\sum \Pr[B_i]$, the sum taken over all $i \in I$ except those i with $i \sim j_l$ for some $1 \le l \le s$. In application, s will be small (otherwise we use Lemma 8.4.1) and μ_s will be close to μ. For some applications, it is convenient to set

$$\nu = \max_{j\in I} \sum_{i\sim j} \Pr[B_i]$$

and note that $\mu_s \ge \mu - s\nu$.

Lemma 8.4.2 *With the above notation and for any integer s*

$$\begin{aligned}\Pr[\textit{there exists a maxdisfam } J, |J| = s] &\le \frac{\mu^s}{s!}e^{-\mu_s}e^{\Delta/2} \\ &\le \frac{\mu^s}{s!}e^{-\mu}e^{s\nu}e^{\Delta/2}\ .\end{aligned}$$

Proof. As in Lemma 8.4.1, we bound this probability by $\sum^*$ of $J = \{j_1, \dots, j_s\}$ being a maxdisfam. For this to occur, J must first be a disfam and then $\wedge^* \overline{B_i}$, where $\wedge^*$ is the conjunction over all $i \in I$ except those with $i \sim j_l$ for some $1 \le l \le s$. We apply Janson's Inequality to give an upper bound to $\Pr[\wedge^* \overline{B_i}]$. The associated values μ^*, Δ^* satisfy

$$\begin{aligned}\mu^* &\ge \mu_s\ , \\ \Delta^* &\le \Delta\end{aligned}$$

the latter because Δ^* has simply fewer addends. Thus

$$\Pr[\overset{*}{\wedge} \overline{B_i}] \leq e^{-\mu_s} e^{\Delta/2}$$

and

$$\sum^{*} \Pr[J \text{ maxdisfam}] \leq e^{-\mu_s} e^{\Delta/2} \sum^{*} \Pr\left[\bigwedge_{j \in J} B_j\right]$$

$$\leq e^{-\mu_s} e^{\Delta/2} \frac{\mu^s}{s!} .$$

■

When $\Delta = o(1)$ and $\nu\mu = o(1)$, or, more generally, $\mu_{3\mu} = \mu + o(1)$, then Lemma 8.4.2 gives a close approximation to the Poisson distribution since

$$\Pr[\text{there exists a maxdisfam } J, |J| = s] \leq (1 + o(1)) \frac{\mu^s}{s!} e^{-\mu}$$

for $s \leq 3\mu$ and the probability is quite small for larger s by Lemma 8.4.1.

8.5 COUNTING EXTENSIONS

We begin with a case that uses the basic large deviation results of Appendix A.

Theorem 8.5.1 *Set $p = [(\ln n)/n]\omega(n)$, where $\omega(n) \to \infty$ arbitrarily slowly. Then in $G(n, p)$ almost always*

$$\deg(x) \sim (n-1)p$$

for all *vertices x.*

This is actually a large deviation result. It suffices to show the following:

Theorem 8.5.2 *Set $p = [(\ln n)/n]\omega(n)$, where $\omega(n) \to \infty$ arbitrarily slowly. Let $x \in G$ be fixed. Fix $\epsilon > 0$. Then*

$$\Pr[|\deg(x) - (n-1)p| > \epsilon(n-1)p] = o(n^{-1}).$$

Proof. As $\deg(x) \sim B(n-1, p)$, that is, it is a binomial random variable with the above parameters, we have from Corollary A.1.14 that

$$\Pr[|\deg(x) - (n-1)p| > \epsilon(n-1)p] < 2e^{-c_\epsilon (n-1)p} = o(n^{-1}),$$

as c_ϵ is fixed and $(n-1)p \gg \ln n$. ■

This result illustrates why logarithmic terms appear so often in the study of random graphs. We want *every* x to have a property, hence we try to get the failure probability down to $o(n^{-1})$. When the Poisson Paradigm applies, the failure probability is roughly an exponential, and hence we want the exponent to be logarithmic. This often leads to a logarithmic term for the edge probability p.

In Section 8.3 we found the threshold function for every vertex to lie on a triangle. It basically occurred when the expected number of extensions of a given vertex to a triangle reached $\ln n$. Now set $N(x)$ to be the number of triangles containing x. Set $\mu = \binom{n-1}{2} p^3 = \mathrm{E}[N(x)]$.

Theorem 8.5.3 *Let p be such that $\mu \gg \ln n$. Then almost always*

$$N(x) \sim \mu$$

for all $x \in G(n,p)$.

As above, this is actually a large deviation result. We actually show the following:

Theorem 8.5.4 *Let p be such that $\mu \gg \ln n$. Let $x \in G$ be fixed. Fix $\epsilon > 0$. Then*

$$\Pr[|N(x) - \mu| > \epsilon\mu] = o(n^{-1}).$$

Proof. We shall prove this under the further assumption $p = n^{-2/3+o(1)}$ (or equivalently, $\mu = n^{o(1)}$), which could be removed by technical methods. We now have, in the notation of Lemmas 8.4.1 and 8.4.2, $\nu\mu, \Delta = o(1)$. Let P denote the Poisson distribution with mean μ. Then

$$\begin{aligned}
&\Pr[\text{there exists a maxdisfam } J, |J| \le \mu(1-\epsilon)] \\
&\quad \le (1+o(1))\Pr[P \le \mu(1-\epsilon)]\ , \\
&\Pr[\text{there exists a maxdisfam } J, \mu(1+\epsilon) \le |J| \le 3\mu] \\
&\quad \le (1+o(1))\Pr[\mu(1+\epsilon) \le P \le 3\mu]\ , \\
&\Pr[\text{there exists a maxdisfam } J, |J| \ge 3\mu] \\
&\quad \le \Pr[\text{there exists a disfam } J, |J| \ge 3\mu] \le \sum_{s=3\mu}^{\infty} \frac{\mu^s}{s!} = O((1-c)^{\mu})\ ,
\end{aligned}$$

where $c > 0$ is an absolute constant. Since $\mu \gg \ln n$, the third term is $o(n^{-1})$. The first and second terms are $o(n^{-1})$ by Theorem A.1.15. With probability $1 - o(n^{-1})$, every maxdisfam J has size between $(1-\epsilon)\mu$ and $(1+\epsilon)\mu$.

Fix one such J. (There *always* is some maximal disfam –even if no B_i held, we could take $J = \emptyset$.) The elements of J are triples xyz that form triangles, hence $N(x) \ge |J| \ge (1-\epsilon)\mu$. The upper bound is ad hoc. The probability that there exist five triangles of the form $xyz_1, xyz_2, xyz_3, xyz_4, xyz_5$ is at most $n^6 p^{11} = o(n^{-1})$. The probability

that there exist triangles xy_iz_i, xy_iz_i', $1 \le i \le 4$, all vertices being distinct, is at most $n^{12}p^{20} = o(n^{-1})$. Consider the graph whose vertices are the triangles xyz, with $\sim$ giving the edge relation. There are $N(x)$ vertices, and the maxdisfam J are the maximal independent sets . In this graph, with probability $1 - o(n^{-1})$, each vertex xyz has degree at most 9, and there is no set of four disjoint edges. This implies that for any J, $|J| \ge N(x) - 27$ and

$$N(x) \le (1+\epsilon)\mu + 27 \le (1+\epsilon')\mu.$$

■

For any graph H with "roots" $x_1, \ldots, x_r$, we can examine in $G(n,p)$ the number of extensions $N(x_1, \ldots, x_r)$ of a given set of r vertices to a copy of H. In Spencer (1990b), some general results are given that generalize Theorems 8.5.2 and 8.5.4. Under fairly wide assumptions (see Exercise 5, Chapter 10), when the expected number μ of extensions satisfies $\mu \gg \ln n$, then almost always all $N(x_1, \ldots, x_r) \sim \mu$.

8.6 COUNTING REPRESENTATIONS

The results of this section shall use the following very basic and very useful result:

Lemma 8.6.1 [The Borel–Cantelli Lemma] *Let $\{A_n\}_{n\in\mathbb{N}}$ be events with*

$$\sum_{n=1}^{\infty} \Pr[A_n] < \infty.$$

Then

$$\Pr\left[\bigwedge_{i=1}^{\infty}\bigvee_{j=i}^{\infty} A_j\right] = 0.$$

That is, almost always A_n is false for all sufficiently large n. In application, we shall aim for $\Pr[A_n] < n^{-c}$ with $c > 1$ in order to apply this lemma.

Again, we begin with a case that involves only the large deviation results of Appendix A. For a given set S of natural numbers, let (for every $n \in \mathbb{N}$) $f(n) = f_S(n)$ denote the number of representations $n = x + y$, $x, y \in S$, $x < y$.

Theorem 8.6.2 Erdős (1956) *There is a set S for which $f(n) = \Theta(\ln n)$. That is, there is a set S and constants c_1, c_2 so that, for all sufficiently large n*

$$c_1 \ln n \le f(n) \le c_2 \ln n.$$

Proof. Define S randomly by

$$\Pr[x \in S] = p_x = \min\left[10\sqrt{\frac{\ln x}{x}}, 1\right].$$

Fix n. Now $f(n)$ is a random variable with mean

$$\mu = \mathrm{E}[f(n)] = \frac{1}{2} \sum_{x+y=n, x \neq y} p_x p_y.$$

Roughly, there are n addends with $p_x p_y > p_n^2 = 100(\ln n)/n$. We have $p_x p_x = \Theta((\ln n)/n)$ except in the regions $x = o(n), y = o(n)$, and care must be taken that those terms do not contribute significantly to μ. Careful asymptotics (and first-year calculus!) yield

$$\mu \sim (50 \ln n) \int_0^1 \frac{dx}{\sqrt{x(1-x)}} = 50\pi \ln n.$$

The negligible effect of the $x = o(n), y = o(n)$ terms reflects the finiteness of the indefinite integral at poles $x = 0$ and $x = 1$. The possible representations $x + y = n$ are mutually independent events, so that from Corollary A.1.14

$$\Pr[|f(n) - \mu| > \epsilon\mu] < 2e^{-\delta\mu}$$

for constants $\epsilon, \delta = \delta(\epsilon)$. To be specific, we can take $\epsilon = 0.9, \delta = 0.1$ and

$$\Pr[|f(n) - \mu| > 0.9\mu] < 2e^{-5\pi \ln n} < n^{-1.1}$$

for n sufficiently large. Take $c_1 < 0.1(50\pi)$ and $c_2 > 1.9(50\pi)$.

Let A_n be the event where $c_1 \ln n \leq f(n) \leq c_2 \ln n$ does *not* hold. We have $\Pr[A_n] < n^{-1.1}$ for n sufficiently large. The Borel–Cantelli Lemma applies, and almost always all A_n fail for n sufficiently large. Thus there exists a specific point in the probability space, that is, a specific set S, for which $c_1 \ln n \leq f(n) \leq c_2 \ln n$ for all sufficiently large n. ■

The development of the infinite probability space used here, and below, has been carefully done in the book *Sequences* by Halberstam and Roth (1983).

The use of the infinite probability space leaves a number of questions about the existential nature of the proof that go beyond the algorithmic. For example, does there exist a recursive set S having the property of Theorem 8.6.3? An affirmative answer is given in Kolountzakis (1999).

Now for a given set S of natural numbers, let $g(n) = g_S(n)$ denote the number of representations $n = x + y + z, x, y, z \in S, x < y < z$. The following result was actually proved for representations of n as the sum of k terms for any fixed k. For simplicity, we present here only the proof for $k = 3$.

Theorem 8.6.3 Erdős and Tetali (1990) *There is a set S for which $g(n) = \Theta(\ln n)$. That is, there is a set S and constants c_1, c_2, so that for all sufficiently large n*

$$c_1 \ln n \leq g(n) \leq c_2 \ln n.$$

Proof. Define S randomly by

$$\Pr[x \in S] = p_x = \min\left[10\Big(\frac{\ln x}{x^2}\Big)^{1/3}, \frac{1}{2}\right].$$

Fix n. Now $g(n)$ is a random variable and

$$\mu = \mathrm{E}[g(n)] = \sum_{x+y+z=n} p_x p_y p_z.$$

Careful asymptotics give

$$\mu \sim \frac{10^3}{6} \ln n \int_{x=0}^{1} \int_{y=0}^{1-x} \frac{dx\ \ dy}{[xy(1-x-y)]^{2/3}} = K \ln n,$$

where K is large. (We may make K arbitrarily large by increasing "10.") We apply Lemma 8.4.2. Here

$$\Delta = \sum p_x p_y p_z p_{y'} p_{z'},$$

the sum over all five-tuples with $x + y + z = x + y' + z' = n$. Roughly, there are n^3 terms, each $\sim p_n^5 = n^{-10/3+o(1)}$, so that the sum is $o(1)$. Again, care must be taken that those terms with one (or more) small variables do not contribute much to the sum. We bound $s \le 3\mu = \Theta(\ln n)$ and consider μ_s. This is the minimal possible $\sum p_x p_y p_z$ over all those x, y, z with $x + y + z = n$ that do not intersect a given set of s representations; let us weaken that and say a given set of $3s$ elements. Again, one needs that the weight of $\sum_{x+y+z=n} p_x p_y p_z$ is not on the edges but "spread" in the center and one shows $\mu_s \sim \mu$. Now, as in Section 8.5, let P denote the Poisson distribution with mean μ. The probability that there exists a maxdisfam J of size less than $\mu(1-\epsilon)$ or between $\mu(1+\epsilon)$ and 3μ is asymptotically the probability that P lies in that range. For moderate ϵ, as K is large, these –as well as the probability of having a disfam of size bigger than 3μ –will be $o(n^{-c})$ with $c > 1$. By the Borel–Cantelli Lemma , almost always all sufficiently large n will have all maxdisfam J of size between $c_1 \ln n$ and $c_2 \ln n$. Then, $g(n) \ge c_1 \ln n$ immediately.

The upper bound is again *ad hoc*. With this p, let $f(n)$ be, as before, the number of representations of n as the sum of two elements of S. We use only that $p_x = x^{-2/3+o(1)}$. We calculate

$$\mathrm{E}[f(n)] = \sum_{x+y=n} (xy)^{-2/3+o(1)} = n^{-1/3+o(1)},$$

again watching the "pole" at 0. Here the possible representations are mutually independent, so

$$\Pr[f(n) \ge 4] \le \mathrm{E}[f(n)]^4/4! = n^{-4/3+o(1)},$$

and by the Borel–Cantelli Lemma almost always $f(n) \le 3$ for all sufficiently large n. But then almost always there is a C so that $f(n) \le C$ for all n. For all sufficiently large n, there is a maxdisfam (with representations as the sum of three terms) of size

less than $c_2 \ln n$. Every triple $x, y, z \in S$ with $x + y + z = n$must contain at least one of these at most $3c_2 \ln n$ points. The number of triples $x, y, z \in S$ with $x + y + z = n$ for a particular x is simply $f(n - x)$, the number of representations $n - x = y + z$ (possibly one less since $y, z \neq x$), and so is at most C. But then, there are at most $C(3c_2 \ln n)$ total representations $n = x + y + z$. ■

8.7 FURTHER INEQUALITIES

Here we discuss some further results that allow one, sometimes, to apply the Poisson Paradigm. Let $B_i, i \in I$ be events in an arbitrary probability space. As in the Lovász Local Lemma of Chapter 5, we say that a symmetric binary relation '$\sim$' on I is a *dependency digraph* if for each $i \in I$ the event B_i is mutually independent of $\{B_j : i \nsim j\}$. [The digraph of Section 5.1 has $E = \{(i,j) : i \sim j\}$.] Suppose the events B_i satisfy the inequalities of Section 8.2:

$$\Pr\left[B_i \,|\, \bigwedge_{j\in J} \overline{B_j}\right] \leq \Pr[B_i]$$

valid for all index sets $J \subset I, i \notin J$ and

$$\Pr\left[B_i \,|\, B_k \wedge \bigwedge_{j\in J} \overline{B_j}\right] \leq \Pr[B_i \,|\, B_k]$$

valid for all index sets $J \subset I, i, k \notin J$. Then, the Janson Inequalities in Theorems 8.1.1 and 8.1.2 and also Lemmas 8.4.1and 8.4.2 hold as stated. The proofs are identical; the above are the only properties of the events B_i that were used.

Suen (1990) (see also Janson (1998) for significant Variations) has given a very general result that allows the approximation of $\Pr[\bigwedge_{i\in I} \overline{B_i}]$ by $M = \prod_{i\in I} \Pr[\overline{B_i}]$. Again, let $\{B_i\}_{i\in I}$ be events in an arbitrary probability space. We say that a binary relation $\sim$ on I is a *superdependency digraph* if the following holds: Let $J_1, J_2 \subset I$ be disjoint subsets so that $j_1 \sim j_2$ for no $j_1 \in J_1, j_2 \in J_2$. Let B^1 be any Boolean combination of the events $\{B_j\}_{j\in J_1}$, and let B^2 be any Boolean combination of the events $\{B_j\}_{j\in J_2}$. Then B^1, B^2 are independent. Note that the '$\sim$' of Section 8.1 is indeed a superdependency digraph.

Theorem 8.7.1 [Suen] *Under the above conditions*

$$\left|\Pr\left[\bigwedge_{i\in I} \overline{B_i}\right] - M\right| \leq M\left[e^{\sum_{i\sim j} y(i,j)} - 1\right],$$

where

$$y(i,j) = (\Pr[B_i \wedge B_j] + \Pr[B_i]\Pr[B_j]) \prod_{l\sim i \text{ or } l\sim j} (1 - \Pr[B_l])^{-1}.$$

We shall not prove Theorem 8.7.1. In many instances, the above product is not large. Suppose it is less than 2 for all $i \sim j$. In that instance

$$\sum_{i \sim j} y(i,j) \leq 2\left(\Delta + \sum_{i \sim j} \Pr[B_i]\Pr[B_j]\right).$$

In many instances, $\sum_{i \sim j} \Pr[B_i]\Pr[B_j]$ is small compared to Δ (as in many instances when $i \sim j$ the events B_i, B_j are positively correlated). When, furthermore, $\Delta = o(1)$, Suen's theorem gives the approximation of $\Pr[\wedge_{i \in I} \overline{B_i}]$ by M. Suen has applied this result to examine the number of induced copies of a fixed graph H in the random $G(n,p)$.

Janson (1990) has given a one-way large deviation result on the X of Section 8.1 that is somewhat simpler to apply than Lemmas 8.4.1 and 8.4.2.

Theorem 8.7.2 Janson *With* $\mu = \mathrm{E}[X]$ *and* $\gamma > 0$ *arbitrary*

$$\Pr[X \leq (1-\gamma)\mu] < e^{-\gamma^2 \mu/[2+2(\Delta/\mu)]}.$$

When $\Delta = o(\mu)$, this bound on the tail approximates that of the normal curve with mean and standard deviation μ. We shall not prove Theorem 8.7.2 here. The proofs of Theorems 8.7.1 and 8.7.2, as well as the original proofs by Janson of Theorems 8.1.1 and 8.1.2, are based on estimations of the Laplace transform of X, bounding $\mathrm{E}[e^{-tX}]$.

8.8 EXERCISES

1. Prove that for every $\epsilon > 0$ there is some $n_0 = n_0(\epsilon)$ so that for every $n > n_0$ there is a graph on n vertices containing every graph on $k \leq (2-\epsilon)\log_2 n$ vertices as an induced subgraph.
2. Find a threshold function for the property: $G(n,p)$ contains at least $n/6$ pairwise vertex disjoint triangles.

THE PROBABILISTIC LENS:
Local Coloring

This result of Erdős (1962) gives further probabilistic evidence that the chromatic number $\chi(G)$ cannot be deduced from local considerations.

Theorem 1 *For all k, there exists $\epsilon > 0$ so that for all sufficiently large n there exist graphs G on n vertices with $\chi(G) > k$ and yet $\chi(G|_S) \leq 3$ for every set S of vertices of size at most ϵn.*

Proof. For a given k let $c, \epsilon > 0$ satisfy (with foresight)

$$c > 2k^2 H(1/k) \ln 2 ,$$
$$\epsilon < e^{-5} 3^3 c^{-3} ,$$

where $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the entropy function. Set $p = c/n$ and let $G \sim G(n,p)$. We show that G almost surely satisfies the two conditions of the theorem.

If $\chi(G) \leq k$, there would be an independent set of size n/k. The expected number of such sets is

$$\binom{n}{n/k}(1-p)^{\binom{n/k}{2}} < 2^{n(H(1/k)+o(1))} e^{-cn/2k^2(1+o(1))},$$

which is $o(1)$ by our condition on c. Hence almost surely $\chi(G) > k$.

Suppose some set S with $t \leq \epsilon n$ vertices required at least four colors. Then, as in the proof of Lemma 7.3.4, there would be a minimal such set S. For any $v \in S$, there would be a three-coloring of $S - \{v\}$. If v had two or fewer neighbors in S, then this could be extended to a three-coloring of S. Hence every $v \in S$ would have degree at

least 3 in $G|_S$ and so $G|_S$ would have at least $3t/2$ edges. The probability that some $t \le \epsilon n$ vertices have at least $3t/2$ edges is less than

$$\sum_{t \le \epsilon n} \binom{n}{t} \binom{\binom{t}{2}}{3t/2} \left(\frac{c}{n}\right)^{3t/2}.$$

We outline the analysis. When $t = O(1)$, the terms are negligible. Otherwise we bound each term from above by

$$\left[\frac{ne}{t}\left(\frac{te}{3}\right)^{3/2}\left(\frac{c}{n}\right)^{3/2}\right]^t \le [e^{5/2}3^{-3/2}c^{3/2}\sqrt{t/n}]^t.$$

Now since $t \le \epsilon n$, the bracketed term is at most $e^{5/2}3^{-3/2}c^{3/2}\epsilon^{1/2}$, which is less than 1 by our condition on ϵ. The full sum is $o(1)$; that is, almost surely no such S exists. ■

Many tempting conjectures are easily *dis*proved by the probabilistic method. If every $n/\ln n$ vertices may be three-colored, then can a graph G on n vertices be four-colored? This result shows that the answer is no.

9

Quasirandomness

> You have brains in your head. You have feet in your shoes. You can steer yourself any direction you choose. You're on your own. And you know what you know. And YOU are the one who'll decide where to go...
> –from *Oh, The Places You'll Go*, by Dr. Seuss

Quasirandomness, like so many profound mathematical concepts, appears to be a contradiction in terms. The random graph $G(n,p)$ is *not* a graph, it is a probability distribution over the family of graphs with n vertices. A specific graph cannot be a random graph anymore than you, the reader, can be a random person. Despite this, we shall give mathematically rigorous notions of quasirandomness.

As shown in the various chapters of this book, the probabilistic method is a powerful tool for establishing the existence of combinatorial structures with certain properties. It is often the case that such an existence proof is not sufficient; we actually prefer an *explicit construction*. This is not only because an explicit construction may shed more light on the corresponding problem but also because it often happens that a random-looking structure is useful for a certain algorithmic procedure; in this case we would like to have an algorithm and not merely to prove that it exists.

The problem of finding explicit constructions may look trivial; after all, since we are mainly dealing with finite cases, once we have a probabilistic proof of existence we can find an explicit example by exhaustive search. Moreover, many of the probabilistic proofs of existence actually show that most members of a properly chosen random space have the desired properties. We may thus expect that it would not be too difficult to find one such member. Although this is true in principle, it is certainly not

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

practical to check all possibilities; it is thus common to define an explicit construction of a combinatorial object as one that can be performed efficiently, say, in time that is polynomial in the parameters of the object.

Let us illustrate this notion by one of the best known open problems in the area of explicit constructions, the problem of constructing explicit *Ramsey graphs* . The first example given in Chapter 1 is the proof of Erdős that for every n there are graphs on n vertices containing neither a clique nor an independent set on $2\log_2 n$ vertices. This proof is an existence proof; but can we actually describe such graphs explicitly? Erdős offered a prize of \$500 for the explicit construction of an infinite family of graphs, in which there is neither a clique nor an independent set of size more than a constant times the logarithm of the number of vertices, for some absolute constant. Of course, we can, in principle, for every fixed n, check all graphs on n vertices until we find a good one, but this does not give an efficient way of producing the desired graphs and hence is not explicit. Although the problem mentioned above received a considerable amount of attention, it is still open. The best known explicit construction appears in Barak, Rao, Shaltiel and Wigderson (2012), improving an elegant earlier construction of Frankl and Wilson (1981). This gives explicit graphs on n vertices which contain neither a clique nor an independent set on more than $2^{\log n^{o(1)}}$ vertices.

Although the problem of constructing explicit Ramsey graphs is still open, there are several other problems for which explicit constructions are known. In this chapter we present a few examples and discuss briefly some of their algorithmic applications. We then describe several seemingly unrelated properties of a graph, which all turn out to be equivalent. All these are properties of the random graph, and it is thus common to call a graph (more precisely, a sequence of graphs) that satisfies these properties *quasirandom*. The equivalence of all these properties enables one to show, in several cases, that certain explicit graphs have many pseudorandom properties by merely showing that they possess one of them. In Section 9.4 we give the Szemerédi Regularity Lemma, which is now recognized as a fundamental result for understanding large graphs. Indeed, Endre Szemerédi was awarded the 2012 Abel Prize in part for the importance of this result. In a rough sense, the Szemerédi Regularity Lemma tells us that *any* large graph can be regarded as a kind of random graph. We culminate this chapter with Section 9.5, introducing the notion of Graphons. These objects reflect the limiting behavior of a sequence of similar looking graphs. They provide a completion of the family of finite graphs, much as the real numbers provide a completion of the rationals.

Whew! Lets get started.

9.1 THE QUADRATIC RESIDUE TOURNAMENTS

Recall that a *tournament* on a set V of n players is an orientation $T = (V, E)$ of the set of edges of the complete graph on the set of vertices V. If (x, y) is a directed edge, we say that x *beats* y. Given a permutation π of the set of players, a (directed) edge (x, y) of the tournament is *consistent* with π if x precedes y in π. If π is viewed as a ranking of the players, then it is reasonable to try and find rankings with as many consistent

arcs as possible. Let $c(\pi, T)$ denote the number of arcs of T that are consistent with π, and define $c(T) = \max(c(\pi, T))$, where the maximum is taken over all permutations π of the set of vertices of T. For every tournament T on n players, if $\pi = 1, 2, \ldots, n$ and $\pi' = n, n-1, \ldots, 1$, then $c(\pi, T) + c(\pi', T) = \binom{n}{2}$. Therefore $c(T) \geq \frac{1}{2}\binom{n}{2}$. In fact, it can be shown that for every such T, $c(T) \geq \frac{1}{2}\binom{n}{2} + \Omega(n^{3/2})$. On the other hand, a simple probabilistic argument shows that there are tournaments T on n players for which $c(T) \leq (1 + o(1))\frac{1}{2}\binom{n}{2}$. (The best known estimate, which gives the right order of magnitude for the largest possible value of the difference of $c(T) - \frac{1}{2}\binom{n}{2}$ is more complicated and was given by de la Vega (1983), where he showed that there are tournaments T on n players for which $c(T) \leq \frac{1}{2}\binom{n}{2} + O(n^{3/2})$.)

Can we describe explicitly tournaments T on n vertices in which $c(T) \leq (1 + o(1))\frac{1}{2}\binom{n}{2}$? This problem was mentioned by Erdős and Moon (1965) and by Spencer (1985b). It turns out that several such constructions can be given. Let us describe one.

Let $p \equiv 3 \pmod 4$ be a prime, and let $T = T_p$ be the tournament whose vertices are all elements of the finite field $GF(p)$ in which (i, j) is a directed edge iff $i - j$ is a quadratic residue. [Since $p \equiv 3 \pmod 4$, -1 is a quadratic nonresidue modulo p and hence T_p is a well-defined tournament.]

Theorem 9.1.1 *For the tournaments T_p described above,*

$$c(T_p) \leq \frac{1}{2}\binom{p}{2} + O(p^{3/2} \log p).$$

In order to prove this theorem we need some preparations. Let χ be the quadratic residue character defined on the elements of the finite field $GF(p)$ by $\chi(y) = y^{(p-1)/2}$. Equivalently, $\chi(y)$ is 1 if y is a nonzero square, 0 if y is 0, and -1 otherwise. Let $D = (d_{ij})_{i,j=0}^{p-1}$ be the $p \times p$ matrix defined by $d_{ij} = \chi(i - j)$.

Fact 1 *For every two distinct j and l, $\sum_{i \in GF(p)} d_{ij} d_{il} = -1$.*

Proof.

$$\sum_i d_{ij} d_{il} = \sum_i \chi(i-j)\chi(i-l) = \sum_{i \neq j,l} \chi(i-j)\chi(i-l)$$
$$= \sum_{i \neq j,l} \chi\left(\frac{i-j}{i-l}\right) = \sum_{i \neq j,l} \chi\left(1 + \frac{l-j}{i-l}\right).$$

As i ranges over all elements of $GF(p)$ besides j and l, the quantity $(1 + (l-j)/(i-l))$ ranges over all elements of $GF(p)$ besides 0 and 1. Since the sum of $\chi(r)$ over all r in $GF(p)$ is 0, this implies that the right-hand side of the last equation is $0 - \chi(0) - \chi(1) = -1$, completing the proof of the fact. ■

For two subsets A and B of $GF(p)$, let $e(A,B)$ denote the number of directed edges of T_p that start in a vertex of A and end in a vertex of B. By the definition of the matrix D, it follows that

$$\sum_{i\in A}\sum_{j\in B} d_{ij} = e(A,B) - e(B,A).$$

The following lemma is proved in Alon (1986b).

Lemma 9.1.2 *For any two subsets A and B of $GF(p)$,*

$$\left|\sum_{i\in A}\sum_{j\in B} d_{ij}\right| \le |A|^{1/2}|B|^{1/2}p^{1/2}.$$

Proof. By the Cauchy–Schwarz Inequality and by the fact above,

$$\begin{aligned}\left(\sum_{i\in A}\sum_{j\in B} d_{ij}\right)^2 &\le |A|\sum_{i\in A}\left(\sum_{j\in B} d_{ij}\right)^2\\ &\le |A|\sum_{i\in GF(p)}\left(\sum_{j\in B} d_{ij}\right)^2\\ &= |A|\sum_{i\in GF(p)}\left(|B| + 2\sum_{j<l\in B} d_{ij}d_{il}\right)\\ &= |A||B|p + 2|A|\sum_{j<l\in B}\sum_{i\in GF(p)} d_{ij}d_{il}\\ &\le |A||B|p,\end{aligned}$$

completing the proof of the lemma. ■

Proof. [Theorem 9.1.1]. Let r be the smallest integer satisfying $2^r \ge p$. Let $\pi = \pi_1, \ldots, \pi_p$ be an arbitrary permutation of the vertices of T_p, and define $\pi' = \pi_p, \ldots, \pi_1$. We must show that $c(\pi, T_p) \le \frac{1}{2}\binom{p}{2} + O(p^{3/2}\log p)$ or, equivalently, that $c(\pi, T_p) - c(\pi', T_p) \le O(p^{3/2}\log p)$. Let a_1 and a_2 be two integers satisfying $p = a_1 + a_2$ and $a_1 \le 2^{r-1}, a_2 \le 2^{r-1}$. Let A_1 be the set of the first a_1 vertices in the permutation π, and let A_2 be the set of the last a_2 vertices in π. By Lemma 9.1.2

$$e(A_1,A_2) - e(A_2,A_1) \le (a_1a_2p)^{1/2} \le 2^{r-1}p^{1/2}.$$

Next, let $a_{11}, a_{12}, a_{21}, a_{22}$ be integers each of which does not exceed 2^{r-2} such that $a_1 = a_{11} + a_{12}$ and $a_2 = a_{21} + a_{22}$. Let A_{11} be the subset of A_1 consisting of those a_{11} elements of A_1 that appear first in π, and let A_{12} be the set of the a_{12} remaining

elements of A_1. The partition of A_2 into the two sets A_{21} and A_{22} is defined similarly. By applying Lemma 9.1.2, we obtain

$$\begin{aligned} & e(A_{11}, A_{12}) - e(A_{12}, A_{11}) + e(A_{21}, A_{22}) - e(A_{22}, A_{21}) \\ & \quad \leq (a_{11}a_{12}p)^{1/2} + (a_{21}a_{22}p)^{1/2} \\ & \quad \leq 2 \cdot 2^{r-2}p^{1/2}. \end{aligned}$$

Continuing in the same manner, we obtain, in the ith step, a partition of the set of vertices into 2^i blocks, each consisting of at most 2^{r-i} consecutive elements in the permutation π. This partition is obtained by splitting each block in the partition corresponding to the previous step into two parts. By applying Lemma 9.1.2 to each such pair $A_{\varepsilon 1}, A_{\varepsilon 2}$ (where ε is a vector of length $i-1$ with $\{1,2\}$-entries) and by summing, we conclude that the sum over all these 2^{i-1} vectors ε of the differences $e(A_{\varepsilon 1}, A_{\varepsilon 2}) - e(A_{\varepsilon 2}, A_{\varepsilon 1})$ does not exceed

$$2^{i-1}2^{r-i}p^{1/2} \leq 2^{r-1}p^{1/2}.$$

Observe that the sum of the left-hand sides of all these inequalities as i ranges from 1 to r is precisely the difference $c(\pi, T_p) - c(\pi', T_p)$. Therefore, by summing we obtain

$$c(\pi, T_p) - c(\pi', T_p) \leq 2^{r-1}p^{1/2}r = O(p^{3/2} \log\, p),$$

completing the proof. ■

We note that any antisymmetric matrix with $\{1,-1\}$-entries in which each two columns are roughly orthogonal can be used to give a construction of a tournament as above. Some related results appear in Frankl, Rödl and Wilson (1988). The tournaments T_p, however, have stronger pseudorandom properties than do some of these other tournaments. For example, for every $k \leq \frac{1}{4}\log\, p$, and for every set S of k vertices of T_p, the number of vertices of T_p that beat all the members of S is $(1+o(1))p/2^k$. This was proved by Graham and Spencer (1971) by applying Weil's famous theorem known as the Riemann hypotheses for curves over finite fields, see Weil (1948). Taking a sufficiently large p, this supplies an explicit construction for the Schütte problem mentioned in Chapter 1.

9.2 EIGENVALUES AND EXPANDERS

A graph $G = (V, E)$ is called an (n, d, c)-*expander* if it has n vertices, the maximum degree of a vertex is d, and, for every set of vertices $W \subset V$ of cardinality $|W| \leq n/2$, the inequality $|N(W)| \geq c|W|$ holds, where $N(W)$ denotes the set of all vertices in $V \setminus W$ adjacent to some vertex in W. We note that sometimes a slightly different definition is used, but the difference is not substantial. Expanders share many of the properties of sparse random graphs and are the subject of an extensive literature.

A family of *linear expanders of density d and expansion c* is a sequence $\{G_i\}_{i=1}^{\infty}$, where G_i is an (n_i, d, c)-expander and n_i tends to infinity as i tends to infinity.

Such a family is the main component of the parallel sorting network of Ajtai, Komlós and Szemerédi (1983) and can be used for constructing certain fault-tolerant linear arrays. It also forms the basic building block used in the construction of graphs with special connectivity properties and a small number of edges. Some other examples of the numerous applications of these graphs to various problems in theoretical computer science can be found, for example, in Alon (1986b) and references therein.

It is not too difficult to prove the existence of a family of linear expanders using probabilistic arguments. This was first done by Pinsker (1973). An explicit construction is much more difficult to find and was first given by Margulis (1973). This construction was later improved by various authors; most known constructions are Cayley graphs of certain groups of matrices, and their expansion properties are proved by estimating the eigenvalues of the adjacency matrices of the graphs and by relying on the close correspondence between the expansion properties of a graph and its spectral properties. This correspondence was first studied, independently, by Tanner (1984) and Alon and Milman (1984). Since it is somewhat simpler for the case of regular graphs, we restrict our attention here to this case.

Let $G = (V, E)$ be a d-regular graph, and let $A = A_G = (a_{uv})_{u,v \in V}$ be its adjacency matrix given by $a_{uv} = 1$ if $uv \in E$ and $a_{uv} = 0$ otherwise. Since G is d-regular, the largest eigenvalue of A is d, corresponding to the all-1 eigenvector. Let $\lambda = \lambda(G)$ denote the second largest eigenvalue of G. For two (not necessarily disjoint) subsets B and C of V, let $e(B, C)$ denote the number of ordered pairs (u, v), where $u \in B$, $v \in C$, and uv is an edge of G. (Note that, if B and C are disjoint, this is simply the number of edges of G that connect a vertex of B with a vertex of C.)

Theorem 9.2.1 *For every partition of the set of vertices V into two disjoint subsets B and C*

$$e(B, C) \geq \frac{(d - \lambda)|B||C|}{n}.$$

Proof. Put $|V| = n$, $b = |B|$, and $c = |C| = n - b$. Let $D = dI$ be the $n \times n$ scalar matrix with the degree of regularity of G on its diagonal. Observe that, for any real vector x of length n (considered as a function $x : V \to \mathbb{R}$), we have

$$\begin{aligned}\langle (D - A)x, x \rangle &= \sum_{u \in V} \left(d(x(u))^2 - \sum_{v : uv \in E} x(v)x(u) \right) \\ &= d \sum_{u \in V} (x(u))^2 - 2 \sum_{uv \in E} x(v)x(u) = \sum_{uv \in E} (x(v) - x(u))^2.\end{aligned}$$

Define, now, a vector x by $x(v) = -c$ if $v \in B$ and $x(v) = b$ if $v \in C$. Note that A and $D - A$ have the same eigenvectors and that the eigenvalues of $D - A$ are precisely $d - \mu$, as μ ranges over all eigenvalues of A. Note, also, that $\sum_{v \in V} x(v) = 0$; that is, x is orthogonal to the constant vector, which is the eigenvector of the smallest

eigenvalue of $D - A$. Since $D - A$ is a symmetric matrix, its eigenvectors are orthogonal to each other and form a basis of the n-dimensional space. It follows that x is a linear combination of the other eigenvectors of $D - A$ and, hence, by the definition of λ and the fact that $d - \lambda$ is the second smallest eigenvalue of $D - A$, we conclude that

$$\langle (D - A)x, x\rangle \geq (d - \lambda)\langle x, x\rangle = (d - \lambda)(bc^2 + cb^2) = (d - \lambda)bcn.$$

By the second paragraph of the proof, the left-hand side of the last inequality is $\sum_{uv \in E}(x(u) - x(v))^2 = e(B, C) \cdot (b + c)^2 = e(B, C) \cdot n^2$. Thus

$$e(B, C) \geq \frac{(d - \lambda)bc}{n},$$

completing the proof. ■

Corollary 9.2.2 *If λ is the second largest eigenvalue of a d-regular graph G with n vertices, then G is an (n, d, c)-expander for $c = (d - \lambda)/2d$.*

Proof. Let W be a set of $w \leq n/2$ vertices of G. By Theorem 9.2.1, there are at least $(d - \lambda)w(n - w)/n \geq (d - \lambda)w/2$ edges from W to its complement. Since no vertex in the complement is adjacent to more than d of these edges, it follows that $|N(W)| \geq (d - \lambda)w/2d$. ■

The estimate for c in the last corollary can in fact be improved to $2(d - \lambda)/(3d - 2\lambda)$, as shown by Alon and Milman (1984). Each of these estimates shows that, if the second largest eigenvalue of G is far from the first, then G is a good expander. The converse of this is also true, although more complicated. This is given in the following result, proved in Alon (1986a), which we state without proof.

Theorem 9.2.3 *If G is a d-regular graph which is an (n, d, c)-expander, then*

$$\lambda(G) \leq d - \frac{c^2}{4 + 2c^2}.$$

The last two results supply an efficient algorithm for approximating the expanding properties of a d-regular graph; we simply compute (or estimate) its second largest eigenvalue. The larger the difference between this eigenvalue and d, the better the expanding properties of G. It is thus natural to ask how far from d this second eigenvalue can be. It is known (see Nilli (1991)) that the second largest eigenvalue of any d-regular graph with diameter k is at least $2\sqrt{d - 1}(1 - O(1/k))$. Therefore, in any infinite family of d-regular graphs, the limsup of the second largest eigenvalue is at least $2\sqrt{d - 1}$. Lubotzky, Phillips and Sarnak (1986), and independently Margulis (1988), gave, for every $d = p + 1$ where p is a prime congruent to 1 modulo 4, explicit constructions of infinite families of d-regular graphs G_i with the second largest eigenvalues $\lambda(G_i) \leq 2\sqrt{d - 1}$. These graphs are Cayley graphs of factor

groups of the group of all 2×2 invertible matrices over a finite field, and their eigenvalues are estimated by applying the results of Eichler and Igusa concerning the Ramanujan conjecture. Eichler's proof relies on Weil's theorem mentioned in the previous section. The nonbipartite graphs G constructed in this manner satisfy a somewhat stronger assertion than $\lambda(G) \leq 2\sqrt{d-1}$. In fact, besides their largest eigenvalue d, they do not have eigenvalues whose absolute value exceeds $2\sqrt{d-1}$. This fact implies some strong pseudorandom properties, as shown in the next results.

Theorem 9.2.4 *Let $G = (V, E)$ be a d-regular graph on n vertices, and suppose the absolute value of each of its eigenvalues but the first one is at most λ. For a vertex $v \in V$ and a subset B of V, denote by $N(v)$ the set of all neighbors of v in G, and let $N_B(v) = N(v) \cap B$ denote the set of all neighbors of v in B. Then, for every subset B of cardinality bn of V*

$$\sum_{v \in V} (|N_B(v)| - bd)^2 \leq \lambda^2 b(1-b)n.$$

Observe that in a random d-regular graph each vertex v would tend to have about bd neighbors in each set of size bn. The above theorem shows that, if λ is much smaller than d, then for most vertices v, $N_B(v)$ is not too far from bd.

Proof. Let A be the adjacency matrix of G, and define a vector $f : V \to \mathbb{R}$ by $f(v) = 1 - b$ for $v \in B$ and $f(v) = -b$ for $v \notin B$. Clearly, $\sum_{v \in V} f(v) = 0$; that is, f is orthogonal to the eigenvector of the largest eigenvalue of A. Therefore

$$\langle Af, Af \rangle \leq \lambda^2 \langle f, f \rangle.$$

The right-hand side of the last inequality is $\lambda^2(bn(1-b)^2 + (1-b)nb^2) = \lambda^2 b(1-b)n$. The left-hand side is

$$\sum_{v \in V} ((1-b)|N_B(v)| - b(d - |N_B(v)|))^2 = \sum_{v \in V} (|N_B(v)| - bd)^2.$$

The desired result follows. ∎

Corollary 9.2.5 *Let $G = (V, E), d, n$, and λ be as in Theorem 9.2.4. Then for every two sets of vertices B and C of G, where $|B| = bn$ and $|C| = cn$, we have*

$$|e(B, C) - cbdn| \leq \lambda \sqrt{bcn}.$$

Proof. By Theorem 9.2.4

$$\sum_{v \in C} (|N_B(v)| - bd)^2 \leq \sum_{v \in V} (|N_B(v)| - bd)^2 \leq \lambda^2 b(1-b)n.$$

Thus, by the Cauchy–Schwarz Inequality

$$|e(B,C) - cbdn| \leq \sum_{v \in C} |N_B(v) - bd| \leq \sqrt{cn} \left(\sum_{v \in C} (|N_B(v)| - bd)^2 \right)^{1/2}$$
$$\leq \sqrt{cn}\lambda\sqrt{b(1-b)n} \leq \lambda\sqrt{bc}\, n. \qquad \blacksquare$$

The special case $B = C$ gives the following result. A slightly stronger estimate is proved in a similar way in Alon and Chung (1988).

Corollary 9.2.6 *Let $G = (V, E)$, d, n, and λ be as in Theorem 9.2.4. Let B be an arbitrary set of bn vertices of G, and let $e(B) = \frac{1}{2}e(B, B)$ be the number of edges in the induced subgraph of G on B. Then*

$$\left| e(B) - \frac{1}{2}b^2 dn \right| \leq \frac{1}{2}\lambda bn.$$

A walk of length l in a graph G is a sequence $v_0, \ldots, v_l$ of vertices of G, where for each $1 \leq i \leq l$, $v_{i-1}v_i$ is an edge of G. Obviously, the total number of walks of length l in a d-regular graph on n vertices is precisely $n \cdot d^l$. Suppose, now , that C is a subset of, say, $n/2$ vertices of G. How many of these walks do not contain any vertex of C? If G is disconnected, it may happen that half of these walks avoid C. However, as shown by Ajtai, Komlós and Szemerédi (1987), there are many fewer such walks if all the eigenvalues of G except the largest are small. This result and some of its extensions have several applications in theoretical computer science, as shown in the above-mentioned paper (see also Cohen and Wigderson (1989)). We conclude this section by stating and proving the result and one of its applications.

Theorem 9.2.7 *Let $G = (V, E)$ be a d-regular graph on n vertices, and suppose that each of its eigenvalues but the first one is at most λ. Let C be a set of cn vertices of G. Then, for every l, the number of walks of length l in G that avoid C does not exceed $(1-c)n((1-c)d + c\lambda)^l$.*

Proof. Let A be the adjacency matrix of G, and let A' be the adjacency matrix of its induced subgraph on the complement of C. We claim that the maximum eigenvalue of A' is at most $(1-c)d + c\lambda$. To prove this claim, we must show that for every vector $f : V \to \mathbb{R}$ satisfying $f(v) = 0$ for each $v \in C$ and $\sum_{v \in V} f(v)^2 = 1$, the inequality $\langle Af, f \rangle \leq (1-c)d + c\lambda$ holds. Let $f_1, f_2, \ldots, f_n$ be an orthonormal basis of eigenvectors of A, where f_i is the eigenvector of λ_i, $\lambda_1 = d$ and each entry of f_1 is $1/\sqrt{n}$. Then, $f = \sum_{i=1}^n c_i f_i$, where $\sum_{i=1}^n c_i^2 = 1$ and

$$c_1 = \sum_{v \in V} \frac{f(v)}{\sqrt{n}} = \sum_{v \in V \setminus C} \frac{f(v)}{\sqrt{n}}$$
$$\leq \left(\sum_{v \in V \setminus C} f(v)^2 \right)^{1/2} \left((1-c)n\frac{1}{n} \right)^{1/2} = \sqrt{1-c},$$

where we used the Cauchy–Schwarz Inequality. Therefore $\sum_{i=2}^{n} c_i^2 \geq c$ and

$$\langle Af, f \rangle = \sum_{i=1}^{n} c_i^2 \lambda_i \leq (1-c)d + c\lambda,$$

supplying the desired estimate for the largest eigenvalue of A'.

Let $\gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_m$ be the eigenvalues of A', where $m = (1-c)n$. By the Perron–Frobenius theorem, it follows that the absolute value of each of them is at most $\gamma_1 \leq (1-c)d + c\lambda$. The total number of walks of length l that avoid C is precisely $< A'^l g, g >$, where g is the all-1 vector indexed by the vertices in $V - C$. By expressing g as a linear combination of the eigenvectors of A', $g = \sum_{i=1}^{m} b_i g_i$, where g_i is the eigenvector of γ_i, we conclude that this number is precisely

$$\sum_{i=1}^{m} b_i^2 \gamma_i^l \leq \gamma_1^l \sum_{i=1}^{m} b_i^2 = m\gamma_1^l \leq m((1-c)d + c\lambda)^l.$$

Substituting $m = (1-c)n$, the desired result follows. ■

A *randomly chosen walk* of length l in a graph G is a walk of length l in G chosen according to a uniform distribution among all walks of that length. Note that, if G is d-regular, such a walk can be chosen by choosing randomly its starting point v_0, and then by choosing, for each $1 \leq i \leq l$, v_i randomly among the d neighbors of v_{i-1}.

Corollary 9.2.8 *Let $G = (V, E)$, d, n, λ, C, and c be as in Theorem 9.2.7, and suppose*

$$(1-c)d + c\lambda \leq \frac{d}{\sqrt{2}}.$$

Then, for every l, the probability that a randomly chosen walk of length l in G avoids C is at most $2^{-l/2}$.

Proof. The number of walks of length l in G that avoid C is at most $(1-c)n((1-c)d + c\lambda)^l \leq nd^l 2^{-l/2}$, by Theorem 9.2.7. Since the total number of walks is nd^l, the desired result follows. ■

The results above are useful for amplification of probabilities in randomized algorithms. Although such an amplification can be achieved for any Monte Carlo algorithm, we prefer, for simplicity, to consider one representative example: the primality testing algorithm of Rabin (1980).

For an odd integer q, define two integers a and b by $q - 1 = 2^a b$, where b is odd. An integer x, $1 \leq x \leq q-1$, is called a witness (for the nonprimality of q) if for the sequence $x_0, \ldots, x_a$ defined by $x_0 = x^b \pmod{q}$ and $x_i = x_{i-1}^2 \pmod{q}$ for $1 \leq i \leq a$ either $x_a \neq 1$ or there is an i such that $x_i \neq -1, 1$ and $x_{i+1} = 1$. One can show that, if q is a prime, then there are no such witnesses for q, whereas if q is an odd nonprime, then at least half of the numbers between 1 and $q-1$ are witnesses for q. (In fact, at least three-fourths are witnesses, as shown by Rabin.) This suggests the following

randomized algorithm for testing whether an odd integer q is a prime (for even integers there is a simpler algorithm!).

Choose, randomly, an integer x between 1 and $q-1$, and check if it is a witness. If it is, report that q is not a prime. Otherwise, report that q is a prime.

Observe that, if q is a prime, the algorithm certainly reports it is a prime, whereas if q is not a prime, the probability that the algorithm makes a mistake and reports it as a prime is at most $1/2$. What if we wish to reduce the probability of making such a mistake? Clearly, we can simply repeat the algorithm. If we repeat it l independent times, then the probability of making an error (i.e., reporting a nonprime as a prime) decreases to $1/2^l$. However, the number of random bits required for this procedure is $l \cdot \log(q-1)$.

Suppose we wish to use fewer random bits. By applying the properties of a randomly chosen walk on an appropriate graph, proved in the last two results, we can obtain the same estimate for the error probability by using only $\log(q-1) + O(l)$ random bits. This is done as follows.

Let G be a d-regular graph with $q-1$ vertices, labeled by all integers between 1 and $q-1$. Suppose G has no eigenvalue, but the first one, which exceeds λ, and suppose that

$$\frac{d+\lambda}{2} \leq \frac{d}{\sqrt{2}}. \tag{9.1}$$

Now choose randomly a walk of length $2l$ in the graph G, and check, for each of the numbers labeling its vertices, if it is a witness. If q is a nonprime, then at least half of the vertices of G are labeled by witnesses. Hence, by Corollary 9.2.8 and by (9.1), the probability that no witness is on the walk is at most $2^{-2l/2} = 2^{-l}$. Thus we obtain the same reduction in the error probability as the one obtained by choosing l independent witnesses. Let us estimate the number of random bits required for choosing such a random walk.

The known constructions of expanders given by Lubotzky et al. (1986) or Margulis (1988) give explicit families of graphs with degree d and with $\lambda \leq 2\sqrt{d-1}$, for each $d = p+1$, where p is a prime congruent to 1 modulo 4. [We note that these graphs will not have exactly $q-1$ vertices but this does not cause any real problem, as we can take a graph with n vertices, where $q-1 \leq n \leq (1+o(1))(q-1)$, and label its ith vertex by $i \pmod{q-1}$. In this case, the number of vertices labeled by witnesses would still be at least $(\frac{1}{2} + o(1))n$.] One can easily check that, for example, $d = 30$ and $\lambda = 2\sqrt{29}$ satisfy (9.1) and thus we can use a 30-regular graph. The number of random bits required for choosing a random walk of length $2l$ in it is less than $\log(q-1) + 10l + 1$, which is much less than the $l \log(q-1)$ bits that are needed in the repetition procedure.

9.3 QUASIRANDOM GRAPHS

In this section we describe several pseudorandom properties of graphs, which, somewhat surprisingly, turn out to be all equivalent. All the properties are the ones satisfied,

almost surely, by a random graph in which every edge is chosen, independently, with probability $1/2$. The equivalence between some of these properties was first proved by several authors; see Thomason (1987), Frankl et al. (1988), and Alon and Chung (1988), but the first paper in which all of them (and some others) appear is the one by Chung, Graham and Wilson (1989). Our presentation here follows that paper. In order to simplify the presentation, we consider only the case of regular graphs. We state the fuller result at the end of the section, leaving the (very similar) arguments as an exercise.

We first need some notations. For two graphs G and H, let $N^*_G(H)$ be the number of labeled occurrences of H as an induced subgraph of G: that is, the number of adjacency-preserving injections $f : V(H) \to V(G)$ whose image is the set of vertices of an induced copy of H in G. Similarly, $N_G(H)$ denotes the number of labeled copies of H as a (not necessarily induced) subgraph of G. Note that $N_G(H) = \sum_L N^*_G(L)$, where L ranges over all graphs on the set of vertices of H obtained from H by adding to it a (possibly empty) set of edges.

Throughout this section, G always denotes a graph with n vertices. We denote the eigenvalues of its adjacency matrix (taken with multiplicities) by $\lambda_1, \ldots, \lambda_n$, where $|\lambda_1| \geq \ldots \geq |\lambda_n|$. (Since we consider in this section only the eigenvalues of G, we simply write λ_1 and not $\lambda_1(G)$.) Recall also the following notation, used in the previous section: for a vertex v of G, $N(v)$ denotes the set of its neighbors in G. If S is a set of vertices of G, $e(S)$ denotes the number of edges in the induced subgraph of G on S. If B and C are two (not necessarily disjoint) subsets of vertices of G, $e(B, C)$ denotes the number of ordered pairs (b, c), where $b \in B$, $c \in C$, and bc is an edge of G. Thus $e(S) = \frac{1}{2}e(S, S)$.

We can now state the pseudorandom properties considered here. All the properties refer to a graph $G = (V, E)$ with n vertices. Throughout the section, we use the $o(\cdot)$-notation, without mentioning the precise behavior of each $o(\cdot)$. Thus occurrences of two $o(1)$, say, need not mean that both are identical but only mean that, if we consider a family of graphs G and let their number of vertices n tend to infinity, then each $o(1)$ tends to 0. That is, the notions below apply to a *sequence* of graphs $G = G_n$ (we generally suppress the subscript n for simplicity of presentation) for which the number of vertices n is going to infinity. It is not necessary that we have all, or even all sufficiently large, n. Indeed, for the example of quadratic residue graphs given at the end of the section, the values n must be primes of the form $4k + 1$.

Property $P_1(s)$: For every graph $H(s)$ on s vertices

$$N^*_G(H(s)) = (1 + o(1))n^s 2^{-\binom{s}{2}}.$$

Property P_2: For the cycle $C(4)$ with four vertices $N_G(C(4)) \leq (1 + o(1))(n/2)^4$.

Property P_3: $|\lambda_2| = o(n)$.

Property P_4: For every set S of vertices of G, $e(S) = \frac{1}{4}|S|^2 + o(n^2)$.

Property P_5: For every two sets of vertices B and C, $e(B,C) = \frac{1}{2}|B||C| + o(n^2)$.
Property P_6: $\sum_{u,v\in V} \left||N(u) \cap N(v)| - n/4\right| = o(n^3)$.

It is easy to check that all the properties above are satisfied, almost surely, by a random graph on n vertices. In this section we show that all these properties are equivalent for a regular graph with n vertices and degree of regularity about $n/2$. The fact that the innocent-looking property P_2 is strong enough to imply for such graphs $P_1(s)$ for every $s \geq 1$ is one of the interesting special cases of this result.

Graphs that satisfy any (and thus all) of the properties above are called *quasirandom*. As noted above, the assumption that G is regular can be dropped (at the expense of slightly modifying property P_2 and slightly complicating the proofs).

Theorem 9.3.1 *Let G be a d-regular graph on n vertices, where $d = (\frac{1}{2} + o(1))n$. If G satisfies any one of the seven properties $P_1(4)$,$P_1(s)$ for all $s \geq 1$, P_2, P_3, P_4, P_5, or P_6, then it satisfies all seven.*

Proof. We show that

$$P_1(4) \Rightarrow P_2 \Rightarrow P_3 \Rightarrow P_4 \Rightarrow P_5 \Rightarrow P_6 \Rightarrow P_1(s) \text{ for all } s \geq 1 \quad (\Rightarrow P_1(4)).$$

1. $P_1(4) \Rightarrow P_2$.
   Suppose G satisfies $P_1(4)$. Then $N_G(C(4)) = \sum_L N^*_G(L)$, as L ranges over the four labeled graphs obtained from a labeled $C(4)$ by adding to it a (possibly empty) set of edges. Since G satisfies $P_1(4)$, $N^*_G(L) = (1 + o(1))n^4 2^{-16}$ for each of these graphs L and hence $N_G(C(4)) = (1 + o(1))n^4 2^{-4}$, showing that G satisfies P_2.
2. $P_2 \Rightarrow P_3$.
   Suppose G satisfies P_2 and let A be its adjacency matrix. The trace of A^4 is precisely $\sum_{i=1}^n \lambda_i^4$. On the other hand, it is easy to see that this trace is precisely the number of (labeled) closed walks of length 4 in G, that is, the number of sequences $v_0, v_1, v_2, v_3, v_4 = v_0$ of vertices of G such that $v_i v_{i+1}$ is an edge for each $0 \leq i \leq 3$. This number is $N_G((C(4))$ plus the number of such sequences in which $v_2 = v_0$, which is nd^2, plus the number of such sequences in which $v_2 \neq v_0$ and $v_3 = v_1$, which is $nd(d-1)$. Thus

$$\sum_{i=1}^n \lambda_i^4 = d^4 + \sum_{i=2}^n \lambda_i^4 = (1 + o(1))(n/2)^4 + \sum_{i=2}^n \lambda_i^4$$
$$= N_G(C(4)) + O(n^3) = (1 + o(1))(n/2)^4.$$

   It follows that $\sum_{i=2}^n \lambda_i^4 = o(n^4)$, and hence that $|\lambda_2| = o(n)$, as needed.

3. $P_3 \Rightarrow P_4$.
   This is an immediate consequence of Corollary 9.2.6.
4. $P_4 \Rightarrow P_5$.
   Suppose G satisfies P_4. We first claim that it satisfies property P_5 for disjoint sets of vertices B and C. Indeed, if B and C are disjoint, then

$$\begin{aligned} e(B,C) &= e(B \cup C) - e(B) - e(C) \\ &= \frac{1}{4}(|B| + |C|)^2 - \frac{1}{4}|B|^2 - \frac{1}{4}|C|^2 + o(n^2) \\ &= \frac{1}{2}|B||C| + o(n^2), \end{aligned}$$

   proving the claim.
   In case B and C are not disjoint, we have

$$e(B,C) = e(B \setminus C, C \setminus B) + e(B \cap C, C \setminus B) + e(B \cap C, B \setminus C) + 2e(B \cap C).$$

   Put $|B| = b$, $|C| = c$, and $|B \cap C| = x$. By the above expression for $e(B,C)$ and by the fact that G satisfies P_4 and P_5 for disjoint B and C, we get

$$\begin{aligned} e(B,C) &= \frac{1}{2}(b-x)(c-x) + \frac{1}{2}x(c-x) + \frac{1}{2}x(b-x) + \frac{2}{4}x^2 + o(n^2) \\ &= \frac{1}{2}bc + o(n^2) = \frac{1}{2}|B||C| + o(n^2), \end{aligned}$$

   showing that G satisfies P_5.
5. $P_5 \Rightarrow P_6$.
   Suppose that G satisfies P_5, and recall that G is d-regular, where $d = (\frac{1}{2} + o(1))n$. Let v be a fixed vertex of G, and let us estimate the sum

$$\sum_{u \in V, u \neq v} \left| |N(u) \cap N(v)| - \frac{n}{4} \right|.$$

   Define

$$B_1 = \left\{ u \in V, u \neq v : \ |N(u) \cap N(v)| \geq \frac{n}{4} \right\}$$

   and similarly

$$B_2 = \left\{ u \in V, u \neq v : \ |N(u) \cap N(v)| < \frac{n}{4} \right\}.$$

   Let C be the set of all neighbors of v in G. Observe that

$$\begin{aligned} \sum_{u \in B_1} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| &= \sum_{u \in B_1} |N(u) \cap N(v)| - |B_1|\frac{n}{4} \\ &= e(B_1, C) - |B_1|\frac{n}{4}. \end{aligned}$$

Since G satisfies P_5, and since $d = (\frac{1}{2} + o(1))n$, the last difference is $\frac{1}{2}|B_1|d + o(n^2) - |B_1|n/4 = o(n^2)$.
A similar argument implies that

$$\sum_{u \in B_2} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| = o(n^2).$$

It follows that, for every vertex v of G,

$$\sum_{u \in V, u \neq v} \left| |N(u) \cap N(v)| - \frac{n}{4} \right| = o(n^2),$$

and by summing over all vertices v we conclude that G satisfies property P_6.

6. $P_6 \Rightarrow P_1(s)$ for all $s \geq 1$.
Suppose $G = (V, E)$ satisfies P_6. For any two distinct vertices u and v of G, let $a(u, v)$ be 1 if $uv \in E$ and 0 otherwise. Also, define $s(u, v) = |\{w \in V : a(u, w) = a(v, w)\}|$. Since G is $d = (\frac{1}{2} + o(1))n$-regular,

$$s(u, v) = 2|N(u) \cap N(v)| + n - 2d = 2|N(u) \cap N(v)| + o(n).$$

Therefore the fact that G satisfies P_6 implies that

$$\sum_{u,v \in V} \left| s(u, v) - \frac{n}{2} \right| = o(n^3). \tag{9.2}$$

Let $H = H(s)$ be an arbitrary fixed graph on s vertices, and put $N_s = N^*_G(H(s))$. We must show that

$$N_s = (1 + o(1))n^s 2^{-\binom{s}{2}}.$$

Denote the vertex set of $H(s)$ by $\{v_1, \ldots, v_s\}$. For each $1 \leq r \leq s$, put $V_r = \{v_1, \ldots, v_r\}$, and let $H(r)$ be the induced subgraph of H on V_r. We prove, by induction on r, that for $N_r = N^*_G(H(r))$,

$$N_r = (1 + o(1))n_{(r)} 2^{-\binom{r}{2}}, \tag{9.3}$$

where $n_{(r)} = n(n-1)\cdots(n-r+1)$.
This is trivial for $r = 1$. Assuming it holds for r, where $1 \leq r < s$, we prove it for $r + 1$. For a vector $\alpha = (\alpha_1, \ldots, \alpha_r)$ of distinct vertices of G, and for a vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_r)$ of $(0, 1)$-entries, define

$$f_r(\alpha, \varepsilon) = |\{v \in V : v \neq \alpha_1, \ldots, \alpha_r \text{ and } a(v, \alpha_j) = \varepsilon_j \text{ for all } 1 \leq j \leq r\}|.$$

Clearly, N_{r+1} is the sum of the N_r quantities $f_r(\alpha, \varepsilon)$ in which $\varepsilon_j = a(v_{r+1}, v_j)$ and α ranges over all N_r induced copies of $H(r)$ in G.

Observe that altogether there are precisely $n_{(r)}2^r$ quantities $f_r(\alpha, \varepsilon)$. It is convenient to view $f_r(\alpha, \varepsilon)$ as a random variable defined on a sample space of $n_{(r)}2^r$ points, each having an equal probability. To complete the proof, we compute the expectation and the variance of this random variable. We show that the variance is so small that most of the quantities $f_r(\alpha, \varepsilon)$ are very close to the expectation, and thus obtain a sufficiently accurate estimate for N_{r+1}, which is the sum of N_r such quantities.

We start with the simple computation of the expectation $\mathrm{E}[f_r]$ of $f_r(\alpha, \varepsilon)$. We have

$$\begin{aligned}\mathrm{E}[f_r] &= \frac{1}{n_{(r)}2^r}\sum_{\alpha,\varepsilon} f_r(\alpha,\varepsilon) = \frac{1}{n_{(r)}2^r}\sum_{\alpha}\sum_{\varepsilon} f_r(\alpha,\varepsilon)\\ &= \frac{1}{n_{(r)}2^r}\sum_{\alpha}(n-r) = \frac{n-r}{2^r},\end{aligned}$$

where we used the fact that every vertex $v \neq \alpha_1, \ldots, \alpha_r$ defines ε uniquely.

Next, we estimate the quantity S_r defined by

$$S_r = \sum_{\alpha,\varepsilon} f_r(\alpha,\varepsilon)(f_r(\alpha,\varepsilon)-1).$$

We claim that

$$S_r = \sum_{u\neq v} s(u,v)_{(r)}. \tag{9.4}$$

To prove this claim, observe that S_r can be interpreted as the number of ordered triples $(\alpha, \varepsilon, (u, v))$, where $\alpha = (\alpha_1, \ldots, \alpha_r)$ is an ordered set of r distinct vertices of G, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_r)$ is a binary vector of length r, and u, v is an ordered pair of additional vertices of G so that

$$a(u,\alpha_k) = a(v,\alpha_k) = \varepsilon_k \text{ for all } 1 \leq k \leq r.$$

For each fixed α and ε, there are precisely $f_r(\alpha, \varepsilon)(f_r(\alpha, \varepsilon) - 1)$ choices for the pair (u, v) and hence S_r counts the number of these triples.

Now, let us compute this number by first choosing u and v. Once u, v are chosen, the additional vertices $\alpha_1, \ldots, \alpha_r$ must all belong to the set $\{w \in V : a(u, w) = a(v, w)\}$. Since the cardinality of this set is $s(u, v)$, it follows that there are $s(u, v)_{(r)}$ choices for $\alpha_1, \ldots \alpha_r$. Once these are chosen, the vector ε is determined and thus (9.4) follows.

We next claim that (9.2) implies

$$\sum_{u\neq v} s(u,v)_{(r)} = (1+o(1))n^{r+2}2^{-r}. \tag{9.5}$$

To prove this claim, define $\varepsilon_{uv} = s(u, v) - n/2$. Observe that, by (9.2), $\sum_{u\neq v}|\varepsilon_{uv}| = o(n^3)$ and $|\varepsilon_{uv}| \leq n/2 \leq n$ for each u, v. Hence, for every fixed $a \geq 1$

$$\sum_{u\neq v}|\varepsilon_{uv}|^a \le n^{a-1}\sum_{u\neq v}|\varepsilon_{uv}| = o(n^{a+2}).$$

This implies that

$$\begin{aligned}
&\sum_{u\neq v} s(u,v)_{(r)}\\
&\quad=\sum_{u\neq v}\left(\frac{n}{2}+\varepsilon_{uv}\right)_{(r)}\\
&\quad=\sum_{k=0}^{r}\sum_{u\neq v}c_k\left(\frac{n}{2}\right)^k\varepsilon_{uv}^{r-k} \qquad \text{(for appropriate constants } c_k\text{)}\\
&\quad=\left(\frac{n}{2}\right)^r n_{(2)}+\sum_{k=0}^{r-1}\sum_{u\neq v}c_k\left(\frac{n}{2}\right)^k\varepsilon_{uv}^{r-k}\\
&\quad\le\left(\frac{n}{2}\right)^r n_{(2)}+\sum_{k=0}^{r-1}\sum_{u\neq v}|c_k|n^k|\varepsilon_{uv}|^{r-k}\\
&\quad\le n^{r+2}2^{-r}+c\sum_{k=0}^{r-1}n^k\sum_{u\neq v}|\varepsilon_{uv}|^{r-k} \qquad \text{(for an appropriate constant } c\text{)}\\
&\quad\le n^{r+2}2^{-r}+c\sum_{k=0}^{r-1}n^k\cdot o(n^{r-k+2})\\
&\quad= n^{r+2}2^{-r}(1+o(1)),
\end{aligned}$$

implying (9.5).
By (9.4) and (9.5), $S_r = (1+o(1))n^{r+2}2^{-r}$. Therefore

$$\begin{aligned}
&\sum_{\alpha,\varepsilon}(f_r(\alpha,\varepsilon)-\mathrm{E}[f_r])^2\\
&\quad=\sum_{\alpha,\varepsilon}f_r^2(\alpha,\varepsilon)-\sum_{\alpha,\varepsilon}\mathrm{E}[f_r]^2,\\
&\sum_{\alpha,\varepsilon}(f_r^2(\alpha,\varepsilon)-f_r(\alpha,\varepsilon))+\sum_{\alpha,\varepsilon}f_r(\alpha,\varepsilon)-n_{(r)}2^r(n-r)^2 2^{-2r}\\
&\quad=S_r+n_{(r)}2^r\mathrm{E}[f_r]-n_{(r)}2^r(n-r)^2 2^{-2r}\\
&\quad=S_r+n_{(r+1)}-n_{(r)}2^r(n-r)^2 2^{-2r}=o(n^{r+2}).
\end{aligned}$$

Recall that N_{r+1} is the summation of N_r quantities of the form $f_r(\alpha,\varepsilon)$. Thus

$$|N_{r+1}-N_r\mathrm{E}[f_r]|^2=\left|\sum_{N_r\ \text{terms}}(f_r(\alpha,\varepsilon)-\mathrm{E}[f_r])\right|^2.$$

By Cauchy–Schwarz , the last expression is at most

$$N_r \sum_{N_r \text{ terms}} (f_r(\alpha,\varepsilon) - \mathrm{E}[f_r])^2 \le N_r \sum_{\alpha,\varepsilon} (f_r(\alpha,\varepsilon) - \mathrm{E}[f_r])^2$$
$$= N_r \cdot o(n^{r+2}) = o(n^{2r+2}).$$

It follows that

$$|N_{r+1} - N_r \mathrm{E}[f_r]| = o(n^{r+1}),$$

and hence, by the induction hypothesis

$$\begin{aligned} N_{r+1} &= N_r \mathrm{E}[f_r] + o(n^{r+1}) \\ &= (1+o(1)) n_{(r)} 2^{-\binom{r}{2}} \cdot (n-r) 2^{-r} + o(n^{r+1}) \\ &= (1+o(1)) n_{(r+1)} 2^{-\binom{r+1}{2}}. \end{aligned}$$

This completes the proof of the induction step and establishes Theorem 9.3.1.

■

There are many examples of families of quasirandom graph sequences. The most widely used is probably the family of Paley graphs G_p defined as follows: For a prime p congruent to 1 modulo 4, let G_p be the graph whose vertices are the integers $0, 1, 2, \ldots, p-1$ in which i and j are adjacent if and only if $i - j$ is a quadratic residue modulo p. The graphs G_p, which are the undirected analogs of the quadratic residue tournaments discussed in Section 9.1, are $(p-1)/2$-regular. For any two distinct vertices i and j of G_p, the number of vertices k that are either adjacent to both i and j or nonadjacent to both is precisely the number of times the quotient $(k-i)/(k-j)$ is a quadratic residue. As k ranges over all numbers between 0 and $p-1$ but i and j, this quotient ranges over all numbers but 1 and 0 and hence it is a quadratic residue precisely $\frac{1}{2}(p-1) - 1$ times. (This is essentially the same assertion as that of the first fact given in the proof of Theorem 9.1.1.) We have thus shown that, for every two vertices i and j of G_p, $s(i,j) = (p-3)/2$, and this, together with the fact that G_p is $(p-1)/2$-regular, easily implies that it satisfies Property P_6. Therefore it is quasirandom. As is the case with the quadratic residue tournaments , G_p satisfies, in fact, some stronger pseudorandom properties that are not satisfied by every quasirandom graph and can be proved by applying Weil's theorem.

Let $p \in (0,1)$. We now give the fuller version of Theorem 9.3.1. We first give a natural generalization of the properties listed earlier.

Property $P_1^p(s)$: For every graph $H(s)$ on s vertices with e edges

$$N_G^*(H(s)) = (1+o(1)) n^s p^e (1-p)^{\binom{s}{2}-e}.$$

Property P_2^p: For the cycle $C(4)$ with four vertices $N_G(C(4)) \le (1+o(1))p^4n^4$.
Property P_3^p: $|\lambda_2| = o(n)$.
Property P_4^p: For every set S of vertices of G, $e(S) = \frac{p}{2}|S|^2 + o(n^2)$.
Property P_5^p: For every two sets of vertices B and C, $e(B,C) = p|B||C| + o(n^2)$.
Property P_6^p: $\sum_{u,v\in V}\left||N(u)\cap N(v)| - p^2n\right| = o(n^3)$.

Theorem 9.3.2 *Let $0 < p < 1$. Let $G = G_n$ be a sequence of graphs on n vertices with $p\binom{n}{2}(1+o(1))$ edges. Assume further that all but $o(n)$ vertices have degree $pn(1+o(1))$. If G satisfies any one of the seven properties $P_1^p(4), P_1^p(s)$ for all $s \ge 1$, P_2^p, P_3^p, P_4^p, P_5^p, and P_6^p, then it satisfies all seven.*

We leave the proof of Theorem 9.3.2, which is quite similar to that of Theorem 9.3.1, as an exercise.

Definition 4 *A sequence of graphs $G = G_n$ satisfying any (and hence all) of the properties given in Theorem 9.3.2 is called a* quasirandom *graph sequence with parameter p.*

9.4 SZEMERÉDI'S REGULARITY LEMMA

In this section we describe a fundamental result, the *Regularity Lemma*, proved by Endre Szemerédi in the 1970s. The original motivation for proving it has been an application in combinatorial number theory, leading, together with several additional deep ideas, to a complete solution of the Erdős–Turán conjecture discussed in Appendix B.2: that is, every set of integers of positive upper density contains arbitrarily long arithmetic progressions. It took some time to realize that the lemma is an extremely powerful tool in extremal graph theory, combinatorics, and theoretical computer science. Stated informally, the regularity lemma asserts that the vertices of *every* large graph can be decomposed into a finite number of parts, so that the edges between almost every pair of parts form a random-looking graph. The power of the lemma lies in the fact it deals with an arbitrary graph, making no assumptions, and yet supplies much useful information about its structure. A detailed survey of the lemma and some of its many variants and fascinating consequences can be found in Komlós and Simonovits (1996).

Let $G = (V, E)$ be a graph. For two disjoint nonempty subsets of vertices $A, B \subset V$, let $e(A,B)$ denote the number of edges of G with one end in A and one in B, and let $d(A,B) = \frac{e(A,B)}{|A||B|}$ denote the *density* of the pair (A,B). For a real $\varepsilon > 0$, a pair (A,B) as above is called ε-regular if, for every $X \subset A$ and $Y \subset B$ that satisfy $|X| \ge \varepsilon|A|$, $|Y| \ge \varepsilon|B|$, the inequality $|d(A,B) - d(X,Y)| \le \varepsilon$ holds. It is not difficult to see that for every fixed positive ε, p, a fixed pair of two sufficiently large disjoint subsets A and B of a random graph $G = G(n,p)$ are very likely to be ε-regular of density roughly p. (This is stated in one of the exercises at the end of the chapter.) Conversely, an ε-regular pair A, B with a sufficiently small positive ε is random-looking in the sense that it shares many properties satisfied by random (bipartite) graphs.

A partition $V = V_0 \cup V_1 \cup \cdots \cup V_k$ of V into pairwise disjoint sets in which V_0 is called the exceptional set is an *equipartition* if $|V_1| = |V_2| = \cdots = |V_k|$. We view the exceptional set as $|V_0|$ distinct parts, each consisting of a single vertex. For two partitions $\mathcal{P}$ and $\mathcal{P}'$ as above, $\mathcal{P}'$ is a *refinement* of $\mathcal{P}$ if every part in $\mathcal{P}$ is a union of some of the parts of $\mathcal{P}'$. By the last comment on the exceptional set, this means, in particular, that, if $\mathcal{P}'$ is obtained from $\mathcal{P}$ by shifting vertices from the other sets in the partition to the exceptional set, then $\mathcal{P}'$ is a refinement of $\mathcal{P}$. An equipartition is called *ε-regular* if $|V_0| \leq \varepsilon|V|$ and all pairs (V_i, V_j) with $1 \leq i < j \leq k$, except at most εk^2 of them, are ε-regular.

Theorem 9.4.1 The Regularity Lemma [Szemerédi (1978)] *For every $\varepsilon > 0$ and every integer t, there exists an integer $T = T(\varepsilon, t)$ so that every graph with at least T vertices has an ε-regular partition $(V_0, V_1, \ldots, V_k)$, where $t \leq k \leq T$.*

The basic idea in the proof is simple. Start with an arbitrary partition of the set of vertices into t disjoint classes of equal sizes (with a few vertices in the exceptional set, if needed, to ensure divisibility by t). Proceed by showing that, as long as the existing partition is not ε-regular, it can be refined in a way that increases the weighted average of the square of the density between a pair of classes of the partition by at least a constant depending only on ε. As this average cannot exceed 1, the process has to terminate after a bounded number of refinement steps. Since in each step we control the growth in the number of parts as well as the number of extra vertices thrown to the exceptional set, the desired result follows. The precise details require some care, and are given in what follows.

Let $G = (V, E)$ be a graph on $|V| = n$ vertices. For two disjoint subsets $U, W \subset V$, define $q(U, W) = \frac{|U||W|}{n^2} d^2(U, W)$. For partitions $\mathcal{U}$ of U and $\mathcal{W}$ of W, define

$$q(\mathcal{U}, \mathcal{W}) = \sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} q(U', W').$$

Finally, for a partition $\mathcal{P}$ of V, with an exceptional set V_0, define $q(\mathcal{P}) = \sum q(U, W)$, where the sum ranges over all unordered pairs of distinct parts U, W in the partition, with each vertex of the exceptional set V_0 forming a singleton part in its own. Therefore, $q(\mathcal{P})$ is a sum of $\binom{k+|V_0|}{2}$ terms of the form $q(U, W)$. The quantity $q(\mathcal{P})$ is called the *index* of the partition $\mathcal{P}$. Since $d^2(U, W) \leq 1$ for all U, W, and since the sum $\sum |U||W|$ over all unordered pairs of distinct parts U, W is at most the number of unordered pairs of vertices, it follows that the index of any partition is smaller than $1/2$.

Lemma 9.4.2

(i) Let U, W be disjoint nonempty subsets of V, and let $\mathcal{U}$ be a partition of U and $\mathcal{W}$ a partition of W. Then $q(\mathcal{U}, \mathcal{W}) \geq q(U, W)$.

(ii) If $\mathcal{P}'$ and $\mathcal{P}$ are partitions of V and $\mathcal{P}'$ is a refinement of $\mathcal{P}$, then $q(\mathcal{P}') \geq q(\mathcal{P})$.

(iii) Suppose $\varepsilon > 0$, and suppose U, W are disjoint nonempty subsets of V and the pair (U, W) is not ε-regular. Then there are partitions $\mathcal{U} = \{U_1, U_2\}$ of U and $\mathcal{W} = \{W_1, W_2\}$ of W so that $q(\mathcal{U}, \mathcal{W}) > q(U, W) + \varepsilon^4 \frac{|U||W|}{n^2}$.

Proof.

(i) Define a random variable Z as follows: Let u be a uniformly chosen random element of U, and let w be a uniformly chosen random element of W. Let $U' \in \mathcal{U}$ and $W' \in \mathcal{W}$ be those members of the partition so that $u \in U', w \in W'$. Then $Z = d(U', W')$.
The expectation of Z is

$$\sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} \frac{|U'||W'|}{|U||W|} d(U', W') = \sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} \frac{|U'||W'|}{|U||W|} \frac{e(U', W')}{|U'||W'|} = d(U, W).$$

By Jensen's Inequality, $\mathrm{E}[Z^2] \geq (\mathrm{E}[Z])^2$, and the desired result follows, as $\mathrm{E}[Z^2] = \frac{n^2}{|U||W|} q(\mathcal{U}, \mathcal{W})$ and $(\mathrm{E}[Z])^2 = d^2(U, W) = \frac{n^2}{|U||W|} q(U, W)$.

(ii) This is an immediate consequence of (i).

(iii) Since the pair (U, W) is not ε-regular, there are subsets $U_1 \subset U, W_1 \subset W$ so that $|U_1| \geq \varepsilon |U|$, $|W_1| \geq \varepsilon |W|$, and $|d(U_1, W_1) - d(U, W)| > \varepsilon$. Put $U_2 = U - U_1, W_2 = W - W_1$, and define the partitions $\mathcal{U} = \{U_1, U_2\}$, $\mathcal{W} = \{W_1, W_2\}$. Let Z be the random variable defined in the proof of part (i). Then, as shown in that proof

$$\mathrm{Var}[Z] = \mathrm{E}[Z^2] - (\mathrm{E}[Z])^2 = \frac{n^2}{|U||W|} (q(\mathcal{U}, \mathcal{W}) - q(U, W)).$$

However, as $\mathrm{E}[Z] = d(U, W)$, it follows that with probability $\frac{|U_1||W_1|}{|U||W|}$, Z deviates from $\mathrm{E}[Z]$ by more than ε, implying that

$$\mathrm{Var}(Z) > \frac{|U_1||W_1|}{|U||W|} \varepsilon^2 \geq \varepsilon^4.$$

This provides the desired result. ∎

Proposition 9.4.3 *Suppose $0 < \varepsilon \leq 1/4$, let $\mathcal{P} = \{V_0, V_1, \ldots, V_k\}$ be an equipartition of V, where V_0 is the exceptional set, $|V_0| \leq \varepsilon n$, and $|V_i| = c$ for all $1 \leq i \leq k$. If $\mathcal{P}$ is not ε-regular, then there exists a refinement $\mathcal{P}' = \{V_0', V_1', \ldots, V_\ell'\}$ of $\mathcal{P}$, in which $k \leq \ell \leq k4^k$, $|V_0'| \leq |V_0| + \frac{n}{2^k}$ all other sets V_i are of the same size, and $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{\varepsilon^5}{2}$.*

Proof. For every pair $1 \leq i < j \leq k$, define a partition $\mathcal{V}_{ij}$ of V_i and $\mathcal{V}_{ji}$ of V_j as follows: If the pair (V_i, V_j) is ε-regular, then the two partitions are trivial. Else, each partition consists of two parts, chosen according to Lemma 9.4.2, part (iii). For each $1 \leq i \leq k$, let $\mathcal{V}_i$ be the partition of V_i obtained by the Venn diagram of all $(k-1)$-partitions $\mathcal{V}_{ij}$. Thus each $\mathcal{V}_i$ has at most 2^{k-1} parts. Let $\mathcal{Q}$ be the partition of V consisting

of all parts of the partitions $\mathcal{V}_i$ together with the original exceptional set V_0. By Lemma 9.4.2, parts (ii) and (iii), and since $\mathcal{P}$ is not ε-regular, we conclude that the index of $\mathcal{Q}$ satisfies

$$q(\mathcal{Q}) \geq q(\mathcal{P}) + \varepsilon k^2 \varepsilon^4 \frac{c^2}{n^2} = q(\mathcal{P}) + \varepsilon^5 \frac{(kc)^2}{n^2} > q(\mathcal{P}) + \frac{\varepsilon^5}{2},$$

where we used the fact that $kc \geq (1-\varepsilon)n \geq 3n/4$. Note that $\mathcal{Q}$ has at most $k2^{k-1}$ parts (besides the exceptional set), but those are not necessarily of equal sizes. Define $b = \lfloor c/4^k \rfloor$ and split every part of $\mathcal{Q}$ arbitrarily into disjoint sets of size b, throwing the remaining vertices in each part, if any, to the exceptional set. This process creates a partition $\mathcal{P}'$ with at most $k4^k$ nonexceptional parts of equal size and a new exceptional set V_0' of size smaller than $|V_0| + k2^{k-1}b < |V_0| + kc/2^k \leq |V_0| + \frac{n}{2^k}$. Moreover, by Lemma 9.4.2, part (ii), the index $q(\mathcal{P}')$ of $\mathcal{P}'$ is at least $q(\mathcal{Q}) > q(\mathcal{P}) + \frac{\varepsilon^5}{2}$, completing the proof. ■

Proof of Theorem 9.4.1. It suffices to prove the lemma for $\varepsilon \leq 1/4$ and t satisfying $2^{t-2} > \frac{1}{\varepsilon^6}$, hence we assume that these inequalities hold. Put $s = \left\lceil \frac{1}{\varepsilon^5} \right\rceil$, and note that for this choice $\frac{1}{2^k} \leq \frac{\varepsilon}{2s}$ for all $k \geq t$. Define $k_0 = t$ and $k_{i+1} = k_i 4^{k_i}$ for all $i \geq 0$. We prove the lemma with $T = k_s$.

Let $G = (V, E)$ be a graph with $|V| = n \geq T$ vertices. Start with an arbitrary partition $\mathcal{P} = \mathcal{P}_0$ of its vertices into $k = k_0 = t$ pairwise disjoint parts, each of size $\lfloor n/t \rfloor$, and let the exceptional set consist of the remaining vertices, if any. Note that their number is less than t, which is (much) smaller than $\varepsilon n/2$. As long as the partition $\mathcal{P}$ we have already defined is not ε-regular, apply Proposition 9.4.3 to refine it to a new equipartition $\mathcal{P}'$ with at most $k4^k$ nonexceptional parts, whose index exceeds that of $\mathcal{P}$ by at least $\frac{\varepsilon^5}{2}$, while the size of the exceptional set increases by at most $\frac{n}{2^k} < \frac{\varepsilon n}{2s}$. As the initial index is nonnegative, and the index never exceeds $1/2$, the process must terminate in at most s steps, yielding an ε-regular partition with at most T nonexceptional parts, and an exceptional set of size smaller than εn. ■

Remark. The proof shows that $T(\varepsilon, \frac{1}{\varepsilon})$ is bounded by a tower of exponents of height roughly $1/\varepsilon^5$. Surprisingly, as shown by Gowers (1997), this tower-type behavior is indeed necessary.

Our next result both gives a good illustration of the near-random nature of ε-regularity and shall play a role in the next section. $N(x)$, as usual, denotes the set of neighbors of x in the graph G. Let H be a graph on vertex set $1, \ldots, s$. Let G be a graph on vertex set V. Let $A_1, \ldots, A_s$ be disjoint subsets of V, each of size m. Let N denote the number of choices of $x_1 \in A_1, \ldots, x_s \in A_s$ such that x_i, x_j are adjacent in G whenever i, j are adjacent in H. (Note: Other x_i, x_j may or may not be adjacent.) Set $p_{ij} = d(A_i, A_j)$.

Theorem 9.4.4 *For all ε, there exists $\gamma = \gamma_H(\varepsilon)$ with the following property: Assume, using the above notation, that (A_i, A_j) is ε-regular for all i, j adjacent in H. Then*

$$\left| Nm^{-s} - \prod_{\{i,j\} \in H} p_{ij} \right| \leq \gamma. \tag{9.6}$$

Further, and critically, we may take γ such that

$$\lim_{\varepsilon \to 0^+} \gamma_H(\varepsilon) = 0. \tag{9.7}$$

The proof has some technicalities, and the reader may take H as a triangle and $p_{12} = p_{13} = p_{23} = \frac{1}{2}$ to get the gist of the argument.

Proof. Set, with foresight, κ such that $(\kappa - \varepsilon)^s \geq \varepsilon$. We say we are in Case 1 if some $p_{ij} \leq \kappa$ (with i, j adjacent in H); otherwise we are in Case 2. We will have $\gamma = \max(\gamma_1, \gamma_2)$ where γ_1, γ_2 handle two cases.

Case 1: Some $p_{ij} \leq \kappa$. Set $\gamma_1 = \kappa$. The product of the p_{ij} over edges $\{i, j\}$ is itself at most κ. There are $p_{ij} m^2 \leq \kappa m^2$ choices of adjacent x_i, x_j and therefore $N \leq \kappa m^2 m^{s-2} = \kappa m^s$. Also $N \geq 0$. Thus (9.6) is satisfied.

Case 2: $p_{ij} \geq \kappa$ for all adjacent i, j. For $1 \leq r \leq s$, we call a choice $x_i \in A_i$, $1 \leq i \leq r$ a partial copy if x_i, x_j are adjacent in G whenever i, j are adjacent in H. Further, we call the choice normal (else abnormal) if the following holds for all $r < l \leq s$: Let U be the set of $u \leq r$ that are adjacent to l in H. Let Y be the intersection of the $N(x_u)$, $u \in U$ and A_l. Then

$$\prod_{u \in U} (p_{ul} - \varepsilon) \leq |Y| m^{-1} \leq \prod_{u \in U} (p_{ul} + \varepsilon). \tag{9.8}$$

Let $x_1, \ldots, x_r$ be a normal partial copy. We say it is destroyed by $x_{r+1} \in A_{r+1}$ if $x_1, \ldots, x_{r+1}$ is a partial copy but is not normal. We claim at most $2s\varepsilon m$ vertices x_{r+1} can destroy $x_1, \ldots, x_r$. How can this occur? Let $l > r + 1$ be adjacent to $r + 1$, and let U, Y be as above (looking only at $x_1, \ldots, x_r$). Then $Y \cap N(x_{r+1})$ would need to be either too big or too small. If more than $2s\varepsilon m$ vertices x_{r+1} destroyed $x_1, \ldots, x_r$, then there would be a set $X \subset A_{r+1}$ of size at least $m\varepsilon$ of such x_{r+1}, all with the same l and with either all $Y \cap N(x_{r+1})$ too big or all too small. Assume the former, the latter being similar. Then $d(X, Y) > p_{r+1,l} + \varepsilon$. But, $|Y| \geq m\varepsilon$ by our choice of κ. From ε-regularity, $|X| \leq m\varepsilon$, as claimed.

The N choices of $x_i \in A_i$, $1 \leq i \leq s$ for which x_i, x_j are adjacent in G whenever i, j are adjacent in H fall into two categories. There are at most $2s^2 \varepsilon m^s$ choices such that x_{r+1} destroys $x_1, \ldots, x_r$ for some r. The other choices are bounded in number between $m^s \prod (p_{ij} - \varepsilon)$ and $m^s \prod (p_{ij} + \varepsilon)$, the products over i, j adjacent in H. Let $f(\varepsilon)$ denote the maximum distance between either of these products and $\prod p_{ij}$. We can then set $\gamma_2 = 2s^2 \varepsilon + f(\varepsilon)$. ∎

9.5 GRAPHONS

As in Section 9.3, we let $N_G(H)$ denote the number of labeled copies of H as a (not necessarily induced) subgraph of G. We set $t(H,G) = N_G(H)n^{-a}$, where H, G have a, n vertices, respectively. This may naturally be interpreted as the proportion of H in G, $0 \le t(H, G) \le 1$ tautologically.

Definition 5 *A sequence of graphs G_n is called a limit sequence if* $\lim_{n\to\infty} t(H, G_n)$ *exists for all finite graphs H.*

Definition 6 *Two limit sequences G_n, G'_n are called equivalent if* $\lim_{n\to\infty} t(H, G_n) = \lim_{n\to\infty} t(H, G'_n)$ *for all finite graphs H. A* graphon *is an equivalence class of limit sequences.*

A graphon is a subtle object, an abstract limit of a convergent (by definition 5) sequences of graphs. (We call a limit sequence G_n a graphon even though, technically, the graphon is the equivalence class.) It is *not* itself an infinite graph, though it may seem like one. It reflects the properties of very large graphs (formally, in a limit sense) of similar nature. The excellent book (Lovász 2012) serves as a general reference to graphons.

Surprisingly, and integral to the strength of this concept, there is a good characterization of graphons. Let $W : [0,1] \times [0,1] \to [0,1]$ be a Lebesgue measurable function with $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$. For each positive integer n, we define a random graph, denoted $G(n, W)$ on vertex set $1, \dots, n$ as follows:

1. Select $x_1, \dots, x_n \in [0,1]$ uniformly and independently.
2. For $i \neq j$, let $\{i,j\}$ be an edge of $G(n, W)$ with probability $W(x_i, x_j)$, and let the events where $\{i,j\}$ are edges be mutually independent.

As an important example, when W is the constant function $W(x,y) = p$, $G(n, W)$ is simply $G(n,p)$. We call W *checkered* if it splits into constant-valued rectangles. More precisely, let $K \ge 1$, let $a_i \ge 0$ for $1 \le i \le K$ with $\sum_{i=1}^{K} a_i = 1$. Decompose $[0,1]$ into intervals I_i, $1 \le i \le K$, of length a_i. Then define $W(x,y) = W(y,x) = p_{ij}$ for $x \in I_i, y \in I_j, x \le y$. For such W, $G(n, W)$ is basically a random multipartite graph with the vertex set split into sets V_i of size $\sim na_i$ and all $\{x,y\}$ with $x \in V_i, y \in V_j$ being adjacent with independent probability p_{ij}.

Let H be a graph on $1, \dots, s$. Set

$$c(H, W) = \int \prod_{\{i,j\}\in H} W(x_i, x_j), \tag{9.9}$$

where the integral is over $x_1, \dots, x_s \in [0,1]$ and the null product is interpreted as one. We leave as an exercise that, with probability 1, the sequence $G(n, W)$ is a limit sequence with

$$\lim_{n\to\infty} t(H, G(n, W)) = c(H, W). \tag{9.10}$$

We say that a graphon G_n is represented by W if

$$\lim_{n\to\infty} t(H, G_n) = c(H, W) \tag{9.11}$$

for every finite H. Observe, from Property $P_1^p(s)$ of Theorem 9.3.2, that a sequence G_n is a graphon represented by the constant function $W(x, y) = p$ if and only if G_n is a quasirandom graph sequence with parameter p, as given by Definition 4.

Theorem 9.5.1 *Every graphon is represented by some W.*

The proof of Theorem 9.5.1 requires some techniques slightly beyond the scope of this chapter. Rather, we prove the following weaker version.

Theorem 9.5.2 *Let $\kappa > 0$, and let a positive integer L be given. Let G_n be an arbitrary graphon. Then there exists a checkered W such that*

$$|\lim_{n\to\infty} t(H, G_n) - c(H, W)| \le \kappa \tag{9.12}$$

for all H with $s \le L$ vertices.

Proof. Let ε be a small positive real and t a large positive integer, as described more fully below. For each G_n, apply the Regularity Lemma (Theorem 9.4.1) to give an ε-regular partition $(V_0, V_1, \dots, V_k)$ with $t \le k \le T = T(\varepsilon, t)$. Take a subsequence on which k is a constant. Further, take a subsequence on which the set of $\{i, j\}$ for which (V_i, V_j) is an ε-regular pair is the same. Further, take a subsequence such that $d(V_i, V_j)$ approaches a limit p_{ij} for all $0 \le i, j \le t$ and such that the proportion of vertices in V_i approaches a limit α_i for all $0 \le i \le t$. Now define a checkered W by splitting $[0, 1]$ into intervals I_i of length α_i, $0 \le i \le t$ and letting W take the constant value p_{ij} on $I_i \times I_j$. Let H be a graph on vertex set $1, \dots, s$ with $s \le L$. We compare $\lim t(H, G_n)$ and $c(H, W)$. Let $\psi : V(H) \to V(G_n)$ with $\psi(i) \in V_{x_i}$. We say $x_1, \dots, x_s \in \{0, 1, \dots, t\}$ is normal (else, abnormal) if the x_i are distinct and nonzero and all (V_{x_i}, V_{x_j}) are ε-regular. The proportion of abnormal ψ is then at most $s\varepsilon$ (some $\psi(i) \in V_0$) plus $\binom{s}{2}\varepsilon$ (some (V_{x_i}, V_{x_j}) are not ε-regular) which is at most $L^2\varepsilon$. We can make this arbitrarily small by adjusting ε. Now suppose $x_1, \dots, x_s$ is normal. Let $N(x_1, \dots, x_s)$ denote the number of choices of $v_i \in V_{x_i}$, $1 \le i \le s$, such that v_i, v_j are adjacent in G whenever i, j are adjacent in H. From Theorem 9.4.4, $N(x_1, \dots, x_s)m^{-L}$ differs from $\prod p_{ij}$ (product over adjacent i, j) by at most γ, where m is the size of each V_{x_i}. Summing over all normal $x_1, \dots, x_s$, the contribution to Nn^{-L} differs from the contribution to $c(H, W)$ by at most γ. From (9.7), we may make γ arbitrarily small by choosing appropriately small ε. The total difference between $t(H, G_n)$ and $c(H, W)$ is then at most $L^2\varepsilon + \gamma$. For any given positive κ, we may find ε so that this is less than κ. ■

Among the applications of graphons is the replacement of asymptotic questions on graphs G_n with analytic questions on functions W. We satisfy ourselves with a typical example.

Let b be the minimal real number so that there exist G_n with $0.7\binom{n}{2} + o(n^2)$ edges and $b\binom{n}{3} + o(n^3)$ triangles. Let b' be the minimum of $\int W(x,y)W(x,z)W(y,z)$ $(x,y,z \in [0,1])$, given that $\int W(x,y) = 0.7$. (We leave as an exercise that the minima b, b' are attained.) Both are tough questions. We will show that they are the *same*, that is, $b = b'$.

The easy part is $b \leq b'$. Let W be such that $c(K_2, W) = 0.7$ and $c(K_3, W) = b'$. Then the random sequence $G_n \sim G(n, W)$ has, with probability unity, $0.7\binom{n}{2} + o(n^2)$ edges and $b'\binom{n}{3} + o(n^3)$ triangles. Hence the minimal possible b has $b \leq b'$.

For the opposite direction, we first give a natural topological result:

Theorem 9.5.3 *Any sequence G_n contains a subsequence which is a limit sequence in the sense of Definition 5.*

Proof. Place all finite graphs into a countable list $H_1, H_2, \ldots$, and set $t_i(G) = t(H_i, G)$. Let SEQ_0 denote the original sequence G_n. As all $t_1(G) \in [0,1]$, we find a subsequence SEQ_1 on which $t_1(G)$ converges. Given SEQ_{i-1}, we find a subsequence of it, denoted SEQ_i, on which $t_i(G)$ converges. Employ diagonalization, letting SEQ_ω be that sequence whose ith term is the ith term of SEQ_i. For each i, SEQ_ω is a subsequence of SEQ_i except for possibly the first $i-1$ terms and hence $t_i(G)$ converges. ∎

Now let G_n be any sequence with $0.7\binom{n}{2} + o(n^2)$ edges and $b\binom{n}{3} + o(n^3)$ triangles. Apply Theorem 9.5.3 to find a limit sequence with the same property. Now apply Theorem 9.5.1 to find W representing that limit sequence. That W has $c(K_2, W) = 0.7$ and $c(K_3, W) = b$. Thus the minimal possible b' has $b' \leq b$.

9.6 EXERCISES

1. By considering a random bipartite three-regular graph on $2n$ vertices obtained by picking three random permutations between the two color classes, prove that there is a $c > 0$ such that for every n there exists a $(2n, 3, c)$-expander.
2. Let $G = (V, E)$ be an (n, d, λ)-graph, suppose n is divisible by k, and let $C : V \to \{1, 2, \ldots, k\}$ be a coloring of V by k colors, so that each color appears precisely n/k times. Prove that there is a vertex of G which has a neighbor of each of the k colors, provided $k\lambda \leq d$.
3. Let $G = (V, E)$ be a graph in which there is at least one edge between any two disjoint sets of size $a + 1$. Prove that for every set Y of $5a$ vertices, there is a set X of at most a vertices, such that for every set Z satisfying $Z \cap (X \cup Y) = \emptyset$ and $|Z| \leq a$, the inequality $|N(Z) \cap Y| \geq 2|Z|$ holds.
4. Prove that for every $\varepsilon > 0$ there exists an $n_0 = n_0(\varepsilon)$ so that for every $(n, n/2, 2\sqrt{n})$- graph $G = (V, E)$ with $n > n_0$, the number of triangles M in G satisfies $|M - n^3/48| \leq \varepsilon n^3$.

5. Let $\varepsilon > 0$, $p \in (0,1)$, $\lambda > 0$. Let $G \sim G(n,p)$, with vertex set V_n. Show that the following property has limiting probability 1 as $n \to \infty$: (A_n, B_n) is ε-regular for *all* disjoint $A_n, B_n \subset V_n$ with $|A| \geq n\lambda$ and $|B| \geq n\lambda$.
6. Combine Turán's theorem with the Regularity Lemma to prove the following result, due to Erdős, Simonovits, and Stone: For every fixed graph H of chromatic number $r > 1$ and every $\varepsilon > 0$, there is an $n_0 = n_0(H, \varepsilon)$ so that if $n > n_0$ then any simple graph with n vertices and at least $(1 - \frac{1}{r-1} + \varepsilon)\binom{n}{2}$ edges contains a copy of H.
7. Let $t'(H,G)$ denote the number of induced copies of H in G, that is, the number of vertex subsets S such that $G|_S$ is isomorphic to H. Show that G_n is a limit sequence if and only if $\lim_{n\to\infty} t'(H, G_n)$ exists for all finite graphs H.
8. Prove Theorem 9.3.2.
9. Prove that the minima b, b' given in Section 9.5 are actually attained.
10. Let $G = G_n$ be a sequence of bipartite graphs with designated parts T_n, B_n each of size n. Let $p \in (0,1)$, and assume $\lim_{n\to\infty} d(T_n, B_n) = p$. Call such a sequence bipartite quasirandom with parameter p if for all $\varepsilon > 0$ the pair (T_n, B_n) is ε-regular for n sufficiently large. State and prove a result analogous to Theorem 9.3.2, giving equivalent notions for bipartite quasirandomness.
11. Prove that for all H, W and $\varepsilon > 0$, there exists $\alpha > 0$ such that

$$\Pr[|t(H, G(n,W)) - c(H,W)| > \varepsilon] < 2e^{-n\alpha}. \tag{9.13}$$

Deduce that with probability 1 the sequence $G(n,W)$ is a limit sequence satisfying (9.10).

THE PROBABILISTIC LENS:
Random Walks

A *vertex-transitive* graph is a graph $G = (V, E)$ such that for any two vertices $u, v \in V$ there is an automorphism of G that maps u into v. A *random walk* of length l in G starting at a vertex v is a randomly chosen sequence $v = v_0, v_1, \ldots, v_l$, where each v_{i+1} is chosen, randomly and independently, among the neighbors of v_i $(0 \leq i < l)$.

The following theorem states that for every vertex-transitive graph G, the probability that a random walk of even length in G ends at its starting point is at least as big as the probability that it ends at any other vertex. Note that the proof requires almost no computation. We note also that the result does not hold for general regular graphs, and the vertex transitivity assumption is necessary.

Theorem 1 *Let $G = (V, E)$ be a vertex-transitive graph. For an integer k and for two (not necessarily distinct) vertices u, v of G, let $P^k(u, v)$ denote the probability that a random walk of length k starting at u ends at v. Then, for every integer k and for every two vertices $u, v \in V$,*

$$P^{2k}(u, u) \geq P^{2k}(u, v).$$

Proof. We need the following simple inequality, sometimes attributed to Chebyshev.

Claim 1 *For every sequence $(a_1, \ldots, a_n)$ of n reals and for any permutation π of $\{1, \ldots, n\}$,*

$$\sum_{i=1}^{n} a_i a_{\pi(i)} \leq \sum_{i=1}^{n} a_i^2.$$

Proof. The inequality follows immediately from the fact that

$$\sum_{i=1}^{n} a_i^2 - \sum_{i=1}^{n} a_i a_{\pi(i)} = \frac{1}{2} \sum_{i=1}^{n} (a_i - a_{\pi(i)})^2 \geq 0.$$

■

Consider, now, a random walk of length $2k$ starting at u. By summing over all the possibilities of the vertex that the walk reaches after k steps, we conclude that for every vertex v

$$P^{2k}(u,v) = \sum_{w \in V} P^k(u,w)P^k(w,v) = \sum_{w \in V} P^k(u,w)P^k(v,w), \tag{1}$$

where the last equality follows from the fact that G is an undirected regular graph.

Since G is vertex-transitive, the two vectors $(P^k(u,w))_{w \in V}$ and $(P^k(v,w))_{w \in V}$ can be obtained from each other by permuting the coordinates. Therefore, by the claim above, the maximum possible value of the sum in the right-hand side of (1) is when $u = v$, completing the proof of the theorem. ■

PART II

TOPICS

10

Random Graphs

> It is six in the morning. The house is asleep. Nice music is playing. I prove and conjecture.
> –Paul Erdős, in a letter to Vera Sós

Let n be a positive integer, $0 \leq p \leq 1$. The random graph $G(n,p)$ is a probability space over the set of graphs on the vertex set $\{1, \ldots, n\}$ determined by

$$\Pr[\{i,j\} \in G] = p$$

with these events mutually independent. This model is often used in the probabilistic method for proving the existence of certain graphs. In this chapter we study the properties of $G(n,p)$ for their own sake.

Random graphs is an active area of research that combines probability theory and graph theory. The subject began in 1960 with the monumental paper "On the Evolution of Random Graphs" by Paul Erdős and Alfred Rényi. The book *Random Graphs* by Bollobás (2001) is the standard source in the field. Another book, also entitled *Random Graphs* by Janson, Łuczak and Ruciński (2000) is also excellent. In this chapter we explore only a few of the many topics in this fascinating area.

There is a compelling dynamic model for random graphs. For all pairs i,j, let $x_{i,j}$ be selected uniformly from $[0,1]$, the choices being mutually independent. Imagine p going from 0 to 1. Originally, all potential edges are "off". The edge from i to j (which we may imagine as a neon light) is turned on when p reaches $x_{i,j}$, and then stays on.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

At $p = 1$, all edges are "on". At time p, the graph of all "on" edges has distribution $G(n, p)$. As p increases, $G(n, p)$ *evolves* from empty to full.

In their original paper, Erdős and Rényi let $G(n, e)$ be the random graph with n vertices and precisely e edges. Again, there is a dynamic model: Begin with no edges and add edges randomly one by one until the graph becomes full. Generally, $G(n, e)$ will have very similar properties as $G(n, p)$ with $p \sim \frac{e}{\binom{n}{2}}$. We will work on the probability model exclusively.

10.1 SUBGRAPHS

The term "random graph" is, strictly speaking, a misnomer. $G(n, p)$ is a probability space over graphs. Given any graph theoretic property A, there will be a probability that $G(n, p)$ satisfies A, which we write $\Pr[G(n, p) \models A]$. When A is monotone, $\Pr[G(n, p) \models A]$ is a monotone function of p. As an instructive example, let A be the event "G is triangle free". Let X be the number of triangles contained in $G(n, p)$. Linearity of Expectation gives

$$\mathrm{E}[X] = \binom{n}{3} p^3 \ .$$

This suggests the parametrization $p = c/n$. Then

$$\lim_{n\to\infty} \mathrm{E}[X] = \lim_{n\to\infty} \binom{n}{3} p^3 = c^3/6 \ .$$

It turns out that the distribution of X is asymptotically Poisson. In particular,

$$\lim_{n\to\infty} \Pr[G(n, p) \models A] = \lim_{n\to\infty} \Pr[X = 0] = e^{-c^3/6} \ .$$

Note that

$$\lim_{c\to 0} e^{-c^3/6} = 1 \ ,$$

$$\lim_{c\to\infty} e^{-c^3/6} = 0 \ .$$

When $p = 10^{-6}/n$, $G(n, p)$ is very unlikely to have triangles, and when $p = 10^6/n$, $G(n, p)$ is very likely to have triangles. In the dynamic view, the first triangles almost always appear at $p = \Theta(1/n)$. If we take a function such as $p(n) = n^{-0.9}$ with $p(n) \gg n^{-1}$, then $G(n, p)$ will almost always have triangles. Occasionally, we will abuse notation and say, for example, that $G(n, n^{-0.9})$ contains a triangle – this meaning that the probability that it contains a triangle approaches 1 as n approaches infinity. Similarly, when $p(n) \ll n^{-1}$, for example, $p(n) = 1/(n \ln n)$, then $G(n, p)$ will almost always not contain a triangle and we abuse notation and say that $G(n, 1/(n \ln n))$ is triangle free. It was a central observation of Erdős and Rényi that many natural graph theoretic properties become true in a very narrow range of p. They made the following key definition:

Definition 7 *$r(n)$ is called a threshold function for a graph theoretic property A if*

1. *When $p(n) \ll r(n)$, $\lim_{n\to\infty} \Pr[G(n,p) \models A] = 0$,*
2. *When $p(n) \gg r(n)$, $\lim_{n\to\infty} \Pr[G(n,p) \models A] = 1$,*

or vice versa.

In our example, $1/n$ is a threshold function for A. Note that the threshold function, when one exists, is not unique. We could equally have said that $10/n$ is a threshold function for A.

Let us approach the problem of $G(n, c/n)$ being triangle free once more. For every set S of three vertices, let B_S be the event that S is a triangle. Then $\Pr[B_S] = p^3$. Then "triangle freeness" is precisely the conjunction $\wedge\overline{B_S}$ over all S. If the B_S were mutually independent, then we *would* have

$$\Pr[\wedge\overline{B_S}] = \prod \Pr[\overline{B_S}] = (1-p^3)^{\binom{n}{3}} \sim e^{-\binom{n}{3}p^3} \to e^{-c^3/6} .$$

The reality is that the B_S are not mutually independent, though when $|S \cap T| \leq 1$, B_S and B_T are mutually independent.

We apply Janson's Inequality, Theorem 8.1.1. In the notation of Section 8.1, $I = \{S \subset V(G) : |S| = 3\}$, and $S \sim T$ if and only if $|S \cap T| = 2$. Here, $\epsilon = p^3 = o(1)$, $\mu = \binom{n}{3} p^3 \sim c^3/6$, and $M = e^{-\mu(1+o(1))} = e^{-c^3/6+o(1)}$. There are $6\binom{n}{4} = O(n^4)$ pairs S, T of triples with $S \sim T$. For each such pair $\Pr[B_S \wedge B_T] = p^5$. Thus

$$\Delta = O(n^4)p^5 = n^{-1+o(1)} = o(1) .$$

When $\Delta = o(1)$, Janson's Inequality sandwiches an asymptotic bound:

$$\lim_{n\to\infty} \Pr[\wedge\overline{B_S}] = \lim_{n\to\infty} M = e^{-c^3/6} .$$

Can we duplicate this success with the property A that G contains no (not necessarily induced) copy of a general given graph H? We use the definitions of balanced and strictly balanced graphs of Section 4.4.

Theorem 10.1.1 *Let H be a* strictly *balanced graph with v vertices, e edges, and a automorphisms. Let $c > 0$ be arbitrary. Let A be the property that G contains no copy of H. Then with $p = cn^{-v/e}$*

$$\lim_{n\to\infty} \Pr[G(n,p) \models A] = \exp[-c^e/a] .$$

Proof. Let A_α, $1 \le \alpha \le \binom{n}{v} v!/a$, range over the edge sets of possible copies of H, and let B_α be the event $G(n,p) \supseteq A_\alpha$. We apply Janson's Inequality. As

$$\lim_{n\to\infty} \mu = \lim_{n\to\infty} \binom{n}{v} v! p^e/a = c^e/a \ ,$$

we find

$$\lim_{n\to\infty} M = \exp\ [-c^e/a] \ .$$

Now we examine (as in Theorem 4.4.2)

$$\Delta = \sum_{\alpha\sim\beta} \Pr[B_\alpha \wedge B_\beta] \ .$$

We split the sum according to the number of *vertices* in the intersection of copies α and β. Suppose they intersect in j vertices. If $j = 0$ or $j = 1$, then $A_\alpha \cap A_\beta = \emptyset$, so that $\alpha \sim \beta$ cannot occur. For $2 \le j \le v$, let f_j be the maximal $|A_\alpha \cap A_\beta|$, where $\alpha \sim \beta$, and α, β intersect in j vertices. As $\alpha \ne \beta$, $f_v < e$. When $2 \le j \le v-1$, the critical observation is that $A_\alpha \cap A_\beta$ is a subgraph of H, and hence, as H is strictly balanced,

$$\frac{f_j}{j} < \frac{e}{v} \ .$$

There are $O(n^{2v-j})$ choices of α, β intersecting in j points, since α, β are determined, except for the order, by $2v - j$ points. For each such α, β

$$\Pr[B_\alpha \wedge B_\beta] = p^{|A_\alpha \cup A_\beta|} = p^{2e-|A_\alpha \cap A_\beta|} \le p^{2e-f_j} \ .$$

Thus

$$\Delta = \sum_{j=2}^{v} O\left(n^{2v-j}\right) O\left(n^{-\frac{v}{e}(2e-f_j)}\right) \ .$$

But

$$2v - j - \frac{v}{e}(2e - f_j) = \frac{vf_j}{e} - j < 0 \ ,$$

so each term is $o(1)$ and hence $\Delta = o(1)$. By Janson's Inequality

$$\lim_{n\to\infty} \Pr[\wedge \overline{B_\alpha}] = \lim_{n\to\infty} M = \exp\ [-c^e/a] \ ,$$

thus completing the proof. ■

10.2 CLIQUE NUMBER

In this section we fix $p = 1/2$ (other values yield similar results) and consider the clique number $\omega(G(n,p))$. For a fixed $c > 0$, let $n, k \to \infty$, so that

$$\binom{n}{k} 2^{-\binom{k}{2}} \to c \ .$$

As a first approximation,

$$n \sim \frac{k}{e\sqrt{2}} \sqrt{2}^k \ ,$$

and

$$k \sim \frac{2 \ln n}{\ln 2} \ .$$

Here, $\mu \to c$ so $M \to e^{-c}$. The Δ term was examined in Section 4.5. For this k, $\Delta = o(E[X]^2)$, and so $\Delta = o(1)$. Therefore

$$\lim_{n,k \to \infty} \Pr[\omega(G(n,p)) < k] = \exp\ [-c] \ .$$

Being more careful, let $n_0(k)$ be the minimum n for which

$$\binom{n}{k} 2^{-\binom{k}{2}} \geq 1 \ .$$

Observe that for this n the left-hand side is $1 + o(1)$. Note that $\binom{n}{k}$ grows, in n, like n^k. For any $\lambda \in (-\infty, +\infty)$, if

$$n = n_0(k) \left[1 + \frac{\lambda + o(1)}{k} \right] ,$$

then

$$\binom{n}{k} 2^{-\binom{k}{2}} = \left[1 + \frac{\lambda + o(1)}{k} \right]^k = e^{\lambda} + o(1) \ ,$$

and so

$$\Pr[\omega(G(n,p)) < k] = e^{-e^{\lambda}} + o(1) \ .$$

As λ ranges from $-\infty$ to $+\infty$, $e^{-e^{\lambda}}$ ranges from 1 to 0. As $n_0(k+1) \sim \sqrt{2} n_0(k)$, the ranges will not "overlap" for different k. More precisely, let K be arbitrarily large and set

$$I_k = \left[n_0(k) \left[1 - \frac{K}{k} \right] , n_0(k) \left[1 + \frac{K}{k} \right] \right] \ .$$

For $k \geq k_0(K)$, $I_{k-1} \cap I_k = \emptyset$. Suppose $n \geq n_0(k_0(K))$. If n lies between the intervals (which occurs for "most" n), which we denote by $I_k < n < I_{k+1}$, then

$$\Pr[\omega(G(n,p)) < k] \leq e^{-e^K} + o(1) ,$$

nearly zero, and

$$\Pr[\omega(G(n,p)) < k+1] \geq e^{-e^{-K}} + o(1) ,$$

nearly 1, so that

$$\Pr[\omega(G(n,p)) = k] \geq e^{-e^{-K}} - e^{-e^K} + o(1) ,$$

nearly 1. When $n \in I_k$, we still have $I_{k-1} < n < I_{k+1}$ so that

$$\Pr[\omega(G(n,p)) = k \text{ or } k-1] \geq e^{-e^{-K}} - e^{-e^K} + o(1) ,$$

nearly 1. As K may be made arbitrarily large, this yields the celebrated two-point concentration theorem on clique number, Corollary 4.5.2 in Section 4.5. Note, however, that for most n the concentration of $\omega(G(n,1/2))$ is actually on a single value!

10.3 CHROMATIC NUMBER

In this section we fix $p = 1/2$ (there are similar results for other p) and let G be the random graph $G(n,1/2)$. We shall find bounds on the chromatic number $\chi(G)$. A different derivation of the main result of this section is presented in Section 7.3. Set

$$f(k) = \binom{n}{k} 2^{-\binom{k}{2}} .$$

Let $k_0 = k_0(n)$ be the value for which

$$f(k_0 - 1) > 1 > f(k_0) .$$

Then $n = \sqrt{2}^{k(1+o(1))}$, so for $k \sim k_0$

$$f(k+1)/f(k) = \frac{n}{k} 2^{-k}(1 + o(1)) = n^{-1+o(1)} .$$

Set

$$k = k(n) = k_0(n) - 4$$

so that

$$f(k) > n^{3+o(1)} .$$

Now we use the Extended Janson Inequality (Theorem 8.1.2) to estimate $\Pr[\omega(G) < k]$. Here $\mu = f(k)$. (Note that Janson's Inequality gives a lower bound of

$2^{-f(k)} = 2^{-n^{3+o(1)}}$ to this probability but this is way off the mark since with probability $2^{-\binom{n}{2}}$ the random G is empty!) The value Δ was examined in Section 4.5, where

$$\frac{\Delta}{\mu^2} = \frac{\Delta^*}{\mu} = \sum_{i=2}^{k-1} g(i) \ .$$

There, $g(2) \sim k^4/n^2$ and $g(k-1) \sim 2kn2^{-k}/\mu$ were the dominating terms. In our instance, $\mu > n^{3+o(1)}$ and $2^{-k} = n^{-2+o(1)}$, so $g(2)$ dominates and

$$\Delta \sim \frac{\mu^2 k^4}{n^2} \ .$$

Hence we bound the *clique* number probability

$$\Pr[\omega(G) < k] < e^{-\mu^2/2\Delta} = e^{-\Theta(n^2/(\ln\ n)^4)}$$

as $k = \Theta(\ln\ n)$. [The possibility that G is empty gives a lower bound so that we may say the probability is $e^{-n^{2+o(1)}}$, though a $o(1)$ in the hyperexponent leaves lots of room.]

Theorem 10.3.1 [Bollobás (1988)] *Almost always*

$$\chi(G) \sim \frac{n}{2\log_2 n} \ .$$

Proof. Let $\alpha(G) = \omega(\overline{G})$ denote, as usual, the independence number of G. The complement of G has the same distribution $G(n, \frac{1}{2})$. Hence $\alpha(G) \leq (2 + o(1))\log_2 n$ almost always. Thus

$$\chi(G) \geq \frac{n}{\alpha(G)} \geq \frac{n}{2\log_2 n}(1 + o(1))$$

almost always.

The reverse inequality was an open question for a full quarter century! Set $m = \lfloor n/\ln\ ^2 n \rfloor$. For any set S of m vertices, the restriction $G|_S$ has the distribution of $G(m, 1/2)$. Let $k = k(m) = k_0(m) - 4$ as above. Note

$$k \sim 2\log_2 m \sim 2\log_2 n \ .$$

Then

$$\Pr[\alpha[G|_S] < k] < e^{-m^{2+o(1)}} \ .$$

There are $\binom{n}{m} < 2^n = 2^{m^{1+o(1)}}$ such sets S. Hence

$$\Pr[\alpha[G|_S] < k \ \text{ for some } m\text{-set } S] < 2^{m^{1+o(1)}} e^{-m^{2+o(1)}} = o(1) \ .$$

That is, almost always *every* m vertices contain a k-element independent set.

Now suppose G has this property. We pull out k-element independent sets and give each a distinct color until there are less than m vertices left. Then we give each point a distinct color. By this procedure

$$\begin{aligned}\chi(G) &\le \left\lceil \frac{n-m}{k} \right\rceil + m \le \frac{n}{k} + m \\ &= \frac{n}{2\log_2 n}(1+o(1)) + o\left(\frac{n}{\log_2 n}\right) \\ &= \frac{n}{2\log_2 n}(1+o(1)) \ ,\end{aligned}$$

and this occurs for almost all G. ■

10.4 ZERO–ONE LAWS

In this section we restrict our attention to graph theoretic properties expressible in the first-order theory of graphs. The language of this theory consists of variables $(x, y, z, \cdots)$, which always represent vertices of a graph, equality and adjacency $(x = y, x \sim y)$, the usual Boolean connectives $(\wedge, \neg, \cdots)$, and universal and existential quantification $(\forall_x, \exists_y)$. Sentences must be finite. As examples, one can express the property of containing a triangle

$$\exists_x \exists_y \exists_z [x \sim y \wedge x \sim z \wedge y \sim z] \ ,$$

having no isolated point

$$\forall_x \exists_y [x \sim y]$$

and having radius at most two

$$\exists_x \forall_y [\neg(y = x) \wedge \neg(y \sim x) \rightarrow \exists_z [z \sim y \wedge z \sim x]] \ .$$

For any property A and any n, p, we consider the probability that the random graph $G(n, p)$ satisfies A, denoted

$$\Pr[G(n,p) \models A] \ .$$

Our objects in this section will be the theorem of Glebskii et al. (1969), independently obtained by Fagin (1976) (Theorem 10.4.1), and that of Shelah and Spencer (1988) (Theorem 10.4.2).

Theorem 10.4.1 *For any fixed p, $0 < p < 1$ and any first-order A,*

$$\lim_{n\to\infty} \Pr[G(n,p) \models A] = 0 \ \text{ or } 1 \ .$$

Theorem 10.4.2 *For any* irrational α, $0 < \alpha < 1$, *setting* $p = p(n) = n^{-\alpha}$, *and for any first-order* A,

$$\lim_{n\to\infty} \Pr[G(n,p) \models A] = 0 \text{ or } 1 .$$

Both proofs are only outlined.

We shall say that a function $p = p(n)$ satisfies the Zero-One Law if the above equality holds for every first-order A.

The Glebskii–Fagin theorem has a natural interpretation when $p = 0.5$ because then $G(n,p)$ gives equal weight to every (labeled) graph. It then says that any first-order property A holds for either almost all graphs or for almost no graphs. The Shelah–Spencer theorem may be interpreted in terms of threshold functions. The general results of Section 10.1 give, as one example, that $p = n^{-2/3}$ is a threshold function for containment of a K_4. That is, when $p \ll n^{-2/3}$, $G(n,p)$ almost surely does not contain a K_4, whereas when $p \gg n^{-2/3}$, it almost surely does contain a K_4. In between, say at $p = n^{-2/3}$, the probability is between 0 and 1, in this case $1 - e^{-1/24}$. The (admittedly rough) notion is that *at* a threshold function the Zero–One Law will not hold, and to say that $p(n)$ satisfies the Zero-One Law is to say that $p(n)$ is not a threshold function – that it is a boring place in the evolution of the random graph, at least through the spectacles of the first-order language. In stark terms: What happens in the evolution of $G(n,p)$ at $p = n^{-\pi/7}$? The answer: Nothing!

Our approach to Zero–One Laws will be through a variant of the Ehrenfeucht Game, which we now define. Let G, H be two vertex disjoint graphs, and t be a positive integer. We define a perfect information game, denoted EHR$[G, H, t]$, with two players, denoted Spoiler and Duplicator. The game has t rounds. Each round has two parts. First, the Spoiler selects either a vertex $x \in V(G)$ or a vertex $y \in V(H)$. He chooses which graph to select the vertex from. Then the Duplicator must select a vertex in the other graph. At the end of the t rounds, t vertices have been selected from each graph. Let $x_1, \ldots, x_t$ be the vertices selected from $V(G)$, and $y_1, \ldots, y_t$ be the vertices selected from $V(H)$, where x_i, y_i are the vertices selected in the ith round. Then Duplicator wins if and only if the induced graphs on the selected vertices are order-isomorphic: that is, if for all $1 \le i < j \le t$,

$$\{x_i, x_j\} \in E(G) \leftrightarrow \{y_i, y_j\} \in E(H) .$$

As there are no hidden moves and no draws, one of the players must have a winning strategy and we will say that this player wins EHR$[G, H, t]$.

Lemma 10.4.3 *For every first-order* A, *there is a* $t = t(A)$ *so that, if* G, H *are any graphs with* $G \models A$ *and* $H \models \neg A$, *then Spoiler wins EHR*$[G, H, t]$.

A detailed proof would require a formal analysis of the first-order language, so we give only an example. Let A be the property $\forall_x \exists_y [x \sim y]$ of not containing an

isolated point, and set $t = 2$. Spoiler begins by selecting an isolated point $y_1 \in V(H)$ which he can do as $H \models \neg A$. Duplicator must pick $x_1 \in V(G)$. As $G \models A$, x_1 is not isolated, so Spoiler may pick $x_2 \in V(G)$ with $x_1 \sim x_2$ and now Duplicator cannot pick a "duplicating" y_2.

Theorem 10.4.4 *A function $p = p(n)$ satisfies the Zero-One Law if and only if for every t, letting $G(n, p(n)), H(m, p(m))$ be independently chosen random graphs on disjoint vertex sets,*

$$\lim_{m,n\to\infty} \Pr[\textit{Duplicator wins } \mathrm{EHR}[G(n,p(n)), H(m,p(m)), t]] = 1 \ .$$

Remark. For any given choice of G, H, somebody must win $\mathrm{EHR}[G, H, t]$. (That is, there is no random play; the play is perfect.) Given this probability distribution over (G, H), there will be a probability that $\mathrm{EHR}[G, H, t]$ will be a win for Duplicator, and this must approach 1.

Proof. We prove only the "if" part. Suppose that $p = p(n)$ did not satisfy the Zero–One Law. Let A satisfy

$$\lim_{n\to\infty} \Pr[G(n, p(n)) \models A] = c \ ,$$

with $0 < c < 1$. Let $t = t(A)$ be as given by the Lemma. With limiting probability $2c(1 - c) > 0$ exactly 1 of $G(n, p(n)), H(m, p(m))$ would satisfy A and thus Spoiler would win, contradicting the assumption. This is not a full proof because, when the Zero–One Law is not satisfied, $\lim_{n\to\infty} \Pr[G(n, p(n)) \models A]$ might not exist. If there is a subsequence n_i on which the limit is $c \in (0, 1)$, we may use the same argument. Otherwise, there will be two subsequences n_i, m_i on which the limit is zero and 1, respectively. Then letting $n, m \to \infty$ through n_i, m_i, respectively, Spoiler will win $\mathrm{EHR}[G, H, t]$ with probability approaching 1. ■

Theorem 10.4.4 provides a bridge from logic to random graphs. To prove that $p = p(n)$ satisfies the Zero-One Law, we now no longer need to know anything about logic – we just have to find a good strategy for the Duplicator.

We say that a graph G has the full level s extension property if for every distinct $u_1, \ldots, u_a, v_1, \ldots, v_b \in G$ with $a + b \le s$, there is an $x \in V(G)$ with $\{x, u_i\} \in E(G)$, $1 \le i \le a$ and $\{x, v_j\} \notin E(G)$, $1 \le j \le b$. Suppose that G, H both have the full level $s - 1$ extension property. Then Duplicator wins $\mathrm{EHR}[G, H, s]$ by the following simple strategy: On the ith round, with $x_1, \ldots, x_{i-1}, y_1, \ldots, y_{i-1}$ already selected, and Spoiler picking, say, x_i, Duplicator simply picks y_i having the same adjacencies to the $y_j, j < i$ as x_i has to the $x_j, j < i$. The full extension property says that such a y_i will surely exist.

Theorem 10.4.5 *For any fixed p, $0 < p < 1$, and any s, $G(n, p)$ almost always has the full level s extension property.*

Proof. For every distinct $u_1, \ldots, u_a, v_1, \ldots, v_b, x \in G$ with $a + b \leq s$, we define $E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}$ to be the event that $\{x, u_i\} \in E(G)$, $1 \leq i \leq a$ and $\{x, v_j\} \notin E(G)$, $1 \leq j \leq b$. Then

$$\Pr[E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}] = p^a(1-p)^b \ .$$

Now define

$$E_{u_1,\ldots,u_a,v_1,\ldots,v_b} = \wedge_x \overline{E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}$$

the conjunction over $x \neq u_1, \ldots, u_a, v_1, \ldots, v_b$. These events are mutually independent over x since they involve different edges. Thus

$$\Pr\left[\wedge_x \overline{E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}\right] = [1 - p^a(1-p)^b]^{n-a-b} \ .$$

Set $\epsilon = \min(p, 1-p)^s$ so that

$$\Pr\left[\wedge_x \overline{E_{u_1,\ldots,u_a,v_1,\ldots,v_b,x}}\right] \leq (1-\epsilon)^{n-s} \ .$$

The key here is that ϵ is a fixed (dependent on p, s) positive number. Set

$$E = \vee E_{u_1,\ldots,u_a,v_1,\ldots,v_b} \ ,$$

the disjunction over all distinct $u_1, \ldots, u_a, v_1, \ldots, v_b \in G$ with $a + b \leq s$. There are less than $s^2 n^s = O(n^s)$ such choices, as we can choose a, b and then the vertices. Thus

$$\Pr[E] \leq s^2 n^s (1-\epsilon)^{n-s} \ .$$

But

$$\lim_{n\to\infty} s^2 n^s (1-\epsilon)^{n-s} = 0$$

and so E almost never holds. Thus $\neg E$, which is precisely the statement that $G(n,p)$ has the full level s extension property, holds almost always. ■

But now we have proven Theorem 10.4.1. For any $p \in (0,1)$ and any fixed s, as $m, n \to \infty$ with probability approaching 1, both $G(n,p)$ and $H(m,p)$ will have the full level s extension property and so Duplicator will win EHR$[G(n,p), H(m,p), s]$.

Why can't Duplicator use this strategy when $p = n^{-\alpha}$? We illustrate the difficulty with a simple example. Let $0.5 < \alpha < 1$, and let Spoiler and Duplicator play a three-move game on G, H. Spoiler thinks of a point $z \in G$ but does not tell Duplicator about it. Instead, he picks $x_1, x_2 \in G$, both adjacent to z. Duplicator simply picks $y_1, y_2 \in H$, either adjacent or not adjacent depending on whether $x_1 \sim x_2$. But now the wily Spoiler picks $x_3 = z$. $H \sim H(m, m^{-\alpha})$ does not have the full level 2 extension property. In particular, most pairs y_1, y_2 do not have a common neighbor. Unless Duplicator was lucky, or shrewd, he then cannot find $y_3 \sim y_1, y_2$ and so he loses. This example does not say that Duplicator will lose with perfect play – indeed, we will

show that he almost always wins with perfect play – it only indicates that the strategy used need be more complex.

We begin our proof of the Zero–One Law, Theorem 10.4.2. Let $\alpha \in (0,1)$, α irrational, be fixed. A *rooted graph* is a pair (R,H) where H is a graph on vertex set, say, $V(H) = \{X_1, \ldots, X_r, Y_1, \ldots, Y_v\}$, and $R = \{X_1, \ldots, X_r\}$ is a specified subset of $V(H)$, called the roots. For example, (R,H) might consist of one vertex Y_1 adjacent to the two roots X_1, X_2. Let $v = v(R,H)$ denote the number of vertices that are not roots, and let $e = e(R,H)$ denote the number of edges, excluding those edges between two roots. We say (R,H) is *dense* if $v - e\alpha < 0$ and *sparse* if $v - e\alpha > 0$. The irrationality of α assures us that all (R,H) are in one of these categories. We call (R,H) *rigid* if, for all S with $R \subseteq S \subset V(H)$, (S,H) is dense. We call (R,H) safe if, for all S with $R \subset S \subseteq V(H)$, $(R, H|_S)$ is sparse. Several elementary properties of these concepts are given as Exercise 4. We sometimes write (R,S) for $(R, H|_S)$ when the graph H is understood.

We think of rooted graphs as on abstract points. In a graph G, we say that vertices $y_1, \ldots, y_v$ form an (R,H) extension of $x_1, \ldots, x_r$ if, whenever X_i is adjacent to Y_j in H, x_i is adjacent to y_j in G and also whenever Y_i and Y_j are adjacent in H, y_i and y_j are adjacent in G. Note that we allow G to have more edges than H and that the edges between the roots "don't count".

Lemma 10.4.6 [Generic Extension] *Let (R,H), as given above, be safe. Let $t \geq 0$ be an arbitrary, but fixed, integer. Then, in $G \sim G(n, n^{-\alpha})$ almost surely for all $x_1, \ldots, x_r$ there exist $y_1, \ldots, y_v$ such that*

(i) $y_1, \ldots, y_v$ form an (R,H) extension of $x_1, \ldots, x_r$.

(ii) x_i, y_j are adjacent in G if and only if X_i, Y_j are adjacent in H and y_i, y_j are adjacent in G if and only if Y_i, Y_j are adjacent in H.

(iii) (For $t > 0$) If $z_1, \ldots, z_u$ with $u \leq t$, form a rigid (R', H') extension over $x_1, \ldots, x_r, y_1, \ldots, y_v$, then there are no adjacencies between any pair z_k, y_j.

Example. Let $\alpha \in \left(\frac{1}{2}, 1\right)$, $t = 2$, and let (R,H) have root X_1, nonroot Y_1, and edge $\{X_1, Y_1\}$. Note that (R', H') consisting of two roots X_1, X_2 with a common neighbor Y_1 has $v = 1, e = 2$ and is rigid. Generic extension in this instance says that every x_1 has a neighbor y_1 such that x_1, y_1 do not have a common neighbor z_1.

Proof. From Exercise 5, almost surely every $x_1, \ldots, x_r$ has $\Theta(n^v p^e)$ (R,H) extensions $y_1, \ldots, y_v$. Our rough notion will be that the number of these $y_1, \ldots, y_v$ that fail to be generic, in any of the bounded number of ways that could occur, would be bounded by a smaller power of n.

Call y special if $y \in \mathrm{cl}_{t+v}(x_1, \ldots, x_r)$ (as defined below), otherwise nonspecial. Let K, from the Finite Closure Lemma 10.4.7 below, be an almost sure bound on the number of special y, uniform over all choices of the x's. Extend (R,H) to (R^+, H^+) by adding K new roots and no new edges. This is still safe and of the same type as (R,H), so again by Exercise 5 almost surely every $x_1, \ldots, x_r, z_1, \ldots, z_K$ has $\Theta(n^v p^e)$

(R^+, H^+) extensions $y_1, \ldots, y_v$. Letting the z's include all the special vertices, we have that almost surely every $x_1, \ldots, x_r$ has $\Theta(n^v p^e)$ (R, H) extensions $y_1, \ldots, y_v$ with all y_i nonspecial. Now we bound from above the number of those nonspecial (R, H) extensions that fail condition 2 or 3.

Consider those extensions (R, H') with an additional edge y_i, y_j or x_i, y_j. This cannot contain a rigid subextension, as that would make some y_i special. Hence by Exercise 4, it must be a safe extension. Applying Exercise 5, there are $\Theta(n^v p^{e+1}) = o(n^v p^e)$ such extensions.

Consider extensions by $y_1, \ldots, y_v$ and $z_1, \ldots, z_u$ as in condition 3 with some z_j, y_k adjacent. We can further assume that the z's form a minimal rigid extension over the x's and y's. Let the z's have type (v_1, e_1) as an extension over the x's and y's, so that $v_1 - e_1\alpha$ is negative. If the y's and z's together formed a safe extension over the x's, there would be $\Theta(n^{v+v_1} p^{e+e_1}) = o(n^v p^e)$ such extensions and hence at most that many choices for the y's. Otherwise, by Exercise 4, there would be a rigid subextension. It could not overlap the nonspecial y's. From the minimality, it must be precisely all of the z's. Given the x's from the Finite Closure Lemma 10.4.7, there are $O(1)$ choices for the z's. Then the y's form a (v, e') extension over the x's and y's with $e' > e$. This extension has no rigid subextensions (again, as the y's are nonspecial) and hence is safe. Again, applying Exercise 5, there are $\Theta(n^v p^{e'})$ such y's for each choice of the z's and so $O(n^v p^{e'}) = o(n^v p^e)$ total choices of such y's.

In all cases, the number of y's that fail conditions 2 or 3 is $o(n^v p^e)$. Hence there exist y's, indeed most choices of nonspecial y's, that are (R, H) extensions and satisfy conditions 2 and 3. ∎

A rigid t-chain in G is a sequence $X = X_0 \subset X_1 \subset \cdots \subset X_K$ with all (X_{i-1}, X_i) rigid and all $|X_{i+1} - X_i| \leq t$. The t-closure of X, denoted by $\mathrm{cl}_t(X)$, is the maximal Y for which there exists a rigid t-chain (of arbitrary length) $X = X_0 \subset X_1 \subset \cdots \subset X_K = Y$. When there are no such rigid t-chains, we define $\mathrm{cl}_t(X) = X$. To see this is well defined, we note (using Exercise 4) that, if $X = X_0 \subset X_1 \subset \cdots \subset X_K = Z$ and $X = X_0 \subset Y_1 \subset \cdots \subset Y_L = Y$ are rigid t-chains, then so is $X = X_0 \subset X_1 \subset \cdots \subset X_K \subset Z \cup Y_1 \subset \cdots \subset Z \cup Y_L = Z \cup Y$. Alternatively, the t-closure $\mathrm{cl}_t(X)$ is the minimal set containing X that has no rigid extensions of $\leq t$ vertices. We say $x_1, \ldots, x_r \in G$, $y_1, \ldots, y_r \in H$ have the same t-type if their t-closures are isomorphic as graphs, the isomorphism sending each x_i to the corresponding y_i.

The t-closure is a critical definition, describing the possible special properties of the roots. Suppose, for example, $\alpha \in \left(\frac{1}{2}, 1\right)$ and consider $\mathrm{cl}_1(x_1, x_2)$. The only rigid extension with $t = 1$ in this range is a nonroot adjacent to two (or more) roots. A sample 1-type would be: x_1, x_2 have common neighbors y_1, y_2 and then x_1, y_1 have common neighbor y_3 and there are no further edges among these vertices and no pairs have common neighbors other than those described. A randomly chosen x_1, x_2 would have the type: x_1, x_2 have no common neighbors and are not adjacent.

We can already describe the nature of Duplicator's strategy. At the end of the rth move, with $x_1, \ldots, x_r$ and $y_1, \ldots, y_r$ having been selected from the two graphs, Duplicator will assure that these sets have the same a_r-type. We shall call this the $(a_1, \ldots, a_t)$ *lookahead strategy*. Here, a_r must depend only on t, the total number of moves in the

game, and α. We shall set $a_t = 0$ so that at the end of the game, if Duplicator can stick to the $(a_1, \dots, a_t)$ lookahead strategy, then he has won. If, however, Spoiler picks, say, x_{r+1} so that there is no corresponding y_{r+1} with $x_1, \dots, x_{r+1}$ and $y_1, \dots, y_{r+1}$ having the same a_{r+1}-type, then the strategy fails and we say that Spoiler wins. The values a_r give the "lookahead" that Duplicator uses, but before defining them we need some preliminary results.

Lemma 10.4.7 (Finite Closure) *Let α, $r > 0$ be fixed. Set ϵ equal to the minimal value of $\frac{e\alpha - v}{v}$ over all integers v, e with $1 \le v \le t$ and $e\alpha - v > 0$. Let K be such that $r - K\epsilon < 0$. Then in $G(n, n^{-\alpha})$, almost surely,*

$$|\, cl_t(X)| \le K + r$$

for all $X \subset G$ *with* $|X| = r$.

Proof. If not, there would be a rigid t-chain $X = X_0 \subset X_1 \subset \cdots \subset X_L = Y$ with $K + r < |Y| < K + r + t$. Letting (X_{i-1}, X_i) have type (v_i, e_i), the restriction of G to Y would have $r + \sum v_i$ vertices and at least $\sum e_i$ edges. But

$$\left(r + \sum v_i\right) - \alpha\left(\sum e_i\right) = r + \sum (v_i - \alpha e_i) \le r - \epsilon \sum v_i < r - K\epsilon < 0,$$

and G almost surely has no such subgraph. ∎

Remark. The bound on $|\, \mathrm{cl}_t(X)|$ given by this proof depends strongly on how closely α may be approximated by rationals of denominator at most t. This is often the case. If, for example, $\frac{1}{2} + \frac{1}{s-1} > \alpha > \frac{1}{2} + \frac{1}{s}$, then a.s. there will be two points $x_1, x_2 \in G(n, n^{-\alpha})$ having s common neighbors so that $|\, \mathrm{cl}_1(x_1, x_2)| \ge s + 2$.

Now we define the $a_1, \dots, a_t$ of the lookahead strategy by reverse induction. We set $a_t = 0$. If at the end of the game Duplicator can assure that the 0-types of $x_1, \dots, x_t$ and $y_1, \dots, y_t$ are the same, then they have the same induced subgraphs and he has won. Suppose, inductively, that $b = a_{r+1}$ has been defined. We define $a = a_r$ to be any integer satisfying

1. $a \ge b$.
2. Almost surely $|\, \mathrm{cl}_b(W)| - r \le a$ for all sets W of size $r + 1$.

Now we need to show that almost surely this strategy works. Let $G_1 \sim G(n, n^{-\alpha})$, $G_2 \sim G(m, m^{-\alpha})$, and Duplicator tries to play the $(a_1, \dots, a_t)$ lookahead strategy on $\mathrm{EHR}(G_1, G_2, t)$.

Consider the $(r+1)$th move. We have $b = a_{r+1}$, $a = a_r$, as above. Points $x_1, \dots, x_r \in G_1$, $y_1, \dots, y_r \in G_2$ have already been selected. Set $X = \{x_1, \dots, x_r\}$, $Y = \{y_1, \dots, y_r\}$ for notational convenience. We assume Duplicator has survived thus far, so that $\mathrm{cl}_a(X) \cong \mathrm{cl}_a(Y)$, the isomorphism sending each x_i to the corresponding y_i. Spoiler picks, say, $x = x_{r+1} \in G_1$. Set $X^+ = X \cup \{x\}$ and $Y^+ = Y \cup \{y\}$, where y is Duplicator's as-yet-undetermined countermove. We distinguish two cases.

We say Spoiler has moved "Inside" if $x \in \mathrm{cl}_a(X)$. Then as $b \le a$, $\mathrm{cl}_b(X^+) \subseteq \mathrm{cl}_a(X)$. Duplicator looks at the isomorphism $\Psi : \mathrm{cl}_a(X) \to \mathrm{cl}_a(Y)$ and selects $y = \Psi(x)$.

We say Spoiler has moved "Outside" if $x \notin \mathrm{cl}_a(X)$. Let NEW be those vertices of $\mathrm{cl}_b(X^+)$ that do not lie in $\mathrm{cl}_a(X)$. NEW $\neq \emptyset$ as $x \in$ NEW. $|\mathrm{NEW}| \le a$ as NEW $\subseteq \mathrm{cl}_b(X^+) - X$. Consider NEW as an (R, H) extension of $\mathrm{cl}_a(X)$. This extension must be safe, as otherwise it would have a rigid subextension NEW$^-$ but that subextension would then be in $\mathrm{cl}_a(X)$. Duplicator now goes to G_2 and, applying the Generic Extension Lemma 10.4.6 with $t = b$, finds an (R, H) extension of $\mathrm{cl}_a(Y)$. That is, he finds an edge-preserving injection $\Psi : \mathrm{cl}_a(X) \cup \mathrm{NEW} \to H$, extending the isomorphism between $\mathrm{cl}_a(X)$ and $\mathrm{cl}_a(Y)$. Duplicator selects $y = \Psi(x)$.

Why does this work? Set NEW$' = \Psi($ NEW$)$ and CORE $= \Psi(\mathrm{cl}_b(X^+))$. We can reach $\mathrm{cl}_b(X^+)$ by a rigid b-chain from X^+, and the isomorphism gives the same chain from Y^+ to CORE so that $\mathrm{cl}_b(Y^+)$ contains CORE. But can it have additional vertices? We use the genericity to say no. Suppose there was a rigid extension MORE over CORE with at most b nonroots. We cannot have MORE entirely inside $\Psi[\mathrm{cl}_a(X) \cup \mathrm{NEW}]$, as then $\Psi^{-1}[$ MORE$]$ would be in $\mathrm{cl}_b(X^+)$ as well. Let MORE$^+$ be the vertices of MORE lying outside $\Psi[\mathrm{cl}_a(X) \cup \mathrm{NEW}]$. MORE$^+$ is then a rigid extension of $\Psi[\mathrm{cl}_a(X) \cup \mathrm{NEW}]$. By genericity, MORE$^+$ would have no adjacencies to NEW$'$ and so would be a rigid extension of $\Psi[\mathrm{cl}_a(X)] = \mathrm{cl}_a(Y)$. As $a \ge b$, the a-closure of a set cannot have rigid extensions with $\le b$ vertices. Hence there is no MORE.

The first move follows the same pattern but is somewhat simpler. Set $b = a_1$, and let a satisfy $a \ge b$ and $a \ge |\mathrm{cl}_b(x)|$ for any x. Spoiler plays $x \in G_1$. (Effectively, there is no Inside move, as $X = \emptyset$ is the set of previous moves and $\mathrm{cl}_a(\emptyset) = \emptyset$.) Duplicator calculates the graph $H = \mathrm{cl}_b(x)$ which has, say, v vertices [including x] and e edges. Since H is a subgraph of G_1, the threshold function for the appearance of H must come before $n^{-\alpha}$. In particular, for every subgraph H' of H with v' vertices and e' edges, we cannot have $v' - \alpha e' < 0$ and therefore must have $v' - \alpha e' > 0$. The conditions of Theorem 4.4.5 then apply, and G_2 almost surely has $\Theta(m^{e-v\alpha})$ copies of H. Consider any graph H^+ consisting of H together with a rigid extension of H with at most b vertices. Such H^+ would have $v + v^+$ vertices and $e + e^+$ edges, with $v^+ - \alpha e^+ < 0$. The expected number of copies of H^+ is then $\Theta(m^{e-v\alpha+(v^+-\alpha e^+)})$, which is $o(m^{e-v\alpha})$. Hence there will be in G_2 a copy of H which is not part of any such H^+. (Effectively, this is a generic extension over the empty set.) Duplicator finds the edge-preserving injection $\Psi : \mathrm{cl}_b(x) \to G_2$, giving such a copy of H, and selects $y = \Psi(x)$.

We have shown that the $(a_1, \ldots, a_t)$ lookahead strategy almost surely results in a win for Duplicator. By Theorem 10.4.4, this implies the Zero–One Law, Theorem 10.4.2.

10.5 EXERCISES

1. Show that there is a graph on n vertices with minimum degree at least $n/2$ in which the size of every dominating set is at least $\Omega(\log\ n)$.

2. Find a threshold function for the property that $G(n,p)$ contains a copy of the graph consisting of a complete graph on four vertices plus an extra vertex joined to one of its vertices.
3. Let X be the number of cycles in the random graph $G(n,p)$, with $p = \frac{c}{n}$. Give an exact formula for $\mathrm{E}[X]$. Find the asymptotics of $\mathrm{E}[X]$ when $c < 1$. Find the asymptotics of $\mathrm{E}[X]$ when $c = 1$.
4. Here we write (R,S) for $(R, H|_S)$, where H is some fixed graph.
   - Let $R \subset S \subset T$. Show that if (R,S), (S,T) are both dense, then so is (R,T). Show that if (R,S), (S,T) are both sparse, then so is (R,T).
   - Let $R \subset S$. Show that if (R,S) is rigid, then $(X \cup R, X \cup S)$ is rigid for any X.
   - $R \subset U$ with (R,U) *not* sparse. Show that there is a T with $R \subset T \subset U$ with (R,T) dense. Show further that there is an S with $R \subset S \subset T$ with (R,S) rigid.
   - Show that *any* (R,T) is either rigid or sparse itself, or there exists S with $R \subset S \subset T$ such that (R,S) is rigid and (S,T) is sparse.
5. We call (R,H) *hinged* if it is safe but there is no S with $R \subset S \subset V(H)$ such that (S,H) is safe. For $x_1, \ldots, x_r \in G$, let $N(x_1, \ldots, x_r)$ denote the number of (R,H) extensions. Set $\mu = \mathrm{E}[N] \sim n^v p^e$.
   - Let (R,H) be hinged, and fix $x_1, \ldots, x_r \in G$. Following the model of Section 8.5, especially Theorem 8.5.4, show that

$$\Pr[|N(x_1, \ldots, x_r) - \mu| > \epsilon\mu] = o(n^{-r}) .$$

   - Deduce that almost surely all $N(x_1, \ldots, x_r) \sim \mu$.
   - Show that $N(x_1, \ldots, x_r) \sim \mu$ holds for any safe (R,H), by decomposing (R,H) into hinged extensions.

THE PROBABILISTIC LENS:
Counting Subgraphs

A graph $G = (V, E)$ on n vertices has 2^n induced subgraphs but some will surely be isomorphic. How many different subgraphs can G have? Here we show that there are graphs G with $2^n(1 - o(1))$ different subgraphs. The argument we give is fairly coarse. It is typical of those situations where a probabilistic approach gives fairly quick answers to questions that are otherwise difficult to approach.

Let G be a random graph on n vertices with edge probability $1/2$. Let $S \subseteq V$, $|S| = t$ be fixed. For any one-to-one $\rho : S \to V$, $\rho \neq id$, let A_ρ be the event where ρ gives a graph isomorphism, that is, for $x, y \in S$, $\{x, y\} \in E \iff \{\rho x, \rho y\} \in E$. Set $M_\rho = \{x \in S : \rho x \neq x\}$. We split the set of ρ by $g = g(\rho) = |M_\rho|$.

Consider the $g(t - g) + \binom{g}{2}$ pairs x, y with $x, y \in S$ and at least one of x, y in M. For all but at most $g/2$ of these pairs, $\{x, y\} \neq \{\rho x, \rho y\}$. (The exceptions are when $\rho x = y, \rho y = x$.) Let E_ρ be the set of pairs $\{x, y\}$ with $\{x, y\} \neq \{\rho x, \rho y\}$. Define a graph H_ρ with vertices E_ρ and vertex $\{x, y\}$ adjacent to $\{\rho x, \rho y\}$. In H_ρ, each vertex has degree at most 2 ($\{x, y\}$ may also be adjacent to $\{\rho^{-1}x, \rho^{-1}y\}$) and so it decomposes into isolated vertices, paths, and circuits. On each such component, there is an independent set of size at least one-third the number of elements, the extreme case being a triangle. Thus there is a set $I_\rho \subseteq E_\rho$ with

$$|I_\rho| \geq |E_\rho| \geq \frac{g(t - g) + \binom{g}{2} - g/2}{3},$$

so that the pairs $\{x, y\}, \{\rho x, \rho y\}$ with $\{x, y\} \in I_\rho$ are all distinct.

For each $\{x, y\} \in I_\rho$, the event $\{x, y\} \in E \iff \{\rho x, \rho y\} \in E$ has probability $1/2$. Moreover, these events are mutually independent over $\{x, y\} \in I_\rho$ since they involve distinct pairs. Thus we bound

$$\Pr[A_\rho] \leq 2^{-|I_\rho|} \leq 2^{-(g(t-g)+\binom{g}{2}-g/2)/3}.$$

For a given g, the function ρ is determined by $\{x : \rho x \neq x\}$ and the values ρx for those x so that there are less than n^{2g} such ρ. We bound

$$\sum_{\rho \neq id} \Pr[A_\rho] = \sum_{g=1}^{t} \sum_{g(\rho)=g} \Pr[A_\rho] \leq \sum_{g=1}^{t} n^{2g} 2^{-(g(t-g)+\binom{g}{2}-g/2)/3} .$$

We make the rough bound

$$g(t-g) + \binom{g}{2} - g/2 = g\left(t - \frac{g}{2} - 1\right) \geq g\left(\frac{t}{2} - 1\right) ,$$

since $g \leq t$. Then

$$\sum_{\rho \neq id} \Pr[A_\rho] \leq \sum_{g=1}^{t} \left[n^2 2^{(-\frac{t}{2}+1)/3}\right]^g .$$

For, again being rough, $t > 50 \ln n$, $2^{\frac{1}{3}-\frac{t}{6}} < n^{-3}$ and $\sum_{\rho \neq id} \Pr[A_\rho] = o(1)$. That is, almost surely there is no isomorphic copy of $G|_S$.

For all $S \subseteq V$ with $|S| > 50 \ln n$, let I_S be the indicator random variable for there being no other subgraph isomorphic to $G|_S$. Set $X = \sum I_S$. Then $\mathrm{E}[I_S] = 1 - o(1)$, so by Linearity of Expectation (there being $2^n(1 - o(1))$ such S)

$$\mathrm{E}[X] = 2^n(1 - o(1)) .$$

Hence there is a specific G with $X > 2^n(1 - o(1))$.

11

The Erdős–Rényi Phase Transition

> She had been born with a map of time in her mind. She pictured other abstractions as well, numbers and the letters of the alphabet, both in English and in Bengali. Numbers and letters were like links on a chain. Months were arrayed as if along an orbit in space.
> –from *The Lowland*, by Jhumpa Lahiri

In their great work *On the Evolution of Random Graphs*, Erdős and Rényi (1960) expressed a special interest in the behavior of $\Gamma_{n,N(n)}$, the random graph with n vertices and $N(n)$ edges, when $N(n)$ was near $\frac{n}{2}$:

> Thus the situation may be summarized as follows: the largest component of $\Gamma_{n,N(n)}$ is of order $\log\ n$ for $\frac{N(n)}{n} \to c < \frac{1}{2}$, of order $n^{2/3}$ for $\frac{N(n)}{n} \to c \sim \frac{1}{2}$, and of order n for $\frac{N(n)}{n} \to c > \frac{1}{2}$. This double "jump" of the size of the largest component when $\frac{N(n)}{n}$ passes the value $\frac{1}{2}$ is one of the most striking facts concerning random graphs.

Striking, indeed. The past half century has certainly confirmed the excitement that Erdős and Rényi (1960) expressed in their discovery.

11.1 AN OVERVIEW

We favor the more modern viewpoint, examining the random graph $G(n,p)$. The behavior of Erdős and Rényi's (1960) $\Gamma_{n,N(n)}$ then corresponds to that of $G(n,p)$ with $p = N(n)/\binom{n}{2}$. We shall assume $p = \Theta(n^{-1})$ throughout this chapter.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

We shall call

$$p = \frac{c}{n} \tag{11.1}$$

the *coarse* parametrization. The value $\frac{1}{2}$ in the Erdős–Rényi formulation corresponds to the value $c = 1$ in our parametrization. Values $c < 1$ and $c > 1$ give $G(n,p)$ that are essentially different. We shall call

$$p = \frac{1}{n} + \lambda n^{-4/3} \tag{11.2}$$

the *fine* parametrization. The importance of this parametrization is not *a priori* at all obvious. Indeed, its "discovery" was one of the great advances in the field. In Section 11.7 we give a heuristic argument why this is the appropriate fine parametrization. Along with the fine parametrization we also define

$$\epsilon = \lambda n^{-1/3} \text{ so that } p = \frac{1+\epsilon}{n} . \tag{11.3}$$

We shall express various results in terms of either λ or ϵ (or both), whichever best illustrates the result. We shall think of ϵ, λ as functions of n. To avoid negative numbers, we shall sometimes parametrize $p = \frac{1-\epsilon}{n}$ with $\epsilon = \lambda n^{-1/3}$. This includes functions such as $p = \frac{1}{n} - 100n^{0.01}n^{-4/3}$. Of course, for n small, this would give $p < 0$ and so would be nonsense. For n sufficiently large, we will have $p \in [0,1]$. As our results are always asymptotic, we shall allow this slight abuse of notation and consider $G(n,p)$ defined only for n appropriately large.

In describing the nature of $G(n,p)$, we shall refer to the complexity of components, as defined below. Observe that complexity zero and one correspond to tree components and unicyclic components, respectively.

Definition 1 *A connected component of a graph G with v vertices and e edges is said to have* complexity $e - v + 1$. *Components with complexity zero or one are called* simple*; components with complexity greater than* 1 *are called* complex.

Let $C(v)$ denote the component containing a given vertex v. Its size $|C(v)|$ has a distribution. From the symmetry of $G(n,p)$, the distribution of all $|C(v)|$ are the same. We shall be concerned with the sizes of the largest components. We shall let C_i denote the ith largest component and L_i denote its number of vertices. Thus $L_1 = \max_v |C(v)|$. We shall be particularly interested in L_1, L_2 and whether they are close together.

The study of $G(n,p)$ when $p = \Theta(n^{-1})$ splits into five regions. We describe them in order of increasing p, thus giving some sense of the evolution.

Very Subcritical. Here we employ the coarse parametrization $p = \frac{c}{n}$ and assume c is a constant with $c < 1$. Example: $p = \frac{1}{2n}$.

1. All components are simple.
2. $L_1 = \Theta(\ln\ n)$.
3. $L_k \sim L_1$ for all fixed k.

Barely Subcritical. Here we employ the fine parametrization. $p = \frac{1-\epsilon}{n}$ with $\epsilon = \lambda n^{-1/3}$. We assume $\epsilon = o(1)$. We assume that $\lambda \to \infty$. Example: $p = \frac{1}{n} - n^{-4/3}n^{0.01}$.

1. All components are simple.
2. $L_1 = \Theta(\epsilon^{-2} \ln\ (\lambda)) = \Theta(n^{2/3}\lambda^{-2} \ln\ (\lambda))$.
3. $L_k \sim L_1$ for all fixed k.

The Critical Window. Here λ is a real constant. Example: $p = \frac{1}{n} - 2n^{-4/3}$. The value $\lambda = 0$, perhaps surprisingly, has no special status. The largest k components (k fixed) all have size $L_k = \Theta(n^{2/3})$. Parametrizing $L_k = c_k n^{2/3}$ and letting d_k denote the complexity of C_k, there is a nontrivial joint distribution for $c_1, \ldots, c_k, d_1, \ldots, d_k$.

Barely Supercritical. Here we employ the fine parametrization. We assume $\epsilon, \lambda > 0$. We assume $\epsilon = o(1)$. We assume that $\lambda \to +\infty$. Example: $p = \frac{1}{n} + n^{-4/3}n^{0.01}$.

1. $L_1 \sim 2\epsilon n = 2\lambda n^{2/3}$.
2. The largest component has complexity approaching infinity.
3. All other components are simple.
4. $L_2 = \Theta(\epsilon^{-2} \ln\ (\lambda)) = \Theta(n^{2/3}\lambda^{-2} \ln\ (\lambda))$.

Note that the ratio L_1/L_2 goes to infinity. For this reason, in this regime we call the largest component the *dominant component*.

Very Supercritical. We employ the coarse parametrization and assume $c > 1$.

1. $L_1 \sim yn$, where $y = y(c)$ is that positive real component satisfying the implicit equation
$$e^{-cy} = 1 - y \tag{11.4}$$
2. The largest component has complexity approaching infinity.
3. All other components are simple.
4. $L_2 = \Theta(\ln\ n)$.

Following the terminology made famous by Erdős and Rényi, we call the largest component the *giant component*.

We shall give arguments for only some of the above statements, and then often in limited form. Other results are given in the exercises. Full arguments for these results, and much more, can be found in the classic texts of Bollobás (2001) and of Janson et al. (2000).

11.2 THREE PROCESSES

We place here in concise form three classes of probability spaces that we shall contrast and analyze. Our goal is to analyze the graph branching model. It is estimated by the binomial branching model, and thence by the Poisson branching model, which has a particularly nice analysis.

- The Poisson branching model
  - Parameter: Nonnegative real c.
  - Underlying space: An infinite sequence Z_t, $t = 1, 2, \ldots$ of independent identically distributed random variables, each having Poisson distribution with mean c.
  - Auxiliary Y_t, $t \geq 0$, given by initial value $Y_0 = 1$ and recursion $Y_t = Y_{t-1} + Z_t - 1$.
  - Auxiliary T: T is that minimal t with $Y_t = 0$. If no such t exists, we write $T = \infty$.
  - Nomenclature: Z_t is the number of nodes born at time t, Y_t is the queue size at time t, and T is the total size.
  - Interpretation: T is the total size of a Galton–Watson process, as described in Section 11.3, using a Poisson distribution with mean c.
- The binomial branching model
  - Parameters: Positive Integer m, Real $p \in [0, 1]$.
  - Underlying space: An infinite sequence Z_t, $t = 1, 2, \ldots$ of independent, identically distributed random variables, each having a binomial distribution $B[m, p]$.
  - Auxiliary Y_t, $t \geq 0$, given by the initial value $Y_0 = 1$ and recursion $Y_t = Y_{t-1} + Z_t - 1$.
  - Auxiliary T: T is that minimal t with $Y_t = 0$. If no such t exists, we write $T = \infty$.
  - Nomenclature: Z_t is the number of nodes born at time t, Y_t is the queue size at time t, and T is the total size.
  - Interpretation: T is the total size of a Galton–Watson process, as described in Section 11.3 using a binomial distribution with parameters m, p.
- The graph branching model
  - Parameters: Positive Integer n, Real $p \in [0, 1]$.
  - Underlying space: A sequence $Z_1, \ldots, Z_n$. Z_t has binomial distribution with parameters N_{t-1}, p, with N_{t-1} as given below.
  - Auxiliary Y_t, $t \geq 0$, given by initial value $Y_0 = 1$ and recursion $Y_t = Y_{t-1} + Z_t - 1$.
  - Auxiliary N_t, $t \geq 0$, given by initial value $N_0 = n - 1$ and recursion $N_t = N_{t-1} - Z_t$. Equivalently: $N_t = n - t - Y_t$.
  - Auxiliary T: T is that minimal t with $Y_t = 0$ or, equivalently, $N_t = n - t$. $1 \leq T \leq n$ always.
  - Nomenclature: Z_t is the number of nodes born at time t, Y_t is the queue size at time t, N_t is the number of neutral vertices at time t, and T is the total size.
  - Interpretation: T is the size of the component $C(v)$ of a given vertex v in $G(n, p)$, as found by the Breadth First Search process described in Section 11.5.

We use the superscripts po (Poisson), bin (binomial), and gr (graph) to distinguish these three processes when necessary.

11.3 THE GALTON–WATSON BRANCHING PROCESS

Let Z be a distribution over the nonnegative integers. The Galton–Watson process begins with a single root node, we can call her Eve. Eve has Z children. Each of her children (if there are any) now independently has Z children. The process continues, each new offspring having an independent number Z children. Let T be the total number of nodes (including Eve herself) created in the process. It is possible that the process goes on forever, in which case we write $T = \infty$.

Our analysis of the Galton–Watson process uses fictional continuation. Let Z_t, $t = 1, 2, \ldots$, be a countable sequence of independent, identically distributed variables, each having distribution Z. This defines our probability space. We think of the children being born in a Breadth First Search manner. That is, Eve has her children that are ordered in some way. Now the children, in order, have children. Each child's children are ordered in some way, and this gives an ordering of Eve's grandchildren. Now the grandchildren have children in order, and the process continues. We count Eve as node number 1, her children have node numbers $2, \ldots, 1 + Z_1$ and, more generally, each node is given a distinct positive integer as its node number. We let Z_t be the number of children of the tth node. Since the Z_t are independent and have distribution Z, this corresponds to the Galton–Watson process. Imagine the tth node having Z_t children and then dying. By time t, we mean the process after the tth node has had her children and died. Let Y_t be the number of living children at time t. We set the initial value $Y_0 = 1$, corresponding to the node Eve. We have the following recursion:

$$Y_t = Y_{t-1} + Z_t - 1 \quad \text{for all } t \geq 1. \tag{11.5}$$

There are two essentially different cases:

- $Y_t > 0$ for all $t \geq 0$. In this case, the Galton–Watson process goes on forever and $T = \infty$.
- $Y_t = 0$ for some $t \geq 0$. In this case, let T be the *least* integer for which $Y_T = 0$. Then the Galton–Watson process stops with the death of the Tth node, where T is the total number of nodes in the process.

Our fictional continuation enables us to consider the Y_t as an infinite random walk, with step size $Z - 1$. When $c < 1$, the walk has negative drift and so tends to minus infinity. When $c > 1$, the walk has positive drift and tends to plus infinity. The process is called *subcritical* when $c < 1$, and it is called *supercritical when* $c > 1$. When $c = 1$, the walk has zero drift and the situation is especially delicate.

The above is quite general. When Z is Poisson or binomial (the only cases of interest to us), this yields the Poisson branching process or the binomial branching process of Section 11.2, respectively.

11.4 ANALYSIS OF THE POISSON BRANCHING PROCESS

In this section we study $T = T_c^{\text{po}}$. We often drop the value c and the superscript po for notational simplicity.

Theorem 11.4.1 *If $c < 1$, T is finite with probability 1. If $c = 1$, T is finite with probability 1. If $c > 1$, then T is infinite with probability $y = y(c)$, where y is unique positive, and real, satisfying (11.4).*

Proof. Suppose $c < 1$. If $T > t$, then $Y_t > 0$ so that $Z_1 + \cdots + Z_t \geq t$. Chernoff bounds give that $\Pr[Y_t > 0] < e^{-kt}$ for a constant k. In particular, $\Pr[Y_t > 0] \to 0$, so that $\Pr[T > t] \to 0$ and T is finite with probability 1.

Suppose $c \geq 1$. Set $z = 1 - y = \Pr[T < \infty]$. Given that Eve has i children, the probability that the branching process is finite is z^i, as all i branches must be finite. Thus

$$z = \sum_{i=0}^{\infty} \Pr[Z_1 = i] z^i = \sum_{i=0}^{\infty} e^{-c} \frac{c^i z^i}{i!} = e^{c(z-1)}.$$

Setting $y = 1 - z$ gives (11.4). For $c = 1$, $e^{-y} > 1 - y$ for $y > 0$, so the solution must be $y = 0$. For $c > 1$, the function $f(y) = 1 - y - e^{-cy}$ has $f(0) = 1$, $f(1) < 0$, and $f'(0) = c - 1 > 0$, so there is a $y \in (0, 1)$ with $f(y) = 0$. Further, as f is convex, there is precisely one y. We have shown that either $\Pr[T < \infty] = 1$ or $\Pr[T < \infty] = 1 - y > 0$. The argument that $\Pr[T < \infty] \neq 1$ (not surprising as the walk has positive drift) is left as an exercise. ■

Theorem 11.4.2 *For any positive real c and any integer k, setting $T = T_c^{po}$,*

$$\Pr[T = k] = \frac{e^{-ck}(ck)^{k-1}}{k!}.$$

We defer the proof of this classic result to Section 11.6 when we will give a probabilistic proof!

When $c = 1$, Stirling's Formula gives

$$\Pr[T_1 = k] = \frac{e^{-k}k^{k-1}}{k!} \sim \frac{1}{\sqrt{2\pi}} k^{-3/2}. \tag{11.6}$$

This perforce approaches zero but it does so only at polynomial speed. In general

$$\Pr[T_c = k] \sim \frac{1}{\sqrt{2\pi}} k^{-3/2} c^{-1} (ce^{1-c})^k.$$

For any $c \neq 1$ (whether larger or smaller than 1), $ce^{1-c} < 1$ and therefore $\Pr[T_c = k]$ approaches zero at exponential speed. This gives a bound on the tail distribution:

$$\Pr[T_c \geq u] < e^{-u(\alpha + o(1))}, \tag{11.7}$$

where $\alpha = c - 1 - \ln c > 0$.

We are particularly interested in the Poisson branching process when c is *near* 1. Let us parametrize

$$c = 1 + \epsilon.$$

When $\epsilon > 0$, $\Pr[T_{1+\epsilon} = \infty]$ is that $y = y(\epsilon) \in (0,1)$ satisfying $f(y) = 1 - y - e^{-(1+\epsilon)y} = 0$. Some fun calculus gives

$$\Pr[T_{1+\epsilon} = \infty] \sim 2\epsilon \text{ as } \epsilon \to 0^+ . \tag{11.8}$$

Suppose $c \to 1^+$ so that $\epsilon \to 0^+$. We have

$$\ln\ (ce^{1-c}) = \ln\ (1+\epsilon) - \epsilon \sim -\frac{\epsilon^2}{2}.$$

Thus

$$\Pr\ [T_{1+\epsilon} = u] \sim \frac{1}{\sqrt{2\pi}} u^{-3/2} \text{ for } u = o(\epsilon^{-2}).$$

Note that $\Pr[T_{1+\epsilon} = u] \sim \Pr[T_1 = u]$ in this range. When u reaches order ϵ^{-2}, there is a change. For $u = A\epsilon^{-2}$ and fixed A

$$\Pr[T_{1+\epsilon} = A\epsilon^{-2}] \sim \frac{1}{\sqrt{2\pi}} \epsilon^3 A^{-3/2} e^{-A/2}.$$

When $A \to \infty$, we absorb smaller factors into the exponential term:

$$\Pr[T_{1+\epsilon} = A\epsilon^{-2}] = \epsilon^3 e^{-(1+o(1))A/2}.$$

When c is slightly less than 1, we can write $c = 1 - \epsilon$, where $\epsilon \to 0^+$. We have $\ln\ (ce^{-c}) \sim -\frac{1}{2}\epsilon^2$, the same as for $c = 1 + \epsilon$. Indeed, when $u = o(\epsilon^{-3})$,

$$\Pr[T_{1-\epsilon} = u] \sim \Pr[T_{1+\epsilon} = u].$$

For $A \to \infty$

$$\Pr[T_{1-\epsilon} = A\epsilon^{-2}] = \epsilon^3 e^{-(1+o(1))A/2}.$$

The Poisson branching processes with means $1 + \epsilon$ and $1 - \epsilon$ look almost the same, with the (important!) distinction that the mean $1 + \epsilon$ process is sometimes infinite while the mean $1 - \epsilon$ process never is.

In short, the Poisson branching process with mean $1 \pm \epsilon$ acts as if it had mean 1 until reaching size on the order ϵ^{-2}. Until then, $\Pr[T_{1\pm\epsilon} = u]$ is dropping at a polynomial rate. Upon reaching order ϵ^{-2}, $\Pr[T_{1\pm\epsilon} = u]$ drops exponentially in u.

We are particularly interested in the tail distribution. For $\epsilon \to 0^+$ and $A \to \infty$,

$$\Pr[T_{1-\epsilon} > A\epsilon^{-2}] < e^{-(1+o(1))A/2}\epsilon. \tag{11.9}$$

The same holds for the *finite* part of $T_{1+\epsilon}$:

$$\Pr[\infty > T_{1+\epsilon} > A\epsilon^{-2}] < e^{-(1+o(1))A/2}\epsilon. \tag{11.10}$$

When $A \to \infty$, this quantity is $o(\epsilon)$, so (11.8) gives

$$\Pr[T_{1+\epsilon} > A\epsilon^{-2}] \sim 2\epsilon \text{ when } \epsilon \to 0^+ \text{ and } A \to \infty\ . \tag{11.11}$$

11.5 THE GRAPH BRANCHING MODEL

Abbreviation

We use BFS as the abbreviation for Breadth First Search. BFS algorithms are a mainstay of computer science and central to our approach.

Let $C(v)$ denote the component, in $G(n,p)$, containing a designated vertex v. We generate $C(v)$ using the (standard) BFS algorithm to find $C(v)$. We begin with root v. In this procedure, all vertices will be live, dead, or neutral. The live vertices will be contained in a queue. Initially, at time zero, v is live, the queue consists of one vertex, v itself, and all other vertices are neutral. At each time t, we remove a live vertex w from the top of the queue (in computer science parlance we "pop the queue") and check all pairs $\{w, w'\}$, w' neutral, for adjacency in G. The popped vertex w is now dead. Those neutral w' (if any) adjacent to w are added to the bottom of the queue and are now live. (They can be placed in any particular order.) The procedure ends when the queue is empty. We let T denote that time. At time T, all vertices are neutral or dead, and the set of dead vertices is precisely the component $C(v)$. That is, $T = |C(v)|$.

Let Z_t denote the number of vertices added to the queue at time t. Let Y_t denote the size of the queue at the conclusion of time t. We set $Y_0 = 1$, reflecting the initial size of the queue. At time t, we remove one vertex and add Z_t vertices to the queue so we have the recursion $Y_t = Y_{t-1} - 1 + Z_t$. Let N_t denote the number of neutral vertices at time t. As Z_t vertices switch from neutral to live at time t, N_t satisfies the recursion $N_0 = n - 1$, $N_t = N_{t-1} - Z_t$. Equivalently, as there are t dead and Y_t live vertices at time t, $N_t = n - t - Y_t$. Z_t is found by checking N_{t-1} pairs for adjacency. As these pairs have not yet been examined, they remain adjacent with independent probability p. That is

$$Z_t \sim B[N_{t-1}, p] \sim B[n - (t-1) - Y_{t-1}, p]. \tag{11.12}$$

The graph branching process of Section 11.2 mirrors the above analysis until time T and then continues until time n. This fictional continuation shall be useful in the analysis of $C(v)$. The graph branching process is similar to a binomial branching process in that Z_t have binomial distributions but dissimilar in that the parameter N_{t-1} in the graph branching process depends on previous values Z_i.

As $N_t = N_{t-1} - Z_t$, (11.12) yields $N_t \sim B[N_{t-1}, 1-p]$. By induction we find the distributions

$$N_t \sim B[n-1, (1-p)^t] \text{ for } 0 \le t \le n.$$

If $T = t$, it is necessary (though not sufficient, due to fictitious continuation) that $N_t = n - t$. This yields the useful inequalities:

Theorem 11.5.1 *In* $G(n,p)$

$$\Pr[|C(v)| = t] \leq \Pr[B[n-1,(1-p)^t] = n-t] \tag{11.13}$$

or, equivalently,

$$\Pr[|C(v)| = t] \leq \Pr[B[n-1,1-(1-p)^t] = t-1]. \tag{11.14}$$

An Alternate Analysis

The following analysis of $C(v)$ on $G(n,p)$ has been explored by van der Hofstad and Spencer (2006). Each $w \neq v$ flips a coin, heads with probability p, repeatedly until getting a head. Let X_w denote that flip on which w gets a head. Suppose $X_w = j$. Then w enters the BFS at time j. (However, it may have missed the boat if the BFS has already terminated.) This reverses the usual randomness; we are here imagining the $w \neq v$ trying to get into the BFS tree, rather than the BFS tree trying to expand by finding neutral vertices. Suppose $t = |C(v)|$. Every $w \neq v$ that is in $C(v)$ must have entered by time t so $X_w \leq t$. Every $w \neq v$ that is not in $C(v)$ had t opportunities to enter $C(v)$ and so $X_w > t$. Thus $\Pr[|C(v)| = t]$ is at most the probability that $X_w \leq t$ for *precisely* $t-1$ $w \neq v$. For each $w \neq v$, $\Pr[X_w = t] = 1-(1-p)^t$ and these events are independent over w, yielding (11.14). In van der Hofstad and Spencer (2006), this analysis is extended to give more precise bounds on $\Pr[|C(v)| = t]$.

11.6 THE GRAPH AND POISSON PROCESSES COMPARED

Set $p = \frac{c}{n}$. A key observation is that $Z_1 \sim B[n-1, \frac{c}{n}]$ approaches (in n) the Poisson distribution with mean c. Further, in a more rough sense, the same holds for Z_t as long as $N_{t-1} \sim o(n)$ or, equivalently, the number of live and dead vertices is $o(n)$. That is, *the generation of* $C(v)$ *mimics the Poisson branching process with mean c as long as the number of vertices found is not too large.* This allows for a very accurate description in the very subcritical regime $c < 1$. But in the very supercritical regime $c > 1$, the relationship between the generation of $C(v)$ and the Poisson branching process breaks down. As the number N_{t-1} of neutral vertices drops, so does the expected number $\mathrm{E}[Z_t]$ of vertices added to the queue. Eventually, the drift of the walk Y_t lowers from positive to negative, and this eventually causes the process to halt. We call this phenomenon *the ecological limitation.* Indeed, there must be an ecological limitation. The Poisson branching process becomes infinite with positive probability, the component $C(v)$ tautologically cannot be greater than n.

Theorem 11.6.1 *For any positive real c and any fixed integer k*

$$\lim_{n\to\infty} \Pr[|C(v)| = k \text{ in } G(n,c/n)] = \Pr[T_c = k].$$

Proof. Let $Z_t^{\text{po}}, T^{\text{po}}$, and $Z_t^{\text{gr}}, T^{\text{gr}}$ denote the values in the Poisson branching process with parameter c and the graph branching process with parameters n, p respectively. Let Γ denote the set of k-tuples $\vec{z} = (z_1, \dots, z_k)$ of nonnegative integers such that the recursion $y_0 = 1, y_t = y_{t-1} + z_t - 1$ has $y_t > 0$ for $t < k$ and $y_k = 0$. Then

$$\Pr[T^{\text{gr}} = k] = \sum \Pr[Z_i^{\text{gr}} = z_i, 1 \le i \le k]$$
$$\Pr[T^{\text{po}} = k] = \sum \Pr[Z_i^{\text{po}} = z_i, 1 \le i \le k],$$

where both sums are over $\vec{z} \in \Gamma$. Fix such a $\vec{z}$.

$$\Pr[Z_i^{\text{gr}} = z_i, 1 \le i \le k] = \prod_{i=1}^{k} \Pr[B[N_{i-1}^{\text{gr}}, p] = z_i].$$

As i, y_{i-1}, z_i are fixed, $Z_{i-1} = n - O(1)$ and $B[Z_{i-1}, p]$ approaches the Poisson distribution. More precisely,

$$\lim_{n\to\infty} \Pr[B[Z_{i-1}, p] = z_i] = \Pr[Z_i^{\text{po}} = z_i].$$

Further, as the products are of a fixed number of terms

$$\lim_{n\to\infty} \Pr[Z_i^{\text{gr}} = z_i, 1 \le i \le k] = \Pr[Z_i^{\text{po}} = z_i, 1 \le i \le k]. \qquad \blacksquare$$

Now we prove Theorem 11.4.2. From Theorem 11.6.1,

$$\Pr[T_c^{\text{po}} = k] = \lim_{n\to\infty} \Pr[|C(v)| = k],$$

where the second probability is in $G(n,p)$ with $p = \frac{c}{n}$ and v is an arbitrary vertex of that graph. There are $\binom{n}{k-1}$ choices for $S := C(v)$. On any particular S, there is probability $O(p^k) = O(n^{-k})$ that $G(n,p)$ has more than $k-1$ edges. If $G(n,p)$ has precisely $k-1$ edges on S, they must form a tree. There are k^{k-2} such trees. Each occurs with probability $p^{k-1}(1-p)^{\binom{k}{2}-k+1} \sim p^{k-1} = c^{k-1}n^{1-k}$. Thus the total probability that $G(n,p)$ restricted to S forms a connected graph is $\sim k^{k-2}c^{k-1}n^{1-k}$. For $S = C(v)$, we must further have no edges between S and its complement; this has probability $(1-p)^{k(n-k)} \sim e^{-ck}$. Thus

$$\Pr[C(v) = k] \sim \binom{n}{k-1} k^{k-2}c^{k-1}n^{1-k}e^{-ck} \to \frac{e^{-ck}(ck)^{k-1}}{k!}$$

as desired.

The graph branching process can be compared to the binomial branching process in both directions. An important cautionary note: the event $T_{n-1,p}^{\text{bin}} \ge u$ in Theorem 11.6.2 (and similarly $T_{n-u,p}^{\text{bin}} \ge u$ in Theorem 11.6.3) includes the possibility

that the binomial branching process is infinite. Indeed, in application this will be the critical term.

Theorem 11.6.2 *For any u*

$$\Pr[T^{gr}_{n,p} \geq u] \leq \Pr[T^{bin}_{n-1,p} \geq u].$$

Proof. We modify the graph branching process by constantly replenishing the supply of neutral vertices. That is, when we pop the vertex w and there are $n-1-s$ neutral vertices, we create s fictional vertices w' and allow w, w' to be adjacent with probability p. This gives a component of size $T^{\text{bin}}_{n-1,p}$; the actual $C(v)$ will be a subset of it. Thus $T^{\text{bin}}_{n-1,p}$ dominates $T^{\text{gr}}_{n,p}$. ■

Theorem 11.6.3 *For any u*

$$\Pr[T^{gr}_{n,p} \geq u] \geq \Pr[T^{bin}_{n-u,p} \geq u].$$

Proof. We halt the graph branching process when the number of found (live plus dead) vertices reaches u. This does not affect the probability of finding at least u vertices. In this truncated graph process, we diminish the number of neutral vertices to $n-u$. That is, when we pop the vertex w and there are $n-1-s \geq n-u$ neutral vertices, we select $n-u$ of them and only allow adjacencies w, w' to them. The truncated graph process dominates this truncated binomial $n-u, p$ process and so has a greater or equal probability of reaching u. ■

The Poisson Approximation

We are working in the range $p = \Theta(n^{-1})$. There, the binomial $B[n-1, p]$ distribution and the Poisson distribution with mean np are very close. The Poisson branching process is precisely understood and, we feel, is the "purest" branching process. Our goal in this chapter is to give the reader a picture for the "why" of the various regimes. To do this, we shall often avoid the technical calculations and simply assume that the binomial $n-1, p$ branching process is very close to the Poisson branching process with mean np.

11.7 THE PARAMETRIZATION EXPLAINED

In the parametrization (11.3) for the critical window, why is the exponent $-\frac{1}{3}$ as opposed to, say, $-\frac{1}{4}$ or $-\frac{2}{7}$, or something completely different? In the experience of the authors, this is the question most frequently asked about the Erdős–Rényi phase transition. Here is a heuristic that may be helpful.

Parametrize $p = \frac{1+\epsilon}{n}$ with $\epsilon = \epsilon(n)$ positive and approaching zero. We look for the following picture: Consider the Poisson branching process $T = T^{\text{po}}_{1+\epsilon}$. It is infinite with

probability $\sim 2\epsilon$; otherwise its probability of exceeding $A\epsilon^{-2}$ drops exponentially in A. The graph branching process mimics the Poisson branching process as long as it is not too successful. The cases when the Poisson branching process is finite are mimicked, yielding components of size up to roughly ϵ^{-2}. The cases when the Poisson branching process is infinite are mimicked by components that "escape" until the ecological limitation sets in. These components all join together. They form a single component, the dominant component, of size $2\epsilon n$.

In order for the above (admittedly rough) picture to hold, there must a distinction between the small components, up to size ϵ^{-2}, and the dominant component of size $2\epsilon n$. That is, we need $2\epsilon n \gg \epsilon^{-2}$. This heuristic leads us to $\epsilon = n^{-1/3}$ as the breakpoint. When $\epsilon \gg n^{-1/3}$, we have the distinction between small and dominant components and are in the supercritical regime. When $\epsilon = O(n^{-1/3})$, there is no effective analogy to the Poisson branching process being infinite, and there is no dominant component.

11.8 THE SUBCRITICAL REGIONS

Let $p = \frac{c}{n}$ with $c < 1$. Theorem 11.6.2 gives

$$\Pr[T^{\text{gr}}_{n,p} \geq u] \leq \Pr[T^{\text{bin}}_{n-1,p} \geq u].$$

With the Poisson approximation, $\Pr[|C(v)| \geq u] \leq (1 + o(1)) \Pr[T_c \geq u]$. From (11.7), this drops exponentially in u. Taking $u = K \ln n$ for appropriately large K, $\Pr[|C(v)| \geq u] < n^{-1.01}$. As this holds for each of the n vertices v, the probability that *any* v has $|C(v)| \geq u$ is less than $nn^{-1.01} \to 0$. That is, $L_1 = O(\ln n)$ with probability tending to 1.

Let us push this argument into the barely subcritical regime $p = \frac{1-\epsilon}{n}$ with $\epsilon = \lambda n^{-1/3}$. Let I_v be the indicator random variable for $C(v)$ having at least u vertices, with u to be determined below. As above, Theorem 11.6.2 and our Poisson approximation give the bound

$$\Pr[|C(v)| \geq u] \leq (1 + o(1)) \Pr[T_{1-\epsilon} \geq u].$$

We now parametrize

$$u = K\epsilon^{-2} \ln \lambda = Kn^{2/3}\lambda^{-2} \ln \lambda.$$

For an appropriately large constant K, the bound (11.9) gives

$$\Pr[T_{1-\epsilon} \geq u] \leq \epsilon e^{-3.1\lambda} = \epsilon\lambda^{-3.1}.$$

Let $X = \sum_v I_v$ be the number of vertices v in components of size at least u, and let Y be the number of components of $G(n,p)$ of size at least u. Linearity of Expectation gives

$$\mathrm{E}[X] = n \,\mathrm{E}[I_v] \leq n\epsilon\lambda^{-3.1} = n^{2/3}\lambda^{-2.1}.$$

As $Y \le Xu^{-1}$,

$$\mathrm{E}[Y] \le u^{-1}\,\mathrm{E}[X] \le K^{-1}\lambda^{-0.1} \to 0\ .$$

With probability approaching 1, $Y = 0$, and so

$$L_1 \le u = K\epsilon^{-2}\ln\ \lambda = Kn^{2/3}\lambda^{-2}\ln\ \lambda\ .$$

11.9 THE SUPERCRITICAL REGIMES

In the supercritical regimes, there are two salient points about the giant or dominant component. First, it exists. Second, it is unique. Neither is trivial.

We start with the very supercritical region, $p = \frac{c}{n}$, with $c > 1$ constant. The ideas here will carry into the barely supercritical region. Let $y = y(c)$ be the positive real solution of the equation $e^{-cy} = 1 - y$. Let δ be an arbitrarily small constant, and let K be an appropriately large constant. Set $S = K\ln\ n$, $L^- = (y - \delta)n$, $L^+ = (y + \delta)n$. Call a component $C(v)$ and its size $|C(v)|$ small if $|C(v)| < S$, giant if $L^- < |C(v)| < L^+$, awkward otherwise.

No Middle Ground. We claim that the probability of having any awkward component is $o(n^{-20})$. (We could make 20 arbitrarily large by changing K.) There are n choices for v and n choices for $t = |C(v)|$. Thus it suffices to show that, for any v and for any awkward t, $\Pr[|C(v)| = t] = o(n^{-18})$. From Theorem 11.5.1 it suffices to bound $\Pr\left[B[n-1, 1-(1-\frac{c}{n})^t] = t-1\right]$. We indicate the technical calculations. When $t = o(n)$, $1-(1-\frac{c}{n})^t \sim \frac{cn}{t}$, and $c > 1$, so $\Pr\left[B[n-1, 1-(1-\frac{c}{n})^t] \le t-1\right]$ is exponentially small in t. As $t \ge K\ln\ n$, this is polynomially small in n. When $t \sim xn$, $1-(1-\frac{c}{n})^t \sim 1-e^{-cx}$. For $x \ne y$, $1-e^{-cx} \ne x$, so the mean of the binomial is not near t and the probability that it is equal to t is exponentially small in n. In all cases, the bounds on $\Pr[|C(v)| = t]$ follow from basic Chernoff bounds.

Escape Probability. Set $\alpha = \Pr[C(v)$ not small]. (When this happens, we like to think that the BFS on $G(n,p)$ starting with root v has escaped an early death.) Theorems 11.6.2 and 11.6.3 sandwich

$$\Pr[T^{\mathrm{bin}}_{n-S,p} \ge S] \le \alpha \le \Pr[T^{\mathrm{bin}}_{n-1,p} \ge S].$$

From our Poisson approximation, both $\Pr[T^{\mathrm{bin}}_{n-S,p} \ge S]$ and $\Pr[T^{\mathrm{bin}}_{n-1,p} \ge S]$ are asymptotic to $\Pr[T_c \ge S]$. Thus $\alpha \sim \Pr[T_c \ge S]$. As c is assumed fixed and $S \to \infty$,

$$\alpha \sim \Pr[T_c \ge S] \sim \Pr[T_c = \infty] = y$$

with y as in (11.4).

Because there is no middle ground, not small is the same as giant. $C(v)$ is giant with probability $\sim y$. Thus the expected number of vertices in giant components is $\sim yn$. Each giant component has a size between $(y-\delta)n$ and $(y+\delta)n$. Our goal is a single giant component of size $\sim yn$. We are almost there. But maybe with probability $1/2$ there are two giant components.

Sprinkling. Set $p_1 = n^{-3/2}$. (Any p_1 with $n^{-2} \ll p_1 \ll n^{-1}$ would do here.) Let $G_1 \sim G(n, p_1)$ be selected independently from $G \sim G(n,p)$ on the same vertex set, and let $G^+ = G \cup G_1$ so that $G^+ \sim G(n, p^+)$ with $p^+ = p + p_1 - pp_1$. (We "sprinkle" the relatively few edges of G_1 on G to make G^+.) Suppose $G(n,p)$ had more than one giant component, and let V_1, V_2 be the vertex sets of two of those components. There are $\Omega(n^2)$ pairs $\{v_1, v_2\}$ with $v_1 \in V_1$, $v_2 \in V_2$. We have selected p_1 large enough so that with probability $1 - o(1)$ at least one of these pairs is in the sprinkling G_1. Adding this edge merges components V_1, V_2 into a component of size at least $2y(1-\delta)n$ in G^+. We have selected p_1 small enough so that $p^+ \sim p = \frac{c}{n}$. The probability that G^+ has a component so large, and hence awkward, is therefore $o(n^{-20})$. Hence the probability that G had more than one giant component is $o(n^{-20})$.

Finally, we make δ arbitrarily small. $G(n,p)$ has an expected number $\sim yn$ of points in giant components and giant components all have size $\sim yn$. Further, by the sprinkling argument, the contribution to this expectation from the possibility of G having more than one giant component is $o(nn^{-20})$, which is negligible. Thus with probability $1 - o(1)$, there is precisely one giant component. This gives the salient features of the very supercritical phase. There is a giant component, so $L_1 \sim yn$. There is only one giant component and no middle ground, so $L_2 \le S = O(\ln\ n)$. The Sprinkling for Complexity argument given below in the Barely Supercritical Phase can be easily modified to show that the giant component has high complexity, indeed, complexity $\Omega(n)$.

The Barely Supercritical Phase. Set $p = \frac{1+\epsilon}{n}$ with $\epsilon = \lambda n^{-1/3}$ and $\lambda \to \infty$. Note $\epsilon^{-2} = \lambda^{-2} n^{2/3} \ll 2\epsilon n$. The analysis of the barely supercritical region becomes more difficult as $\lambda = \lambda(n)$ approaches infinity more slowly. We shall add the simplifying assumption that $\lambda \gg \ln\ n$. Further, we shall find somewhat weaker bounds than stated on L_2.

Bollobás (1984) in 1984 showed the existence of the dominant component when $\lambda > K\sqrt{\ln\ n}$, K constant. That paper was the first indication of the appropriate scaling for the critical window. Łuczak (1990) tightened the result to "best possible," showing that if $\lambda \to +\infty$ then the dominant component exists.

Let δ be an arbitrarily small constant, and let K be an appropriately large constant. Set $S = K\epsilon^{-2} \ln\ n$, $L^- = (1-\delta)2\epsilon n$, and $L^+ = (1+\delta)2\epsilon n$. Call a component $C(v)$ and its size $|C(v)|$ small if $|C(v)| < S$, dominant if $L^- < |C(v)| < L^+$, and awkward otherwise.

No Middle Ground. We claim that the probability of having any awkward component is $o(n^{-20})$. (We could make 20 arbitrarily large by changing K.) There are n choices for v and n choices for $t = |C(v)|$. Thus it suffices to show that, for any v and for any awkward t, $\Pr[|C(v)| = t] = o(n^{-18})$. Again,

we bound $\Pr[B[n-1, 1-(1-p)^t] = t-1]$. We indicate the technical calculations. Let μ and σ^2 denote the mean and variance of the binomial. Then $\mu = (n-1)(1-(1-p)^t)$ and, in this range, $\sigma^2 \sim \mu$. When $t = o(n\epsilon)$, we estimate $1-(1-p)^t$ by $pt = t + t\epsilon$ so that μ is estimated by $npt \sim (1+\epsilon)t$. Then $\mu - t \sim t\epsilon$ and $\sigma^2 \sim t$. This probability is roughly $\exp[-(t\epsilon)^2/2t] = \exp[-t\epsilon^2/2]$. As $t \geq S$, the exponent is $o(n^{-18})$ for $K > 36$. (To push S down to $K\epsilon^{-2}\ln(\lambda)$ requires a finer bound on $\Pr[|C(v)| = t]$.) Now suppose $t \sim xn\epsilon$, where $x \neq 2$. The ecological limitation now has an effect, and we estimate $1-(1-p)^t$ by $pt - \frac{1}{2}p^2t^2$ so

$$\mu - t \sim t\epsilon - \frac{1}{2}t^2n^{-2} \sim (n\epsilon)(x - \frac{1}{2}x^2).$$

(Observe that, when $x = 2$, the mean of the binomial is very close to t and so we do not get a small bound on $\Pr[|C(v)| = t]$. This is natural when we consider that there will be a dominant component of size $\sim 2\epsilon n$.) Again, $\sigma^2 \sim t$, so the probability is $\exp[-\Omega((n\epsilon)^2/t)]$, which is extremely small. When $t \gg n\epsilon$, the probability is even smaller.

Escape Probability. Set $\alpha = \Pr[C(v) \text{ not small}]$. Theorems 11.6.2 and 11.6.3 sandwich

$$\Pr\left[T^{\text{bin}}_{n-S,p} \geq S\right] \leq \alpha \leq \Pr\left[T^{\text{bin}}_{n-1,p} \geq S\right].$$

The Poisson approximation for $T^{\text{bin}}_{n-1,p}$ is $T_{1+\epsilon}$. As $S \gg \epsilon^{-2}$, bound (11.11) gives

$$\alpha \leq \Pr[T_{1+\epsilon} \geq S] \sim \Pr[T_{1+\epsilon} = \infty] \sim 2\epsilon\ .$$

Replacing $n-1$ by $n-S$ lowers the mean by $\sim Sn^{-1}$. But $Sn^{-1}/\epsilon \sim (\ln n)/(n\epsilon^3) = \lambda^{-3}\ln n$, and we have made λ large enough that this is $o(1)$. That is, $Sn^{-1} = o(\epsilon)$. Therefore, $T^{\text{bin}}_{n-S,p}$ is approximated by $T_{1+\epsilon-o(\epsilon)}$ and

$$\alpha \geq \Pr[T_{1+\epsilon+o(\epsilon)} \geq S] \sim \Pr[T_{1+\epsilon+o(\epsilon)} = \infty] \sim 2\epsilon\ .$$

α has been sandwiched and $\alpha \sim 2\epsilon$.

Because there is no middle ground, not small is the same as dominant. $C(v)$ is dominant with probability $\sim 2\epsilon$. Thus the expected number of vertices in the dominant components is $\sim 2n\epsilon$. Each giant component has size between $(1-\delta)2n\epsilon$ and $(1+\delta)2n\epsilon$. As in the very supercritical case, we need worry about having more than one giant component.

Sprinkling. Set $p_1 = n^{-4/3}$. Let $G_1 \sim G(n, p_1)$ be selected independently from $G \sim G(n,p)$ on the same vertex set and let $G^+ = G \cup G_1$ so that $G^+ \sim G(n, p^+)$ with $p^+ = p + p_1 - pp_1 = 1 + \epsilon + o(\epsilon)$. Suppose $G(n,p)$ had more than one giant component and let V_1, V_2 be the vertex sets of two of those components. There are $\gg n^{4/3}$ pairs $\{v_1, v_2\}$ with $v_1 \in V_1, v_2 \in V_2$. With probability $1 + o(1)$, at least one of these pairs is in the sprinkling G_1. Adding this edge merges components V_1, V_2 into a component of size at least $(1-\delta)\ 4\epsilon n$ in G^+. The probability that G^+ has such a large, and hence awkward, component is $o(n^{-20})$. Thus the probability that

G had two (or more) dominant components is $o(n^{-20})$. Taking δ arbitrarily small, as in the supercritical case, G has with probability $1 - o(1)$ precisely one dominant component. Thus $L_1 \sim 2n\epsilon$ and, as there is no middle ground, $L_2 \leq K\epsilon^{-2} \ln\ n$.

Sprinkling for Complexity. Take $p_1 = (1 + \frac{\epsilon}{2})/n$ and $p_2 \sim \frac{\epsilon}{2}/n$, so that $p_1 + p_2 - p_1p_2 = (1 + \epsilon)/n$. Let $G_1 \sim G(n, p_1)$, $G_2 \sim G(n, p_2)$, and $G_3 = G_1 \cup G_2$ so that $G_3 \sim G(n, (1 + \epsilon)/n)$. G_1, G_3 will have dominant components V_1, V_3 of sizes $\sim n\epsilon$ and $\sim 2n\epsilon$. As G_3 has "no middle ground" in its component sizes, $V_1 \subseteq V_3$. Now the sprinkling G_2 adds $\sim p_2 \binom{n\epsilon}{2} \sim n\epsilon^3/2$ edges internal to V_1. Thus V_3 will have complexity at least $n\epsilon^3/2 = \lambda^3/2$, which approaches infinity.

11.10 THE CRITICAL WINDOW

We now fix a real λ and set $p = \frac{1}{n} + \lambda n^{-4/3}$. There has been massive study of this critical window, Łuczak (1990) and the monumental Janson et al. (1993) being only two examples. Calculations in this regime are remarkably delicate.

Fix $c > 0$, and let X be the number of tree components of size $k = cn^{2/3}$. Then

$$\mathrm{E}[X] = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-(k-1)}.$$

Watch the terms cancel!

$$\binom{n}{k} = \frac{(n)_k}{k!} \sim \frac{n^k e^k}{k^k \sqrt{2\pi k}} \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right).$$

For $i < k$,

$$-\ln\ \left(1 - \frac{i}{n}\right) = \frac{i}{n} + \frac{i^2}{2n^2} + O\left(\frac{i^3}{n^3}\right)\ ,$$

so that

$$\sum_{i=1}^{k-1} -\ln\left(1 - \frac{i}{n}\right) = \frac{k^2}{2n} + \frac{k^3}{6n^2} + o(1) = \frac{k^2}{2n} + \frac{c^3}{6} + o(1)\ .$$

Also $p^{k-1} = n^{1-k}(1 + \lambda n^{-1/3})^{k-1}$ and expanding $\ln(1 + \epsilon) = \epsilon - \frac{1}{2}\epsilon^2 + O(\epsilon^3)$:

$$(k-1)\ln(1 + \lambda n^{-1/3}) = k\lambda n^{-1/3} - \frac{1}{2}c\lambda^2 + o(1)\ .$$

Also

$$\ln(1-p) = -p + O(n^{-2}) = -\frac{1}{n} - \frac{\lambda}{n^{4/3}} + O(n^{-2})$$

and

$$k(n-k) + \binom{k}{2} - (k-1) = kn - \frac{k^2}{2} + O(n^{2/3})\ ,$$

so that

$$[k(n-k)+\binom{k}{2}-(k-1)]\ln(1-p)=-k+\frac{k^2}{2n}-\frac{\lambda k}{n^{1/3}}+\frac{\lambda c^2}{2}+o(1)$$

and

$$\mathrm{E}[X]\sim\frac{n^k k^{k-2}}{k^k\sqrt{2\pi k}n^{k-1}}e^A=nk^{-5/2}(2\pi)^{-1/2}e^A\ .$$

Here $A=k-\frac{k^2}{2n}-\frac{c^3}{6}+\frac{\lambda k}{n^{1/3}}-\frac{\lambda^2 c}{2}-k+\frac{k^2}{2n}-\frac{\lambda k}{n^{1/3}}+\frac{\lambda c^2}{2}$. The k and n terms cancel, and we can give A the intriguing form

$$A=A(c)=\frac{(\lambda-c)^3-\lambda^3}{6}\ .$$

Writing k in terms of n then yields

$$\mathrm{E}[X]\sim n^{-2/3}e^{A(c)}c^{-5/2}(2\pi)^{-1/2}\ .$$

For any particular such k, $\mathrm{E}[X]\to 0$ but, if we sum k between $cn^{2/3}$ and $(c+dc)n^{2/3}$, we multiply by $n^{2/3}dc$. Going to the limit gives an integral. For any fixed a,b,λ, let X be the number of tree components of size between $an^{2/3}$ and $bn^{2/3}$. Then

$$\lim_{n\to\infty}\mathrm{E}[X]=\int_a^b e^{A(c)}c^{-5/2}(2\pi)^{-1/2}dc\ .$$

The large components are not all trees. Wright (1977) proved that for fixed $l\geq 0$ there are asymptotically $c_l k^{k-2+\frac{3}{2}l}$ connected graphs on k points with $k-1+l$ edges, where $c_0=1$, $c_1=\sqrt{\pi/8}$ and c_l was given by a specific recurrence. Asymptotically in l, $c_l=l^{-l/2(1+o(1))}$. The calculation for $X^{(l)}$, the number of such components on k vertices, leads to extra factors of $c_l k^{\frac{3}{2}l}$ and n^{-l}, which gives $c_l c^{\frac{3}{2}l}$. For fixed a,b,λ,l, the number $X^{(l)}$ of components of size between $an^{2/3}$ and $bn^{2/3}$ with $l-1$ more edges than vertices satisfies

$$\lim_{n\to\infty}\mathrm{E}[X^{(l)}]=\int_a^b e^{A(c)}c^{-5/2}(2\pi)^{-1/2}\left(c_l c^{\frac{3}{2}l}\right)\,dc,$$

and letting X^* be the total number of components of size between $an^{2/3}$ and $bn^{2/3}$

$$\lim_{n\to\infty}\mathrm{E}[X^*]=\int_a^b e^{A(c)}c^{-5/2}(2\pi)^{-1/2}g(c)\,dc\ ,$$

where

$$g(c)=\sum_{l=0}^{\infty}c_l c^{\frac{3}{2}l}\ ,$$

a sum convergent for all c. A component of size $\sim cn^{2/3}$ will have probability $c_l c^{\frac{3}{2}l}/g(c)$ of having complexity l, independent of λ. As $\lim_{c\to 0} g(c) = 1$, most components of size $\epsilon n^{2/3}$, $\epsilon \ll 1$, are trees, but as c gets bigger, the distribution on l moves inexorably higher.

An Overview

For any fixed λ, the sizes of the largest components are of the form $cn^{2/3}$ with a distribution over the constant. This distribution has as its support the positive reals. Thus, for example, for $\lambda = -4$, there is some positive limiting probability that the largest component is bigger than $10n^{2/3}$, and for $\lambda = +4$ there is some positive limiting probability that the largest component is smaller than $0.1n^{2/3}$, though both these probabilities are miniscule. The $c^{-5/2}$ term dominates the integral as $c \to 0^+$, reflecting the notion that for any fixed λ there should be many components of size near $\epsilon n^{2/3}$ for $\epsilon = \epsilon(\lambda)$ appropriately small. When λ is large negative (e.g., $\lambda = -4$), the largest component is likely to be $\epsilon n^{2/3}$, ϵ small, and there will be many components of nearly that size. The non-tree components will be a negligible fraction of the tree components. When λ is large positive (e.g., $\lambda = +4$), the dominant component will have begun to emerge. The largest component is likely to be $\sim 2\lambda n^{2/3}$ and of moderately high (not zero or one) complexity, and the second largest component will be considerably smaller and simple.

Now consider the evolution of $G(n, p)$ in terms of λ. Suppose that at a given λ there are components of size $c_1 n^{2/3}$ and $c_2 n^{2/3}$. When we move from λ to $\lambda + d\lambda$, there is a probability $c_1 c_2 d\lambda$ that they will merge. Components have a peculiar gravitation in which the probability of merging is proportional to their sizes. With probability $(c_1^2/2)d\lambda$, there will be a new internal edge in a component of size $c_1 n^{2/3}$ so that large components rarely remain trees. Simultaneously, big components are eating up other vertices.

With $\lambda = -4$, say, we have feudalism. Many small components (castles) are each vying to be the largest. As λ increases, the components increase in size and a few large components (nations) emerge. An already large France has much better chances of becoming larger than a smaller Andorra. The largest components tend strongly to merge, and by $\lambda = +4$, it is very likely that a dominant component, a Roman Empire, has emerged. With high probability, this component is nevermore challenged for supremacy but continues absorbing smaller components until full connectivity – One World – is achieved.

11.11 ANALOGIES TO CLASSICAL PERCOLATION THEORY

The study of percolation has involved the intense efforts of both mathematicians and physicists for many years. A central object of that study has been bond percolation on Z^d, as described below. Here we explore, without proofs, the fruitful analogies between that percolation and the Erdős–Rényi phase transition. Grimmett (1999) *Percolation* is a classic text in this field, and we shall follow its treatment.

Let $d \geq 2$. (All parameters below shall depend on the choice of d.) Let Z^d, as usual, represent the set of $\vec{a} = (a_1, \ldots, a_d)$ with a_i integers. The d-dimensional cubic lattice, written L^d, is that graph with vertices Z^d, two vertices $\vec{a}, \vec{b}$ being adjacent if they agree on $d-1$ coordinates and differ by one on the other coordinate. Let $p \in [0,1]$. The random subgraph $L^d(p)$ contains each edge of L^d (and no others) with independent probability p. We let $C(\vec{a})$ denote the connected component of $L^d(p)$ containing the vertex $\vec{a}$. We generally examine $C(\vec{0})$ as, by symmetry, all $C(\vec{a})$ look the same. (In Grimmett (1999) and elsewhere the edges of L^d are called bonds and they are open with probability p and closed otherwise. The word "cluster" is used in place of connected component.) Naturally, as p becomes larger, $L^d(p)$ will have more adjacencies. There is a critical probability, denoted by p_c, at which $L^d(p)$ undergoes a macroscopic change.

- For $p < p_c$, the subcritical region, all connected components are finite.
- For $p > p_c$, the supercritical region, there is precisely one infinite component.
- For $p = p_c$, at the critical point, the situation is particularly delicate, as discussed below.

The constant probabilities of bond percolation correspond to probabilities parametrized $p = c/n$ in the Erdős–Rényi $G(n,p)$. The value $c = 1$ is then the critical probability in the Erdős–Rényi model.

- The infinite component in the bond percolation model is analogous to giant components, components of size $\Omega(n)$, in the Erdős–Rényi model.
- The finite components in the bond percolation model are analogous to components of size $O(\ln\ n)$ in the Erdős–Rényi model.

The uniqueness of the infinite component in bond percolation was an open question (though the physicists "knew" it was true!) for many years. It was proven by Aizenman, Kesten and Newman (1987), the Book Proof is given by Burton and Keane (1989). This corresponds to the uniqueness of the giant component in $G(n,p)$.

In the bond percolation model there are only three choices for p, that is, it can be less than, greater than, or equal to p_c. The barely subcritical and barely supercritical phases of the Erdős–Rényi model correspond to an asymptotic study of the bond percolation model as p approaches p_c from below and from above, respectively. This study is done through the use of critical exponents as described below.

Set $\theta(p) = \Pr[C(\vec{0})\ \text{is infinite}]$. For $p < p_c$, $\theta(p) = 0$, as there are no infinite components with probability 1. For $p > p_c$, $\theta(p) > 0$. This corresponds to the infinite component having positive density, strengthening the analogy to the giant components of the Erdős–Rényi model. When p is barely greater than p_c, there will be an infinite component but its density will be very small. The critical exponent β is that real number so that

$$\theta(p) = (p - p_c)^{\beta + o(1)} \quad \text{as } p \to p_c^+.$$

(As mathematicians we are aware that $\theta(p)$ *could* behave erratically as $p \to p_c^+$ and β might not exist. This holds for all critical exponents we discuss. For a physicist, there is no doubt that the critical exponents do exist, and they can tell you the values to a few decimal places!) Analogously, in the Erdős–Rényi model $\theta(c)$ is the proportion of points in the giant component, that $y = y(c) > 0$ satisfying (11.4). From (11.8), $y(1+\epsilon) \sim 2\epsilon$ as $\epsilon \to 0^+$. Therefore, $\beta = 1$.

The susceptibility, denoted by $\chi(p)$ (not to be confused with chromatic number) is given by $\chi(p) = \mathrm{E}[|C(\vec{0})|]$. For $p > p_c$, $\chi(p) = \infty$ as with positive probability $C(\vec{0})$ is infinite. For $p < p_c$, $\chi(p)$ is finite and $\chi(p) \to \infty$ as $p \to p_c^-$. That the susceptibility approaches infinity at the same critical value for which an infinite component appears is not at all obvious, and was one of the great developments of the field, due independently to Alon and Boppana (1987) and Men'shikov (1986). When p is barely less than p_c, $\chi(p)$ will be finite but large. The critical number γ is that real number so that

$$\chi(p) = (p_c - p)^{-\gamma + o(1)} \text{ as } p \to p_c^-.$$

Analogously, in the Erdős–Rényi model we examine $\mathrm{E}[|C(v)|]$ in $G(n, \frac{1-\epsilon}{n})$. In the subcritical region, this is well mirrored by $T_{1-\epsilon}$, the total size of a subcritical Poisson branching process. We find $\mathrm{E}[T_{1-\epsilon}]$ by looking at each generation. There is one root Eve, who has an expected number $1-\epsilon$ children. They behave similarly, and so Eve has an expected number $(1-\epsilon)^2$ of grandchildren. This continues, there are an expected number $(1-\epsilon)^i$ nodes in the ith generation so that

$$\mathrm{E}[T_{1-\epsilon}] = \sum_{i=0}^{\infty} (1-\epsilon)^i = \epsilon^{-1}$$

precisely. Therefore $\gamma = 1$.

While $\chi(p)$ is infinite in the supercritical region, we can examine the "finite portion" of $L^d(p)$. The finite susceptibility χ^f is given by

$$\chi^f(p) = \mathrm{E}\left[|C(\vec{0})| \text{ conditional on } C(\vec{0}) \text{ being finite}\right].$$

When p is barely greater than p_c, $\chi^f(p)$ will be finite but large. The critical number γ' is that real number so that

$$\chi^f(p) = (p - p_c)^{-\gamma' + o(1)} \text{ as } p \to p_c^+.$$

The Erdős–Rényi analog is $\mathrm{E}[|C(v)|]$ in $G(n, \frac{1+\epsilon}{n})$, conditional on v not being in the giant component. In $G(n, \frac{1+\epsilon}{n})$, $|C(v)|$ has basically distribution $T^{\mathrm{po}}_{1+\epsilon}$, with the value $T^{\mathrm{po}}_{1+\epsilon} = \infty$ corresponding to being in the giant component. The finite analog then corresponds to $T^{\mathrm{po}}_{1+\epsilon}$, conditional on it being finite. The probability $T^{\mathrm{po}}_{1+\epsilon}$ is finite, and approaches 1 as $\epsilon \to 0^+$. The Poisson branching processes $T^{\mathrm{po}}_{1+\epsilon}$,$T^{\mathrm{po}}_{1-\epsilon}$ have nearly the same finite distribution. Conditioning on v not being in the giant component, $|C(v)|$ then behaves like $T^{\mathrm{po}}_{1-\epsilon}$. Therefore $\gamma' = 1$.

At the critical value $p = p_c$, all components are finite. The distribution of $|C(\vec{0})|$ will have a heavy tail. The critical number δ is that real number so that at $p = p_c$

$$\Pr[|C(\vec{0})| \geq s] = s^{-1/\delta + o(1)} \text{ as } s \to \infty.$$

For the Erdős–Rényi analog we consider $|C(v)|$ in $G(n, 1/n)$. One needs be cautious about the double limit. For any *fixed* s,

$$\lim_{n\to\infty} \Pr[|C(v)| \geq s] = \Pr\left[T_1^{\text{po}} \geq s\right] = \Theta(s^{-1/2})$$

from (11.6). Therefore $\delta = 2$.

We further examine the gap exponent, denoted by Δ. In the subcritical region the distribution of $|C(\vec{0})|$ drops off exponentially. For each $k \geq 1$, it has a finite kth moment. The hypothetical quantity Δ is such that

$$\frac{\mathrm{E}[|C(\vec{0})|^{k+1}]}{\mathrm{E}[|C(\vec{0})|^k]} = (p_c - p)^{-\Delta + o(1)}.$$

The belief is that Δ does not depend on the choice of k. In the supercritical region, the belief is that the same asymptotics hold when the infinite component is erased. More precisely, the belief is that

$$\frac{\mathrm{E}[|C(\vec{0})|^{k+1} \text{ given } C(\vec{0}) \text{ is finite}]}{\mathrm{E}[|C(\vec{0})|^{k} \text{ given } C(\vec{0}) \text{ is finite}]} = (p_c - p)^{-\Delta + o(1)}$$

for all $k \geq 1$. In the Erdős–Rényi analog, the distribution of $C(v)$ in $G(n, \frac{1-\epsilon}{n})$ mirrors that of $T_{1-\epsilon}^{\text{po}}$. (The supercritical $G(n, \frac{1+\epsilon}{n})$, with its giant component erased, behaves similarly.) From Section 11.4, $\Pr[T_{1-\epsilon}^{\text{po}} = s]$ drops like $s^{-3/2}$ until k reaches $\Theta(\epsilon^{-2})$ when it begins its exponential drop-off. The region of exponential drop-off has negligible effect on the finite moments. The kth moment of $T_{1-\epsilon}^{\text{po}}$ is basically the sum of $s^{-3/2}s^k$ for $s = O(\epsilon^{-2})$, which is of order $(\epsilon^{-2})^{k+\frac{1}{2}}$, or ϵ^{-2k-1}. The ratio of the $(k+1)$th and kth moments is then $\Theta(\epsilon^{-2})$. Therefore $\Delta = 2$.

For bond percolation in Z^d, define the triangle function

$$T(p) = \sum_{\vec{x},\vec{y}\in Z^d} \Pr[\vec{0} \leftrightarrow \vec{x}] \Pr[\vec{0} \leftrightarrow \vec{y}] \Pr[\vec{x} \leftrightarrow \vec{y}],$$

where $\vec{x} \leftrightarrow \vec{y}$ means that $\vec{x}, \vec{y}$ lie in the same component. In Aizenman and Newman (1984), the following condition was introduced:

$$\text{Triangle condition: } T(p_c) < \infty.$$

They showed that, when the triangle condition holds, some of the conditions for mean field theory (as discussed below) are valid. For the Erdős–Rényi percolation analog we fix a vertex v of $G(n,p)$ and define the discrete triangle function

$$T(p) = \sum_{x,y} \Pr[v \leftrightarrow x] \Pr[v \leftrightarrow y] \Pr[v \leftrightarrow y].$$

The critical probability p_c is replaced by $p = n^{-1}$. Finiteness is replaced by the boundedness giving the

$$\text{Discrete triangle condition: } T(p) = O(1).$$

The contribution to $T(p)$ when two or three of v, x, y are equal is easily bounded, leaving the contribution from all triples v, x, y of distinct vertices. As all pairs behave the same and there are $(n-1)(n-2) \sim n^2$ terms,

$$T(p) \sim O(1) + n^2 \Pr[v \leftrightarrow x]^3$$

and

$$\Pr[v \leftrightarrow x] = \sum_t \Pr[|C(v)| = t] \frac{t-1}{n-1} \sim n^{-1} \sum_t t \Pr[|C(v)| = t].$$

We know that $\Pr[|C(v)| = t]$ behaves like $t^{-3/2}$ until t reaches $\Theta(n^{2/3})$ and then drops off exponentially. Ignore constants

$$\sum_t t \Pr[|C(v)| = t] = \Theta\left(\sum_{t=\Theta(n^{2/3})} t t^{-3/2} \right) = \Theta((n^{2/3})^{1/2}).$$

Now $\Pr[v \leftrightarrow x] = \Theta(n^{-2/3})$. (Basically, the main contribution to $\Pr[v \leftrightarrow x]$ comes when v lies in a component of size $\Theta(n^{2/3})$, even though that rarely occurs.) The triangle condition does hold as

$$T(p) = O(1) + O(n^2)\Theta(n^{-2/3})^3 = O(1).$$

The discrete triangle condition does not hold in the barely supercritical region. There, $\Pr[v \leftrightarrow x]$ is dominated by the probability that both v, w lie in the dominant component. As the dominant component has size $\gg n^{2/3}$, $\Pr[v \leftrightarrow x] \gg n^{-2/3}$, and $T(p) \gg 1$. This is not mere serendipity. Rather, the boundedness of $T(p)$ provides a natural boundary between the critical window and the barely supercritical region for discrete random structures. This connection is explored in depth in Borgs et al. (2005) and the more recent lecture notes Slade (2006).

Hara and Slade (1990) (see also the survey Hara and Slade (1994)) proved that the triangle condition holds in the bond percolation model for sufficiently high dimensions d. (More precisely, they showed that $T(p)$ could be made very small by

taking p slightly less than p_c. Their argument works for $d \geq 19$ and for all $d > 6$ with a somewhat different model. It is strongly believed that the condition does hold for all $d > 6$.) Building on that, they found that the critical exponent values $\beta = 1$, $\gamma = \gamma' = 2$, $\delta = 2$, $\Delta = 2$ hold for those d. Mathematical physicists have a term *mean field*, which, quoting Grimmett, "permits several interpretations depending on context." A commonly held requirement is that the critical exponents have the values given above. Thus bond percolation for $d \geq 19$ is regarded as exhibiting mean field behavior. Using the analogs described above, it seems reasonable to say that the Erdős–Rényi model exhibits mean field behavior.

11.12 EXERCISES

1. Consider the Poisson branching model with mean $c = 1$ and root Eve. For $n \geq 3$, let A_n be the event where Eve has precisely two children, Dana and Fan, and that the total tree size $T = n$. Let X be the size of the subtree with root Dana. For each $j \geq 1$, find $\lim_{n\to\infty} \Pr[X = j \mid A_n]$. Find an asymptotic formula for $\Pr\left[\frac{n}{3} < X < \frac{2n}{3}\right]$.
2. Consider the binomial branching model with parameters m, p, and $mp > 1$. Set $y = y(m, p) = \Pr[T = \infty]$. Give an implicit equation for y analogous to (11.4). With m fixed set $mp = (1 + \epsilon)$. Find $\lim_{\epsilon\to 0^+} \frac{y(m,p)}{\epsilon}$.
3. Let $c > 1$. Let Z_i, $i = 1, 2, \ldots$, be independent Poisson variables with mean c. For $a > 1$ consider the walk defined by the initial condition $Y_1 = a$ and recursion $Y_t = Y_{t-1} + Z_i - 1$ for $t \geq 2$. Use Chernoff bounds to show $\lim_{a\to\infty} \sum_{t\geq 2} \Pr[Y_t \leq 0] = 0$. Use this to show that the walk defined by initial condition $Y_0 = 1$ and recursion $Y_t = Y_{t-1} + Z_t - 1$ for $t \geq 1$ has a positive probability of being positive for all t.
4. An Openended Computer Experiment. Begin with vertices $1, \ldots, n$ ($n = 10^6$ is very quick when done right) and no edges. In each round, pick two random vertices and add an edge between them. Use a Union-Find algorithm to keep track of the components and the component sizes. Parametrize round number E by $E/\binom{n}{2} = \frac{1}{n} + \lambda n^{-4/3}$ and concentrate on the region $-4 \leq \lambda \leq +4$. Update the 10 largest component sizes, noting particularly when two of the ten largest components merge. Watch the barely subcritical picture at $\lambda = -4$ turn into a barely supercritical picture at $\lambda = +4$ as the bulk of the moderate size components merge to form a dominant component.

THE PROBABILISTIC LENS:
Long paths in the supercritical regime

As described in the last chapter, when $p = \frac{1+\epsilon}{n}$, the graph $G(n,p)$ contains, almost surely, a linear size connected component. Ajtai, Komlós, and Szemerédi proved that in fact it almost surely contains a path of linear length.

Theorem 1 [Ajtai, Komlós and Szemerédi (1981)] *For a fixed small real $\epsilon > 0$ and $p = \frac{1+\epsilon}{n}$, the random graph $G = G(n,p)$ contains, almost surely, a path of length at least $\frac{\epsilon^2 n}{12}$.*

The constant $\frac{1}{12}$ can be easily improved, and we make no attempt to optimize it here.

We describe here a beautiful, short proof, due to Krivelevich and Sudakov (2013). The idea is to run a Depth First Search (DFS) algorithm on the graph, generating it while running the algorithm. The DFS algorithm is a standard graph-searching procedure. Let $[n] = \{1, 2, \ldots, n\}$ be the (ordered) set of vertices of the graph. The algorithm maintains three pairwise disjoint sets of vertices, B (Black), P (Gray), and W (White) whose union is $[n]$, where B is the set of vertices whose exploration is already complete, P is a path consisting of the vertices that have been visited already but whose exploration has not ended, and W are the vertices that have not been revealed yet. At the beginning, $W = [n]$ is the set of all vertices of the graph, and $B = P = \emptyset$, and at the end, $B = [n]$ and $W = P = \emptyset$. At each round of the algorithm, while $W \cup P \neq \emptyset$, if the path P is empty, the algorithm selects the first vertex in W and shifts it to P, creating a path of one vertex. If P is nonempty, let v be its last vertex. The algorithm queries, in order, the existence of edges from v to W and picks the first

neighbor of v in W, if there is such a neighbor, appending it to the end of the path. If there is no such neighbor, v is shifted to B. To complete the exploration of the graph, when $P = W = \emptyset$ and $B = [n]$, the algorithm queries all remaining pairs of vertices in B which have not been queried before.

The crucial properties of the algorithm are that during its process P is always a path, all pairs of vertices from B to W have been queried, and there are no edges between B and W, and that in each round a vertex moves, either from the path P to B, or from W to the path P.

We set $N_0 = \frac{\epsilon n^2}{2}$ and analyze the algorithm until N_0 queries have been made. To do so, consider a run of DFS on the random graph $G(n,p)$ on the set of vertices $[n]$, where $N = \binom{n}{2}$, $p = \frac{1+\epsilon}{n}$ and the algorithm is fed with a sequence of i.i.d. Bernoulli variables $X_1, X_2, \ldots, X_N$, each being 1 with probability p. When the algorithm makes its query number i about a pair of vertices, it gets a positive answer (i.e., an edge) if $X_i = 1$ and a negative answer (a non-edge) if $X_i = 0$. This clearly generates a random graph $G(n,p)$ and enables us to establish the existence of a long path by analyzing the properties of the sequence $\{X_i\}_{i=1}^{N}$. The only property of the sequence we will use is that for N_0 as above, the sum $\sum_{i=1}^{N_0} X_i$ is, almost surely, not much smaller than its expected value. Without trying to optimize the constants, we require

$$\sum_{i=1}^{N_0} X_i \geq pN_0 - \frac{\epsilon^2 n}{12} = \frac{\epsilon n}{2} + \frac{5\epsilon n^2}{12}.$$

By Chernoff's or Chebyshev's Inequality, this occurs almost surely provided $\epsilon^3 n$ tends to infinity, (indeed the argument works also when $\epsilon = \epsilon(n)$ approaches zero appropriately slowly). We stress the fact that the only assumption about the sequence (X_i) is that the number of 1's it contains among its first N_0 terms is not far less than its expectation. This is a pretty mild assumption, but as the proof below shows it suffices to ensure the existence of a long path.

Suppose, for contradiction, that the graph contains no path of length $\frac{\epsilon^2 n}{12}$. In that case, during the course of the algorithm the size of P is always smaller than $\frac{\epsilon^2 n}{12}$. We first claim that after N_0 queries, $|B| < n/3$. Indeed, otherwise sometime before that $|B|$ has been of size exactly $n/3$, as vertices join B one by one. Since $|P| \leq \frac{\epsilon^2 n}{12} < n/3$, at that point $|W| = |[n] - B - P| \geq n/3$ and hence, as all pairs of vertices between B and W have already been queried, the number of queries up to that point exceeds $n/3 \cdot n/3 = n^2/9 > N_0$, which is a contradiction. Therefore, indeed, after N_0 queries B is of size smaller than $n/3$. On the other hand, since every positive answer shifts a vertex from W to P, the assumption about the sequence (X_i) implies that after N_0 queries $|B \cup P| \geq \frac{\epsilon n}{2} + \frac{5\epsilon^2 n}{12}$. As the size of P is smaller than $\frac{\epsilon^2 n}{12}$, it follows that

$$|B| \geq \frac{\epsilon n}{2} + \frac{5\epsilon^2 n}{12} - \frac{\epsilon^2 n}{12} = \frac{\epsilon n}{2} + \frac{\epsilon^2 n}{3}.$$

As $|W| = n - |B| - |P| \geq n - |B| - \frac{\epsilon^2 n}{12}$ and $|B| < n/3$, this implies that

$$|B||W| \geq \left(\frac{\epsilon n}{2} + \frac{\epsilon^2 n}{3}\right)\left(n - \frac{\epsilon n}{2} - \frac{\epsilon^2 n}{3} - \frac{\epsilon^2 n}{12}\right)$$
$$= \frac{\epsilon n^2}{2} + \frac{\epsilon^2 n^2}{3} - \frac{\epsilon^2 n^2}{4} - O(\epsilon^3)n^2 > \frac{\epsilon n^2}{2} = N_0.$$

This is impossible, as the number of queries is only N_0 and yet all pairs of vertices from B to W have been queried already. This establishes the existence of the required path.

12

Circuit Complexity

> It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again; the never-satisfied man is so strange – if he has completed a structure then it is not in order to dwell in it peacefully, but in order to begin another. I imagine the world conqueror must feel thus, who, after one kingdom is scarcely conquered, stretches out his arms for another.
> –Karl Friedrich Gauss

12.1 PRELIMINARIES

A *Boolean function* $f = f(x_1, \ldots, x_n)$ on the n variables $x_1, x_2, \ldots, x_n$ is simply a function $f : \{0,1\}^n \to \{0,1\}$. In particular, $0, 1, x_1 \wedge \cdots \wedge x_n, x_1 \vee \cdots \vee x_n, x_1 \oplus \cdots \oplus x_n$ denote, as usual, the two constant functions, the And function (whose value is 1 iff $x_i = 1$ for all i), the Or function (whose value is 0 iff $x_i = 0$ for all i), and the Parity function (whose value is 0 iff an even number of variables x_i is 1), respectively. For a function f, we let $\overline{f} = f \oplus 1$ denote its complement *Not f*. The functions x_i and $\overline{x}_i$ are called *atoms*. In this section we consider the problem of computing various Boolean functions efficiently. A *circuit* is a directed, acyclic graph, with a special vertex with no outgoing edges called the Output vertex. Every vertex is labeled by a Boolean function of its immediate parents, and the vertices with no parents (i.e., those with no ingoing edges) are labeled either by

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

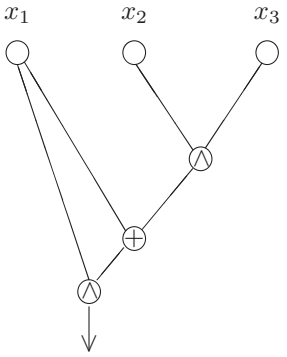


Figure 12.1 A Boolean circuit.

one of the variables x_i or by a constant 0 or 1. For every assignment of binary values to each variable x_i, one can compute, recursively, the corresponding value of each vertex of the circuit by applying the corresponding function labeling it to the already computed values of its parents. We say that the circuit *computes* the function $f = f(x_1, \ldots, x_n)$ if for each $x_i \in \{0, 1\}$, the corresponding value of the output vertex of the circuit equals $f(x_1, \ldots, x_n)$. For example Figure 12.1 presents a circuit computing $f(x_1, x_2, x_3) = (x_1 \oplus (x_2 \wedge x_3)) \wedge x_1$.

If every fanout in a circuit is at most 1 (i.e., the corresponding graph is a tree), the circuit is called a *formula*. If every fanin in a circuit is at most 2, the circuit is called a *binary circuit* . Therefore the circuit in Figure 12.1 is binary , but it is not a formula. The *size* of a circuit is the number of vertices in it, and its *depth* is the maximum length (number of edges) of a directed path in it. The *binary circuit complexity* of a Boolean function is the size of the smallest binary circuit computing it. An easy counting argument shows that for large n the binary circuit complexity of almost all the functions of n variables is at least $(1 + o(1))2^n/n$. This is because the number of binary circuits of size s on n variables can be shown to be less than $(c(s + n))^s$, whereas the total number of Boolean functions on n variables is 2^{2^n}. On the other hand, there is no known nonlinear, not to mention exponential (in n), lower bound for the binary circuit complexity of any "explicit" function. By "explicit" here we mean an **NP**-function, that is, one of a family $\{f_{n_i}\}_{i \geq 1}$ of Boolean functions, where f_{n_i} has n_i variables, $n_i \to \infty$, and there is a nondeterministic Turing machine which, given n_i and $x_1, \ldots, x_{n_i}$, can decide in (nondeterministic) polynomial time (in n_i) if $f_{n_i}(x_1, \ldots, x_{n_i}) = 1$. (An example for such a family is the $\frac{n}{2}$-clique function; here $n_i = \binom{i}{2}$, the n_i variables $x_1, \ldots, x_{n_i}$ represent the edges of a graph on i vertices, and $f_{n_i}(x_1, \ldots, x_{n_i}) = 1$ iff the corresponding graph contains a clique on at least $i/2$ vertices). Any nonpolynomial lower bound for the binary circuit complexity of an explicit function would imply (among other things) that P $\neq$ NP and thus solve the arguably most important open problem in theoretical

computer science. Unfortunately, at the moment, the best known lower bound for the binary circuit complexity of an explicit function of n variables is only $3n$, (see Blum (1984), Paul (1977)). However, several nontrivial lower bounds are known when we impose certain restrictions on the structure of the circuits. Most of the known proofs of these bounds rely heavily on probabilistic methods. In this chapter we describe some of these results. We note that there are many additional beautiful known results about circuit complexity ; see, for example, Wegener (1987) and Karchmer and Wigderson (1990). Those included here are not only among the crucial ones but also represent the elegant methods used in this field. Since most results in this chapter are asymptotic, we assume, throughout the chapter, whenever needed, that the number of variables we have is sufficiently large.

12.2 RANDOM RESTRICTIONS AND BOUNDED-DEPTH CIRCUITS

Let us call a Boolean function G a t-And-Or if it can be written as an And of an arbitrary number of functions, each being an Or of at most t atoms, that is, $G = G_1 \wedge \cdots \wedge G_w$, where $G_i = y_{i1} \vee \cdots \vee y_{ia_i}$, $a_i \leq t$ and each y_j is an atom. Similarly, we call a Boolean function an s-Or-And, if it can be written as an Or of And gates, each containing at most s atoms. A *minterm* of a function is a minimal assignment of values to some of the variables that forces the function to be 1. Its *size* is the number of variables whose values are set. Notice that if each of the minterms of a function is of size at most s, then it is an s-Or-And (the converse is not true in general, but will not be needed here). A *restriction* is a map ρ of the set of indices $\{1, \ldots, n\}$ to the set $\{0, 1, * \}$. The restriction of the function $G = G(x_1, \ldots, x_n)$ by ρ, denoted by $G|_\rho$, is the Boolean function obtained from G by setting the value of each x_i for $i \in \rho^{-1}\{0, 1\}$ to $\rho(i)$, and leaving each x_j for $j \in \rho^{-1}(*)$ as a variable. Thus, for example, if $G(x_1, x_2, x_3) = (x_1 \wedge x_2) \vee x_3$ and $\rho(1) = 0$, $\rho(2) = \rho(3) = *$, then $G|_\rho = x_3$. For $0 \leq p \leq 1$, a *random p-restriction* is a random restriction ρ defined by choosing, for each $1 \leq i \leq n$ independently, the value of $\rho(i)$ according to the following distribution:

$$\Pr[\rho(i) = *] = p, \ \Pr[\rho(i) = 0] = \Pr[\rho(i) = 1] = (1 - p)/2 \ . \tag{12.1}$$

Improving results of Furst, Saxe and Sipser (1984), Ajtai (1983) and Yao (1985), Håstad (1988) proved the following result, which is very useful in establishing lower bounds for bounded-depth circuits.

Lemma 12.2.1 [The Switching Lemma] *Let $G = G(x_1, \ldots, x_n)$ be a t-And-Or, that is, $G = G_1 \wedge G_2 \wedge \cdots \wedge G_w$, where each G_i is an Or of at most t atoms. Let ρ be the random restriction defined by (12.1). Then*

$$Pr[G|_\rho \ \textit{is not an } (s-1)\textit{-Or-And}]$$
$$\leq \ Pr[G|_\rho \ \textit{has a minterm of size } \geq s] \leq (5pt)^s \ .$$

Proof. Let E_s be the event that $G|_\rho$ has a minterm of size at least s. To bound $\Pr[E_s]$, we prove a stronger result; for any Boolean function F

$$\Pr[E_s|F|_\rho \equiv 1] \leq (5pt)^s \ . \tag{12.2}$$

Here we agree that, if the condition is unsatisfied, then the conditional probability is 0. Lemma 12.2.1 is obtained from (12.2) by taking $F \equiv 1$. We prove (12.2) by induction on w. For $w = 0, G \equiv 1$ and there is nothing to prove. Assuming (12.2) holds whenever the number of G_i is less than w, we prove it for w. Put $G = G_1 \wedge G^*$, where $G^* = G_2 \wedge \cdots \wedge G_w$, and let E_s^* be the event where $G^*|_\rho$ has a minterm of size at least s. By interchanging, if necessary, some of the variables with their complements, we may assume, for convenience, that $G_1 = \bigvee_{i \in T} x_i$, where $|T| \leq t$. Either $G_1|_\rho \equiv 1$ or $G_1|_\rho \not\equiv 1$. In the former case, E_s holds if and only if E_s^* holds and, hence, by induction

$$\Pr[E_s|F|_\rho \equiv 1, G_1|_\rho \equiv 1] = \Pr[E_s^*|(f \wedge G_1)|_\rho \equiv 1] \leq (5\ pt)^s \ . \tag{12.3}$$

The case $G_1|_\rho \not\equiv 1$ requires more work. In this case, any minterm of $G|_\rho$ must assign a value 1 to at least one x_i, for $i \in T$. For a nonempty $Y \subseteq T$ and for a function $\sigma : Y \to \{0, 1\}$ which is not identically 0, let $E_s(Y, \sigma)$ be the event where $G|_\rho$ has a minterm of size at least s which assigns the value $\sigma(i)$ to x_i for each $i \in Y$ and does not assign any additional values to variables x_j with $j \in T$. By the preceding remark

$$\Pr[E_s|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \leq \sum_{Y,\sigma} \Pr[E_s(Y, \sigma)|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \ . \tag{12.4}$$

Observe that the condition $G_1|_\rho \not\equiv 1$ means precisely that $\rho(i) \in \{0, * \}$ for all $i \in T$ and, hence, for each $i \in T$

$$\Pr[\rho(i) =* \ |G_1|_\rho \not\equiv 1] = \frac{p}{p + (1-p)/2} = 2p/(1+p) \ .$$

Thus, if $|Y| = y$,

$$\Pr[\rho(Y) =* \ |G_1|_\rho \not\equiv 1] \leq \left(\frac{2p}{1+p}\right)^y \ .$$

The further condition $F|_\rho \equiv 1$ can only decrease this probability. This can be shown using the FKG inequality (see Chapter 6). It can also be shown directly as follows: For any fixed $\rho' : N - Y \to \{0, 1, * \}$, where $N = \{1, \ldots, n\}$, we claim that

$$\Pr[\rho(Y) =* \ |F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho|_{N-Y} = \rho'] \leq \left(\frac{2p}{1+p}\right)^y \ .$$

Indeed, the given ρ' has a unique extension ρ with $\rho(Y) =*$. If that ρ does not satisfy the above conditions, then the conditional probability is zero. If it does, then so do

all extensions ρ with $\rho(i) \in \{0, * \}$ for $i \in Y$, and so the inequality holds in this case too. As this holds for all fixed ρ', we conclude that indeed

$$\Pr[\rho(Y) =* \, |F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \leq \left(\frac{2p}{1+p}\right)^y \leq (2p)^y \, . \tag{12.5}$$

Let $\rho' : T \to \{0, * \}$ satisfy $\rho(Y) =*$, and consider all possible restrictions ρ satisfying $\rho|_T = \rho'$. Under this condition, ρ may be considered as a random restriction on $N - T$. The event $F|_\rho \equiv 1$ reduces to the event $\mathrm{F}\,|_{\rho|_{N-T}} \equiv 1$, where F is the And of all functions obtained from F by substituting the values of x_i according to ρ' for those $i \in T$ with $\rho'(i) = 0$, and by taking all possibilities for all the other variables x_j for $j \in T$. If the event $E_s(Y, \sigma)$ occurs, then $G^*|_{\rho\sigma}$ has a minterm of size at least $s - y$ that does not contain any variable x_i with $i \in T - Y$. But this happens if and only if $\mathrm{G}\,|_{\rho|_{N-T}}$ has a minterm of size at least $s - y$, where G is the function obtained from G^* by substituting the values of x_j for $j \in Y$ according to σ, the values of x_i for $i \in T - Y$ and $\rho'(i) = 0$ according to ρ' and by removing all the variables x_k with $k \in T - Y$ and $\rho'(k) =*$. Denoting this event by $\tilde{E}_{s-y}$, we can apply induction and obtain

$$\Pr[E_s(Y, \sigma)|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho|_T = \rho'] \leq \Pr[\tilde{E}_{s-y}|\mathrm{F}\,|_\rho \equiv 1] \leq (5pt)^{s-y} \, .$$

Since any ρ with $F|_\rho \equiv 1$, $G_1|_\rho \equiv 1$, $\rho(Y) =*$ must have $\rho|_T = \rho'$ for some ρ' of this form, and since the event $E_s(Y, \sigma)$ may occur only if $\rho(Y) =*$, we conclude that

$$\Pr[E_s(Y, \sigma)|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho(Y) =*] \leq (5pt)^{s-y} \, ,$$

and, by (12.5)

$$\begin{aligned} &\Pr[E_s(Y, \sigma)|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \\ &= \Pr[\rho(Y) =* \, |F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \\ &\quad \cdot \Pr[E_s(Y, \sigma)|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho(Y) =*] \\ &\leq (2p)^y (5pt)^{s-y} \, . \end{aligned}$$

Substituting in (12.4) and using the fact that $|T| \leq t$ and that

$$\sum_{y=1}^{t} (2^y - 1)2^y/(5^y y!) \leq \frac{2}{5} + \sum_{y=2}^{\infty} \frac{(4/5)^y}{y!} = \frac{2}{5} + e^{4/5} - 1 - \frac{4}{5} < 1$$

we obtain

$$\begin{aligned} &\Pr[E_s|F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \\ &\leq \sum_{y=1}^{|T|} \binom{|T|}{y} (2^y - 1)(2p)^y (5pt)^{s-y} \leq (5pt)^s \sum_{y=1}^{t} \frac{t^y}{y!} (2^y - 1)\left(\frac{2}{5t}\right)^y \\ &= (5pt)^s \sum_{y=1}^{t} (2^y - 1) \cdot \frac{2^y}{5^y \cdot y!} \leq (5pt)^s \, . \end{aligned}$$

This, together with (12.3), gives

$$\Pr[E_s|F|_\rho \equiv 1] \leq (5pt)^s$$

completing the induction and the proof. ■

By taking the complement of the function G in Lemma 12.2.1 and applying De Morgan's rules, one clearly obtains its dual form: If G is a t-Or-And and ρ is the random restriction given by (12.1) then

$$\Pr[G|_\rho \text{ is not an } (s-1)\text{-And-Or}] \leq (5pt)^s .$$

We now describe an application of the Switching Lemma, which supplies a lower bound to the size of circuits of small depth that compute the parity function $x_1 \oplus \cdots \oplus x_n$. We consider circuits in which the vertices are arranged in levels, namely those in the first level are atoms (i.e., variables or their complements), and each other gate is either an *Or* or an *And* of an arbitrary number of vertices from the previous level. We assume that the gates in each level are either all *And* gates or all *Or* gates, and that the levels alternate between *And* levels and *Or* levels. A circuit of this form is called a $C(s, s', d, t)$-circuit if it contains at most s gates, at most s' of which are above the second level, its depth is at most d, and the fanin of each gate in its second level is at most t. Thus, for example, the circuit that computes the complement of the parity function by computing an *Or* of the 2^{n-1} terms $x_1^{\epsilon_1} \wedge \cdots \wedge x_n^{\epsilon_n}$, where $(\epsilon_1, \ldots, \epsilon_n)$ ranges over all even binary vectors and $x_i^{\epsilon_i} = x_i \oplus \epsilon_i \oplus 1$, is a $C(2^{n-1}+1, 1, 2, n)$-circuit.

Theorem 12.2.2 *Let $f = f(x_1, \ldots, x_n)$ be a function, and let C be a $C(\infty, s, d, t)$-circuit computing f, where $s \cdot \left(\frac{1}{2}\right)^t \leq 0.5$. Then either f or its complement $\overline{f}$ has a minterm of size at most $n - \frac{n}{2\cdot(10t)^{d-2}} + t$.*

Proof. Let us apply to C, repeatedly, $d-2$ times a random $1/(10t)$-restriction. Each of these random restrictions, when applied to any bottom subcircuit of depth 2, transforms it by Lemma 12.2.1 with probability at least $1 - \left(\frac{1}{2}\right)^t$ from a t-Or-And to a t-And-Or (or conversely). If all these transformations succeed, we can merge the new *And* gates with these from the level above them and obtain a circuit with a smaller depth. As the total size of the circuit is at most s, and $s\left(\frac{1}{2}\right)^t \leq 0.5$, we conclude that with probability at least $1/2$, all transformations succeed and C is transformed into a $C(\infty, 1, 2, t)$-circuit. Each variable x_i, independently, is still a variable (i.e., has not been assigned a value) with probability $\frac{1}{(10t)^{d-2}}$. Thus, the number of remaining variables is a binomial random variable with expectation $\frac{n}{(10t)^{d-2}}$ and a little smaller variance. By the standard estimates for binomial distributions (see Appendix A) the probability that at least $\frac{n}{2\cdot(10t)^{d-2}}$ variables are still variables is more than $1/2$. Therefore, with positive probability, at most $n - \frac{n}{2\cdot(10t)^{d-2}}$ of the variables have been fixed

and the resulting restriction of f has a $C(\infty, 1, 2, t)$-circuit, that is, its value can be fixed by assigning values to at most t additional variables. This completes the proof. ■

Corollary 12.2.3 *For any $d \geq 2$, there is no*

$$C\left(\infty, \frac{1}{2} \cdot 2^{\frac{1}{10}n^{1/(d-1)}}, d, \frac{1}{10}n^{1/(d-1)}\right) \text{-circuit}$$

that computes the parity function $f(x_1, \ldots, x_n) = x_1 \oplus \cdots \oplus x_n$.

Proof. Assuming there is such a circuit, we obtain, by Theorem 12.2.2, that the value of f can be fixed by assigning values to at most $n - \frac{1}{2}n^{1/(d-1)} + \frac{1}{10}n^{1/(d-1)} < n$ variables. This is false, and hence there is no such circuit. ■

The estimate in Corollary 12.2.3 is, in fact, nearly best possible. Since every $C(s, s', d, t)$-circuit can be transformed into a $C((t+1)s, s, d+1, 2)$-circuit (by replacing each atom by an *Or* or *And* of two copies of itself), Corollary 12.2.3 easily implies that the depth d of any $C(s, s', d, t)$-circuit of polynomial size that computes the parity of n bits is at least $\Omega(\log n / \log\log n)$. This lower bound is also optimal.

12.3 MORE ON BOUNDED-DEPTH CIRCUITS

In the previous section we saw that the parity function is hard to compute in small depth using *And*, *Or*, and *Not* gates. It turns out that, even if we allow the use of parity gates (in addition to the *And*, *Or*, and *Not* gates), there are still some relatively simple functions that are hard to compute. Such a result was first proved by Razborov (1987). His method was modified and strengthened by Smolensky (1987). For an integer $k \geq 2$, let $\text{Mod}_k(x_1, x_2, \ldots, x_n)$ be the Boolean function whose value is 1 iff $\Sigma x_i \not\equiv 0 \ (mod\ k)$. Smolensky showed that, for every two powers p and q of distinct primes, the function Mod_p cannot be computed in a bounded-depth polynomial-size circuit that uses *And*, *Or*, *Not*, and Mod_q gates. Here we present the special case of this result in which $q = 3$ and $p = 2$.

Let C be an arbitrary circuit of depth d and size s consisting of *And*, *Or*, *Not*, and Mod_3 gates. A crucial fact, due to Razborov, is the assertion that the output of C can be approximated quite well (depending on d and s) by a polynomial of relatively small degree over $GF(3)$. This is proved by applying the probabilistic method as follows: Let us replace each gate of the circuit C by an approximate polynomial operation, according to the following rules that guarantee that in each vertex in the new circuit we compute a polynomial over $GF(3)$, whose values are all 0 or 1 (whenever the input is a 0-1 input).

(i) Each *Not* gate $\overline{y}$ is replaced by the polynomial gate $(1 - y)$.

(ii) Each Mod_3 gate $\text{Mod}_3(y_1, \ldots, y_m)$ is replaced by the polynomial gate $(y_1 + y_2 + \cdots + y_m)^2$.

The rule for replacement of *Or* and *And* gates is a little more complicated. Observe that in the two previous cases (i) and (ii) there was no approximation; the new gates compute precisely what the old ones did, for all possible Boolean values of the variables. This can, in principle, be done here too. An *And* gate $y_1 \wedge \cdots \wedge y_m$ should simply be replaced by the product $y_1 \cdots y_m$. An *Or* gate $y_1 \vee \cdots \vee y_m$ can then be computed by de Morgan's rules. Since $y_1 \vee \cdots \vee y_m = (\overline{y}_1 \wedge \cdots \wedge \overline{y}_m)$ and $\overline{y}$ is realized by $(1 - y)$, this would give

$$1 - (1 - y_1)(1 - y_2) \cdots (1 - y_m) \quad . \tag{12.6}$$

The trouble is that this procedure would increase the degree of our polynomials too much. Hence, we need to be a little more tricky. Let ℓ be an integer, to be chosen later. Given an *Or* gate $y_1 \vee \cdots \vee y_m$, we choose ℓ random subsets $I_1, \ldots, I_\ell$ of $\{1, \ldots, m\}$, where for each $1 \leq i \leq \ell$ and for each $1 \leq j \leq m$ independently $\Pr[j \in I_i] = 1/2$. Observe that for each fixed i, $1 \leq i \leq \ell$, the sum $(\sum_{j \in I_i} y_j)^2$ over $GF(3)$ is certainly 0 if $y_1 \vee \cdots \vee y_m = 0$, and is 1 with probability at least $1/2$ if $y_1 \vee \cdots \vee y_m = 1$. Hence, if we compute the *Or* function of the ℓ expressions $(\sum_{j \in I_i} y_j)^2$, $(1 \leq i \leq \ell)$, this function is 0 if $y_1 \vee \cdots \vee y_m = 0$ and is 1 with probability at least $1 - (1/2)^\ell$ if $y_1 \vee \cdots \vee y_m = 1$. We thus compute the *Or* and write it as a polynomial, in the way explained in (12.6). This gives

$$1 - \prod_{i=1}^{\ell} \left(1 - \left(\sum_{j \in I_i} y_j \right)^2 \right) . \tag{12.7}$$

Therefore, in our new circuit we replace each *Or* gate by an approximation polynomial gate of the form described in (12.7). Once we have an approximation to an *Or* gate, we can obtain the corresponding one for an *And* gate by applying de Morgan rules. Since $y_1 \wedge \cdots \wedge y_m = (\overline{y}_1 \vee \cdots \vee \overline{y}_m)$, we replace each *And* gate of the form $y_1 \wedge \cdots \wedge y_m$ by

$$\prod_{i=1}^{\ell} \left(1 - \left[\sum_{j \in I_i} (1 - y_j) \right]^2 \right) . \tag{12.8}$$

Observe that the polynomials in (12.7) and (12.8) are both of degree at most 2ℓ.

Given the original circuit C of depth d and size s, we can now replace all its gates by our approximating polynomial gates and get a new circuit CP, which depends on all the random choices made in each replacement of each of the *And*/*Or* gates. The new circuit CP computes a polynomial $P(x_1, \ldots, x_n)$ of degree at most $(2\ell)^d$. Moreover, for each fixed Boolean value of $x_1, x_2, \ldots, x_n$, the probability that all the new gates compute exactly what the corresponding gates in C computed is at least $1 - s/2^\ell$. Therefore, the expected number of inputs on which $P(x_1, \ldots, x_n)$ is equal to the output of C is at least $2^n(1 - s/2^\ell)$. We have thus proved the following:

Lemma 12.3.1 *For any circuit C of depth d and size s on n Boolean variables that uses Not, Or, And, and* Mod_3 *gates, and for any integer* ℓ*, there is a polynomial*

$P = P(x_1, \ldots, x_n)$ of degree at most $(2\ell)^d$ over $GF(3)$ whose value is equal to the output of C on at least $2^n(1 - s/2^\ell)$ inputs.

In order to apply this lemma for obtaining lower bounds for the size of any circuit of the above type that computes the parity function, we need the following additional combinatorial result.

Lemma 12.3.2 *For $n \geq 20$, there is no polynomial $P(x_1, \ldots, x_n)$ over $GF(3)$ of degree at most $\sqrt{n}$ which is equal to the parity of $x_1, \ldots, x_n$ for a set S of at least $0.9 \cdot 2^n$ distinct binary vectors $(x_1, \ldots, x_n)$.*

Proof. Suppose this is false, and suppose $S \subset \{0,1\}^n$, $|S| \geq 0.9 \cdot 2^n$ and $P(x_1, \ldots, x_n) = x_1 \oplus \cdots \oplus x_n$ for all $(x_1, \ldots, x_n) \in S$. Define a polynomial $Q = Q(y_1, \ldots, y_n)$ by $Q = Q(y_1, \ldots, y_n) = P(y_1 + 2, \ldots, y_n + 2) - 2$, and $T = \{(y_1, \ldots, y_n) \in \{1, -1\}^n : (y_1 + 2, \ldots, y_n + 2) \in S\}$, where all additions are in $GF(3)$. Clearly, Q has degree at most $\sqrt{n}$, and $Q(y_1, \ldots, y_n) = \prod_{i=1}^{n} y_i$ for all $(y_1, \ldots, y_n) \in T$. Let now $G = G(y_1, \ldots, y_n) : T \to GF(3)$ be an arbitrary function. Extend it in an arbitrary way to a function from $(GF(3))^n \to GF(3)$, and write this function as a polynomial in n variables. [Trivially, any function from $(GF(3))^n \to GF(3)$ is a polynomial. This follows from the fact that it is a linear combination of functions of the form $\prod_{i=1}^{n}(y_i - \epsilon_i)(y_i - \epsilon_i - 1)$, where $\epsilon_i \in GF(3)$]. Replace each occurrence of y_i^2 in this polynomial by 1 to obtain a multilinear polynomial $\tilde{G}$ which agrees with G on T. Now replace each monomial $\prod_{i \in U} y_i$, where $|U| > \frac{n}{2} + \frac{\sqrt{n}}{2}$ by $\prod_{i \notin U} y_i \cdot Q(y_1, \ldots, y_n)$, and replace this new polynomial by a multilinear one, $\tilde{G}'$, again by replacing each y_i^2 by 1. Since for $y_i \in \{\pm 1\}$, $\prod_{i \notin U} y_i \cdot \prod_{i=1}^{n} y_i = \prod_{i \in U} y_i$, $\tilde{G}'$ is equal to G on T and its degree is at most $\frac{n}{2} + \frac{\sqrt{n}}{2}$. However, the number of possible $\tilde{G}'$ is $3^{\sum_{i=0}^{\frac{n}{2}+\frac{\sqrt{n}}{2}} \binom{n}{i}} < 3^{0.88 \cdot 2^n}$, whereas the number of possible G is $3^{|T|} \geq 3^{0.9 \cdot 2^n}$. This is impossible, and hence the assertion of the lemma holds. ∎

Corollary 12.3.3 *There is no circuit of depth d and size $s \leq \frac{1}{10} 2^{\frac{1}{2} n^{1/2d}}$ computing the parity of $x_1, x_2, \ldots, x_n$ using Not, And, Or, and Mod_3 gates.*

Proof. Suppose this is false, and let C be such a circuit. Put $\ell = \frac{1}{2} \cdot n^{1/2d}$. By Lemma 12.3.1, there is a polynomial $P = P(x_1, \ldots, x_n)$ over $GF(3)$, whose degree is at most $(2\ell)^d = \sqrt{n}$, which is equal to the parity of $x_1, \ldots, x_n$ on at least $2^n(1 - \frac{s}{2^{\frac{1}{2} n^{1/2d}}}) \geq 0.9 \cdot 2^n$ inputs. This contradicts Lemma 12.3.2, and hence completes the proof. ∎

12.4 MONOTONE CIRCUITS

A Boolean function $f = f(x_1, \dots, x_n)$ is *monotone* if $f(x_1, \dots, x_n) = 1$ and $x_i \le y_i$ imply $f(y_1, \dots, y_n) = 1$. A *binary monotone circuit* is a binary circuit that contains only binary *And* and *Or* gates. It is easy to see that a function is monotone if and only if there is a binary monotone circuit that computes it. The *monotone complexity* of a monotone function is the smallest size of a binary monotone circuit that computes it. Until 1985, the largest known lower bound for the monotone complexity of a monotone **NP**-function of n variables was $4n$. This was considerably improved in the fundamental paper of Razborov (1985), where a bound of $n^{\Omega(\log n)}$ to the Clique$_k$-function (which is 1 iff a given graph contains a clique of size k) was established. Shortly afterwards, Andreev (1985) used similar methods to obtain an exponential lower bound to a somewhat unnatural **NP**-function. Alon and Boppana (1987) strengthened the combinatorial arguments of Razborov and proved an exponential lower bound for the monotone circuit complexity of the clique function. In this section we describe a special case of this bound by showing that there are no linear size monotone circuits that decide whether a given graph contains a triangle. Although this result is much weaker than the ones stated above, it illustrates nicely all the probabilistic considerations in the more complicated proofs and avoids some of the combinatorial subtleties, whose detailed proofs can be found in the above-mentioned papers.

Put $n = \binom{m}{2}$, and let $x_1, x_2, \dots, x_n$ be n Boolean variables representing the edges of a graph on the set of vertices $\{1, 2, \dots, m\}$. Let $T = T(x_1, \dots, x_n)$ be the monotone Boolean function whose value is 1 if the corresponding graph contains a triangle. Clearly, there is a binary monotone circuit of size $O(m^3)$ computing T. Thus, the following theorem is tight, up to a polylogarithmic factor:

Theorem 12.4.1 *The monotone circuit complexity of T is at least $\Omega(m^3/\log^4 m)$.*

Before we present the proof of this theorem, we introduce some notations and prove a simple lemma. For any Boolean function $f = f(x_1, \dots, x_n)$, define $A(f) = \{(x_1, \dots, x_n) \in \{0,1\}^n : f(x_1, \dots, x_n) = 1\}$. Clearly, $A(f \vee g) = A(f) \cup A(g)$ and $A(f \wedge g) = A(f) \cap A(g)$. Let C be a monotone circuit of size s computing the function $f = f(x_1, \dots, x_n)$. Clearly, C supplies a monotone straight-line program of length s computing f, that is, a sequence of functions $x_1, x_2, \dots, x_n, f_1, \dots, f_s$, where $f_s = f$ and each f_i, for $1 \le i \le s$, is either an *Or* or an *And* of two of the previous functions. By applying the operation A, we obtain a sequence $A(C)$ of subsets of $\{0,1\}^n : A_{-n} = A_{x_n}, \dots, A_{-1} = A_{x_1}, A_1, \dots, A_s$, where $A_{x_i} = A(x_i), A_s = A(f)$, and each A_i, for $1 \le i \le s$, is either a union or an intersection of two of the previous subsets. Let us replace the sequence $A(C)$ by an *approximating sequence* $M(C) : M_{-n} = M_{x_n} = A_{x_n}, \dots, M_{-1} = M_{x_1} = A_{x_1}, M_1, \dots, M_s$ defined by replacing the union and intersection operations in $A(C)$ by the approximating operations $\sqcup$ and $\sqcap$, respectively. The exact definition of these two operations will be given later, in such a way that, for all admissible M and L, the inclusions

$$M \sqcup L \supseteq M \cup L \quad \text{and} \quad M \sqcap L \subseteq M \cap L \tag{12.9}$$

will hold. Thus $M_{x_i} = A_{x_i}$ for all $1 \le i \le n$, and if for some $1 \le j \le s$ we have $A_j = A_\ell \cup A_k$, then $M_j = M_\ell \sqcup M_k$, whereas if $A_j = A_\ell \cap A_k$, then $M_j = M_\ell \sqcap M_k$. In the former case, put $\delta^j_\sqcup = M_j - (M_\ell \cup M_k)$ and $\delta^j_\sqcap = \phi$, and in the latter case put $\delta^j_\sqcap = (M_\ell \cap M_k) - M_j$ and $\delta^j_\sqcup = \phi$.

Lemma 12.4.2 *For all members M_i of $M(C)$*

$$A_i - \left(\bigcup_{j \le i} \delta^j_\sqcap \right) \subseteq M_i \subseteq A_i \cup \bigcup_{j \le i} \delta^j_\sqcup \quad . \tag{12.10}$$

Proof. We apply induction on i. For $i < 0$ $M_i = A_i$, and thus (12.10) holds. Assuming (12.10) holds for all M_j with $j < i$, we prove it for i. If $A_i = A_\ell \cup A_k$, then, by the induction hypothesis

$$M_i = M_\ell \cup M_k \cup \delta^i_\sqcup \subseteq A_\ell \cup A_k \cup \bigcup_{j \le i} \delta^j_\sqcup = A_i \cup \bigcup_{j \le i} \delta^j_\sqcup$$

and

$$M_i = M_\ell \sqcup M_k \supseteq M_\ell \cup M_k \supseteq \left(A_\ell - \left(\bigcup_{j \le \ell} \delta^j_\sqcap \right) \right) \cup \left(A_k - \left(\bigcup_{j \le k} \delta^j_\sqcap \right) \right)$$

$$\supseteq A_i - \left(\bigcup_{j \le i} \delta^j_\sqcap \right)$$

as needed. If $A_i = A_\ell \cap A_k$, the proof is similar. ■

Lemma 12.4.2 holds for any choice of the operations $\sqcup$ and $\sqcap$ that satisfies (12.9). In order to prove Theorem 12.4.1, we define these operations as follows: Put $r = 100 \log^2 m$. For any set R of at most r edges on $V = \{1, 2, \ldots, m\}$, let $\lceil R \rceil$ denote the set of all graphs on V containing at least one edge of R. In particular, $\lceil \phi \rceil$ is the empty set. We also let $\lceil * \rceil$ denote the set of all graphs. The elements of $M(C)$ will all have the form $\lceil R \rceil$ or $\lceil * \rceil$. Note that $A_{x_i} = M_{x_i}$ is simply the set $\lceil R \rceil$, where R is a singleton containing the appropriate single edge. For two sets R_1 and R_2 of at most r edges each, we define $\lceil R_1 \rceil \sqcap \lceil R_2 \rceil = \lceil R_1 \cap R_2 \rceil$, $\lceil R_1 \rceil \sqcap \lceil * \rceil = \lceil R_1 \rceil$, and $\lceil * \rceil \sqcap \lceil * \rceil = \lceil * \rceil$. Similarly, if $|R_1 \cup R_2| \le r$, we define $\lceil R_1 \rceil \sqcup \lceil R_2 \rceil = \lceil R_1 \cup R_2 \rceil$, whereas if $|R_1 \cup R_2| > r$, then $\lceil R_1 \rceil \sqcup \lceil R_2 \rceil = \lceil * \rceil$. Finally $\lceil * \rceil \sqcup \lceil R_1 \rceil = \lceil * \rceil \sqcup \lceil * \rceil = \lceil * \rceil$.

Proof [Theorem 12.4.1] We now prove Theorem 12.4.1 by showing that there is no monotone circuit of size $s < \binom{m}{3}/2r^2$ computing the function T. Indeed, suppose this is false, and let C be such a circuit. Let $M(C) = M_{x_n}, \ldots, M_{x_1}, M_1, \ldots, M_s$ be an approximating sequence of length s obtained from C as described above. By Lemma 12.4.2,

$$A(T) - \left(\bigcup_{j \le s} \delta^j_\sqcap \right) \subseteq M_s \subseteq A(T) \cup \bigcup_{j \le s} \delta^j_\sqcup \quad . \tag{12.11}$$

We consider two possible cases.

Case 1 $M_s = \lceil R \rceil$, where $|R| \leq r$.

Let us choose a random triangle Δ on $\{1, 2, \ldots, m\}$. Clearly

$$\Pr[\Delta \in M_s] \leq \frac{r \cdot (m-2)}{\binom{m}{3}} < \frac{1}{2} .$$

Moreover, for each fixed j, $j \leq s$

$$\Pr[\Delta \in \delta^j_{\sqcap}] \leq \frac{r^2}{\binom{m}{3}} .$$

This is because, if $\delta^j_{\sqcap} \neq \phi$, then $\delta^j_{\sqcap} = (\lceil R_1 \rceil \cap \lceil R_2 \rceil) - \lceil R_1 \cap R_2 \rceil$ for some two sets of edges R_1, R_2, each of cardinality at most r. The only triangles in this difference are those containing an edge from R_1 and another edge from R_2 (and no edge of both). Since there are at most r^2 such triangles, the last inequality follows. Since $s < \binom{m}{3}/2r^2$, the last two inequalities imply that $\Pr[\Delta \notin M_s \text{ and } \Delta \notin \bigcup_{j \leq s} \delta^j_{\sqcap}] > 0$ and thus there is such a triangle Δ. Since this triangle belongs to $A(T)$, this contradicts (12.11), showing that Case 1 is impossible.

Case 2 $M_s = \lceil * \rceil$.

Let B be a random spanning complete bipartite graph on $V = \{1, 2, \ldots, m\}$ obtained by coloring each vertex in V randomly and independently by 0 or 1 and taking all edges connecting vertices with distinct colors. Since M_s is the set of all graphs, $B \in M_s$. Also $B \notin A(T)$, as it contains no triangle. We claim that for every fixed $j, j \leq s$,

$$\Pr[B \in \delta^j_{\sqcup}] \leq 2^{-\sqrt{r}/2} < \frac{1}{m^5} . \tag{12.12}$$

Indeed, if $\delta^j_{\sqcup} \neq \phi$, then $\delta^j_{\sqcup} = \lceil * \rceil - (\lceil R_1 \rceil \cup \lceil R_2 \rceil)$, where $|R_1 \cup R_2| > r$. Consider the graph whose set of edges is $R_1 \cup R_2$. Let d be its maximum degree. By Vizing's Theorem, the set of its edges can be partitioned into at most $d + 1$ matchings. Thus either $d > \frac{\sqrt{r}}{2}$ or the size of the maximum matching in this graph is at least $\sqrt{r}/2$. It follows that our graph contains a set of $k = \sqrt{r}/2$ edges $e_1, \ldots, e_k$, which form either a star or a matching. In each of these two cases, $\Pr[e_i \in B] = \frac{1}{2}$ and these events are mutually independent. Hence

$$\Pr[B \notin \lceil R_1 \rceil \cup \lceil R_2 \rceil] \leq 2^{-\sqrt{r}/2}$$

implying (12.12). Note that a similar estimate can be established without Vizing's Theorem by observing that B does not belong to $(\lceil R_1 \rceil \cup \lceil R_2 \rceil)$ if and only if the vertices in any connected component of the graph whose edges are $R_1 \cup R_2$ belong to the same color class of B.

Since $s < \binom{m}{3}/2r^2 < m^5$, Inequality (12.12) implies that there is a bipartite B such that $B \in M_s, B \notin A(T)$, and $B \notin \bigcup_{j \leq s} \delta^j_{\sqcup}$. This contradicts (12.11), and shows that Case 2 is impossible and hence completes the proof of Theorem 12.4.1. ■

12.5 FORMULAE

Recall that a formula is a circuit in which every fanout is at most 1. Unlike in the case of circuits, there are known superlinear lower bounds for the minimum size of formulae computing various explicit **NP**-functions over the full binary basis. For a Boolean function $f = f(x_1, \ldots, x_n)$, let us denote by $L(f)$ the minimum number of *And* and *Or* gates in a formula that uses *And*, *Or*, and *Not* gates and computes f. By the de Morgan rules, we may assume that all *Not* gates appear in the first level of this formula. We conclude this chapter with a simple result of Subbotovskaya (1961), which implies that for the parity function $f = x_1 \oplus \cdots \oplus x_n$, $L(f) \geq \Omega(n^{3/2})$. This bound has been improved later by Khrapchenko (1971) to $L(f) = n^2 - 1$. However, we present here only the weaker $\Omega(n^{3/2})$ lower bound, not only because it demonstrates, once more, the power of relatively simple probabilistic arguments but also because a modification of this proof enabled Andreev (1987) to obtain an $\Omega(n^{5/2}/(\log n)^{O(1)})$ lower bound for $L(g)$ for another **NP**-function $g = g(x_1, \ldots, x_n)$. Håstad (1998), later improved this lower bound to $\Omega(n^{3-o(1)})$. This is at present the largest known lower bound for the formula complexity of an **NP**-function of n variables over a complete basis.

The method of Subbotovskaya (1961) is based on random restrictions similar to the ones used in Section 12.2. The main lemma is the following:

Lemma 12.5.1 *Let $f = f(x_1, \ldots, x_n)$ be a nonatom Boolean function of n variables. Then, there is an i, $1 \leq i \leq n$, and an $\epsilon \in \{0, 1\}$ such that for the function $g = f(x_1, \ldots, x_{i-1}, \epsilon, x_{i+1}, \ldots, x_n)$ of $n - 1$ variables obtained from f by substituting $x_i = \epsilon$ the following inequality holds:*

$$(L(g) + 1) \leq \left(1 - \frac{3}{2n}\right)(L(f) + 1) \leq \left(1 - \frac{1}{n}\right)^{3/2}(L(f) + 1)\,.$$

Proof. Fix a formula F computing f with $l = L(f)$ *And* and *Or* gates. F can be represented by a binary tree each of whose $l + 1$ leaves is labeled by an atom x_i or $\overline{x}_i$. Let us choose, randomly, a variable x_i, $1 \leq i \leq n$ according to a uniform distribution, and assign to it a random binary value $\epsilon \in \{0, 1\}$. When we substitute the values ϵ and $1 - \epsilon$ to x_i and $\overline{x}_i$, respectively, the number of leaves in F is reduced; the expected number of leaves omitted in this manner is $(l + 1)/n$. However, further reduction may occur. Indeed, suppose a leaf is labeled x_i and it feeds, say, an *And* gate $x_i \wedge H$ in F. Observe that we may assume that the variable x_i does not appear in the subformula H, as otherwise F can be simplified by substituting $x_i = 1$ in H. If $x_i = \epsilon = 0$, then H can be deleted once we substitute the value for x_i, thus further decreasing the number of leaves. Since the behavior of this effect is similar for an *Or* gate (and also for $\overline{x}_i$ instead of x_i), it follows that the expected number of additional leaves omitted is at least $(l + 1)/2n$. Hence the expected number of remaining leaves in the simplified formula is at most $(l + 1)[1 - \frac{3}{2n}]$, as claimed. ∎

By repeatedly applying Lemma 12.5.1, we obtain the following:

Corollary 12.5.2 *If $f = f(x_1, \ldots, x_n)$ and $L(f) \leq (\frac{n}{k})^{3/2} - 1$, then one can assign values to $n - k$ variables so that the resulting function g is an atom.*

Proof. Repeated application of Lemma 12.5.1 $n-k$ times yields a g with

$$(L(g)+1) \leq \prod_{i=k+1}^{n} \left(1-\frac{1}{i}\right)^{3/2} (L(f)+1) = (k/n)^{3/2}(L(f)+1) \leq 1 \ .$$

Hence g is either x_i or $\overline{x}_i$ for some i. ■

Corollary 12.5.3 *For the parity function* $f = x_1 \oplus \cdots \oplus x_n$,

$$L(f) > \left(\frac{n}{2}\right)^{3/2} - 1 \ .$$

12.6 EXERCISES

1. Show that there exists a constant c such that the number of binary Boolean circuits of size s on n variables is at most $(c(s+n))^s$.
2. Let f be a Boolean formula in the n variables $x_1, x_2, \ldots, x_n$, where f is an *And* of an arbitrary (finite) number of clauses, each clause is an *Or* of 10 literals, where each literal is either a variable or its negation, and suppose each variable appears (negated or unnegated) in at most 10 clauses. Prove that f is satisfiable.
3. (*) Prove that there is a bounded-depth, polynomial-size, monotone circuit of n Boolean inputs $x_1, x_2, \ldots, x_n$, computing a function f whose value is 1 if $\sum_{i=1}^{n} x_i \geq n/2 + n/\log n$, and is 0 if $\sum_{i=1}^{n} x_i \leq n/2 - n/\log n$.

THE PROBABILISTIC LENS:
Maximal Antichains

A family $\mathcal{F}$ of subsets of $\{1, \ldots, n\}$ is called an *antichain* if no set of $\mathcal{F}$ is contained in another.

Theorem 1 *Let $\mathcal{F}$ be an antichain. Then*

$$\sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \leq 1 \, .$$

Proof. Let σ be a uniformly chosen permutation of $\{1, \ldots, n\}$, and set

$$\mathcal{C}_\sigma = \{\{\sigma(j) \, : \, 1 \leq j \leq i\} \, : \, 0 \leq i \leq n\}.$$

(The cases $i = 0, n$ give $\emptyset, \{1, \ldots, n\} \in \mathcal{C}$, respectively.) Define a random variable

$$X = |\mathcal{F} \cap \mathcal{C}_\sigma| \, .$$

We decompose

$$X = \sum_{A \in \mathcal{F}} X_A \, ,$$

where X_A is the indicator random variable for $A \in \mathcal{C}$. Then

$$\mathrm{E}[X_A] = \Pr[A \in \mathcal{C}_\sigma] = \frac{1}{\binom{n}{|A|}} \, ,$$

since C_σ contains precisely one set of size $|A|$, which is distributed uniformly among the $|A|$-sets. By Linearity of Expectation

$$\mathrm{E}[X] = \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}}.$$

For *any* σ, C_σ forms a chain –every pair of sets is comparable. Since $\mathcal{F}$ is an antichain, we *must* have $X = |\mathcal{F} \cap C_\sigma| \leq 1$. Thus $\mathrm{E}[X] \leq 1$. ■

Corollary 2 (Sperner's Theorem) *Let $\mathcal{F}$ be an antichain. Then*

$$|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

Proof. The function $\binom{n}{x}$ is maximized at $x = \lfloor n/2 \rfloor$ so that

$$1 \geq \sum_{A \in \mathcal{F}} \frac{1}{\binom{n}{|A|}} \geq \frac{|\mathcal{F}|}{\binom{n}{\lfloor n/2 \rfloor}}.$$ ■

13

Discrepancy

The mystery, as well as the glory, of mathematics lies not so much in the fact that abstract theories do turn out to be useful in solving problems but in that wonder of wonders, the fact that a theory meant for solving one type of problem is often the only way of solving problems of entirely different kinds, problems for which the theory was not intended. These coincidences occur so frequently that they must belong to the essence of mathematics.
–Gian-Carlo Rota

13.1 BASICS

Suppose we are given a finite family of finite sets. Our object is to color the underlying points red and blue so that all of the sets have nearly the same number of red and blue points. It may be that our cause is hopeless – if the family consists of all subsets of a given set Ω, then regardless of the coloring, some set, either the red or the blue points, will have size at least half that of Ω and be monochromatic. In the other extreme, should the sets of the family be disjoint, then it is trivial to color so that all sets have the same number of red and blue points or, at worst if the cardinality is odd, the number of red and blue points differs by only 1. The discrepancy will measure how good a coloring we may find.

To be formal, let a family $\mathcal{A}$ of subsets of Ω be given. Rather than using red and blue, we consider colorings as maps

$$\chi : \Omega \to \{-1, +1\} .$$

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

For any $A \subset \Omega$, we set

$$\chi(A) = \sum_{a \in A} \chi(a) \ .$$

Define the discrepancy of $\mathcal{A}$ with respect to χ by

$$\mathrm{disc}(\mathcal{A}, \chi) = \max_{A \in \mathcal{A}} |\chi(A)|$$

and the discrepancy of $\mathcal{A}$ by

$$\mathrm{disc}(\mathcal{A}) = \min_{\chi : \Omega \to \{-1,+1\}} \mathrm{disc}(\mathcal{A}, \chi) \ .$$

Other equivalent definitions of discrepancy reveal its geometric aspects. Let $\mathcal{A} = \{S_1, \ldots, S_m\}$, $\Omega = \{1, \ldots, n\}$, and let $B = [b_{ij}]$ be the $m \times n$ incidence matrix: $b_{ij} = 1$ if $j \in S_i$, otherwise $b_{ij} = 0$. A coloring χ may be associated with the vector $u = (\chi(1), \ldots, \chi(n)) \in \{-1, +1\}^n$ so that $Bu^T = (\chi(S_1), \ldots, \chi(S_m))$ and

$$\mathrm{disc}(\mathcal{A}) = \min_{u \in \{-1,+1\}^n} |Bu^T|_\infty,$$

where $|v|_\infty$ is the L^∞-norm, the maximal absolute value of the coordinates. Similarly, letting v_j denote the jth column vector of B (the profile of point j),

$$\mathrm{disc}(\mathcal{A}) = \min |\pm v_1 \pm \cdots \pm v_n|_\infty,$$

where the minimum ranges over all 2^n choices of sign.

We will generally be concerned with upper bounds to the discrepancy. Unravelling the definitions, $\mathrm{disc}(\mathcal{A}) \le K$ if and only if there *exists* a coloring χ for which $|\chi(A)| \le K$ for all $A \in \mathcal{A}$. Naturally, we try the random coloring.

Theorem 13.1.1 *Let $\mathcal{A}$ be a family of n subsets of an m-set Ω. Then*

$$\mathrm{disc}(\mathcal{A}) \le \sqrt{2m \ln\ (2n)} \ .$$

Proof. Let $\chi : \Omega \to \{-1, +1\}$ be random. For $A \subset \Omega$, let X_A be the indicator random variable for $|\chi(A)| > \alpha$, where we set $\alpha = \sqrt{2m \ln\ (2n)}$. If $|A| = a$, then $\chi(A)$ has distribution S_a, so by Theorem A.1.1

$$\mathrm{E}[X_A] = \Pr[|\chi(A)| > \alpha] < 2e^{-\alpha^2/2a} \le 2e^{-\alpha^2/2m} = 1/n$$

by our propitious choice of α. Let X be the number of $A \in \mathcal{A}$ with $|\chi(A)| > \alpha$, so that

$$X = \sum_{A \in \mathcal{A}} X_A$$

and Linearity of Expectation gives

$$\mathrm{E}[X] = \sum_{A \in \mathcal{A}} \mathrm{E}[X_A] < |\mathcal{A}|(1/n) = 1 \ .$$

Thus for some χ we must have $X = 0$. This means $\mathrm{disc}(\mathcal{A}, \chi) \leq \alpha$ and, therefore, $\mathrm{disc}(\mathcal{A}) \leq \alpha$. ■

13.2 SIX STANDARD DEVIATIONS SUFFICE

When $\mathcal{A}$ has both n sets and n points, Theorem 13.1.1 gives

$$disc(\mathcal{A}) = O(\sqrt{n \ln\ (n)}). \tag{13.1}$$

This was improved by the second author in Spencer (1985a).

Theorem 13.2.1 *Let $\mathcal{A}$ be a family of n subsets of an n element set Ω. Then*

$$disc(\mathcal{A}) \leq 6\sqrt{n}.$$

With $\chi : \Omega \to \{-1, +1\}$ random, $A \in \mathcal{A}$, $\chi(A)$ has zero mean and standard deviation at most $\sqrt{n}$. If $|\chi(A)| > 6\sqrt{n}$, then $\chi(A)$ is at least six standard deviations off the mean. The probability of this occurring is a very small but fixed positive constant and the number of sets A is going to infinity. In fact, a random χ almost always will *not* work. The specific constant 6 (actually 5.32) was the result of detailed calculations that could certainly be further improved and will not concern us here. Rather, we show Theorem 13.2.1 with some constant K replacing 6. The initial argument (found in earlier editions of this work) did not yield an efficient algorithm for finding the desired coloring χ. Indeed, for many years the second author conjectured that no such algorithm would exist. Bansal (2010) gave the first algorithmic argument for Theorem 13.2.1. Here we follow the approach of Lovett and Meka (2012). Their argument is a virtual cornucopia of modern probabilistic methods, and we give the basic ideas and leave many of the details to the exercises. We begin by generalizing the problem to vectors.

Theorem 13.2.2 *Let $\vec{r}_i \in R^n$, $1 \leq i \leq n$, with all $|\vec{r}_i|_\infty \leq 1$. Let $\vec{z} = (z_1, \ldots, z_n)$, with all $z_j \in [-1, +1]$. Then there exists $\vec{x} = (x_1, \ldots, x_n)$ with all $x_j \in \{-1, +1\}$ such that*

$$|\vec{r}_i \cdot (\vec{x} - \vec{z})| \leq K\sqrt{n} \tag{13.2}$$

for all $1 \leq i \leq n$. Here K is an absolute constant.

When $\mathcal{A}$ is a family of n subsets $A_1, \ldots, A_n$ of $\Omega = \{1, \ldots, n\}$, consider the $n \times n$ incidence matrix A, $a_{ij} = 1$ if $j \in A_i$, else $a_{ij} = 0$. Let $\vec{r}_i$ be the ith row of A, and set $\vec{z} = 0$. The $\vec{x} = (x_1, \ldots, x_n)$ given by Theorem 13.2.2 gives the coloring $\chi(j) = x_j$ with the properties of Theorem 13.2.1.

During the proof, the vector $\vec{x}$ shall *move* inside the cube $[-1,+1]^n$. We refer to this general technique as a *floating colors* method. It will initially have value $\vec{x} = \vec{z}$ so that (13.2) is trivially satisfied. When a coordinate x_i comes close to ± 1, it will be frozen. For definiteness, we set

$$\epsilon = n^{-1} \tag{13.3}$$

and say x_i is *near the border* if $1 - \epsilon \le |x_i| \le 1$. We call such *i frozen*, and all other *i floating*.

We reduce Theorem 13.2.2 to the following:

Theorem 13.2.3 *Let $\vec{r}_i \in R^n$, $1 \le i \le n$ with all $|\vec{r}_i|_\infty \le 1$. Let $\vec{z} = (z_1, \ldots, z_n)$ with all $z_j \in [-1,+1]$. Then there exists $\vec{x} = (x_1, \ldots, x_n)$ with all x_j near the border such that*

$$|\vec{r}_i \cdot (\vec{x} - \vec{z})| \le K\sqrt{n} \tag{13.4}$$

for all $1 \le i \le n$. Here K is an absolute constant.

With $\vec{x}$ given by Theorem 13.2.3, one can then simply round each x_i to either -1 or $+1$, whichever is closer. The values $\vec{r}_i \cdot (\vec{x} - \vec{z})$ are then changed by at most $n\epsilon = 1$, which is $o(\sqrt{n})$, thus giving Theorem 13.2.2.

We find $\vec{x}$ in *phases*. Phase t ends when at most $n2^{-t}$ of the x_i are not near the border. As Phase 1 contains the basic ideas of the argument, we state it separately.

Theorem 13.2.4 *Let $\vec{r}_i \in R^n$, $1 \le i \le n$ with all $|\vec{r}_i|_\infty \le 1$. Let $\vec{z} = (z_1, \ldots, z_n)$ with all $z_j \in [-1,+1]$. Then there exists $\vec{x} = (x_1, \ldots, x_n)$ with at least $n/2$ of the x_j near the border such that*

$$|\vec{r}_i \cdot (\vec{x} - \vec{z})| \le K_1\sqrt{n} \tag{13.5}$$

for all $1 \le i \le n$. Here K_1 is an absolute constant.

Set $\vec{u}_i = n^{-1/2}\vec{r}_i$. We will use (in Phase 1) only that the Euclidean norm of $\vec{u}_i$ is at most 1. We initially set $\vec{x} = \vec{z}$. We move $\vec{x}$ in *steps* until at least half the coordinates are near the border. For $1 \le i \le n$, set

$$L_i = \vec{u}_i \cdot (\vec{x} - \vec{z}). \tag{13.6}$$

Call *i dangerous* if $|L_i|$ is one of the $\frac{n}{4}$ largest of the values $|L_s|$, $1 \le s \le n$. In case of ties, select precisely $\frac{n}{4}$ values i. We emphasize that, as $\vec{x}$ moves, the $\frac{n}{4}$ dangerous i can and will change.

We define a vector space $V \subset R^n$, which will describe the allowable directions in which $\vec{x}$ may move. V is $\vec{y} = (y_1, \ldots, y_n)$ satisfying the following linear conditions:

1. If x_i is near the border then $y_i = 0$.
2. $\vec{y} \cdot (\vec{x} - \vec{z}) = 0$.
3. $\vec{y} \cdot \vec{u}_i = 0$ for all dangerous i.

The number of linear conditions is less than $\frac{n}{2} + 1 + \frac{n}{4}$. Let d denote the dimension of V, $d \geq \frac{n}{4}$. We let $\vec{y}$ be a standard multidimensional Gaussian on V. That is, let $\vec{b_1}, \ldots, \vec{b_d}$ be an orthonormal basis for V and set

$$\vec{y} = d^{-1/2}[n_1\vec{b_1} + \cdots + n_d\vec{b_d}], \tag{13.7}$$

where the n_j are independent, each with the standard normal distribution.

We shall use the directionless property of the Gaussian. Let $\vec{a} \in V$. Then $\vec{y} \cdot \vec{a}$ has a Gaussian distribution with mean 0 and variance $d^{-1}|\vec{a}|^2$. Suppose $\vec{b} \in R^n$. We can decompose $\vec{b} = \vec{a} + \vec{c}$ with $\vec{a} \in V$, $\vec{c} \in V^\perp$. Then $\vec{y} \cdot \vec{b} = \vec{y} \cdot \vec{a}$. Thus $\vec{y} \cdot \vec{b}$ has a Gaussian distribution with mean 0 and variance at most $d^{-1}|\vec{b}|^2$.

We now move $\vec{x}$ a small distance in the direction $\vec{y}$. Set, for definiteness,

$$\delta = n^{-10}. \tag{13.8}$$

A single step then consists of resetting

$$\vec{x} \leftarrow \vec{x} + \delta\vec{y}. \tag{13.9}$$

While the Lovett–Meka algorithm is discrete, as the δ of (13.8) becomes small, one may think of $\vec{x}$ as moving in a controlled Brownian motion, with the vector space V of permissible directions always changing.

A step fails if some $|x_i| > 1$. When x_i is near the border, $y_i = 0$ and so x_i does not change. If x_i is not near the border, it would need to change by at least ϵ in one step. Let $\vec{U_i}$ denote the vector with 1 in the ith position, zero elsewhere. In one step, the change in x_i is $\delta\vec{y} \cdot \vec{U_i}$, which is Gaussian with mean zero and variance at most $d^{-1}\delta^2$. With the values ϵ, δ, the probability that the change in x_i is more than ϵ is then exponentially small. There are only n choices of i, and we shall see that there are only polynomially many steps. Thus with probability $1 - o(1)$ no step fails. The chi-squared distribution (see Exercises) $n_1^2 + \cdots + n_d^2$ is tightly concentrated around its mean d. Thus $|\delta\vec{y}|^2$ is at least $(1 - o(1))\delta^2$ throughout Phase 1. At each step, $|\vec{x} - \vec{z}|^2$ is being increased by this amount. As they both lie in $[-1, +1]^n$, $|\vec{x} - \vec{z}|^2 \leq 4n$. Letting T denote the number of steps in Phase 1, we deduce $T \leq (1 + o(1))4n\delta^{-2}$.

Fix $1 \leq i \leq n$. Let $L_i(t)$ denote the value of L_i given in (13.6) after the tth step, with initial value $L_i(0) = 0$. With $\vec{y}$, the Gaussian selected at the tth step is

$$L_i(t) = L_i(t-1) + \delta\vec{u_i} \cdot \vec{y}. \tag{13.10}$$

Hence L_i will change by a Gaussian with variance $\tau^2 \leq \delta^2$. $L_i(t)$ then form a martingale. We apply the martingale inequality (13.17) in the Exercises. Here

$$\sigma^2 = T\delta^2 \leq (1 + o(1))4n\delta^{-2}d^{-1}\delta^2 \leq (1 + o(1))4n/d \leq 16 + o(1) \tag{13.11}$$

so $\sigma = 4 + o(1)$. Thus

$$\Pr[\max_{0 \leq t \leq T} |L_i(t)| > K(1 + o(1))] < 2e^{-K^2/32}(1 + o(1)). \tag{13.12}$$

Theorem 13.2.4 is shown by selecting $K_1 = K$ such that

$$2e^{-K^2/32} < 0.05. \tag{13.13}$$

Each L_i has probability less than 0.05 of ever becoming bigger than K in absolute value. By Linearity of Expectation, the expected number of such i is less than $0.05n$. The randomized Phase 1 is a success if one never has $|x_i| > 1$. It ends in at most $(1 + o(1))4n\delta^{-2}$ steps *and* there are fewer than $0.1n$ values i such that $|L_i|$ ever becomes bigger than K. The last occurs with probability at least 0.5 and so Phase 1 is a success with probability at least $0.5 - o(1)$.

Suppose we have success. Have we improved the situation over a standard randomized selection of the x_i? It seems that we still have a positive proportion of *outliers* to deal with. But look again! At each step the dangerous i had their L_i unchanged since the move $\vec{v}$, being in V, was orthogonal to $\vec{u}_i$. As less than $0.1n$ of the L_i ever have $|L_i| \geq K$, it must be that whenever an $|L_i|$ becomes at least K it will become, and stay, dangerous and so $|L_i|$ will remain the same throughout the remainder of Phase 1. The single move in which $|L_i|$ exceeds K is miniscule, so that after it $|L_i|$ is only $K + o(1)$. Therefore, at the end of the process *all* of the $|L_i| \leq K + o(1)$, completing the argument.

We outline the remainder of the argument for Theorem 13.2.3. The $\vec{x}$ at the end of Phase $t - 1$ becomes the initial $\vec{z}$ of Phase t. (When the number of floating variables reaches $O(n \ln^{-1/2} n)$, we can switch to a more standard random choice of the x_i. See the Exercises.) In Phase t, we begin with $n2^{-t} \leq m \leq n2^{1-t}$ floating variables so that $n \leq m2^t$. Ignore the nonfloating variables so that we consider $\vec{r}_i \in R^m$. As all coefficients lie in $[-1, +1]$, so we may bound $|\vec{r}_i|^2 \leq m$. We set $\vec{u}_i = m^{-1/2}\vec{r}_i$, so that $|\vec{u}_i| \leq 1$. We modify Theorem 13.2.4 as follows:

Theorem 13.2.5 *Let $n \leq m2^t$. Let $\vec{r}_i \in R^m$, $1 \leq i \leq n$, with all $|\vec{r}_i|_\infty \leq 1$. Let $\vec{z} = (z_1, \ldots, z_m)$, with all $z_j \in [-1, +1]$. Then there exists $\vec{x} = (x_1, \ldots, x_n)$ with at least $m/2$ of the x_j near the border such that*

$$|\vec{r}_i \cdot (\vec{x} - \vec{z})| \leq K_t \sqrt{m} \tag{13.14}$$

for all $1 \leq i \leq n$. Here K_t is an absolute constant.

We define $L_i = \vec{u}_i \cdot (\vec{x} - \vec{z})$ as in (13.6). Now i is dangerous if $|L_i|$ is one of the $n2^{-t-2} \leq \frac{m}{4}$ largest values. The large deviation bound (13.12) for the L_i is still valid, but now, instead of (13.13) we define $K = K_t$ such that

$$2e^{-K^2/2} \leq 0.05 \cdot 2^{-t}. \tag{13.15}$$

Now the expected number of i, $1 \leq i \leq n$, for which $|L_i(t)| \geq K$ ever occurs is less than $0.05n2^{-t} \leq 0.05m$. The remainder of the argument is as before.

From (13.15) we may set $K_t = \sqrt{c_1 + c_2 \ln t} = O(\sqrt{\ln t})$. As $m \leq n2^{-t+1}$, in Phase t all $|L_i| \leq K_t^* \sqrt{n}$ with $K_t^* = 2^{(1-t)/2}\sqrt{c_1 + c_2 \ln t} = O(2^{-t/2}\sqrt{\ln t})$.

Finally, we glue all the phases together. For each i, using the original definition (13.6) of L_i, the absolute value of the change in L_i in Phase t is at most K_t^*. But $\sum_{t=1}^{\infty} K_t^*$ converges to some K –basically the $2^{-t/2}$ gain by having fewer variables outweighs the $\sqrt{\ln t}$ loss by having more vectors than variables – and hence at the end of the process all $|L_i| \le K$.

13.3 LINEAR AND HEREDITARY DISCREPANCY

We now suppose that $\mathcal{A}$ has more points than sets. We write $\mathcal{A} = \{A_1, \dots, A_n\}$ and $\Omega = \{1, \dots, m\}$ and assume $m > n$. Note that $\operatorname{disc}(\mathcal{A}) \le K$ is equivalent to the existence of a set S; namely $S = \{j : \chi(j) = +1\}$, with $|S \cap A|$ within $K/2$ of $|A|/2$ for all $A \in \mathcal{A}$. We define the *linear discrepancy* $\operatorname{lindisc}(\mathcal{A})$ by

$$\operatorname{lindisc}(\mathcal{A}) = \max_{p_1,\dots,p_m \in [0,1]} \min_{\epsilon_1,\dots,\epsilon_m \in \{0,1\}} \max_{A \in \mathcal{A}} \left| \sum_{i \in A} (\epsilon_i - p_i) \right| .$$

The upper bound $\operatorname{lindisc}(\mathcal{A}) \le K$ means that, given any $p_1, \dots, p_m$, there is a "simultaneous roundoff" $\epsilon_1, \dots, \epsilon_m$ so that, with $S = \{j : \epsilon_j = 1\}$, $|S \cap A|$ is within K of the weighted sum $\sum_{j \in A} p_j$ for all $A \in \mathcal{A}$. Taking all $p_j = \frac{1}{2}$, the upper bound implies $\operatorname{disc}(\mathcal{A}) \le 2K$. But $\operatorname{lindisc}(\mathcal{A}) \le K$ is much stronger. It implies, taking all $p_j = \frac{1}{3}$, the existence of an S with all $|S \cap A|$ within K of $|A|/3$, and much more. Linear discrepancy and its companion hereditary discrepancy defined below have been developed in Lovász, Spencer and Vesztergombi (1986). For $X \subset \Omega$, let $\mathcal{A}|_X$ denote the restriction of $\mathcal{A}$ to X, that is, the family $\{A \cap X : A \in \mathcal{A}\}$. The next result "reduces" the bounding of $\operatorname{disc}(\mathcal{A})$ when there are more points than sets to the bounding of $\operatorname{lindisc}(\mathcal{A})$ when the points do not outnumber the sets.

Theorem 13.3.1 *Let $\mathcal{A}$ be a family of n sets on m points with $m \ge n$. Suppose that $\operatorname{lindisc}(\mathcal{A}|_X) \le K$ for every subset X of at most n points. Then $\operatorname{lindisc}(\mathcal{A}) \le K$.*

Proof. Let $p_1, \dots, p_m \in [0, 1]$ be given. We define a reduction process. Call index j fixed if $p_j \in \{0, 1\}$, otherwise call it floating, and let F denote the set of floating indices. If $|F| \le n$, then halt. Otherwise, let $y_j, j \in F$, be a nonzero solution to the homogeneous system

$$\sum_{j \in A \cap F} y_j = 0, \qquad A \in \mathcal{A}.$$

Such a solution exists since there are more variables ($|F|$) than equations (n) and may be found by standard techniques of linear algebra. Now set

$$p_j' = p_j + \lambda y_j, \quad j \in F,$$
$$p_j' = p_j, \qquad\quad j \notin F,$$

where we let λ be the real number of the least absolute value so that for some $j \in F$ the value p_j' becomes zero or 1. Critically

$$\sum_{j\in A} p_j' = \sum_{j\in A} p_j + \lambda \sum_{j\in A\cap F} y_j = \sum_{j\in A} p_j \qquad (*)$$

for all $A \in \mathcal{A}$. Now iterate this process with the new p_j'. At each iteration, at least one floating j becomes fixed and so the process eventually halts at some $p_1^*, \dots, p_m^*$. Let X be the set of floating j at this point. Then $|X| \le n$. By assumption, there exist $\epsilon_j, j \in X$ so that

$$\left| \sum_{j\in A\cap X} p_j^* - \epsilon_j \right| \le K, \quad A \in \mathcal{A}.$$

For $j \notin X$, set $\epsilon_j = p_j^*$. As (*) holds at each iteration,

$$\sum_{j\in A} p_j^* = \sum_{j\in A} p_j$$

and hence

$$\left| \sum_{j\in A} (p_j - \epsilon_j) \right| = \left| \sum_{j\in A} (p_j - p_j^*) + \sum_{j\in A\cap X} (p_j^* - \epsilon_j) \right| \le K$$

for all $A \in \mathcal{A}$. ■

We now define the *hereditary discrepancy* herdisc($\mathcal{A}$) by

$$\text{herdisc}(\mathcal{A}) = \max_{X\subseteq\Omega} \text{disc}(\mathcal{A}|_X).$$

Example. Let $\Omega = \{1, \dots, n\}$, and let $\mathcal{A}$ consist of all intervals $[i,j] = \{i, i+1, \dots, j\}$ with $1 \le i \le j \le n$. Then disc($\mathcal{A}$) = 1, as we may color Ω alternately $+1$ and -1. But also herdisc($\mathcal{A}$) = 1. For given any $X \subseteq \Omega$, say with elements $x_1 < x_2 < \cdots < x_r$, we may color X alternately by $\chi(x_k) = (-1)^k$. For any set $[i,j] \in \mathcal{A}$, the elements of $[i,j] \cap X$ are alternately colored. ■

Theorem 13.3.2 lindisc($\mathcal{A}$) $\le$ herdisc($\mathcal{A}$).

Proof. Set $K =$ herdisc($\mathcal{A}$). Let $\mathcal{A}$ be defined on $\Omega = \{1, \dots, m\}$, and let $p_1, \dots, p_m \in [0,1]$ be given. First let us assume that all p_i have finite expansions when written in base 2. Let T be the minimal integer so that all $p_i 2^T \in Z$. Let J be the set of i for which p_i has a 1 in the Tth digit of its binary expansion, that is, so that $p_i 2^{T-1} \notin Z$. As disc($\mathcal{A}|_J$) $\le K$, there exist $\epsilon_j \in \{-1, +1\}$, so that

$$\left| \sum_{j\in J\cap A} \epsilon_j \right| \le K$$

for all $A \in \mathcal{A}$. Write $p_j = p_j^{(T)}$. Now set

$$p_j^{(T-1)} = \begin{cases} p_j^{(T)} & \text{if } j \notin J, \\ p_j^{(T)} + \epsilon_j 2^{-T} & \text{if } j \in J. \end{cases}$$

That is, the $p_j^{(T-1)}$ are the "roundoffs" of the $p_j^{(T)}$ in the Tth place. Note that all $p_j^{(T-1)} 2^{-(T-1)} \in Z$. For any $A \in \mathcal{A}$,

$$\left| \sum_{j \in A} p_j^{(T-1)} - p_j^{(T)} \right| = \left| \sum_{j \in J \cap A} 2^{-T} \epsilon_j \right| \leq 2^{-T} K \ .$$

Iterate this procedure, finding $p_j^{(T-2)}, \ldots, p_j^{(1)}, p_j^{(0)}$. All $p_j^{(0)} 2^{-0} \in Z$ so all $p_j^{(0)} \in \{0, 1\}$ and

$$\left| \sum_{j \in A} p_j^{(0)} - p_j^{(T)} \right| \leq \sum_{i=1}^{T} \left| \sum_{j \in A} p_j^{(i-1)} - p_j^{(i)} \right| \leq \sum_{i=1}^{T} 2^{-i} K \leq K \ ,$$

as desired.

What about general $p_1, \ldots, p_m \in [0, 1]$? We can flip and say that, at least to a computer scientist, all real numbers have finite binary expansions. More rigorously, the function

$$f(p_1, \ldots, p_m) = \min_{\epsilon_1, \ldots, \epsilon_m \in \{0,1\}} \max_{A \in \mathcal{A}} \left| \sum_{i \in A} (\epsilon_i - p_i) \right|$$

is the finite minimum of finite maxima of continuous functions and thus is continuous. The set of $(p_1, \ldots, p_m) \in [0, 1]^m$ with all $p_i 2^T \in Z$ for some T is a dense subset of $[0, 1]^m$. As $f \leq K$ on this dense set, $f \leq K$ for all $(p_1, \ldots, p_m) \in [0, 1]^m$. ∎

Corollary 13.3.3 *Let $\mathcal{A}$ be a family of n sets on m points. Suppose* $\operatorname{disc}(\mathcal{A}|_X) \leq K$ *for every subset X with at most n points. Then* $\operatorname{disc}(\mathcal{A}) \leq 2K$.

Proof. For every $X \subseteq \Omega$ with $|X| \leq n$, $\operatorname{herdisc}(\mathcal{A}|_X) \leq K$ so by Theorem 13.3.2, $\operatorname{lindisc}(\mathcal{A}|_X) \leq K$. By Theorem 13.3.1, $\operatorname{lindisc}(\mathcal{A}) \leq K$. But

$$\operatorname{disc}(\mathcal{A}) \leq 2 \ \operatorname{lindisc}(\mathcal{A}) \leq 2K.$$

∎

Corollary 13.3.4 *For any family $\mathcal{A}$ of n sets of arbitrary size*

$$\operatorname{disc}(\mathcal{A}) \leq 12\sqrt{n} \ .$$

Proof. Apply Theorem 13.2.1 and Corollary 13.3.3. ∎

13.4 LOWER BOUNDS

We now give two quite different proofs that, up to a constant factor, Corollary 13.3.4 is the best possible. A Hadamard matrix is a square matrix $H = (h_{ij})$ with all $h_{ij} \in \{-1, +1\}$ and with row vectors mutually orthogonal (and hence with column vectors mutually orthogonal). Let H be a Hadamard matrix of order n, and let $v = (v_1, \ldots, v_n)$, $v_i \in \{-1, +1\}$. Then

$$Hv = v_1 c_1 + \cdots + v_n c_n,$$

where c_i denotes the ith column vector of H. Writing $Hv = (L_1, \ldots, L_n)$ and letting $|c|$ denote the usual Euclidean norm

$$L_1^2 + \cdots + L_n^2 = |Hv|^2 = v_1^2 |c_1|^2 + \cdots + v_n^2 |c_n|^2 = n + \cdots + n = n^2$$

since c_i are mutually orthogonal. Hence some $L_i^2 \geq n$ and thus

$$|Hv|_\infty = \max(|L_1|, \ldots, |L_n|) \geq \sqrt{n}\ .$$

Now we transfer this result to one on families of sets. Let H be a Hadamard matrix of order n with first row and first column all 1s. (Any Hadamard matrix can be so "normalized" by multiplying appropriate rows and columns by -1.) Let J denote the all-1s matrix of order n. Let $v_1, \ldots, L_1, \cdots$ be as above. Then

$$L_1 + \cdots + L_n = \sum_{i,j=1}^{n} v_j h_{ij} = \sum_{j=1}^{n} v_j \sum_{i=1}^{n} h_{ij} = n v_1 = \pm n,$$

since the first column sums to n but the other columns, being orthogonal to it, sum to zero. Set $\lambda = v_1 + \cdots + v_n$ so that $Jv = (\lambda, \ldots, \lambda)$ and

$$(H + J)v = (L_1 + \lambda, \ldots, L_n + \lambda).$$

We calculate

$$|(H + J)v|^2 = \sum_{i=1}^{n} (L_i + \lambda)^2 = \sum_{i=1}^{n} (L_i^2 + 2\lambda L_i + \lambda^2) = n^2 \pm 2n\lambda + n\lambda^2.$$

Assume n is even. (Hadamard matrices do not exist for odd n, except for $n = 1$.) Then λ is an even integer. The quadratic (in λ) $n^2 \pm 2n\lambda + n\lambda^2$ has a minimum at ∓ 1 implying under the restriction of being an even integer, its minimum is at $\lambda = 0, \mp 2$, implying

$$|(H + J)v|^2 \geq n^2.$$

Again, some coordinate must be at least $\sqrt{n}$. Setting $H^* = \frac{H+J}{2}$

$$|H^* v|_\infty \geq \sqrt{n}/2.$$

Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ be any family of subsets of $\Omega = \{1, \ldots, n\}$, and let M denote the corresponding $m \times n$ incidence matrix. A coloring $\chi : \Omega \to \{-1, +1\}$ corresponds to a vector $v = (\chi(1), \ldots, \chi(n)) \in \{-1, +1\}^n$. Then

$$\mathrm{disc}(\mathcal{A}, \chi) = |Mv|_\infty$$

and

$$\mathrm{disc}(\mathcal{A}) = \min_{v \in \{-1,+1\}^n} |Mv|_\infty.$$

In our case, H^* has entries 0, 1. Thus we have following:

Theorem 13.4.1 *If a Hadamard matrix exists of order $n > 1$ then there exists a family $\mathcal{A}$ consisting of n subsets of an n-set with*

$$\mathrm{disc}(\mathcal{A}) \geq \sqrt{n}/2.$$

While it is not known precisely for which n a Hadamard matrix exists (the Hadamard conjecture is that they exist for $n = 1, 2$ and all multiples of 4; see, for example, Hall (1986)), it is known that the orders of Hadamard matrices are dense in the sense that, for all ϵ if n is sufficiently large, there will exist a Hadamard matrix of order between n and $n(1 - \epsilon)$. This result suffices to extend Theorem 13.4.1 to an asymptotic result on all n.

Our second argument for the existence of $\mathcal{A}$ with high discrepancy involves turning the probabilistic argument "on its head." Let $v = (v_1, \ldots, v_n)$, $v_j = \pm 1$ be fixed. Let A be the set of indices j with $v_j = +1$ and B those with $v_j = -1$. Set $a = |A|$, $b = n - a = |B|$. Let $\vec{w} = (w_1, \ldots, w_n) \in \{0, 1\}^n$. For any $-b \leq t \leq a$, $\vec{w} \cdot \vec{v} = t$ if and only if the number of $j \in A$ with $w_j = 1$ plus the number of $j \in B$ with $w_j = 0$ is $b + t$. Thus when $\vec{w}$ is chosen uniformly, $\Pr[\vec{w} \cdot \vec{v} = t] = \Pr[\ BIN[n, \frac{1}{2}] = b + t]$. Suppose $n = 2k + 1$, the case n even being similar. Then $\Pr[|\vec{w} \cdot \vec{v}| \leq u] = \Pr[b - u \leq\ BIN[n, \frac{1}{2}] \leq b + u]$. As the binomial distribution is symmetric about $k + 1$ and decreasing as one moves away from k, $\Pr[|\vec{w} \cdot \vec{v}| \leq u] \leq \Pr[|\ BIN[n, \frac{1}{2}] - (k + 1)| \leq u]$. Let λ_0 be positive real so that $\Pr[|N| \leq \lambda_0] = \frac{1}{2}$, and N the standard Gaussian. Let $0 < \lambda < \lambda_0$. For $u \sim \lambda\sqrt{n}/2$, the Central Limit Theorem gives $\lim_{n\to\infty} \Pr[|\ BIN[n, \frac{1}{2}] - \frac{n}{2}| \leq u] = \Pr[|N| \leq \lambda] < \frac{1}{2}$ so that $\Pr[|\ BIN[n, \frac{1}{2}] - \frac{n}{2}| \leq u] < \frac{1}{2}$ for sufficiently large n.

Now let M be a random 0, 1 matrix of order n. Let $u \sim \lambda\sqrt{n}/2$ as above. Let $\vec{r}_1, \ldots, \vec{r}_n$ denote the row vectors of M. For each $\vec{v} \in \{-1, +1\}^n$ and each $1 \leq i \leq n$, $\Pr[|\vec{v} \cdot \vec{r}_i| \leq u] < \frac{1}{2}$. As $\vec{r}_i$ are mutually independent, $\Pr[|\vec{v} \cdot \vec{r}_i| \leq u \ \textit{for all}\ 1 \leq i \leq n] < 2^{-n}$. Hence the expected number of $\vec{v}$ such that $|\vec{v} \cdot \vec{r}_i| \leq u$ for all $1 \leq i \leq n$ is less than 1. With positive probability, M has the property that there is no such $\vec{v}$. Let M be such a matrix. The corresponding family of sets $\mathcal{A}$ then has discrepancy greater than u. As λ can be chosen arbitrarily close to λ_0, there are $\mathcal{A}$ with discrepancy at least $(\lambda_0 + o(1))\sqrt{n}/2$.

13.5 THE BECK–FIALA THEOREM

For any family $\mathcal{A}$, let $\deg(\mathcal{A})$ denote the maximal number of sets containing any particular point. The following result due to Beck and Fiala (1981) uses only methods from linear algebra and thus is technically outside the scope we have set for this book. We include it both for the sheer beauty of the proof and because the result itself is very much in the spirit of this chapter.

Theorem 13.5.1 *Let $\mathcal{A}$ be a finite family of finite sets, with no restriction on either the number of sets or on the cardinality of the sets, and with* $\deg(\mathcal{A}) \leq t$. *Then*

$$\operatorname{disc}(\mathcal{A}) \leq 2t - 1.$$

Proof. For convenience, write $\mathcal{A} = \{A_1, \ldots, A_m\}$ with all $A_i \subseteq \Omega = \{1, \ldots, n\}$. To each $j \in \Omega$, there is assigned a value x_j which will change as the proof progresses. Initially all $x_j = 0$. At the end, all $x_j = \pm 1$. We will have $-1 \leq x_j \leq +1$ at all times, and once $x_j = \pm 1$, it "sticks" there and that becomes its final value. A set S_i has value $\sum_{j \in S_i} x_j$. At any time j is called fixed if $x_j = \pm 1$; otherwise it is *floating*. A set S_i is safe if it has fewer than t floating points; otherwise it is *active*. Note, crucially, that as points are in at most t sets and active sets contain more than t floating points, there must be fewer active sets than floating points.

We insist at all times that all active sets have value zero. This holds initially since all sets have value zero. Suppose this condition holds at some stage. Consider x_j a variable for each floating j and a constant for each fixed j. The condition that S_i has value zero then becomes a linear equation in these variables. This is an underdetermined system, as there are fewer linear conditions (active sets) than variables (floating points). Hence we may find a line, parameterized

$$x'_j = x_j + \lambda y_j, \quad j \text{ floating}$$

on which the active sets retain value zero. Let λ be the smallest value for which some x'_j becomes ± 1 and replace each x_j by x'_j. (Geometrically, follow the line until reaching the boundary of the cube in the space over the floating variables.) This process has left fixed variables fixed and so safe sets stayed safe sets (though active sets may have become safe) and so the condition still holds. In addition, at least one previously floating j has become fixed.

We iterate the above procedure until all j have become fixed. (Towards the end, we may have no active sets, at which time we may simply set the floating x_j to ± 1 arbitrarily.) Now consider any set S_i. Initially it had value zero, and it retained the value zero while it contained at least t floating points. Consider the time when it first becomes safe, say $1, \ldots, l$ were its floating points. At this moment its value is zero. The variables $y_1, \ldots, y_l$ can now change by less than 2 to their final value since all values are in $[-1, +1]$. Thus, in total, they may change by less than $2t$. Hence the final value of S_i is less than $2t$ and, as it is an integer, it is at most $2t - 1$. ∎

Conjecture 13.5.2 *If* $\deg(\mathcal{A}) \leq t$, *then* $\operatorname{disc}(\mathcal{A}) \leq K\sqrt{t}$, *with K an absolute constant.*

This conjecture seems to call for a melding of probabilistic methods and linear algebra. The constructions of t sets on t points, described in Section 13.4, show that, if true, this conjecture would be the best possible.

13.6 EXERCISES

1. Let $\mathcal{A}$ be a family of n subsets of $\Omega = \{1, \ldots, m\}$ with m even. Let $\chi(i)$, $1 \leq i \leq \frac{m}{2}$, be independent and uniform in $\{-1, +1\}$ and set $\chi(i + \frac{m}{2}) = -\chi(i)$ for $1 \leq i \leq \frac{m}{2}$. Using this notion of random coloring, improve Theorem 13.1.1 by showing

$$\text{disc}(\mathcal{A}) \leq \sqrt{m \ln (2n)}\,.$$

   Show that this can be improved even further by splitting Ω randomly into $\frac{m}{2}$ disjoint pairs.

2. Let $\vec{v}_1, \ldots, \vec{v}_s \in R^n$. Let $x_1, \ldots, x_s \in [-1, +1]$ such that $\sum_{i=1}^{s} x_i \vec{v}_i = \vec{0}$ and such that $x_i \in \{-1, +1\}$ for all but at most n values of i. Let $\vec{v}_{s+1} \in R^n$. Use the linear ideas of Section 13.5 to find $x'_1, \ldots, x'_s, x'_{s+1}$ with the following properties:
   - $\sum_{i=1}^{s+1} x'_i \vec{v}_i = \vec{0}$.
   - All $x'_i \in [-1, +1]$.
   - $x'_i \in \{-1, +1\}$ for all but at most n values of i.
   - $x'_i = x_i$ whenever $x_i \in \{-1, +1\}$.

   Use the above to prove the following result of Bárány and Grinberg: Let $|\cdot|$ be an arbitrary norm in R^n. Let $\vec{v}_1, \ldots, \vec{v}_s \in R^n$ with all $|v_i| \leq 1$. Then there exist $x_1, \ldots, x_s \in \{-1, +1\}$ such that

$$\left| \sum_{i=1}^{t} x_i \vec{v}_i \right| \leq 2n$$

   for all $1 \leq t \leq s$.

3. Let $n_1, \ldots, n_d$ be mutually independent standard Gaussians, and set $Y = \sum_{i=1}^{d} n_i^2$. (Y is known as the *chi-squared* distribution.) Set $\mu = E[Y] = d$. Show that for all positive ϵ there exists a positive $c = c(\epsilon)$ so that $\Pr[Y \leq \mu(1 - \epsilon)] < e^{-cd}$.

4. Let $\sigma, \sigma_1, \ldots, \sigma_T > 0$ with $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$. Let $0 = X_0, X_1, \ldots, X_T$ be a martingale with $X_i - X_{i-1}$ having a Gaussian distribution with mean zero and variance τ_i^2, with $\tau_i^2 \leq \sigma_i^2$. (Note: τ_i can depend on the previous history. In particular, we allow $\tau_i = 0$.) Let $a > 0$. Show that

$$\Pr[X_T > a\sigma] < e^{-a^2/2} \tag{13.16}$$

   and further that

$$\Pr[\max_{0 \leq t \leq T} X_t > a\sigma] < e^{-a^2/2}. \tag{13.17}$$

5. Let $n > cm\sqrt{\ln\ m}$, c an appropriately large constant. For $1 \le i \le n$, let $\vec{u}_i \in R^m$ with $|\vec{u}_i| \le 1$. Let $\vec{z} = (z_1, \ldots, z_m) \in [-1, +1]^m$. Define a random $\vec{x} = (x_1, \ldots, x_m) \in \{-1, +1\}^m$, with $\Pr[x_i = +1] = (1 + z_i)/2$, $\Pr[x_i = -1] = (1 - z_i)/2$ and the x_i mutually independent. Prove that, with probability $1 - o(1)$, $|\vec{u}_i \cdot (\vec{x} - \vec{z})| \le \sqrt{n}$ for all $1 \le i \le n$. (This result allows us to apply randomized rounding when $m \le n2^{1-t}$ becomes small enough.)
6. Replace (13.3) with $\epsilon = \ln^{-1/2} n$. Show that Theorem 13.2.3 still implies Theorem 13.2.2 by appropriately randomly rounding the x_i with $|x_i| \ge 1 - \epsilon$. Now show δ in (13.8) can be increased to $O(\sqrt{n}\ \ln^{-1} n)$. Show that the number of steps in Phase 1 can then be made $O(\ln^2 n)$. (While this does not affect the mathematical proof, it does increase the speed of the algorithm. However, the biggest time factor in the implementation of the algorithm is the generation of the Gaussian $\vec{g}$, which in turn depends on an orthonormal basis for V.)

THE PROBABILISTIC LENS:
Unbalancing Lights

For any $m \times n$ matrix $B = (b_{ij})$ with coefficients $b_{ij} = \pm 1$, set

$$F[B] = \max_{x_i, y_j = \pm 1} \sum_{i=1}^{m} \sum_{j=1}^{n} x_i y_j b_{ij} \ .$$

As in Section 2.5, we may interpret B as an $m \times n$ array of lights, each either on ($b_{ij} = +1$) or off ($b_{ij} = -1$). For each row and each column, there is a switch that, when pulled, changes all lights in that line from on to off or from off to on. Then $F[B]$ gives the maximal achievable number of lights on minus lights off. In Section 2.5 we found a lower bound for $F[B]$ when $m = n$. Here we set $n = 2^m$ and find the precise best possible lower bound.

With $n = 2^m$ let A be an $m \times n$ matrix with coefficients ± 1 containing every possible column vector precisely once. We claim $F[A]$ is the minimal value of $F[B]$ over all $m \times n$ matrices B.

For any given B, let $x_1, \ldots, x_m = \pm 1$ be independently and uniformly chosen and set

$$X_j = \sum_{i=1}^{m} x_i b_{ij},$$
$$X = |X_1| + \cdots + |X_n|,$$

so that

$$F[B] = \max_{y_j = \pm 1} \max_{x_i = \pm 1} \sum_{j=1}^{n} y_j X_j = \max_{x_i = \pm 1} \sum_{j=1}^{n} |X_j| = \max X.$$

Regardless of the b_{ij}, X_i has distribution S_m so that $\mathrm{E}[|X_i|] = \mathrm{E}[|S_m|]$ and, by Linearity of Expectation,

$$\mathrm{E}[X] = n\,\mathrm{E}[|S_m|]\ .$$

With $B = A$, any choices of $x_1, \ldots, x_m = \pm 1$ have the effect of permuting the columns – the matrix $(x_i a_{ij})$ also has every column vector precisely once – so that $X = |X_1| + \cdots + |X_m|$ is a constant. Note that $\mathrm{E}[X]$ is independent of B. In general, fixing $\mathrm{E}[X] = \mu$, the minimal possible value for $\max X$ is achieved when X is the constant μ. Thus $F[B]$ is minimized with $B = A$.

14

Geometry

If creativity were anything but random, someone would have figured out the algorithm by now.
–Dilbert (Scott Adams)

Suppose we choose randomly n points $P_1, \ldots, P_n$ on the unit circle, according to a uniform distribution. What is the probability that the origin lies in the convex hull of these points? There is a surprisingly simple (yet clever) way to compute this probability. Let us first choose n random pairs of antipodal points $Q_1, Q_{n+1} = -Q_1, Q_2, Q_{n+2} = -Q_2, \ldots, Q_n, Q_{2n} = -Q_n$ according to a uniform distribution. Notice that with probability 1 these pairs are all distinct. Next we choose each P_i to be either Q_i or its antipodal $Q_{n+i} = -Q_i$, where each choice is equally likely. Clearly, this corresponds to a random choice of the points P_i. The probability that the origin does *not* belong to the convex hull of the points P_i, given the (distinct) points Q_j, is precisely $\frac{x}{2^n}$, where x is the number of subsets of the points Q_j contained in an open half plane determined by a line through the origin, which does not pass through any of the points Q_j. It is easy to see that $x = 2n$. This is because, if we renumber the points Q_j so that their cyclic order on the circle is $Q_1, \ldots, Q_n, Q_{n+1}, \ldots, Q_{2n}$, and $Q_{n+i} = -Q_i$, then the subsets contained in such half planes are precisely $\{Q_i, \ldots, Q_{n+i-1}\}$, where the indices are reduced modulo $2n$. Therefore, the probability that the origin is in the convex hull of n randomly chosen points on the unit circle is precisely $1 - \frac{2n}{2^n}$. Observe that the same result holds if we replace the unit circle by any centrally symmetric bounded

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

planar domain with center 0, and the argument can be easily generalized to higher dimensions.

This result, due to Wendel (1962), shows how in some cases a clever idea can replace a tedious computation. It also demonstrates the connection between probability and geometry. The probabilistic method has been recently used extensively for deriving results in discrete and computational geometry. Some of these results are described in this chapter.

14.1 THE GREATEST ANGLE AMONG POINTS IN EUCLIDEAN SPACES

There are several striking examples, in different areas of combinatorics, where the probabilistic method supplies very simple counterexamples to long-standing conjectures. Here is an example, due to Erdős and Füredi (1983).

Theorem 14.1.1 *For every $d \geq 1$, there is a set of at least $\left\lfloor \frac{1}{2}\left(\frac{2}{\sqrt{3}}\right)^d \right\rfloor$ points in the d-dimensional Euclidean space R^d, such that all angles determined by three points from the set are strictly less than $\pi/2$.*

This theorem disproves an old conjecture of Danzer and Grünbaum (1962) that the maximum cardinality of such a set is at most $2d - 1$. We note that, as proved by Danzer and Grunbaum, the maximum cardinality of a set of points in R^d in which all angles are at most $\pi/2$ is 2^d.

Proof [Theorem 14.1.1] We select the points of a set X in R^d from the vertices of the d-dimensional cube. As usual, we view the vertices of the cube, which are 0, 1-vectors of length d, as the characteristic vectors of subsets of a d-element set; that is, each 0, 1-vector a of length d is associated with the set $A = \{i : 1 \leq i \leq d, a_i = 1\}$. A simple consequence of Pythagoras' theorem gives that the three vertices a, b, and c of the d-cube, corresponding to the sets A, B, and C, respectively, determine a right angle at c if and only if

$$A \cap B \subset C \subset A \cup B. \tag{14.1}$$

As the angles determined by triples of points of the d-cube are always at most $\pi/2$, it suffices to construct a set X of cardinality at least the one stated in the theorem no three distinct members of which satisfy (14.1).

Define $m = \lfloor (2/\sqrt{3})^d/2 \rfloor$, and choose, randomly and independently, $2m$ d-dimensional $\{0, 1\}$-vectors $a_1, \ldots, a_{2m}$, where each coordinate of each of the vectors independently is chosen to be either 0 or 1 with equal probability. For every fixed triple a, b, and c of the chosen points, the probability that the corresponding sets satisfy (14.1) is precisely $(3/4)^d$. This is because (14.1) simply means that for each i, $1 \leq i \leq d$, neither $a_i = b_i = 0, c_i = 1$ nor $a_i = b_i = 1, c_i = 0$ hold. Therefore, the probability that for three fixed indices i, j, and k, our chosen points, a_i, a_j, a_k form

a right angle at a_k is $(3/4)^d$. Since there are $\binom{2m}{3}$ 3 possible triples that can produce such angles, the expected number of right angles is

$$\binom{2m}{3} 3(3/4)^d \leq m \ ,$$

where the last inequality follows from the choice of m. Thus there is a choice of a set X of $2m$ points in which the number of right angles is at most m. By deleting one point from each such angle, we obtain a set of at least $2m - m = m$ points in which all angles are strictly less than $\pi/2$. Notice that the remaining points are all distinct since (14.1) is trivially satisfied if $A = C$. This completes the proof. ■

It is worth noting that, as observed by Erdős and Füredi (1983), the proof above can be easily modified to give the following:

Theorem 14.1.2 *For every $\epsilon > 0$, there is a $\delta > 0$ such that for every $d \geq 1$ there is a set of at least $(1 + \delta)^d$ points in R^d so that all the angles determined by three distinct points from the set are at most $\pi/3 + \epsilon$.*

We omit the detailed proof of this result.

14.2 EMPTY TRIANGLES DETERMINED BY POINTS IN THE PLANE

For a finite set X of points in a general position in the plane, let $f(X)$ denote the number of *empty* triangles determined by triples of points of X, that is, the number of triangles determined by points of X that contain no other point of X. Katchalski and Meir (1988) studied the minimum possible value of $f(X)$ for a set X of n points. Define $f(n) = \min\{f(X)\}$, as X ranges over all planar sets of n points in general position (i.e., containing no three colinear points). They proved that

$$\binom{n-1}{2} \leq f(n) < 200n^2 \ .$$

These bounds were improved by Bárány and Füredi (1987), who showed that as n grows

$$(1 + o(1))n^2 \leq f(n) \leq (1 + o(1))2n^2 \ .$$

The construction that establishes the upper bound is probabilistic and is given in the following theorem. See also Valtr (1995) for a slightly better result.

Theorem 14.2.1 *Let $I_1, I_2, \ldots, I_n$ be parallel unit intervals in the plane, where*

$$I_i = \{(x, y) : x = i, 0 \leq y \leq 1\} \ .$$

For each i, let us choose a point p_i randomly and independently from I_i according to a uniform distribution. Let X be the set consisting of these n randomly chosen points. Then the expected number of empty triangles in X is at most $2n^2 + O(n \log n)$.

Clearly, with probability 1, X is a set of points in general position, and hence the above theorem shows that $f(n) \leq 2n^2 + O(n \log n)$.

Proof. We first estimate the probability that the triangle determined by the points p_i, p_{i+a}, and p_{i+k} is empty, for some fixed i, a, and $k = a + b \geq 3$. Let $A = (i, x)$, $B = (i + a, y)$, and $C = (i + k, z)$ be the points p_i, p_{i+a}, and p_{i+k}, respectively. Let m be the distance between B and the intersection point of the segment AC with the interval I_{i+a}. Since each of the points p_j for $i < j < i + k$ is chosen randomly according to a uniform distribution on I_j, it follows that the probability that the triangle determined by A, B, and C is empty is precisely

$$\begin{aligned}
&\left(1 - \frac{m}{a}\right)\left(1 - 2\frac{m}{a}\right) \cdots \left(1 - (a-1)\frac{m}{a}\right)\left(1 - (b-1)\frac{m}{b}\right) \cdots \left(1 - \frac{m}{b}\right) \\
&\leq \exp\ \left(-\frac{m}{a} - 2\frac{m}{a} \cdot \cdots - (a-1)\frac{m}{a} - (b-1)\frac{m}{b} \cdots - \frac{m}{b}\right) \\
&= \exp\ \left(-\binom{a}{2}\frac{m}{a} - \binom{b}{2}\frac{m}{b}\right) = \exp\ \left(-(k-2)\frac{m}{2}\right)\ .
\end{aligned}$$

For every fixed choice of A and C, when the point $p_{i+a} = B$ is chosen randomly, the probability that its distance m from the intersection of the segment AC with the interval I_{i+a} is at most d is clearly at most $2d$, for all $d \geq 0$. Therefore, the probability that the triangle determined by p_i, p_{i+a}, and p_{i+k} is empty is at most

$$2\int_{m \geq 0} \exp\ (-(k-2)m/2) \quad dm = 4/(k-2)\ .$$

It follows that the expected value of the total number of empty triangles is at most

$$\begin{aligned}
&n - 2 + \sum_{1 \leq i \leq n-3} \sum_{3 \leq k \leq n-i} \sum_{1 \leq a \leq k-1} 4/(k-2) \\
&= n - 2 + \sum_{3 \leq k \leq n-1} (n-k)\frac{4(k-1)}{k-2} \\
&= n - 2 + \sum_{3 \leq k \leq n-1} (n-k)4/(k-2) + 4 \sum_{3 \leq k \leq n-1} (n-k) \\
&= 2n^2 + O(n \log\ n)\ .
\end{aligned}$$

This completes the proof. ∎

The result above can be extended to higher dimensions by applying a similar probabilistic construction. A set X of n points in the d-dimensional Euclidean space is called *independent* if no $d + 1$ of the points lie on a hyperplane. A simplex determined by $d + 1$ of the points is called *empty* if it contains no other point of X. Let $f_d(X)$ denote the number of empty simplices of X, and define $f_d(n) = \min f_d(X)$, as X ranges over all independent sets of n points in R^d. Katchalski and Meir (1988) showed

that $f_d(n) \geq \binom{n-1}{d}$. The following theorem of Bárány and Füredi shows that, here again, a probabilistic construction gives a matching upper bound, up to a constant factor (which depends on the dimension). We omit the detailed proof.

Theorem 14.2.2 *There exists a constant $K = K(d)$, such that for every convex, bounded set $A \subset R^d$ with nonempty interior, if X is a random set of n points obtained by n random and independent choices of points of A picked with uniform distribution, then the expected number of empty simplices of X is at most $K\binom{n}{d}$.*

14.3 GEOMETRICAL REALIZATIONS OF SIGN MATRICES

Let $A = (a_{i,j})$ be an m by n matrix with $+1, -1$-entries. We say that A is *realizable* in R^d if there are m hyperplanes $H_1, \ldots, H_m$ in R^d passing through the origin and n points $P_1, \ldots, P_n$ in R^d, so that for all i and j, P_j lies in the positive side of H_i if $a_{i,j} = +1$, and in the negative side if $a_{i,j} = -1$. Let $d(A)$ denote the minimum dimension d such that A is realizable in R^d, and define $d(m, n) = \max(d(A))$, as A ranges over all m by n matrices with $+1, -1$-entries. Since $d(m, n) = d(n, m)$, we can consider only the case $m \geq n$.

The problem of determining or estimating $d(m, n)$, and in particular $d(n, n)$, was raised by Paturi and Simon (1984). This problem was motivated by an attempt to estimate the maximum possible "unbounded-error probabilistic communication complexity" of Boolean functions. Alon, Frankl and Rödl (1985) proved that, as n grows, $n/32 \leq d(n, n) \leq (\frac{1}{2} + o(1))n$. Both the upper and the lower bounds are proved by combining probabilistic arguments with certain other ideas. In the next theorem we prove the upper bound, which is probably closer to the truth.

Theorem 14.3.1 *For all $m \geq n$,*

$$d(m, n) \leq (n + 1)/2 + \sqrt{\frac{n-1}{2} \log\ m}.$$

For the proof, we need a definition and two lemmas. For a vector $\mathbf{a} = (a_1, \ldots, a_n)$ of $+1, -1$-entries, the number of sign changes in $\mathbf{a}$ is the number of indices i, $1 \leq i \leq n-1$ such that $a_i = -a_{i+1}$. For a matrix A of $+1, -1$-entries, denote by $s(A)$ the maximum number of sign changes in a row of A.

Lemma 14.3.2 *For any matrix A of $+1, -1$-entries, $d(A) \leq s(A) + 1$.*

Proof. Let $A = (a_{i,j})$ be an $m \times n$ matrix of $+1, -1$ entries, and suppose $s = s(A)$. Let $t_1 < t_2 < \cdots < t_n$ be arbitrary reals, and define n points $P_1, P_2, \ldots, P_n$ in R^{s+1} by $P_j = (1, t_j, t_j^2, \ldots, t_j^s)$. These points, whose last s coordinates represent points on the d-dimensional moment-curve, will be the points used in the realization of A. To complete the proof we have to show that each row of A can be realized by a suitable hyperplane through the origin. This is proved by applying some of the known properties of the moment-curve as follows: Consider the sign vector representing an

arbitrary row of A. Suppose this vector has r sign changes, where, of course, $r \leq s$. Suppose the sign changes in this vector occur between the coordinates i_j and $i_j + 1$, for $1 \leq j \leq r$. Choose arbitrary reals $y_1, \ldots, y_r$, where $t_{i_j} < y_j < t_{i_j+1}$ for $1 \leq j \leq r$. Consider the polynomial $P(t) = \prod_{j=1}^{r}(t - y_j)$. Since its degree is at most s, there are real numbers a_j such that $P(t) = \sum_{j=0}^{s} a_j t^j$. Let H be the hyperplane in R^{s+1} defined by $H = \{(x_0, x_1, \ldots, x_s) \in R^{s+1} : \sum_{j=0}^{s} a_j x_j = 0\}$. Clearly, the point $P_j = (1, t_j, \ldots, t_j^s)$ is on the positive side of this hyperplane if $P(t_j) > 0$, and is on its negative side if $P(t_j) < 0$. Since the polynomial P changes sign only in the values y_j, it follows that the hyperplane H separates the points $P_1, \ldots, P_n$ according to the sign pattern of the corresponding row of A. Hence, by choosing the orientation of H appropriately, we conclude that A is realizable in R^{s+1}, completing the proof of the lemma. ■

Lemma 14.3.3 *For every $m \times n$ matrix A of $+1, -1$-entries, there is a matrix B obtained from A by multiplying some of the columns of A by -1, such that $s(B) \leq (n-1)/2 + \sqrt{\frac{n-1}{2} \log m}$.*

Proof. For each column of A, randomly and independently, choose a number $\epsilon \in \{+1, -1\}$, where each of the two choices is equally likely, and multiply this column by ϵ. Let B be the random sign-matrix obtained in this way. Consider an arbitrary fixed row of B. One can easily check that the random variable describing the number of sign changes in this row is a binomial random variable with parameters $n - 1$ and $p = 1/2$. This is because, no matter what the entries of A in this row are, the row of B is a totally random row of $-1, 1$ entries. By the standard estimates for binomial distributions, described in Appendix A, the probability that this number is greater than $(n-1)/2 + \sqrt{\frac{n-1}{2} \log m}$ is smaller than $1/m$. Therefore, with positive probability the number of sign changes in each of the m rows is at most that large, completing the proof. ■

Proof [Theorem 14.3.1] Let A be an arbitrary $m \times n$ matrix of $+1, -1$-entries. By Lemma 14.3.3, there is a matrix B obtained from A by replacing some of its columns by their inverses, such that $s(B) \leq (n-1)/2 + \sqrt{\frac{n-1}{2} \log m}$. Observe that $d(A) = d(B)$, since any realization of one of these matrices by points and hyperplanes through the origin gives a realization of the other one by replacing the points corresponding to the altered columns by their antipodal points. Therefore, by Lemma 14.3.2

$$d(A) = d(B) \leq s(B) + 1 \leq (n+1)/2 + \sqrt{\frac{n-1}{2} \log m}.$$

This completes the proof. ■

It is worth noting that by applying the (general) six-standard-deviations theorem stated at the end of Section 13.2, the estimate in Lemma 14.3.3 (and hence in Theorem 14.3.1) can be improved to $n/2 + O(\sqrt{n \log (m/n)})$. It can be also shown that if n and m grow so that m/n^2 tends to infinity and $(\log_2 m)/n$ tends to 0, then for almost all $m \times n$ matrices A of $+1, -1$-entries $d(A) = (\frac{1}{2} + o(1))n$.

14.4 ϵ-NETS AND VC-DIMENSIONS OF RANGE SPACES

What is the minimum number $f = f(n, \epsilon)$ such that every set X of n points in the plane contains a subset S of at most f points such that every triangle containing at least ϵn points of X contains at least one point of S? As we shall see in this section, there is an absolute constant c such that $f(n, \epsilon) \leq \frac{c}{\epsilon} \log(1/\epsilon)$, and this estimate holds for every n. This somewhat surprising result is a very special case of a general theorem of Vapnik and Chervonenkis (1971), which has been extended by Haussler and Welzl (1987), and which has many interesting applications in computational geometry and in statistics. In order to describe this result, we need a few definitions. A *range space* S is a pair (X, R), where X is a (finite or infinite) set and R is a (finite or infinite) family of subsets of X. The members of X are called *points* and those of R are called *ranges*. If A is a subset of X, then $P_R(A) = \{r \cap A : r \in R\}$ is the *projection* of R on A. In case this projection contains all subsets of A, we say that A is shattered. The Vapnik–Chervonenkis dimension (or *VC-dimension*) of S, denoted by $\mathrm{VC}(S)$, is the maximum cardinality of a shattered subset of X. If there are arbitrarily large shattered subsets, then $\mathrm{VC}(S) = \infty$.

The number of ranges in any finite range space with a given number of points and a given VC-dimension cannot be too large. For integers $n \geq 0$ and $d \geq 0$, define a function $g(d, n)$ by

$$g(d, n) = \sum_{i=0}^{d} \binom{n}{i}.$$

Observe that for all $n, d \geq 1$, $g(d, n) = g(d, n-1) + g(d-1, n-1)$. The following combinatorial lemma was proved, independently, by Sauer (1972) and Perles and Shelah, and in a slightly weaker form by Vapnik and Chervonenkis (1971).

Lemma 14.4.1 *If (X, R) is a range space of VC-dimension d with $|X| = n$ points, then $|R| \leq g(d, n)$.*

Proof. We apply induction on $n + d$. The assertion is trivially true for $d = 0$ and $n = 0$. Assuming it holds for n and $d - 1$ and for $n - 1$ and $d - 1$, we prove it for n and d. Let $S = (X, R)$ be a range space of VC-dimension d on n points. Suppose $x \in X$, and consider the two range spaces $S - x$ and $S \setminus x$ defined as follows. $S - x = (X - \{x\}, R - x)$, where $R - x = \{r - \{x\} : r \in R\}$. $S \setminus x = (X - \{x\}, R \setminus x)$, where $R \setminus x = \{r \in R : x \notin r, r \cup \{x\} \in R\}$. Clearly, the VC-dimension of $S - x$ is at most d. It is also easy to see that the VC-dimension of $S \setminus x$ is at most $d - 1$. Therefore, by the induction hypothesis

$$|R| = |R - x| + |R \setminus x| \leq g(d, n-1) + g(d-1, n-1) = g(d, n),$$

completing the proof. ∎

It is easy to check that the estimate given in the above lemma is sharp for all possible values of n and d. If (X, R) is a range space of VC-dimension d and $A \subset X$,

then the VC-dimension of $(A, P_R(A))$ is clearly at most d. Therefore, the last lemma implies the following:

Corollary 14.4.2 *If (X, R) is a range space of VC-dimension d, then for every finite subset A of X, $|P_R(A)| \leq g(d, |A|)$.*

There are many range spaces with finite VC-dimension that arise naturally in discrete and computational geometry. One such example is the space $S = (R^d, H)$, whose points are all the points in the d-dimensional Euclidean space, and whose set of ranges is the set of all (open) half-spaces. Any set of $d + 1$ affinely independent points is shattered in this space, and, by Radon's theorem, no set of $d + 2$ points is shattered. Therefore $\mathrm{VC}(S) = d + 1$. As shown by Dudley (1978), if (X, R) has a finite VC-dimension, so does (X, R_k), where R_k is the set of all Boolean combinations formed from at most k ranges in R. In particular, the following statement is a simple consequence of Corollary 14.4.2:

Corollary 14.4.3 *Let (X, R) be a range space of VC-dimension $d \geq 2$, and let (X, R_h) be the range space on X in which $R_h = \{(r_1 \cap \cdots \cap r_h) : r_1, \ldots, r_h \in R\}$. Then $\mathrm{VC}(X, R_h) \leq 2dh \log(dh)$.*

Proof. Let A be an arbitrary subset of cardinality n of X. By Corollary 14.4.2, $|P_R(A)| \leq g(d, n) \leq n^d$. Since each member of $P_{R_h}(A)$ is an intersection of h members of $P_R(A)$, it follows that $|P_{R_h}(A)| \leq \binom{g(d,n)}{h} \leq n^{dh}$. Therefore, if $n^{dh} < 2^n$, then A cannot be shattered. But this inequality holds for $n \geq 2dh \log(dh)$, since $dh \geq 4$. ■

As shown above, the range space whose set of points is R^d and whose set of ranges is the set of all half spaces has VC-dimension $d + 1$. This and the last corollary imply that the range space (R^d, C_h), where C_h is the set of all convex d-polytopes with h facets, has a VC-dimension which does not exceed $2(d + 1)h \log((d + 1)h)$.

An interesting property of range spaces with a finite VC-dimension is the fact that each finite subset of such a set contains relatively small good samples in the sense described below. Let (X, R) be a range space, and let A be a finite subset of X. For $0 \leq \epsilon \leq 1$, a subset $B \subset A$ is an *ϵ-sample* for A if for any range $r \in R$ the inequality

$$\|A \cap r|/|A| - |B \cap r|/|B\| \leq \epsilon$$

holds. Similarly, a subset $N \subset A$ is an *ϵ-net* for A if any range $r \in R$ satisfying $|r \cap A| > \epsilon|A|$ contains at least one point of N.

Notice that every ϵ-sample for A is also an ϵ-net and that the converse is not true. However, both notions define subsets of A, which represent approximately some of the behavior of A with respect to the ranges. Our objective is to show the existence of small ϵ-nets or ϵ-samples for finite sets in some range spaces. Observe that, if (X, R) is a range space with an infinite VC-dimension, then for every n there is a shattered subset A of X of cardinality n. It is obvious that any ϵ-net (and hence certainly any

ϵ-sample) for such an A must contain at least $(1-\epsilon)n$ points, that is, it must contain almost all points of A. Therefore, in infinite VC-dimension, there are no small nets or samples. However, it turns out that in finite VC-dimension there are always very small nets and samples. The following theorem was proved by Vapnik and Chervonenkis (1971).

Theorem 14.4.4 *There is a positive constant c such that, if (X, R) is any range space of VC-dimension at most d, $A \subset X$ is a finite subset and $\epsilon, \delta > 0$, then a random subset B of cardinality s of A where s is at least the minimum between $|A|$ and*

$$\frac{c}{\epsilon^2}\left(d \log \frac{d}{\epsilon} + \log \frac{1}{\delta}\right)$$

is an ϵ-sample for A with probability at least $1-\delta$.

Using similar ideas, Haussler and Welzl (1987) proved the following Theorem:

Theorem 14.4.5 *Let (X, R) be a range space of VC-dimension d, let A be a finite subset of X, and suppose $0 < \epsilon, \delta < 1$. Let N be a set obtained by m random independent draws from A, where*

$$m \geq \max\left(\frac{4}{\epsilon}\log\frac{2}{\delta}, \frac{8d}{\epsilon}\log\frac{8d}{\epsilon}\right). \tag{14.2}$$

Then N is an ϵ-net for A with probability at least $1-\delta$.

Therefore, if A is a finite subset of a range space of finite VC-dimension d, then for any $\epsilon > 0$, A contains ϵ-nets as well as ϵ-samples whose size is at most some function of ϵ and d, independent of the cardinality of A! The result about the triangles mentioned in the first paragraph of this section thus follows from Theorem 14.4.5, together with the observation following Corollary 14.4.3 which implies that the range space whose ranges are all triangles in the plane has a finite VC-dimension. We note that, as shown by Pach and Woeginger (1990), there are cases in which for fixed δ, the dependence of m in $1/\epsilon$ cannot be linear. See also Komlós, Pach and Woeginger (1992) for a tight form of the last theorem.

The proofs of Theorems 14.4.4 and 14.4.5 are very similar. Since the computation in the proof of Theorem 14.4.5 is simpler, we describe here only the proof of this theorem, and encourage the reader to try and make the required modifications that yield a proof for Theorem 14.4.4.

Proof [Theorem 14.4.5] Let (X, R) be a range space with VC-dimension d, and let A be a subset of X of cardinality $|A| = n$. Suppose m satisfies (14.2), and let $N = (x_1, \ldots, x_m)$ be obtained by m independent random choices of elements of A. (The elements in N are not necessarily distinct, of course). Let E_1 be the following event:

$$E_1 = \{\exists r \in R : |r \cap A| > \epsilon n, r \cap N = \emptyset\}\ .$$

To complete the proof we must show that the probability of E_1 is at most δ. To this end, we make an additional random choice and define another event as follows: Independently of our previous choice, we let $T = (y_1, \ldots, y_m)$ be obtained by m independent random choices of elements of A. Let E_2 be the event defined by

$$E_2 = \left\{ \exists r \in R : |r \cap A| > \epsilon n, r \cap N = \emptyset, |r \cap T| \geq \frac{\epsilon m}{2} \right\} .$$

(Since the elements of T are not necessarily distinct, the notation $|r \cap T|$ means here $|\{i : 1 \leq i \leq m, y_i \in r\}|$. The quantities $|r \cap N|$ and $|r \cap (N \cup T)|$ are similarly defined).

Claim 14.4.6 $\Pr[E_2] \geq \frac{1}{2} \Pr[E_1]$.

Proof. It suffices to prove that the conditional probability $\Pr[E_2|E_1]$ is at least $1/2$. Suppose that the event E_1 occurs. Then there is an $r \in R$ such that $|r \cap A| > \epsilon n$ and $r \cap N = \emptyset$. The conditional probability above is clearly at least the probability that for this specific r, $|r \cap T| \geq \frac{\epsilon m}{2}$. However, $|r \cap T|$ is a binomial random variable with expectation pm, and variance $(1-p)pm \leq pm$, where $p = |r \cap A|/|A| \geq \epsilon$. Hence, by Chebyshev's inequality,

$$\Pr\left[|r \cap T| < \frac{\epsilon m}{2}\right] \leq \Pr\left[|r \cap T| < \frac{pm}{2}\right] \leq \frac{pm}{(pm/2)^2} \leq \frac{4}{\epsilon m} \leq \frac{1}{2} ,$$

where the last inequality follows from (14.2). Thus, the assertion of Claim 14.4.6 is correct. ■

Claim 14.4.7

$$\Pr[E_2] \leq g(d, 2m) 2^{-\frac{\epsilon m}{2}} .$$

Proof. The random choice of N and T can be described in the following way, which is equivalent to the previous one. First one chooses $N \cup T = (z_1, \ldots, z_{2m})$ by making $2m$ random independent choices of elements of A, and then one chooses randomly precisely m of the elements z_i to be the set N (the remaining elements z_j form the set T, of course). For each range $r \in R$ satisfying $|r \cap A| > \epsilon n$, let E_r be the event where $|r \cap T| > \frac{\epsilon m}{2}$ and $r \cap N = \emptyset$. A crucial fact is that, if $r, r' \in R$ are two ranges, $|r \cap A| > \epsilon n$ and $|r' \cap A| > \epsilon n$ and if $r \cap (N \cup T) = r' \cap (N \cup T)$, then the two events E_r and $E_{r'}$, when both are conditioned on the choice of $N \cup T$, are identical. This is because the occurrence of E_r depends only on the intersection $r \cap (N \cup T)$. Therefore, for any fixed choice of $N \cup T$, the number of distinct events E_r does not exceed the number of different sets in the projection $P_{N \cup T}(R)$. Since the VC-dimension of X is d, Corollary 14.4.2 implies that this number does not exceed $g(d, 2m)$.

Let us now estimate the probability of a fixed event of the form E_r, given the choice of $N \cup T$. This probability is at most

$$\Pr\left[r \cap N = \emptyset \mid |r \cap (N \cup T)| > \frac{\epsilon m}{2}\right] .$$

Define $s = |r \cap (N \cup T)|$. Since the choice of N among the elements of $N \cup T$ is independent of the choice of $N \cup T$, the last conditional probability is precisely

$$\frac{(2m-s)(2m-s-1)\cdots(m-s+1)}{2m(2m-1)\cdots(m+1)} = \frac{m(m-1)\cdots(m-s+1)}{2m(2m-1)\cdots(2m-s+1)} \leq 2^{-s} \leq 2^{-\epsilon m/2} .$$

Since there are at most $g(d, 2m)$ potential distinct events E_r, it follows that the probability that at least one of them occurs given the choice of $N \cup T$ is at most $g(d, 2m)2^{-\epsilon m/2}$. Since this estimate holds conditioned on every possible choice of $N \cup T$, it follows that the probability of the event E_2 is at most $g(d, 2m)2^{-\epsilon m/2}$. This establishes Claim 14.4.7. ■

By Claims 14.4.6 and 14.4.7, $\Pr[E_1] \leq 2g(d, 2m)2^{-\epsilon m/2}$. To complete the proof of the theorem it remains to show that, if m satisfies inequality (14.2), then

$$2g(d, 2m)2^{-\frac{\epsilon m}{2}} \leq \delta .$$

We describe the proof for $d \geq 2$. The computation for $d = 1$ is easier. Since $g(d, 2m) \leq (2m)^d$, it suffices to show that

$$2(2m)^d \leq \delta 2^{\frac{\epsilon m}{2}} ,$$

that is

$$\frac{\epsilon m}{2} \geq d \log (2m) + \log \frac{2}{\delta} .$$

From (14.2) it follows that

$$\frac{\epsilon m}{4} \geq \log \frac{2}{\delta} ,$$

and hence it suffices to show that

$$\frac{\epsilon m}{4} \geq d \log (2m) .$$

The validity of the last inequality for some value of m implies its validity for any bigger m, and hence it suffices to check that it is satisfied for $m = \frac{8d}{\epsilon} \log \frac{8d}{\epsilon}$, that is

$$2d \log \frac{8d}{\epsilon} \geq d \log \left(\frac{16d}{\epsilon} \log \frac{8d}{\epsilon} \right) .$$

The last inequality is equivalent to $\frac{4d}{\epsilon} \geq \log \frac{8d}{\epsilon}$, which is certainly true. This completes the proof of the theorem. ■

Theorems 14.4.4 and 14.4.5 have been used for constructing efficient data structures for various problems in computational geometry. A trivial example is just the observation that Theorem 14.4.4 implies the following: for every $\epsilon > 0$, there is a

constant $c = c(\epsilon)$ such that for every n and every set A of n points in the plane there is a data structure of size $c(\epsilon)$ that enables us to estimate, given any triangle in the plane, the number of points of A in this triangle up to an additive error of ϵn. This is done simply by storing the coordinates of a set of points that form an ϵ-sample for A considered as a subset of the range space whose ranges are all planar triangles. More sophisticated data structures whose construction relies on the above two theorems can be found in the paper by Haussler and Welzl (1987).

14.5 DUAL SHATTER FUNCTIONS AND DISCREPANCY

The *dual shatter function* h of a range space $S = (X, R)$ is the function h mapping integers to integers, defined by letting $h(g)$ denote the maximum, over all possible choices of g members of R, of the number of atoms in the Venn diagram of these members. It is not too difficult to prove that, if the VC-dimension of S is d, then $h(g) \leq O(g^{2^{d+1}-1})$, but in geometric applications it is usually better to bound this function directly.

In Matoušek, Welzl and Wernisch (1993) it is proved that, if the dual shatter function of a range space $S = (X, R)$ satisfies $h(g) \leq O(g^t)$, A is any set of n points in the range space, and $\mathcal{F}$ is the projection $P_R(A)$ of R on A, then the discrepancy of $\mathcal{F}$ satisfies

$$\mathrm{disc}(\mathcal{F}) \leq O(n^{\frac{1}{2}-\frac{1}{2t}}\sqrt{\log\ n}). \tag{14.3}$$

This supplies nontrivial estimates in various geometric situations, improving the trivial bound that follows from Theorem 13.1.1 of Chapter 13. In most of these geometric applications, it is widely believed that the $\sqrt{\log\ n}$ factor can be omitted. In the abstract setting, however, this factor cannot be omitted, as proved in Matoušek (1997) (for $t = 2, 3$) and later in Alon, Rónyai and Szabó (1999) for all t.

The proof of (14.3) is based on a beautiful result of Chazelle and Welzl (1989) and its improvement by Haussler (1995). It is somewhat simpler to prove the result with an extra logarithmic factor, and this is the proof we present here. See Pach and Agarwal (1995), for some additional information.

Let $\mathcal{F}$ be a family of subsets of a finite set A. In what follows, we consider graphs whose edges are (unordered) pairs of points of A. For $F \in \mathcal{F}$ and $x, y \in A$, the edge xy stabs F if F contains exactly one of the two points x and y. The following theorem is proved in Chazelle and Welzl (1989). An improvement by a logarithmic factor appears in Haussler (1995).

Theorem 14.5.1 *Let $(A, \mathcal{F})$ be a finite range space, where $|A| = n$, and suppose that its dual shatter function h satisfies $h(g) \leq cg^t$ for some fixed $c, t > 0$. Then, there is a $C = C(c, t)$ and a Hamilton path on A, such that each member F of $\mathcal{F}$ is stabbed by at most $Cn^{1-1/t}\log\ n$ edges of the path.*

To prove the above theorem, we need the following lemma.

Lemma 14.5.2 *Let $(A, \mathcal{F}), n, h, t$, and c be as above, let B be a finite subset of $p > 1$ points of A, and let $\mathcal{G}$ be a collection of m (not necessarily distinct) members of $\mathcal{F}$. Then there are two distinct points x, y in B, such that the edge xy stabs at most $\frac{bm \log p}{p^{1/t}}$ members of $\mathcal{G}$, where $b = b(c)$.*

Proof. We may and will assume that p is larger than $c + 1$. Let g be the largest integer such that $cg^t \leq p - 1$, that is, $g = \left\lfloor \left(\frac{p-1}{c} \right)^{1/t} \right\rfloor$. Let L be a random collection of g members of $\mathcal{G}$, each picked randomly and independently (with possible repetitions) from among all m members of $\mathcal{G}$ with uniform distribution. The Venn diagram of all members of L partitions B into at most $h(g) \leq cg^t < p$ atoms, and hence there are two distinct points x, y of B that lie in the same atom. To complete the proof, it suffices to show that with positive probability, for each pair of points of B that stabs more than $\frac{bm \log p}{p^{1/t}}$ members of $\mathcal{G}$, at least one of these members lies in L (and hence the pair does not lie in an atom of the corresponding Venn diagram.) There are $\binom{p}{2}$ such pairs, and for each of them the probability that L contains no member of $\mathcal{G}$ it stabs is at most

$$\left(1 - \frac{b \log p}{p^{1/t}}\right)^g \leq e^{-\frac{b \log p}{p^{1/t}} \left\lfloor \left(\frac{p-1}{c}\right)^{1/t} \right\rfloor},$$

which is less than $1/p^2$ for an appropriately chosen constant $b = b(c)$. This completes the proof. ■

Proof [Theorem 14.5.1] Note, first, that if d is the VC-dimension of the given space, then there is a shattered set D of size d. It is not difficult to see that there are $g = \lceil \log_2 d \rceil$ sets among those shattering D, so that no two points of D lie in the same atom of their Venn diagram. Therefore, $d \leq c(\lceil \log_2 d \rceil)^t$, implying that $d \leq 2^{c' t \log t}$, where $c' = c'(c)$. By Lemma 14.4.1, this implies that the total number of ranges in R is at most $n^{2^{c' t \log t}}$.

We next prove that there is a spanning tree of A satisfying the assertion of Theorem 14.5.1, and then show how to replace it by a Hamilton path. By Lemma 14.5.2 with $B_0 = A, p_0 = n$, and $\mathcal{G}_0 = \mathcal{F}, m_0 = |\mathcal{G}_0| \quad (\leq n^{2^{c' t \log t}})$, we conclude that there is a pair $x_0 y_0$ of points in A such that the edge $x_0 y_0$ does not stab more than $\frac{b \log n}{n^{1/t}} m_0$ members of $\mathcal{G}$. Let $\mathcal{G}_1$ be the collection obtained from $\mathcal{G}$ by duplicating all members of $\mathcal{G}$ that are stabbed by $x_0 y_0$, and define $B_1 = B - x_0, p_1 = n - 1$, $m_1 = |\mathcal{G}_1| \leq m_0(1 + \frac{b \log n}{n^{1/t}})$. Applying Lemma 14.5.2 again, this time to B_1 and $\mathcal{G}_1$, we obtain another pair $x_1 y_1$, define $B_2 = B_1 - x_1, p_2 = p_1 - 1 = n - 2$, and let $\mathcal{G}_2$ be the collection obtained from $\mathcal{G}_1$ by duplicating all members of $\mathcal{G}_1$ stabbed by $x_1 y_1$, $m_2 = |\mathcal{G}_2|$. By the assertion of the lemma, $m_2 \leq m_1(1 + \frac{b \log n}{(n-1)^{1/t}})$. Proceeding in this manner, we get a sequence $x_0 y_0, x_1 y_1, \ldots, x_{n-1} y_{n-1}$ of edges of a graph on A, a sequence of subsets $B_0 = A, B_1, \ldots, B_{n-1}$, where each B_i is obtained from the previous one by omitting the point x_{i-1}, and a sequence of collections $\mathcal{G}_0, \mathcal{G}_1, \ldots, \mathcal{G}_{n-1}$,

where

$$|\mathcal{G}_{n-1}| \le m_0 \prod_{i=0}^{n-1}(1+\frac{b\log\ n}{(n-i)^{1/t}})$$
$$\le n^{2^{c't\log\ t}} e^{b\log\ n\sum_{i=0}^{n-1}(n-i)^{-1/t}} \le 2^{b'n^{1-1/t}\log\ n}$$

for an appropriate $b' = b'(c,t)$.

Note, now, that the edges x_iy_i form a spanning tree on the set A. The crucial observation is the fact that, if a member of $\mathcal{F}$ is stabbed by s of the edges, then it is being duplicated s times during the above process that generates $\mathcal{G}_{n-1}$, implying that $2^s \le |\mathcal{G}_{n-1}|$ and hence that $s \le b'n^{1-1/t}\log\ n$. It remains to replace the spanning tree by a Hamiltonian path. To do so, replace each edge of the tree by two parallel edges, and take an Euler tour in the resulting graph (in which all degrees are even). This is a sequence $x_0, x_1, x_2, \ldots, x_{2n-2} = x_0$ of points of A such that each adjacent pair of elements of the sequence is an edge of the tree, and each edge appears twice this way. The subsequence of the above one obtained by keeping only the first appearance of each point of A is a Hamilton path, and it is easy to check that each member of $\mathcal{F}$ is stabbed by at most $2b'n^{1-1/t}\log\ n$ of its edges, completing the proof.

The following result is a simple consequence of Theorem 14.5.1. As mentioned above, its assertion can be improved by a factor of $\sqrt{\log\ n}$.

Theorem 14.5.3 *Let $(A,\mathcal{F})$ be a finite range space, where $|A| = n$, and suppose that its dual shatter function h satisfies $h(g) \le cg^t$ for some fixed $c, t > 0$. Then, there is a $C' = C'(c,t)$ such that the discrepancy of $\mathcal{F}$ satisfies*

$$\mathrm{disc}(\mathcal{F}) \le C'n^{\frac{1}{2}-\frac{1}{2t}}\log\ n.$$

Proof. Without loss of generality, assume that the number of points of A is even (otherwise, simply omit a point). By Theorem 14.5.1, there is a Hamiltonian path $x_1x_2\cdots x_n$ on these points such that each member of $\mathcal{F}$ is stabbed by at most $Cn^{1-1/t}\log\ n$ edges of the path. Let $f : A \to \{-1, 1\}$ be a random coloring of A, where for each i, $1 \le i \le n/2$, randomly and independently, either $f(x_{2i-1}) = 1, f(x_{2i}) = -1$ or $f(x_{2i-1}) = -1, f(x_{2i}) = 1$, the two choices being equally likely. Fix a member $F \in \mathcal{F}$, and note that the contribution of each pair $x_{2i-1}x_{2i}$ to the sum $\sum_{x_j\in F} f(x_j)$ is zero if the edge $x_{2i-1}x_{2i}$ does not stab F, and it is either $+1$ or -1 otherwise. It thus follows that this sum has, in the notation of Theorem A.1.1, the distribution S_r for some $r \le Cn^{1-1/t}\log\ n$. Thus, the probability that it is, in absolute value, at least α can be bounded by $2e^{-\alpha^2/2r}$. As shown in the first paragraph of the proof of Theorem 14.5.1, the total number of members of $\mathcal{F}$ does not exceed $n^{2^{c't\log\ t}}$, and thus the probability that there exists a member $F \in \mathcal{F}$ for which the sum $\sum_{x_j\in F} f(x_j)$ exceeds $C'n^{\frac{1}{2}-\frac{1}{2t}}\log\ n$ is less than 1 for an appropriately chosen constant $C' = C'(c,t)$. ■

The range space whose set of points is an arbitrary set of points in the plane, and whose ranges are all disks in the plane, has a dual shatter function $O(g^2)$. The above theorem thus shows that it is possible to color any set of n points in the plane red and blue, such that the absolute value of the difference between the number of red points and the number of blue points inside any disk would not exceed $n^{1/4+o(1)}$. Similar results can be proved for many other geometric range spaces.

14.6 EXERCISES

1. Let A be a set of n points in the plane, and let $\mathcal{F}$ be the set of all intersections of A with an open triangle in the plane. Prove that the discrepancy of $\mathcal{F}$ does not exceed $n^{1/4+o(1)}$.
2. Prove that n distinct points in the plane determine at most $O(n^{4/3})$ unit distances.

THE PROBABILISTIC LENS:
Efficient Packing

Let $C \subset R^n$ be bounded with Riemann measure $\mu = \mu(C) > 0$. Let $N(C, x)$ denote the maximal number of disjoint translates of C that may be packed in a cube of side x, and define the packing constant

$$\delta(C) = \mu(C) \lim_{x \to \infty} N(C, x) x^{-n} ,$$

the maximal proportion of space that may be packed by copies of C. The following result improves the one described in Section 3.4.

Theorem 1 *Let C be bounded, convex, and centrally symmetric about the origin. Then,*

$$\delta(C) \geq 2^{-(n-1)} .$$

Proof. Fix $\epsilon > 0$. Normalize so that $\mu = \mu(C) = 2 - \epsilon$. For any real z, let C_z denote the "slab" of $(z_1, \ldots, z_{n-1}) \in R^{n-1}$ such that $(z_1, \ldots, z_{n-1}, z) \in C$, and let $\mu(C_z)$ be the usual $n-1$-dimensional measure of C_z. Riemann measurability implies

$$\lim_{\gamma \to 0} \sum_{m \in Z} \mu(C_{m\gamma}) \gamma = \mu(C) .$$

Let K be an integer sufficiently large so that

$$\sum_{m \in Z} \mu(C_{mK^{-(n-1)}}) K^{-(n-1)} < 2$$

and further that all points of C have all coordinates less than $K/2$.

For $1 \le i \le n-1$, let $v_i \in R^n$ be that vector with all coordinates zero except K as the ith coordinate. Let

$$v = (z_1, \ldots, z_{n-1}, K^{-(n-1)}) ,$$

where $z_1, \ldots, z_{n-1}$, are chosen uniformly and independently from the real interval $[0, K)$. Let Λ_v denote the lattice generated by the v's, that is

$$\begin{aligned}\Lambda_v &= \{m_1 v_1 + \cdots + m_{n-1} v_{n-1} + mv : m_1, \ldots, m_{n-1}, m \in Z\} \\ &= \{(mz_1 + m_1 K, \ldots, mz_{n-1} + m_{n-1} K, mK^{-(n-1)} : m_1, \ldots, m_{n-1}, m \in Z\}.\end{aligned}$$

Let $\theta(x)$ denote the unique $x' \in (-\frac{K}{2}, \frac{K}{2}]$ so that $x - mK = x'$ for some $m \in Z$. For $m \in Z$, let A_m be the event where some $m_1 v_1 + \cdots + m_{n-1} v_{n-1} + mv \in C$. Since all coordinates of all points of C are less than $K/2$, A_m occurs if and only if

$$(\theta(mz_1), \ldots, \theta(mz_{n-1}), mK^{-(n-1)}) \in C ,$$

which occurs if and only if $(\theta(mz_1), \ldots, \theta(mz_{n-1})) \in C_{mK^{-(n-1)}}$. The independence and uniformity of the z_i over $[0, K)$ implies the independence and uniformity of the $\theta(z_i)$ over $(-\frac{K}{2}, \frac{K}{2}]$, and so

$$\Pr[A_m] = K^{-(n-1)} \mu(C_{mK^{-(n-1)}}) .$$

Summing over positive m and employing the central symmetry,

$$\sum_{m>0} \Pr[A_m] < \frac{1}{2} \sum_{m \in Z} K^{-(n-1)} \mu(C_{mK^{-(n-1)}}) < \frac{1}{2} 2 = 1.$$

Hence there *exists* v with all A_m, $m > 0$ not holding. By the central symmetry, A_m and A_{-m} are the same event so no A_m, $m \neq 0$ holds. When $m = 0$, the points $m_1 v_1 + \cdots + m_{n-1} v_{n-1} = K(m_1, \ldots, m_{n-1}, 0)$ all lie outside C except the origin. For this v

$$\Lambda_v \cap C = \{0\} .$$

Consider the set of translates $C + 2w$, $w \in \Lambda_v$. Suppose

$$z = c_1 + 2w_1 = c_2 + 2w_2 \text{ with } c_1, c_2 \in C, w_1, w_2 \in \Lambda_v .$$

Then $(c_1 - c_2)/2 = w_2 - w_1$. From convexity and central symmetry, $(c_1 - c_2)/2 \in C$. As $w_2 - w_1 \in \Lambda_v$, it is zero and hence $c_1 = c_2$ and $w_1 = w_2$. That is, the translates form a packing of R^n. As $\det(2\Lambda_v) = 2^n \det(\Lambda_v) = 2^n$, this packing has density $2^{-n}\mu = 2^{-n}(2 - \epsilon)$. As $\epsilon > 0$ was arbitrary, $\delta(C) \ge 2^{-(n-1)}$. ■

15

Codes, Games, and Entropy

Mathematics was a natural gift, like the northern lights. It was not mixed up with anything else in the world, not with papers, prizes, colleagues, and diplomas.

–from *Too Much Happiness*, by Alice Munro

15.1 CODES

Suppose we want to send a message, here considered a string of bits, across a noisy channel. There is a probability p that any bit sent will be received incorrectly. The value p is a parameter of the channel and cannot be changed. We assume that p is both the probability that a sent zero is received as a one and that a sent one is received as a zero. Sent bits are always received, but perhaps incorrectly. We further assume that the events where the bits are received incorrectly are mutually independent. The case $p = 0.1$ will provide a typical example.

How can we improve the reliability of the system? One simple way is to send each bit three times. When the three bits are received, we use majority rule to decode. The probability of incorrect decoding is then $3p^2(1-p) + p^3 = 0.028$ in our instance. We have sacrificed speed – the rate of transmission of this method is $1/3$ – and gained accuracy in return. If we send each bit five times and use majority rule to decode, the probability of incorrect decoding drops to 0.00856 but the rate of transmission also drops to $1/5$. Clearly, we may make the probability of incorrect decoding as low as needed, but seemingly with the tradeoff that the rate of transmission tends to zero.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

It is the fundamental theorem of information theory – due to Claude Shannon – that this tradeoff is not necessary: there are codes with rate of transmission approaching a positive constant (dependent on p) with probability of incorrect transmission approaching zero.

A *Coding Scheme* consists of positive integers m, n, a function $f : \{0,1\}^m \to \{0,1\}^n$ called the encoding function, and a function $g : \{0,1\}^n \to \{0,1\}^m$ called the decoding function. The notion is that a message (or segment of message) $x \in \{0,1\}^m$ will be encoded and sent as $f(x)$ and a received message $y \in \{0,1\}^n$ will be decoded as $g(y)$. The rate of transmission of such a scheme is defined as m/n. Let $E = (e_1, \ldots, e_n)$ be a random string defined by $\Pr[e_i = 1] = p$, $\Pr[e_i = 0] = 1 - p$, the values e_i being mutually independent. We define the probability of correct transmission as $\Pr[g(f(x) + E) = x]$. Here, x is assumed to be uniformly distributed over $\{0,1\}^m$ and independent of E, and $+$ is mod 2 vector addition.

A crucial role is played by the *entropy function*

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

defined for $p \in (0,1)$. For any fixed p, the entropy function appears in the asymptotic formula

$$\binom{n}{pn} = \frac{n^n e^{-n}}{(pn)^{pn} e^{-pn} ((1-p)n)^{(1-p)n} e^{-(1-p)n}} (1 + o(1))^n = 2^{n(H(p)+o(1))}.$$

For $p \in (0, 0.5)$, we further bound

$$\sum_{i \le pn} \binom{n}{i} \le (1 + pn) \binom{n}{pn} = 2^{n(H(p)+o(1))}.$$

Theorem 15.1.1 [Shannon's Theorem] *Let $p \in (0, 0.5)$ be fixed. For $\epsilon > 0$ arbitrarily small, there exists a coding Scheme with rate of transmission greater than $1 - H(p) - \epsilon$ and probability of incorrect transmission less than ϵ.*

Remark. It is not difficult to show that, for every such p, any coding scheme whose rate of transmission exceeds $1 - H(p) + \epsilon$ must have a significant error probability. Indeed, if $f(x)$, the image of x, is transmitted, then with high probability, the obtained output y is of distance $(1 + o(1))pn$ from $f(x)$. Hence, if there are 2^m input words, the total size of all typical outputs is about $2^m \cdot \binom{n}{pn} = 2^{m+(1+o(1))H(p)n}$. If this quantity is much bigger than 2^n, then there must be significant overlaps between the output sets of different input words, making the decoding likely to err.

Proof. Let $\delta > 0$ be such that $p + \delta < 0.5$, and $H(p + \delta) < H(p) + \epsilon/2$. For n large, set $m = n(1 - H(p) - \epsilon)$, guaranteeing the rate of transmission. Let $f : \{0,1\}^m \to \{0,1\}^n$ be a random function – each $f(x)$ uniformly and independently chosen. Given f, define the decoding function $g : \{0,1\}^n \to \{0,1\}^m$ by setting $g(y) = x$ if x is the unique vector in $\{0,1\}^m$ whose image, $f(x)$, is within $n(p + \delta)$ of y. We measure

distance by the Hamming metric ρ: $\rho(y, y')$ is the number of coordinates in which y, y' differ. If there is no such x, or more than one such x, then we shall consider the decoding to be incorrect.

There are two ways in which the decoding can be incorrect. Possibly, $f(x) + E$ is not within $n(p + \delta)$ of $f(x)$. The distance from $f(x) + E$ to $f(x)$ is simply the number of 1's in E which has binomial distribution $B(n, p)$ and so this occurs with probability $o(1)$, in fact, with exponentially small probability. The only other possibility is that there is some $x' \neq x$ with $f(x') \in S$ where S is the set of y' within $n(p + \delta)$ of $f(x) + E$. Conditioning on the values $f(x), E, f(x')$ is still uniformly distributed over $\{0, 1\}^n$ and hence this occurs with probability $|S|2^{-n}$ for any particular x' and thus with total probability at most

$$2^m |S| 2^{-n} < 2^{-n(\frac{\epsilon}{2} + o(1))} = o(1).$$

The total probability for incorrect decoding from both sources is thus $o(1)$ and, in fact, exponentially small. For n sufficiently large, this is less than ϵ.

The average over all choices of f, x of the probability of incorrect decoding is less than ϵ. Therefore there exists a specific f (hence a specific coding scheme) with probability of incorrect coding less than ϵ. ■

Shannon's theorem, dealing with the intensely practical subject of communications, puts the shortcomings of the probabilistic approach in sharp contrast. Where is the coding scheme? Supposing that a coding scheme may be found, how can encoding and decoding be rapidly processed? A group code is a coding scheme in which the map $f : \{0, 1\}^m \to \{0, 1\}^n$ is linear, that is, $f(0) = 0$ and $f(x + x') = f(x) + f(x')$, all calculations modulo 2. Group codes are of particular interest, in part because of the ease of encoding.

Theorem 15.1.2 *Let $p \in (0, 0.5)$ be fixed. For $\epsilon > 0$ arbitrarily small, there exists a group code with rate of transmission greater than $1 - H(p) - \epsilon$ and probability of incorrect transmission less than ϵ.*

Proof. For $1 \leq i \leq m$, let $u_i \in \{0, 1\}^m$ be that vector with a 1 in position i, all other entries being zero. Let $f(u_1), \ldots, f(u_m)$ be chosen randomly and independently, and then extend f by setting

$$f(\epsilon_1 u_1 + \ldots + \epsilon_m u_m) = \epsilon_1 f(u_1) + \ldots + \epsilon_m u_m.$$

We follow the proof of Shannon's theorem until bounding the probability that $f(x) + E$ lies within $n(p + \delta)$ of $f(x)$. Set $z = x - x' = \epsilon_1 u_1 + \ldots + \epsilon_m u_m$, again all modulo 2. As $x \neq x'$, $z \neq 0$. Reorder for convenience so that $\epsilon_m = 1$. By linearity, $f(z) = f(x) - f(x')$, so we bound $\Pr[f(z) \in S]$ where S is the set of vectors within $n(p + \delta)$ of E. Fixing E and the $f(u_i)$, $i < m$, $f(z)$ still has an additive term $f(u_m)$ that is uniform and independent. Hence $f(z)$ is distributed uniformly. Thus

$$\Pr[f(z) \in S] = |S| 2^{-n},$$

and the remainder of the proof is as in Shannon's theorem. ■

15.2 LIAR GAME

Paul is trying to find a number $x \in \{1, \ldots, n\}$ from a recalcitrant and mendacious Carole. He may ask q questions of the form "Is $x \in S$?," where S can be any subset of the possibilities. The questions are asked sequentially, and Paul's choice of his ith question can depend on previous responses. Carole is allowed to lie – but she can lie at most k times. For which n, q, k can Paul determine the number?

When $k = 0$, Paul can win exactly when $n \leq 2^q$. The values $n = 100, q = 10, k = 1$ make for an amusing parlor game. Carole is hardly a passive observer; she may play an adversary strategy. By that we mean she does not select an x in advance but answers consistently with at least one x. At the end of the game, if her answers were consistent with more than one x, then she has won. The game, called the (n, q, k)-Liar Game, is now a perfect information game with no hidden moves and no draws. Hence either Paul or Carole has a perfect winning strategy. But who?

We describe an equivalent game, the ChipLiar game. There is a board with positions $0, 1, \ldots, k$. There are n chips labeled $1, \ldots, n$, which are initially at position k. There are q rounds. On each round Paul selects a set S of the chips. Carole can *either* move every chip not in S one position to the left *or* move every chip in S one position to the left. (Here position $i - 1$ is one position to the left of position i. Chips moved one position to the left from position 0 are removed from the board.) At the end of the q rounds, Carole wins if there is more than one chip remaining on the board and Paul wins if there is one or zero chip remaining on the board. Basically, chip i at position j represents that the answer $x = i$ has already received $k - j$ lies; Paul selecting S represents his asking if $x \in S$; Carole moving the chips not in S represents a Yes answer, moving the chips in S represents a No answer. (In the ChipLiar game, Carole can remove all chips from the board, while in the Liar game Carole must play consistently with at least one x. But when Carole removes all chips from the board, she automatically has lost and hence this difference does not affect the determination of the winner.)

In the ChipLiar game, there is no reason to place all chips at position k at the start. More generally, for $x_0, \ldots, x_k \geq 0$, we define the $(x_0, \ldots, x_k), q$-ChipLiar Game to be the above q round game with initial position consisting of x_i chips at position i. This, in turn, corresponds to a Liar Game in which there are x_i possibilities for which Carole is constrained to lie at most i times.

Let us define $B(q, j)$ as the probability that in q flips of a fair coin there are at most j heads. Of course, we have the exact formula

$$B(q, j) = 2^{-q} \sum_{i=0}^{j} \binom{q}{i}.$$

Theorem 15.2.1 *If*

$$\sum_{i=0}^{k} x_i B(q, i) > 1,$$

then Carole wins the $(x_0, \ldots, x_k), q$*-ChipLiar Game.*

Corollary 15.2.2 *If*

$$n > \frac{2^q}{\sum_{i=0}^{k} \binom{q}{i}},$$

then Carole wins the (n, q, k)*-Liar Game.*

Proof [Theorem 15.2.1]. Fix a strategy for Paul. Now Carole plays randomly! That is, at each round, after Paul has selected a set S of chips, Carole flips a coin – if it comes up heads she moves every chip not in S one position to the left, and if it comes up tails she moves every chip in S one position to the left. For each chip c, let I_c be the indicator random variable for c remaining on the board at the end of the game. Set $X = \sum I_c$, the number of chips remaining on the board at the end of the game. Consider a single chip c. Each round Paul may have chosen $c \in S$ or $c \notin S$, but in either case c is moved to the left with probability $1/2$. Suppose c starts at position i. It remains on the board at the end of the game if and only if in the q rounds it has been moved to the left at most i times. Then $\mathrm{E}[I_c]$, the probability of this occurring, is precisely $B(q, i)$. By Linearity of Expectation, $\mathrm{E}[X] = \sum_{i=0}^{k} x_i B(q, i)$. The assumption of the theorem gives $\mathrm{E}[X] > 1$. But then $X > 1$ must occur with positive probability. That is, Carole must win with positive probability.

No strategy of Paul allows him to always win. But this is a perfect information game with no draws, so someone has a perfect strategy that always wins. That someone isn't Paul, so it must be Carole. ■

The above proof certainly illustrated the magical element of the probabilistic method. Carole has a winning strategy but what is it? The general notion of moving from a probabilistic existence proof to an explicit construction is called derandomization and will be dealt with in detail in the next chapter. Here we can give an explicit strategy. With l moves remaining in the game and y_i chips on position i, define the weight of the position as $\sum_i y_i B(l, i)$ – note this is $\mathrm{E}[Y]$ where Y is the number of chips that would remain on the board should Carole play the rest of the game randomly. Carole's explicit strategy is to always move so as to maximize the weight.

Consider any position with weight W and any move S by Paul. Let W^y, W^n be the new weights should Carole move all chips not in S or all chips in S, respectively. We claim $W = \frac{1}{2}(W^y + W^n)$. One argument is that, by linearity, this identity reduces to the case of one chip and it then follows from the identity $B(l, j) = \frac{1}{2}(B(l-1, j) + B(l-1, j-1))$. But we need not actually do any calculation. Carole's playing randomly can be thought of as first flipping a coin to decide on her first move and then playing randomly so that $\mathrm{E}[Y]$ is the average of the two conditional expectations.

At the start of the game, by assumption, the weight is bigger than 1. Carole's explicit strategy assures that the weight does not decrease so at the end of the game the weight is bigger than 1. But at the end of the game the weight is the number of chips remaining. Being bigger than 1, Carole has won the game.

The converse of the theorem, and even that of corollary, is false. Consider the Liar Game with $n = 5$, $q = 5$ questions, and $k = 1$ possible lie. In the ChipLiar version this is the $(0, 5)$, 5-ChipLiar game. Here $B(5, 1) = 6/32$ and $5(6/32) \leq 1$. Still, Carole

wins with perfect play. The problem is that Paul has no good first move. Suppose he selects two chips as S (asks "Is $x \leq 2$?" in the Liar Game). Then Carole moves the two chips one to the left (responds Yes) leaving the position $(2, 3)$ with four questions remaining. As $2B(4, 0) + 3B(4, 1) = 17/16 > 1$, Carole will now win. It is easy to check that all other moves of Paul fail. The difficulty here is that Paul was in a position with weight $W \leq 1$ but was unable to find a move such that $W^y \leq 1$ and $W^n \leq 1$.

15.3 TENURE GAME

Paul, Chair of Department, is trying to promote one of his faculty to tenure but standing in his way is a recalcitrant and mean-spirited Carole, the Provost. There are k pre-tenure levels, labeled $1, \ldots, k$, level 1 the highest, and a level 0 representing tenure. For our purposes, each faculty member is represented by a chip. The $(x_1, \ldots, x_k)$-Tenure Game begins with x_i chips at level i for $1 \leq i \leq k$ and no chips on level zero. Each year Paul presents a set S of chips to Carole. Carole may either

- promote all chips in S and fire the others *or*
- promote all chips not in S and fire those in S.

Promote, as used above, means to move from level i to level $i - 1$. Fired means just that: removing the chip from the game. If a chip reaches level 0, then Paul is the winner. The draconian promote or perish provision ensures that the game will end within k years with either Paul winning or Carole having successfully eliminated all chips.

Theorem 15.3.1 *If $\sum_i x_i 2^{-i} < 1$, then Carole wins the $(x_1, \ldots, x_k)$-Tenure Game.*

Proof. Fix a strategy for Paul. Now Carole plays randomly! That is, at each round, after Paul has selected a set S of chips, Carole flips a coin – if it comes up heads she moves every chip not in S one position to the left, and if it comes up tails she moves every chip in S one position to the left. For each chip c, let I_c be the indicator random variable for c reaching level 0. Set $X = \sum I_c$, the number of chips reaching level 0 at the end of the game. Consider a single chip c. Each round Paul may have chosen $c \in S$ or $c \notin S$, but in either case c is moved to the left with probability $1/2$. Suppose c starts at position i. It remains on the board at the end of the game if and only if the first i coin flips of Carole led to promotions for c. Then $\mathrm{E}[I_c]$, the probability of this occurring, is precisely 2^{-i}. By Linearity of Expectation, $\mathrm{E}[X] = \sum_{i=1}^{k} x_i 2^{-i}$. The assumption of the theorem gives $\mathrm{E}[X] < 1$. But then $X < 1$ must occur with positive probability. That is, Carole must win with positive probability.

No strategy of Paul allows him to always win. But this is a perfect information game with no draws, so someone has a perfect strategy that always wins. That someone isn't Paul, so it must be Carole. ■

As with the Liar Game, we may derandomize the above argument to give an explicit strategy for Carole. With y_i chips on position i, define the weight of the position as $\sum_i y_i 2^{-i}$ – note this is $\mathrm{E}[Y]$, where Y is the number of chips that would reach

level 0 should Carole play the rest of the game randomly. Carole's explicit strategy is to always move so as to minimize the weight. Consider any position with weight W and any move S by Paul. Let W^y, W^n be the new weights should Carole move all chips not in S or all chips in S, respectively. As in the Liar Game, $W = \frac{1}{2}(W^y + W^n)$. At the start of the game, by assumption, the weight is less than 1. Carole's explicit strategy ensures that the weight does not increase so at all times the weight is smaller than 1. A chip at level 0 would add one to the weight by itself, so that this never occurs and hence Carole wins.

In the Liar Game, the sufficient condition for Carole to win was not necessary because Paul did not always have an appropriately splitting move. Here, however, we have an amusing lemma.

Lemma 15.3.2 *If a set of chips has weight at least* 1*, it may be split into two parts, each of weight at least* $1/2$*.*

Proof. There must be two chips at some position i, otherwise the weight is less than 1. If there are two chips at position 1, simply split them. If there are two chips at position $i > 1$, glue them together and consider them as one superchip at position $i - 1$. Then the proof follows by induction on the number of chips. ■

Theorem 15.3.3 *If* $\sum x_i 2^{-i} \geq 1$*, then Paul wins the* $(x_1, \ldots, x_k)$*-Tenure Game.*

Proof. The initial weight is at least 1. Applying the lemma, Paul splits the chips into two parts, each of weight at least $1/2$, and sets S equal one of the parts. Carole moves all chips in one part one position to the left, doubling their weight, leaving a new position of weight at least 1. Thus the weight never goes below 1. Therefore the game cannot end with all chips having been removed (which would have weight zero) and so it must end with a win for Paul. ■

15.4 BALANCING VECTOR GAME

The balancing vector game is a perfect information game with two players, Pusher and Chooser. There is a parameter $n \geq 1$, and we shall be concerned with asymptotics in n. There are n rounds, each involving vectors in R^n. There is a position vector $P \in R^n$, initially set at 0. Each round has two parts. First, Pusher picks $v \in \{-1, +1\}^n$. Then Chooser either resets P to $P + v$ or to $P - v$. At the end of the nth round, the payoff to Pusher is $|P|_\infty$, the maximal absolute value of the coordinates of P. Let VAL(n) denote the value of this game to Pusher, that is, the maximum payoff Pusher can ensure when both players play optimally. Let S_n denote, as usual, the sum of n independent uniform $\{1, -1\}$ random variables.

Theorem 15.4.1 *If* $\Pr[|S_n| > \alpha] < n^{-1}$*, then* $\mathrm{VAL}(n) \leq \alpha$*.*

Proof. Consider the game a win for Pusher if the final $|P|_\infty > \alpha$. Suppose Chooser announces that she will flip a fair coin each round to determine whether to reset P as

$P + v$ or $P - v$. Let x_i be the ith coordinate for the final value of the position vector P. Let W_i be the event $|x_i| > \alpha$ and $W = \vee W_i$ so that W is the event of Pusher winning. Regardless of Pusher's strategy, x_i has distribution S_n so that

$$\Pr[W] \leq \sum_{i=1}^{n} \Pr[|S_n| > \alpha] < 1.$$

Pusher cannot always win, so Chooser always wins. ■

Corollary 15.4.2 $\mathrm{VAL}(n) = O(\sqrt{n \ln n})$.

To give a lower bound on $\mathrm{VAL}(n)$, one wants to find a strategy for Pusher that wins against any Chooser. It is not sufficient to find a strategy that does well against a randomly playing Chooser–the Chooser is an adversary. Still, the notion of a randomly playing Chooser motivates the following result.

Theorem 15.4.3 *If* $\Pr[|S_n| > \alpha] > cn^{-1/2}$, *where* c *is an absolute constant, then* $\mathrm{VAL}(n) > \alpha$.

Corollary 15.4.4 $\mathrm{VAL}(n) = \Omega(\sqrt{n \ln n})$ *and hence* $\mathrm{VAL}(n) = \Theta(\sqrt{n \ln n})$.

Proof [Theorem 15.4.3]. Define, for $x \in Z, 0 \leq i \leq n$,

$$w_i(x) = \Pr\left[|x + S_{n-i}| > \alpha\right].$$

For $P = (x_1, \ldots, x_n)$, set $w_i(P) = \sum_{1 \leq j \leq n} w_i(x_j)$. When P is the position vector at the end of the ith round, $w_i(P)$ may be interpreted as the expected number of coordinates with absolute value greater than α at the end of the game, assuming random play by Chooser. At the beginning of the game, $w_0(P) = w_0(0) > c\sqrt{n}$ by assumption. Given position P at the end of round i, Pusher's strategy will be to select $v \in \{-1, +1\}^n$ so that $w_{i+1}(P - v)$ and $w_{i+1}(P + v)$ are close together.

The distribution $x + S_{n-i}$ splits into $x + 1 + S_{n-i-1}$ and $x - 1 + S_{n-i-1}$ depending on the first coin flip so that for any i, x,

$$w_i(x) = \frac{1}{2}\left[w_{i+1}(x+1) + w_{i+1}(x-1)\right].$$

Set $P = (x_1, \ldots, x_n)$, $v = (v_1, \ldots, v_n)$. For $1 \leq j \leq n$, set

$$\Delta_j = w_{i+1}(x_j + 1) - w_{i+1}(x_j - 1)$$

so that

$$w_{i+1}(P + v) - w_{i+1}(P - v) = \sum_{j=1}^{n} v_j \Delta_j,$$

and, for $\epsilon = \pm 1$,

$$w_{i+1}(P + \epsilon v) = w_i(P) + \frac{1}{2}\epsilon \sum_{j=1}^{n} v_j \Delta_j.$$

Now we bound $|\Delta_j|$. Observe that

$$\Delta_j = \Pr[S_{n-i-1} = y] - \Pr[S_{n-i-1} = z],$$

where y is the unique integer of the same parity as $n - i - 1$ in the interval $(\alpha - (x_j + 1), \alpha - (x_j - 1)]$ and z the same in $(-\alpha - (x_j + 1), -\alpha - (x_j - 1))$. Let us set

$$g(m) = \max_s \Pr[S_m = s] = \binom{m}{\lfloor m/2 \rfloor} 2^{-m} \sim \sqrt{\frac{2}{\pi m}},$$

so that $|\Delta_j| \le g(n - i - 1)$ for all j.

A simple strategy for Pusher is then to reorder the coordinates so that $|\Delta_1| \ge \cdots \ge |\Delta_n|$ and then select $v_1, \ldots, v_n \in \{-1, +1\}$ sequentially, giving $v_i \Delta_i$ the opposite sign of $v_1\Delta_1 + \ldots + v_{i-1}\Delta_{i-1}$. (When $i = 1$ or the sum is zero, choose v_i arbitrarily.) This ensures

$$|v_1\Delta_1 + \ldots + v_n\Delta_n| \le |\Delta_1| \le g(n - i - 1) .$$

Let P^i denote the position vector at the end of the ith round and v Pusher's choice for the $(i + 1)$th round. Then regardless of Chooser's choice of $\epsilon = \pm 1$,

$$w_{i+1}(P^{i+1}) = w_{i+1}(P^i + \epsilon v) \ge w_i(P^i) - \frac{1}{2}\left|\sum_{j=1}^{n} v_j \Delta_j\right| \ge w_i(P^i) - \frac{1}{2} g(n - i - 1).$$

Thus

$$w_n(P^n) \ge w_0(P^0) - \frac{1}{2}\sum_{i=0}^{n-1} g(n - i - 1).$$

Simple asymptotics give that the above sum is asymptotic to $(8n/\pi)^{1/2}$. Choosing $c > (2/\pi)^{1/2}$, $w_n(P^n) > 0$. But $w_n(P^n)$ is simply the *number* of coordinates with absolute value greater than α in the final $P = P^n$. This Pusher strategy ensures that there is more than zero, hence at least one such coordinate, and therefore Pusher wins. ■

15.5 NONADAPTIVE ALGORITHMS

Let us modify the balancing game of Section 15.4 by requiring the vectors selected by Pusher to have coordinates zero and 1 rather than plus and minus 1. Let $\mathrm{VAL}^*(n)$ denote the value of the modified game. One can use the bounds on $\mathrm{VAL}(n)$ to show $\mathrm{VAL}^*(n) = \Theta(\sqrt{n \ln n})$.

In Chapter 13 we showed that any family of n sets $S_1, \ldots, S_n$ on n points $1, \ldots, n$ has discrepancy $O(\sqrt{n})$; that is, there is a coloring $\chi : \{1, \ldots, n\} \to \{-1, +1\}$ so that all $|\chi(S_i)| \leq c\sqrt{n}$. The proof of this result does not yield an effective algorithm for finding such a coloring, and indeed it is not known if there is a polynomial time algorithm to do so. Suppose one asks for a *nonadaptive* or *online* algorithm in the following sense. Instead of being presented the entire data of $S_1, \ldots, S_n$ at once, one is presented with the points sequentially. At the jth "round", the algorithm looks at point j – more specifically, at which sets S_i contain j or, equivalently, at the jth column of the incidence matrix. At that stage, the algorithm must decide how to color j, and, once colored, the coloring cannot be changed. How small can we assure $\max|\chi(S_i)|$ with such an algorithm? We may think of the points as being presented by an adversary. Thinking of the points as their associated column vectors, Pusher as the Worst Case adversary and Chooser as the algorithm, the best such an algorithm can do is precisely $\mathrm{VAL}^*(n)$.

The requirement that an algorithm be nonadaptive is both stronger and weaker than the requirement that an algorithm take polynomial time. Still, this lends support to the conjecture that there is no polynomial time algorithm for finding a coloring with all $|\chi(S_i)| \leq c\sqrt{n}$.

15.6 HALF LIAR GAME

We modify the Liar Game of Section 15.2 by limiting Carole's mendacity. If the correct answer is Yes, then Carole is now required to answer Yes. She may answer Yes when the correct answer is No, and that would count as one of her k lies. Let $A_k(q)$ denote the maximal n for which Paul wins the Half Liar game with n values, q queries, and a maximum of k lies.

Theorem 15.6.1 [Dumitriu and Spencer (2004)] *For each fixed $k \geq 1$,*

$$A_k(q) \sim \frac{2^{q+k}}{\binom{q}{k}}.$$

While the methods below extend to arbitrary k, we give the proof only for the case $k = 1$. This case was first given by Cicalese and Mundici (2000). Let us fix a winning strategy for Paul with $n = A_1(q)$. This may be described by a binary decision tree of depth q. For each value i, $1 \leq i \leq n$, let $\sigma_i = (x_{i1}, \ldots, x_{iq}) \in \{Y, N\}^q$ be the string of truthful responses to Paul's queries with that value. Let S_i be the set of possible response strings given by Carole with that value. For each $x_{ij} = N$, Carole may lie on the jth question, otherwise responding truthfully. Thus $|S_i| = W(\sigma_i)$, where we define the weight $W(\sigma)$ to be 1 plus the number of N's in the string σ. We cannot have any common $\sigma \in S_i \cap S_j$ because then Carole could respond with σ and Paul would not

be able to distinguish i, j. Thus

$$\sum_{i=1}^{n} W(\sigma_i) \leq 2^q.$$

For a given u, call σ Carole-friendly if $W(\sigma) \leq 1 + \frac{q-u}{2}$, otherwise Paul-friendly. There are at most $2^q \Pr[S_q \leq -u]$ Carole-friendly σ. From the inequality, there are at most $2^{q+1}(q-u)^{-1}$ Paul-friendly σ. Thus

$$n \leq 2^q \Pr[S_q \leq -u] + \frac{2^{q+1}}{q-u}.$$

The optimization of u is left as an exercise, but even taking a suboptimal $u = \lfloor q^{2/3} \rfloor$ gives

$$A_1(q) = n \leq (1 + o(1)) \frac{2^{q+1}}{q}.$$

For larger n, Paul cannot have a winning strategy and thus Carole must have a winning adversary strategy. Intriguingly, this argument does not yield an explicit strategy for Carole.

In the other direction, let $\epsilon > 0$ be fixed and small and set $n = \lfloor (1-\epsilon) 2^{q+1} q^{-1} \rfloor$. We will give a strategy for Paul. For $r \geq 1$, let M_r denote those $\sigma \in \{Y, N\}^r$ with at least $\frac{r-u}{2}$ N's. For definiteness, take $u = \lfloor r^{2/3} \rfloor$. Then $f(r) \sim 2^r$. We first massage n. Pick r with (say) $10\epsilon^{-1} \leq n/f(r) \leq 21\epsilon^{-1}$, set $A = \lceil n/f(r) \rceil$, and boost n to $n = Af(r)$. As the boost (which makes things only harder for Paul) was by a factor less than $1 + (10/\epsilon)^{-1}$, the new n still has $n \leq (1 - \frac{\epsilon}{2}) 2^{q+1} q^{-1}$.

Paul associates the $n = f(r)A$ values with pairs (σ, j), $\sigma \in M_r$, $1 \leq j \leq A$. For his first r queries, he asks for the coordinates of σ. Carole responds τ, which can differ from the truthful σ in at most one coordinate. Thus τ has at most $\frac{r+u}{2} + 1$ Y's. (Basically, these r queries are nearly even splits and force Carole to answer No nearly half the time.) What does Paul know at this moment? If Carole has not lied, the answer must be one of the A values (τ, j). If Carole has lied, the answer must be one of the at most $A(\frac{r+u+2}{2})$ values (τ^+, j), where τ^+ is derived from τ by shifting a Y to an N.

Set $s = q - r$, the number of remaining queries. As A is bounded and $2^r A \sim f(r)A = n = \Theta(2^q q^{-1})$ we have $r = q - \log_2 q - O(1)$. In particular, $r \sim q$, the first r queries were the preponderance of the queries. Then

$$A \leq (1 + o(1)) n 2^{-r} \leq (1 - \frac{\epsilon}{2} + o(1)) 2^{q+1-r} r$$

and

$$A \frac{r+u+2}{2} \sim A \frac{r}{2} \leq (1 - \frac{\epsilon}{2} + o(1)) 2^s.$$

Paul may now give further ground and allows Carole to lie in either direction for the remaining s questions. This is the $(x_0, x_1), s$-ChipLiar game with $x_0 = A$ and

$x_1 \leq (1 - \frac{\epsilon}{2} + o(1))2^s$. The endgame strategy required at this point is given in the exercises.

15.7 ENTROPY

Let X be a random variable taking values in some range S, and let $P(X = x)$ denote the probability that the value of X is x. The *binary entropy* of X, denoted by $H(X)$, is defined by

$$H(X) = \sum_{x \in S} P(X = x) \log_2 \left(\frac{1}{P(X = x)} \right).$$

If Y is another random variable, taking values in T, and (X, Y) is the random variable taking values in $S \times T$ according to the joint distribution of X and Y, then the *conditional entropy* of X given Y is

$$H(X|Y) = H(X, Y) - H(Y).$$

In this section we prove some simple properties of the entropy function, and describe several surprising combinatorial and geometric applications. Intuitively, the entropy of a random variable measures the amount of information it encodes. This provides an intuitive explanation to the four parts of the next simple lemma. The formal proof, given below, uses the properties of the functions $\log z$ and $z \log z$, where – here and in the rest of this section – all logarithms are to the base 2.

Lemma 15.7.1 *Let X, Y, and Z be three random variables taking values in S, T, and U, respectively. Then*

(i) $H(X) \leq \log_2 |S|$.
(ii) $H(X, Y) \geq H(X)$.
(iii) $H(X, Y) \leq H(X) + H(Y)$.
(iv) $H(X|Y, Z) \leq H(X|Y)$.

Proof.
(i) Since the function $\log z$ is concave, it follows, by Jensen's Inequality, that

$$H(X) = \sum_{i \in S} P(X = i) \log \left(\frac{1}{P(X = i)} \right)$$
$$\leq \log \left(\sum_{i \in S} P(X = i) \frac{1}{P(X = i)} \right) = \log |S|.$$

(ii) By the monotonicity of $\log z$ for all $z > 0$,

$$H(X, Y) = \sum_{i \in S} \sum_{j \in T} P(X = i, Y = j) \log \left(\frac{1}{P(X = i, Y = j)} \right)$$

$$\geq \sum_{i\in S}\sum_{j\in T} P(X=i,Y=j)\log\left(\frac{1}{P(X=i)}\right)$$
$$= \sum_{i\in S} P(X=i)\log\left(\frac{1}{P(X=i)}\right) = H(X).$$

(iii) By definition

$$H(X)+H(Y)-H(X,Y)$$
$$= \sum_{i\in S}\sum_{j\in T} P(X=i,Y=j)\log\left(\frac{P(X=i,Y=j)}{P(X=i)P(Y=j)}\right)$$
$$= \sum_{i\in S}\sum_{j\in T} P(X=i)P(Y=j)f(z_{ij})\ ,$$

where $f(z)=z\log z$ and $z_{ij}=\dfrac{P(X=i,Y=j)}{P(X=i)P(Y=j)}$. Since $f(z)$ is convex, it follows, by Jensen's Inequality, that the last quantity is at least

$$f\left(\sum_{i\in S}\sum_{j\in T} P(X=i)P(Y=j)z_{ij}\right) = f(1) = 0.$$

(iv) Note that

$$H(X|Y) = H(X,Y)-H(Y)$$
$$= \sum_{i\in S}\sum_{j\in T} P(X=i,Y=j)\log\left(\frac{P(Y=j)}{P(X=i,Y=j)}\right).$$

Similarly

$$H(X|Y,Z)$$
$$= \sum_{i\in S}\sum_{j\in T}\sum_{k\in U} P(X=i,Y=j,Z=k)\log\left(\frac{P(Y=j,Z=k)}{P(X=i,Y=j,Z=k)}\right).$$

Therefore,

$$H(X|Y)-H(X|Y,Z)$$
$$= \sum_{i\in S}\sum_{j\in T}\sum_{k\in U} P(X=i,Y=j,Z=k)$$
$$\cdot\log\left(\frac{P(Y=j)P(X=i,Y=j,Z=k)}{P(X=i,Y=j)P(Y=j,Z=k)}\right)$$
$$= \sum_{i\in S}\sum_{j\in T}\sum_{k\in U}\frac{P(X=i,Y=j)P(Y=j,Z=k)}{P(Y=j)}f(z_{ijk}),$$

where $f(z) = z \log z$ and

$$z_{ijk} = \frac{P(Y=j)P(X=i, Y=j, Z=k)}{P(X=i, Y=j)P(Y=j, Z=k)}.$$

By the convexity of $f(z)$, and since

$$\sum_{i\in S}\sum_{j\in T}\sum_{k\in U} \frac{P(X=i, Y=j)P(Y=j, Z=k)}{P(Y=j)} = 1\ ,$$

it follows that the above quantity is at least

$$f\left(\sum_{i\in S}\sum_{j\in T}\sum_{k\in U} \frac{P(X=i, Y=j)P(Y=j, Z=k)}{P(Y=j)} z_{ijk}\right) = f(1) = 0.$$

■

The following simple but useful fact that the entropy is subadditive has already been applied in Section 13.2.

Proposition 15.7.2 *Let $X = (X_1, \ldots, X_n)$ be a random variable taking values in the set $S = S_1 \times S_2 \times \ldots \times S_n$, where each of the coordinates X_i of X is a random variable taking values in S_i. Then*

$$H(X) \leq \sum_{i=1}^{n} H(X_i).$$

Proof. This follows by induction from Lemma 15.7.1, part (iii). ■

The above proposition is used in Kleitman, Shearer and Sturtevant (1981) to derive several interesting applications in extremal finite set theory, including an upper estimate for the maximum possible cardinality of a family of k-sets in which the intersection of no two is contained in a third. The basic idea in Kleitman et al. (1981) can be illustrated by the following very simple corollary of the last proposition.

Corollary 15.7.3 *Let $\mathcal{F}$ be a family of subsets of $\{1, 2, \ldots, n\}$, and let p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then*

$$|\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)}\ ,$$

where $H(y) = -y\log_2 y - (1-y)\log_2(1-y)$.

Proof. Associate each set $F \in \mathcal{F}$ with its characteristic vector $v(F)$, which is a binary vector of length n. Let $X = (X_1, \ldots, X_n)$ be the random variable taking values in $\{0,1\}^n$, where $P(X = v(F)) = 1/|\mathcal{F}|$ for all $F \in \mathcal{F}$. Clearly, $H(X) = |\mathcal{F}|(\frac{1}{|\mathcal{F}|}\log |\mathcal{F}|) = \log |\mathcal{F}|$, and since here $H(X_i) = H(p_i)$ for all $1 \leq i \leq n$, the result follows from Proposition 15.7.2. ■

The following interesting extension of Proposition 15.7.2 has been proved by Shearer; see Chung et al. (1986). As in that proposition, let $X = (X_1, \ldots, X_n)$ be a

random variable taking values in the set $S = S_1 \times S_2 \times \ldots \times S_n$, where each X_i is a random variable taking values in S_i. For a subset I of $\{1, 2, \ldots, n\}$, let $X(I)$ denote the random variable $(X_i)_{i \in I}$.

Proposition 15.7.4 *Let $X = (X_1, \ldots, X_n)$ and S be as above. If $\mathcal{G}$ is a family of subsets of $\{1, \ldots, n\}$ and each $i \in \{1, \ldots, n\}$ belongs to at least k members of $\mathcal{G}$, then*

$$kH(X) \leq \sum_{G \in \mathcal{G}} H(X(G)).$$

Proof. We apply induction on k. For $k = 1$, replace each set $G \in \mathcal{G}$ by a subset of it to obtain a family $\mathcal{G}'$ whose members form a partition of $\{1, \ldots, n\}$. By Lemma 15.7.1, part (ii), $\sum_{G \in \mathcal{G}} H(X(G)) \geq \sum_{G' \in \mathcal{G}'} H(X(G'))$, and by Lemma 15.7.1, part (iii), $\sum_{G' \in \mathcal{G}'} H(X(G')) \geq H(X)$, supplying the desired result for $k = 1$.

Assuming the result holds for $k - 1$, we prove it for $k \quad (\geq 2)$. If there is a $G \in \mathcal{G}$ with $G = \{1, \ldots, n\}$, the result follows from the induction hypothesis. Otherwise, let G_1, G_2 be two members of $\mathcal{G}$. By applying part (iv) of Lemma 15.7.1, we conclude that

$$H(X(G_1 \setminus G_2) | X(G_1 \cap G_2), X(G_2 \setminus G_1)) \leq H(X(G_1 \setminus G_2) | X(G_1 \cap G_2)),$$

implying that

$$H(X(G_1 \cup G_2)) - H(X(G_2)) \leq H(X(G_1)) - H(X(G_1 \cap G_2)).$$

Therefore, $H(X(G_1 \cup G_2)) + H(X(G_1 \cap G_2)) \leq H(X(G_1)) + H(X(G_2))$. It follows that, if we modify $\mathcal{G}$ by replacing G_1 and G_2 by their union and intersection, then the sum $\sum_{G \in \mathcal{G}} H(X(G))$ can only decrease. After a finite number of such modifications, we can reach the case in which one of the sets in $\mathcal{G}$ is $\{1, \ldots, n\}$, and as this case has already been proved, this completes the proof. ■

Corollary 15.7.5 *Let $\mathcal{F}$ be a family of vectors in $S_1 \times S_2 \ldots \times S_n$. Let $\mathcal{G} = \{G_1, G_2, \ldots, G_m\}$ be a collection of subsets of $N = \{1, 2, \ldots, n\}$, and suppose that each element $i \in N$ belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$, let $\mathcal{F}_i$ be the set of all projections of the members of $\mathcal{F}$ on G_i. Then*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

Proof. Let $X = (X_1, \ldots, X_n)$ be the random variable taking values in $\mathcal{F}$, where $P(X = F) = \frac{1}{|\mathcal{F}|}$ for all $F \in \mathcal{F}$. By Proposition 15.7.4

$$kH(X) \leq \sum_{i=1}^{m} H(X(G_i)).$$

But $H(X) = \log_2 |\mathcal{F}|$, whereas by Lemma 15.7.1, part (i), $H(X(G_i)) \leq \log_2 |\mathcal{F}_i|$, implying the desired result. ■

Since the volume of every d-dimensional measurable set in R^n can be approximated by the volume of an appropriate approximation of it by standard aligned boxes in a fine enough grid, the last result has the following geometric application, proved in Loomis and Whitney (1949) in a different manner:

Corollary 15.7.6 *Let B be a measurable body in the n-dimensional Euclidean space, let* $\mathrm{Vol}(B)$ *denote its (n-dimensional) volume, and let* $\mathrm{Vol}(B_i)$ *denote the $(n-1)$-dimensional volume of the projection of B on the hyperplane spanned by all coordinates besides the ith one. Then*

$$(\mathrm{Vol}(B))^{n-1} \leq \prod_{i=1}^{n} \mathrm{Vol}(B_i).$$

■

If $S_i = \{0, 1\}$ for all i in Corollary 15.7.5, we get the following statement about set systems.

Corollary 15.7.7 [Chung et al. (1986)] *Let N be a finite set, and let $\mathcal{F}$ be a family of subsets of N. Let $\mathcal{G} = \{G_1, \ldots, G_m\}$ be a collection of subsets of N, and suppose that each element of N belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$, define $\mathcal{F}_i = \{F \cap G_i : F \in \mathcal{F}\}$. Then*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

■

We close the section with the following application of the last result, given in Chung et al. (1986).

Corollary 15.7.8 *Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, \ldots, t\}$, and suppose that for any two members of $\mathcal{F}$ there is a triangle contained in both of them. Then*

$$|\mathcal{F}| < \frac{1}{4} 2^{\binom{t}{2}}.$$

Proof. Let N be the set of all $\binom{t}{2}$ unordered pairs of vertices in $T = \{1, 2 \ldots, t\}$, and consider $\mathcal{F}$ as a family of subsets of N. Let $\mathcal{G}$ be the family of all subsets of N consisting of the edge sets of unions of two vertex-disjoint, nearly equal complete graphs in T. Let

$$s = \binom{\lceil t/2 \rceil}{2} + \binom{\lfloor t/2 \rfloor}{2}$$

denote the number of edges of such a union, and let m denote the total number of members in $\mathcal{G}$. By symmetry, each edge in N lies in precisely $k = \frac{sm}{\binom{t}{2}}$ members of $\mathcal{G}$. The crucial point is that every two graphs in $\mathcal{F}$ must have at least one common edge in

each $G \in \mathcal{G}$, since their intersection contains a triangle (and there are no triangles in the complement of G.) Therefore, in the notation of Corollary 15.7.7, the cardinality of each $\mathcal{F}_i$ is at most 2^{s-1}. We thus conclude that

$$|\mathcal{F}|^{\frac{sm}{\binom{t}{2}}} \leq (2^{s-1})^m,$$

implying that

$$|\mathcal{F}| \leq 2^{\binom{t}{2}-\binom{t}{2}/s},$$

and the desired result follows, as $s < \binom{t}{2}/2$. ■

Simonovits and Sós conjectured that, if $\mathcal{F}$ satisfies the assumptions of the last corollary, then, in fact,

$$|\mathcal{F}| \leq \frac{1}{8}2^{\binom{t}{2}},$$

which is tight. This, together with several extensions, has recently been proved in Ellis, Filmus and Friedgut (2012) by a sophisticated combination of algebraic and analytic tools. It seems plausible to conjecture that there is some absolute constant $\epsilon > 0$, such that for any fixed graph H which is not a star-forest (i.e., a forest each connected component of which is a star), the following holds: Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, \ldots, t\}$, and suppose that for any two members of $\mathcal{F}$ there is a copy of H contained in both of them. Then

$$|\mathcal{F}| < \left(\frac{1}{2} - \epsilon\right)2^{\binom{t}{2}}.$$

This question, suggested by the first author, remains open, though it is not difficult to show that it is true for every H of chromatic number at least 3, and that the conclusion fails for every star-forest H.

15.8 EXERCISES

1. Suppose that in the $(x_1, \ldots, x_k)$ tenure game of Section 15.3 the object of Paul is to maximize the number of faculty receiving tenure while the object of Carole is to minimize that number. Let v be that number with perfect play. Prove $v = \left\lfloor \sum_{i=1}^{k} x_i 2^{-i} \right\rfloor$.
2. Let $A_1, \ldots, A_n \subseteq \{1, \ldots, m\}$ with $\sum_{i=1}^{n} 2^{-|A_i|} < 1$. Paul and Carole alternately select distinct vertices from $\{1, \ldots, m\}$, Paul having the first move, until all vertices have been selected. Carole wins if she has selected all the vertices of some A_i. Paul wins if Carole does not win. Give a winning strategy for Paul.

3. Let $\mathcal{F}$ be a family of graphs on the labeled set of vertices $\{1, 2, \ldots, 2t\}$, and suppose that for any two members of $\mathcal{F}$ there is a perfect matching of t edges contained in both of them. Prove that
$$|\mathcal{F}| \leq 2^{\binom{2t}{2}-t}.$$

4. (Han's inequality). Let $X = (X_1, \ldots, X_m)$ be a random variable, and let $H(X)$ denote its entropy. For a subset I of $\{1, 2, \ldots, m\}$, let $X(I)$ denote the random variable $(X_i)_{i\in I}$. For $1 \leq q \leq m$, define
$$H_q(X) = \frac{1}{\binom{m-1}{q-1}} \sum_{Q\subset\{1,\ldots,m\},|Q|=q} H(X(Q)).$$
Prove that
$$H_1(X) \geq H_2(X) \geq \ldots \geq H_m(X) = H(X).$$

5. Let $X_i = \pm 1$, $1 \leq i \leq n$, be uniform and independent and let $S_n = \sum_{i=1}^n X_i$. Let $0 \leq p \leq \frac{1}{2}$. Prove
$$\Pr[S_n \geq (1-2p)n] \leq 2^{H(p)n}2^{-n}$$
by computing precisely the Chernoff bound $\min_{\lambda\geq 0} \mathrm{E}[e^{\lambda S_n}]e^{-\lambda(1-2p)n}$. (The case $p = 0$ will require a slight adjustment in the method, though the end result is the same.)

6. Find, asymptotically, that $u = u(q)$ which minimizes $2^q\Pr[S_q \leq -u] + 2^{q+1}(q-u)^{-1}$ and express the minimal value in the form $2^{q+1}q^{-1} + (1 + o(1))g(q)$.

7. Show that for A fixed and r sufficiently large, Paul wins the $(2^r - (r+1)A, A)$, r-ChipLiar game.

THE PROBABILISTIC LENS:
An Extremal Graph

Let T (top) and B (bottom) be disjoint sets of size m, and let G be a bipartite graph, with all edges between T and B. Suppose G contains no 4-cycle. How many edges can G have? This is a question from extremal graph theory. Surprisingly, for some m we may give the precise answer.

Suppose $m = n^2 + n + 1$ and that a projective plane P of order n (and hence containing m points) exists. Identify T with the points of P and B with the lines of P, and define $G = G_P$ by letting $t \in T$ be adjacent to $b \in B$ if and only if point t is on line b in P. As two points cannot lie on two lines, G_P contains no 4-cycle. We claim that such a G_P has the largest number of edges of any G containing no 4-cycle and further that any G containing no 4-cycle and having that many edges can be written in the form $G = G_P$.

Suppose G contains no 4-cycle. Let $b_1, b_2 \in B$ be a uniformly selected pair of distinct elements. For $t \in T$, let $D(t)$ be the set of $b \in B$ adjacent to t and $d(t) = |D(t)|$ the degree of t. Let I_t be the indicator random variable for t being adjacent to b_1, b_2. Then

$$\mathrm{E}[I_t] = \Pr[b_1, b_2 \in D(t)] = \binom{d(t)}{2} \Big/ \binom{m}{2}.$$

Now set

$$X = \sum_{t \in T} I_t,$$

the number of $t \in T$ adjacent to b_1, b_2. Then $X \leq 1$; that is, all b_1, b_2 have at most one common neighbor. ($X \leq 1$ is actually equivalent to G containing no 4-cycle.)

Linearity of Expectation gives

$$\mathrm{E}[X] = \sum_{t \in T} \mathrm{E}[I_t] = \sum_{t \in T} \binom{d(t)}{2} \Big/ \binom{m}{2}.$$

Let $\overline{d} = m^{-1}\sum_{t\in T} d(t)$ be the average degree. Convexity of the function $\binom{y}{2}$ gives

$$\sum_{t\in T} \binom{d(t)}{2} \Big/ \binom{m}{2} \geq m \binom{\overline{d}}{2} \Big/ \binom{m}{2}$$

with equality if and only if all $t \in T$ have the same degree. Now

$$1 \geq \max X \geq \mathrm{E}[X] \geq m \binom{\overline{d}}{2} \Big/ \binom{m}{2}.$$

When $G = G_P$, all $d(x) = \overline{d}$ (every line has $n + 1$ points) and X is identically 1 (two points determine precisely one line) so that the above inequalities are all equalities and

$$1 = m \binom{\overline{d}}{2} \Big/ \binom{m}{2}.$$

Any graph with more edges would have a strictly larger $\overline{d}$ so that $1 \geq m \binom{\overline{d}}{2} \Big/ \binom{m}{2}$ would fail and the graph would contain a 4-cycle.

Suppose further G has the same number of edges as G_P and contains no 4-cycle. The inequalities then must be equalities and so $X = 1$ always. Define a geometry with points T and lines given by the neighbor sets of $b \in B$. As $X = 1$, any two points determine a unique line. Reversing the roles of T, B, one also has that any two lines must determine a unique point. Thus G is generated from a projective plane.

16

Derandomization

> There are two ways of reading: one of them deepens and intensifies what one already knows; the other one takes new facts, new views to weave into ones life.
> –Doris Lessing

As mentioned in Chapter 1, the probabilistic method supplies, in many cases, effective randomized algorithms for various algorithmic problems. In some cases, these algorithms can be derandomized and converted into deterministic ones. In this chapter we discuss some examples.

16.1 THE METHOD OF CONDITIONAL PROBABILITIES

An easy application of the basic probabilistic method implies the following statement, which is a special case of Theorem 2.3.1.

Proposition 16.1.1 *For every integer n, there exists a coloring of the edges of the complete graph K_n by two colors so that the total number of monochromatic copies of K_4 is at most $\binom{n}{4} \cdot 2^{-5}$.*

Indeed, $\binom{n}{4} \cdot 2^{-5}$ is the expected number of monochromatic copies of K_4 in a random 2-edge-coloring of K_n, and hence a coloring as above exists.

Can we actually find *deterministically* such a coloring in time which is polynomial in n? Let us describe a procedure that does it, and is a special case of a general technique called the *method of conditional probabilities.*

We first need to define a weight function for any partially colored K_n. Given a coloring of some of the edges of K_n by red and blue, we define, for each copy K of K_4

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

in K_n, a weight $w(K)$ as follows: If at least one edge of K is colored red and at least one edge is colored blue, then $w(K) = 0$. If no edge of K is colored, then $w(K) = 2^{-5}$, and if $r \geq 1$ edges of K are colored, all with the same color, then $w(K) = 2^{r-6}$. Also define the total weight W of the partially colored K_n as the sum $\sum w(K)$, as K ranges over all copies of K_4 in K_n. Observe that the weight of each copy K of K_4 is precisely the probability that it will be monochromatic, if all the presently uncolored edges of K_n will be assigned randomly and independently one of the two colors red or blue. Hence, by Linearity of Expectation, the total weight W is simply the expected number of monochromatic copies of K_4 in such a random extension of the partial coloring of K_n to a full coloring.

We can now describe the procedure for finding a coloring as in Proposition 16.1.1. Order the $\binom{n}{2}$ edges of K_n arbitrarily, and construct the desired two-coloring by coloring each edge either red or blue in turn. Suppose $e_1, \ldots, e_{i-1}$ have already been colored, and we now have to color e_i. Let W be the weight of K_n, as defined above, with respect to the given partial coloring c of $e_1, \ldots, e_{i-1}$. Similarly, let W_{red} be the weight of K_n with respect to the partial coloring obtained from c by coloring e_i red, and let W_{blue} be the weight of K_n with respect to the partial coloring obtained from c by coloring e_i blue. By the definition of W (and as follows from its interpretation as an expected value)

$$W = \frac{W_{\text{red}} + W_{\text{blue}}}{2}.$$

The color of e_i is now chosen so as to minimize the resulting weight; that is, if $W_{\text{red}} \leq W_{\text{blue}}$, then we color e_i red, otherwise we color it blue. By the above inequality, the weight function never increases during the algorithm. Since at the beginning its value is exactly $\binom{n}{4} 2^{-5}$, its value at the end is at most this quantity. However, at the end all edges are colored, and the weight is precisely the number of monochromatic copies of K_4. Thus the procedure above produces, deterministically and in polynomial time, a 2-edge-coloring of K_n, satisfying the conclusion of Proposition 16.1.1.

Let us describe, now, the method of conditional probabilities in a more general setting. An instance of this method is due, implicitly, to Erdős and Selfridge (1973), and more explicit examples appear in Spencer (1987) and in Raghavan (1988). Suppose we have a probability space, and assume, for simplicity, that it is symmetric and contains 2^l points, denoted by the binary vectors of length l. Let $A_1, \ldots, A_s$ be a collection of events, and suppose that $\sum_{i=1}^{s} \Pr[A_i] = k$. Thus, k is the expected value of the number of events A_i that hold, and hence there is a point $(\epsilon_1, \ldots, \epsilon_l)$ in the space in which at most k events hold. Our objective is to find such a point deterministically.

For each choice of $(\epsilon_1, \ldots, \epsilon_{j-1})$, and for each event A_i, the conditional probability

$$\Pr[A_i \,|\, \epsilon_1, \ldots, \epsilon_{j-1}]$$

of the event A_i given the values of $\epsilon_1, \ldots, \epsilon_{j-1}$ is clearly the average of the two conditional probabilities corresponding to the two possible choices for ϵ_j. That is

$$\Pr[A_i \,|\, \epsilon_1, \ldots, \epsilon_{j-1}] = \frac{\Pr[A_i \,|\, \epsilon_1, \ldots, \epsilon_{j-1}, 0] + \Pr[A_i \,|\, \epsilon_1, \ldots, \epsilon_{j-1}, 1]}{2}.$$

Consequently

$$\begin{aligned}
&\sum_{i=1}^{s} \Pr\left[A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}\right] \\
&= \frac{\sum_{i=1}^{s} \Pr\left[A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 0\right] + \sum_{i=1}^{s} \Pr\left[A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 1\right]}{2} \\
&\geq \min\left\{ \sum_{i=1}^{s} \Pr\left[A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 0\right], \sum_{i=1}^{s} \Pr\left[A_i \mid \epsilon_1, \ldots, \epsilon_{j-1}, 1\right] \right\} .
\end{aligned}$$

Therefore, if the values of ϵ_j are chosen, each one in turn, so as to minimize the value of $\sum_{i=1}^{s} \Pr[A_i \mid \epsilon_1, \ldots, \epsilon_j]$, then the value of this sum cannot increase. Since this sum is k at the beginning, it follows that it is at most k at the end. But at the end each ϵ_j is fixed, and hence the value of this sum is precisely the number of events A_i that hold at the point $(\epsilon_1, \ldots, \epsilon_l)$, showing that our procedure works.

Note that the assumptions that the probability space is symmetric and that it has 2^l points can be relaxed. The procedure above is efficient provided l is not too large (as is usually the case in combinatorial examples), and, more importantly, provided the conditional probabilities $\Pr[A_i \mid \epsilon_1, \ldots, \epsilon_j]$ can be computed efficiently for each of the events A_i and for each possible value of $\epsilon_1, \ldots, \epsilon_j$. This is, indeed, the case in the example considered in Proposition 16.1.1. However, there are many interesting examples where this is not the case. A trick that can be useful in such cases is the introduction of *pessimistic estimators*, introduced by Raghavan (1988). Consider, again, the symmetric probability space with 2^l points described above, and the events $A_1, \ldots, A_s$ in it. Suppose that for each event A_i, and for each $0 \leq j \leq l$, we have a function $f_j^i(\epsilon_1, \ldots, \epsilon_j)$, which can be efficiently computed. Assume, also, that

$$f_{j-1}^i(\epsilon_1, \ldots, \epsilon_{j-1}) \geq \frac{f_j^i(\epsilon_1, \ldots, \epsilon_{j-1}, 0) + f_j^i(\epsilon_1, \ldots, \epsilon_{j-1}, 1)}{2} , \tag{16.1}$$

and that f_j^i is an upper bound on the conditional probabilities for the event A_i, that is

$$f_j^i(\epsilon_1, \ldots, \epsilon_j) \geq \Pr[A_i \mid \epsilon_1, \ldots, \epsilon_j] . \tag{16.2}$$

Clearly, the same inequalities hold for the sums over i. In this case, if in the beginning $\sum_{i=1}^{s} f_0^i \leq t$, and we choose the values of the ϵ_j so as to minimize the sum $\sum_{i=1}^{s} f_j^i(\epsilon_1, \ldots, \epsilon_j)$ in each step, we get in the end a point $(\epsilon_1, \ldots, \epsilon_l)$ for which the sum $\sum_{i=1}^{s} f_l^i(\epsilon_1, \ldots, \epsilon_l) \leq t$. The number of events A_i that hold in this point is at most t. The functions f_j^i in the argument above are called pessimistic estimators.

This enables us to obtain efficient algorithms in some cases where there is no known efficient way of computing the required conditional probabilities. The following theorem is an example; it is related to some of the results in Chapters 13 and 15.

Theorem 16.1.2 *Let $(a_{ij})_{i,j=1}^{n}$ be an $n \times n$ matrix of reals, where $-1 \leq a_{ij} \leq 1$ for all i, j. Then one can find, in polynomial time, $\epsilon_1, \ldots, \epsilon_n \in \{-1, 1\}$ such that for every i, $1 \leq i \leq n$, the inequality $|\sum_{j=1}^{n} \epsilon_j a_{ij}| \leq \sqrt{2n \ln(2n)}$ holds.*

Proof. Consider the symmetric probability space on the 2^n points corresponding to the 2^n possible vectors $(\epsilon_1, \ldots, \epsilon_n) \in \{-1, 1\}^n$. Define $\beta = \sqrt{2n \ln(2n)}$, and let A_i be the event $|\sum_{j=1}^{n} \epsilon_j a_{ij}| > \beta$. We next show that the method of conditional probabilities with appropriate pessimistic estimators enables us to find efficiently a point of the space in which no event A_i holds.

Define $\alpha = \beta/n$, and let $G(x)$ be the function

$$G(x) = \cosh(\alpha x) = \frac{e^{\alpha x} + e^{-\alpha x}}{2}.$$

By comparing the terms of the corresponding Taylor series, it is easy to see that for every real x

$$G(x) \leq e^{\frac{\alpha^2 x^2}{2}}$$

with strict inequality if both x and α are not 0. It is also simple to check that for every real x and y

$$G(x)G(y) = \frac{G(x+y) + G(x-y)}{2}.$$

We can now define the functions f_p^i which will form our pessimistic estimators. For each $1 \leq i \leq n$ and for each $\epsilon_1, \ldots, \epsilon_p \in \{-1, 1\}$, we define

$$f_p^i(\epsilon_1, \ldots, \epsilon_p) = 2e^{-\alpha\beta} G\left(\sum_{j=1}^{p} \epsilon_j a_{ij}\right) \prod_{j=p+1}^{n} G(a_{ij}).$$

Obviously, these functions can be efficiently computed. It remains to check that they satisfy the conditions described in (16.1) and (16.2), and that the sum $\sum_{i=1}^{n} f_0^i$ is less than 1. This is proved in the following claims.

Claim 16.1.3 *For every $1 \leq i \leq n$ and every $\epsilon_1, \ldots, \epsilon_{p-1} \in \{-1, 1\}$*

$$f_{p-1}^i(\epsilon_1, \ldots, \epsilon_{p-1}) = \frac{f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, -1) + f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, 1)}{2}.$$

Proof. Put $v = \sum_{j=1}^{p-1} \epsilon_j a_{ij}$. By the definition of f_p^i and by the properties of G

$$\begin{aligned} f_{p-1}^i(\epsilon_1, \ldots, \epsilon_{p-1}) &= 2e^{-\alpha\beta} G(v) G(a_{ip}) \prod_{j=p+1}^{n} G(a_{ij}) \\ &= 2e^{-\alpha\beta} \frac{G(v - a_{ip}) + G(v + a_{ip})}{2} \prod_{j=p+1}^{n} G(a_{ij}) \\ &= \frac{f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, -1) + f_p^i(\epsilon_1, \ldots, \epsilon_{p-1}, 1)}{2}, \end{aligned}$$

completing the proof of the claim. ∎

Claim 16.1.4 *For every* $1 \le i \le n$ *and every* $\epsilon_1, \ldots, \epsilon_{p-1} \in \{-1, 1\}$

$$f^i_{p-1}(\epsilon_1, \ldots, \epsilon_{p-1}) \ge \Pr[A_i \mid \epsilon_1, \ldots, \epsilon_{p-1}] \,.$$

Proof. Define v as in the proof of Claim 16.1.3. Then

$$\begin{aligned}
\Pr[A_i \mid \epsilon_1, \ldots, \epsilon_{p-1}] &= \Pr\left[v + \sum_{j \ge p} \epsilon_j a_{ij} > \beta\right] + \Pr\left[-v - \sum_{j \ge p} \epsilon_j a_{ij} > \beta\right] \\
&= \Pr\left[e^{\alpha(v + \sum_{j \ge p} \epsilon_j a_{ij})} > e^{\alpha\beta}\right] \\
&\quad + \Pr\left[e^{-\alpha(v + \sum_{j \ge p} \epsilon_j a_{ij})} > e^{\alpha\beta}\right] \\
&\le e^{\alpha v} e^{-\alpha\beta}\, \mathrm{E}\left[e^{\alpha(\sum_{j \ge p} \epsilon_j a_{ij})}\right] \\
&\quad + e^{-\alpha v} e^{-\alpha\beta}\, \mathrm{E}\left[e^{-\alpha(\sum_{j \ge p} \epsilon_j a_{ij})}\right] \\
&= 2e^{-\alpha\beta} G(v) \prod_{j \ge p} G(a_{ij}) = f^i_{p-1}(\epsilon_1, \ldots, \epsilon_{p-1}) \,.
\end{aligned}$$

This completes the proof of Claim 16.1.4. ∎

To establish the theorem, it remains to show that $\sum_{i=1}^n f^i_0 < 1$. Indeed, by the properties of G and by the choice of α and β

$$\begin{aligned}
\sum_{i=1}^n f^i_0 &= \sum_{i=1}^n 2e^{-\alpha\beta} \prod_{j=1}^n G(a_{ij}) \\
&\le \sum_{i=1}^n 2e^{-\alpha\beta} \prod_{j=1}^n e^{\frac{\alpha^2 a_{ij}^2}{2}} \\
&\le \sum_{i=1}^n 2e^{-\alpha\beta} e^{\frac{\alpha^2 n}{2}} \\
&= 2ne^{\frac{\alpha^2 n}{2} - \alpha\beta} = 2ne^{-\frac{\alpha^2 n}{2}} = 1 \,.
\end{aligned}$$

Moreover, the first inequality is strict unless $a_{ij} = 0$ for all i, j, whereas the second is strict unless $a_{ij}^2 = 1$ for all i, j. This completes the proof of the theorem. ∎

16.2 *d*-WISE INDEPENDENT RANDOM VARIABLES IN SMALL SAMPLE SPACES

The complexity class NC is, roughly speaking, the class of all problems that can be solved in time which is polylogarithmic (in the size of the input) using a polynomial

number of parallel processors. Several models of computation, which form the theoretical abstraction of the parallel computer, have been used in considering this class. The most common one is the EREW (=Exclusive Read, Exclusive Write) PRAM, in which different processors are not allowed to read from or write into the same memory cell simultaneously. See Karp and Ramachandran (1990) for more details.

Let n denote the size of the input. There are several simple tasks that can be easily performed in NC. For example, it is possible to copy the content of a cell c into $m = n^{O(1)}$ cells in time $O(\log n)$, using, say, m processors. To do so, consider a complete binary tree with m leaves, and associate each of its internal vertices with a processor. At first, the processor corresponding to the root of the tree reads from c and writes its content in two cells, corresponding to its two children. Next, each of these two, in parallel, reads from its cell and writes its content in two cells corresponding to its two children. In general, at the ith step all the processors whose distance from the root of the tree is $i - 1$, in parallel, read the content of c previously stored in their cells and write it twice. The procedure clearly ends in time $O(\log m)$, as claimed. [In fact, it can be shown that $O(m/\log m)$ processors suffice for this task but we do not try to optimize this number here.]

A similar technique can be used for computing the sum of m numbers with m processors in time $O(\log m)$; we consider the numbers as if they lie on the leaves of a complete binary tree with m leaves, and in the ith step each one of the processors whose distance from the leaves is i computes, in parallel, the sum of the two numbers previously computed by its children. The root will clearly have, in such a way, the desired sum in time $O(\log m)$.

Let us now return to the edge-coloring problem of the complete graph K_n discussed in Proposition 16.1.1. By the remarks above, the problem of *checking* whether in a given edge-coloring there are at most $\binom{n}{4} 2^{-5}$ monochromatic copies of K_4 is in NC, that is, this checking can be done in time $(\log n)^{O(1)}$ – (in fact, in time $O(\log n)$) – using $n^{O(1)}$ processors. Indeed, we can first copy the given coloring $\binom{n}{4}$ times. Then we assign a processor for each copy of K_4 in K_n, and this processor checks whether its copy is monochromatic or not (all these checkings can be done in parallel, since we have enough copies of the coloring). Finally, we sum the number of processors whose copies are monochromatic. Clearly, we can complete the work in time $O(\log n)$ using $n^{O(1)}$ parallel processors.

Thus we can *check*, in NC, whether a given coloring of K_n satisfies the assertion of Proposition 16.1.1. Can we *find* such a coloring deterministically in NC? The method described in the previous section does not suffice, as the edges have been colored one by one, so the procedure is sequential and requires time $\Omega(n^2)$. However, it turns out that in fact we can find, in NC, a coloring with the desired properties by applying a method that relies on a technique first suggested by Joffe (1974), and later developed by many researchers. This method is a general technique for converting randomized algorithms whose analysis depends only on d-wise rather than fully independent random choices (for some constant d) into deterministic (and in many cases also parallel) ones. Our approach here follows that of Alon, Babai and Itai (1986), but for simplicity

we only consider here the case of random variables that take the two values 0, 1 with equal probability.

The basic idea is to replace an exponentially large sample space by one of polynomial size. If a random variable on such a space takes a certain value with positive probability, then we can find a point in the sample space in which this happens simply by deterministically checking all the points. This can be done with no loss of time by using a polynomial number of parallel processors. Note that for the edge-coloring problem considered in Proposition 16.1.1, 6-wise independence of the random variables corresponding to the colors of the edges would suffice – since this already gives a probability of 2^{-5} for each copy of K_4 to be monochromatic, and hence gives the required expected value of monochromatic copies. Therefore, for this specific example it suffices to construct a sample space of size $n^{O(1)}$ and $\binom{n}{2}$ random variables in it, each taking the values 0 and 1 with probability $1/2$, such that each 6 of the random variables are independent.

Small sample spaces with many d-wise independent 0, 1-random variables in them can be constructed from any linear error-correcting code with appropriate parameters. The construction we describe here is based on the binary BCH codes (see, e.g., MacWilliams and Sloane (1977)).

Theorem 16.2.1 *Suppose $n = 2^k - 1$ and $d = 2t + 1$. Then there exists a symmetric probability space Ω of size $2(n+1)^t$ and d-wise independent random variables $y_1, \ldots, y_n$ over Ω each of which takes the values 0 and 1 with probability $1/2$.*

The space and the variables are explicitly constructed, given a representation of the field $F = GF(2^k)$ as a k-dimensional algebra over $GF(2)$.

Proof. Let $x_1, \ldots, x_n$ be the n nonzero elements of F, represented as column vectors of length k over $GF(2)$. Let H be the following $1 + kt \times n$ matrix over $GF(2)$:

$$H = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^3 & x_2^3 & \cdots & x_n^3 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{2t-1} & x_2^{2t-1} & \cdots & x_n^{2t-1} \end{pmatrix}$$

This is the parity check matrix of the extended binary BCH code of length n and designed distance $2t + 2$. It is well known that any $d = 2t + 1$ columns of H are linearly independent over $GF(2)$. For completeness, we present the proof in the next lemma.

Lemma 16.2.2 *Any set of $d = 2t + 1$ columns of H is linearly independent over $GF(2)$.*

Proof. Let $J \subset \{1, 2, \ldots, n\}$ be a subset of cardinality $|J| = 2t + 1$ of the set of indices of the columns of H. Suppose that $\sum_{j \in J} z_j H_j = 0$, where H_j denotes the jth

column of H and $z_j \in GF(2)$. To complete the proof we must show that $z_j = 0$ for all $j \in J$. By the assumption,

$$\sum_{j \in J} z_j x_j^i = 0 \tag{16.3}$$

for $i = 0$ and for every odd i satisfying $1 \leq i \leq 2t-1$. Suppose, now, that $a = 2^b \cdot l$, where $l \leq 2t-1$ is an odd number. By squaring (16.3) for $i = l$ b times, using the fact that in characteristic 2, $(u+v)^2 = u^2 + v^2$, and the fact that since each z_j is either 0 or 1, the equality $z_j = z_j^2$ holds for all j, we conclude that (16.3) holds for $i = a$. Consequently, (16.3) holds for all i, $0 \leq i \leq 2t$. This is a homogeneous system of $2t+1$ linear equations in $2t+1$ variables. The matrix of the coefficients is a Vandermonde matrix, which is nonsingular. Thus, the only solution is the trivial one $z_j = 0$ for all $j \in J$, completing the proof of the lemma. ■

Returning to the proof of the theorem, we define $\Omega = \{1, 2, \ldots, 2(n+1)^t\}$, and let $A = (a_{ij}), i \in \Omega, 1 \leq j \leq n$ be the $(0,1)$-matrix whose $2(n+1)^t = 2^{kt+1}$ rows are all the linear combinations (over $GF(2)$) of the rows of H. The sample space Ω is now endowed with the uniform probability measure, and the random variable y_j is defined by the formula $y_j(i) = a_{ij}$ for all $i \in \Omega, 1 \leq j \leq n$.

It remains to show that the variables y_j are d-wise independent, and that each of them takes the values 0 or 1 with equal probability. For this we have to show that, for every set J of up to d columns of A, the rows of the $|\Omega|$ by $|J|$ submatrix $A_J = (a_{ij}), i \in \Omega, j \in J$ take on each of the $2^{|J|}$ $(0,1)$-vectors of length $|J|$ equally often. However, by Lemma 16.2.2, the columns of the corresponding submatrix H_J of H are linearly independent. The number of rows of A_J that are equal to any given vector is precisely the number of linear combinations of the rows of H_J that are equal to this vector. This number is the number of solutions of a system of $|J|$ linearly independent linear equations in $kt+1$ variables, which is, of course, $2^{kt+1-|J|}$, independent of the vector of free coefficients. This completes the proof of the theorem. ■

Theorem 16.2.1 supplies an efficient way of constructing, for every fixed d and every n, a sample space of size $O(n^{\lfloor d/2 \rfloor})$ and n d-wise independent random variables in it, each taking the values 0 or 1 with equal probability. In particular, we can use such a space of size $O\left(\binom{n}{2}^3\right) = O(n^6)$ for finding a coloring as in Proposition 16.1.1 in NC. Several other applications of Theorem 16.2.1 appear in the paper by Alon et al. (1986).

It is natural to ask whether the size $O(n^{\lfloor d/2 \rfloor})$ can be improved. We next show that this size is optimal, up to a constant factor (depending on d).

Let us call a random variable *almost constant* if it attains a single value with probability 1. Let $m(n,d)$ denote the function defined by

$$m(n,d) = \sum_{j=0}^{d/2} \binom{n}{j} \text{ if } d \text{ is even.}$$

and

$$m(n,d) = \sum_{j=0}^{(d-1)/2} \binom{n}{j} + \binom{n-1}{(d-1)/2} \text{ if } d \text{ is odd.}$$

Observe that for every fixed d, $m(n,d) = \Omega(n^{\lfloor d/2 \rfloor})$.

Proposition 16.2.3 *If the random variables $y_1, \ldots, y_n$ over the sample space Ω are d-wise independent and none of them is almost constant, then $|\Omega| \geq m(n,d)$.*

Note that we assume here neither that Ω is a symmetric space nor that the variables y_j are $(0,1)$-variables.

Proof. Clearly, we may assume that the expected value of each y_j is 0 (since otherwise we can replace y_j by $y_j - \mathrm{E}[y_j]$). For each subset S of $\{1, \ldots, n\}$, define $\alpha_S = \prod_{j \in S} y_j$. Observe that, since no y_j is almost constant and since the variables are d-wise independent,

$$\mathrm{E}[\alpha_S \alpha_S] = \prod_{j \in S} \mathrm{Var}[y_j] > 0 \tag{16.4}$$

for all S satisfying $|S| \leq d$. Similarly, for all S and T satisfying $|S \cup T| \leq d$ and $S \neq T$, we have

$$\mathrm{E}[\alpha_s \alpha_T] = \prod_{j \in S \cap T} \mathrm{Var}[y_j] \prod_{j \in S \cup T \setminus (S \cap T)} \mathrm{E}[y_j] = 0 \,. \tag{16.5}$$

Let $S_1, \ldots, S_m$, where $m = m(n,d)$, be subsets of $\{1, \ldots, n\}$ such that the union of each two is of size at most d. [Take all subsets of size at most $d/2$, and if d is odd, add all the subsets of size $(d+1)/2$ containing 1.]

To complete the proof, we show that the m functions α_{S_j} (considered as real vectors of length $|\Omega|$) are linearly independent. This implies that $|\Omega| \geq m = m(n,d)$, as stated in the proposition.

To prove linear independence, suppose $\sum_{j=1}^{m} c_j \alpha_{S_j} = 0$. Multiplying by α_{S_i} and computing the expected values, we obtain, by (16.5),

$$0 = \sum_{j=1}^{m} c_j \, \mathrm{E}[\alpha_{S_j} \alpha_{S_i}] = c_i \, \mathrm{E}[\alpha_{S_i} \alpha_{S_i}] \,.$$

This implies, by (16.4), that $c_i = 0$ for all i. The required linear independence follows, completing the proof. ∎

The last proposition shows that the size of a sample space with n d-wise independent nontrivial random variables can be polynomial in n only when d is fixed. However, as shown by Naor and Naor (1990), if we only require the random variables to be *almost* d-wise independent, the size can be polynomial even when

$d = \Omega(\log n)$. Such sample spaces and random variables, which can be constructed explicitly in several ways, have various interesting applications in which almost d-wise independence suffices. More details appear in Naor and Naor (1990) and in Alon et al. (1990).

16.3 EXERCISES

1. Let $A_1, \ldots, A_n \subseteq \{1, \ldots, m\}$, with $\sum_{i=1}^{n} 2^{1-|A_i|} < 1$. Prove there exists a two-coloring $\chi : \{1, \ldots, m\} \to \{0, 1\}$ with no A_i monochromatic. With $m = n$, give a deterministic algorithm to find such a χ in polynomial time.
2. Describe a deterministic algorithm that, given n, constructs, in time polynomial in n, a family $\mathcal{F}$ of n^{10} subsets of the set $N = \{1, 2, \ldots, n\}$, where each $F \in \mathcal{F}$ is of size at most $10 \log n$ and for every family $\mathcal{G}$ of n subsets, each of cardinality $n/2$ of N, there is an $F \in \mathcal{F}$ that intersects all members of $\mathcal{G}$.

THE PROBABILISTIC LENS: Crossing Numbers, Incidences, Sums and Products

In this lens we start with a simple result in graph theory, whose proof is probabilistic, and then describe some of its fascinating consequences in combinatorial geometry and combinatorial number theory. Some versions of most of these seemingly unrelated consequences have been proved earlier, in a far more complicated manner. Before the discovery of the new proofs shown here, the only clue that there might be a connection between all of them has been the fact that Endre Szemerédi is one of the coauthors of each of the papers providing the first proofs.

An *embedding* of a graph $G = (V, E)$ in the plane is a planar representation of it, where each vertex is represented by a point in the plane, and each edge uv is represented by a curve connecting the points corresponding to the vertices u and v. The crossing number of such an embedding is the number of pairs of intersecting curves that correspond to pairs of edges with no common endpoints. The *crossing number* $\mathrm{cr}(G)$ of G is the minimum possible crossing number in an embedding of it in the plane. The following theorem was proved by Ajtai et al. (1982) and, independently, by Leighton. Here we describe a very short probabilistic proof.

Theorem 1 *The crossing number of any simple graph $G = (V, E)$ with $|E| \geq 4|V|$ is at least $\frac{|E|^3}{64|V|^2}$.*

Proof. By Euler's formula, any simple planar graph with $n \geq 3$ vertices has at most $3n - 6$ edges, implying that any simple planar graph with n vertices has at most $3n$ edges. Therefore, the crossing number of any simple graph with n vertices and m

edges is at least $m - 3n$. Let $G = (V, E)$ be a graph with $|E| \geq 4|V|$ embedded in the plane with $t = \text{cr}(G)$ crossings. Let H be the random induced subgraph of G obtained by picking each vertex of G, randomly and independently, to be a vertex of H with probability p (where p will be chosen later). The expected number of vertices of H is $p|V|$, the expected number of its edges is $p^2|E|$, and the expected number of crossings in its given embedding is $p^4 t$, implying that the expected value of its crossing number is at most $p^4 t$. Therefore, $p^4 t \geq p^2|E| - 3p|V|$, implying that

$$\text{cr}(G) = t \geq \frac{|E|}{p^2} - 3\frac{|V|}{p^3}.$$

Without trying to optimize the constant factor, substitute $p = 4|V|/|E| \quad (\leq 1)$, to get the desired result. ■

Székely (1997) noticed that this result can be applied to obtain a surprisingly simple proof of a result of Szemerédi and Trotter in combinatorial geometry. The original proof is far more complicated.

Theorem 2 *Let P be a set of n distinct points in the plane, and let L be a set of m distinct lines. Then, the number of incidences between the members of P and those of L (i.e., the number of pairs (p, l) with $p \in P, l \in L$ and $p \in l$) is at most $c(m^{2/3}n^{2/3} + m + n)$, for some absolute constant c.*

Proof. We may and will assume that every line in L is incident with at least one of the points of P. Denote the number of incidences by I. Let $G = (V, E)$ be the graph whose vertices are all members of P, where two are adjacent if and only if they are consecutive points of P on some line in L. Clearly, $|V| = n$ and $|E| = I - m$. Note that G is already given embedded in the plane, where the edges are represented by segments of the corresponding lines in L. In this embedding, every crossing is an intersection point of two members of L, implying that $\text{cr}(G) \leq \binom{m}{2} \leq m^2/2$. By Theorem 1, either $I - m = |E| < 4|V| = 4n$, that is, $I \leq m + 4n$, or

$$\frac{m^2}{2} \geq \text{cr}(G) \geq \frac{(I-m)^3}{64n^2},$$

implying that $I \leq (32)^{1/3}m^{2/3}n^{2/3} + m$. In both cases, $I \leq 4(m^{2/3}n^{2/3} + m + n)$, completing the proof. ■

An analogous argument shows that the maximum possible number of incidences between a set of n points and a set of m unit circles in the plane does not exceed $O(m^{2/3}n^{2/3} + m + n)$, and this implies that the number of unit distances determined by a set of n points in the plane is at most $O(n^{4/3})$. While the above upper bound for the number of incidences of points and lines is sharp, up to a constant factor, an old conjecture of Erdős asserts that the maximum possible number of unit distances determined by a set of n points in the plane is at most $c_\epsilon n^{1+\epsilon}$ for any $\epsilon > 0$. The $O(n^{4/3})$

estimate is, however, the best known upper bound, and was first proved by Spencer, Szemerédi, and Trotter in a far more complicated way.

Elekes (1997) found several applications of Theorem 2 in additive number theory. Here, too, the proofs are amazingly simple. Here is a representative result.

Theorem 3 *For any three sets A,B, and C of s real numbers each,*

$$|A \cdot B + C| = |\{ab + c : a \in A, b \in B, c \in C\}| \geq \Omega(s^{3/2}).$$

Proof. Put $R = A \cdot B + C$, $|R| = r$ and define

$$P = \{(a, t) : a \in A, t \in R\}, \quad L = \{y = bx + c : b \in B, c \in C\}.$$

Thus P is a set of $n = sr$ points in the plane, L is a set of $m = s^2$ lines in the plane, and each line $y = bx + c$ in L is incident with s points of P, that is, with all the points $\{(a, ab + c) : a \in A\}$. Therefore, by Theorem 2, $s^3 \leq 4(s^{4/3}(sr)^{2/3} + sr + s^2)$, implying that $r \geq \Omega(s^{3/2})$, as needed. ■

The same method implies that, for every set A of n reals, either $|A + A| \geq \Omega(n^{5/4})$ or $|A \cdot A| \geq n^{5/4}$, greatly improving and simplifying a result of Erdős and Szemerédi.

17

Graph Property Testing

'Call the first witness,' said the King; and the White Rabbit blew three blasts on the trumpet, and called out, 'First witness!'
–from *Alice in Wonderland*, by Lewis Carroll

17.1 PROPERTY TESTING

Property testers are fast randomized algorithms for distinguishing between combinatorial structures that satisfy a certain property, and ones that are far from satisfying it. The basic algorithmic task in this area is to design a randomized algorithm, which given a combinatorial structure S, can distinguish with high probability between the case that S satisfies a prescribed property $\mathcal{P}$ and the case that S is ε-far from satisfying $\mathcal{P}$. Here, S is said to be ε-far from satisfying $\mathcal{P}$ if an ε-fraction of its representation should be modified in order to turn it to a structure that satisfies $\mathcal{P}$. The main objective is to design randomized algorithms, which look at a very small portion of the input, and using this information distinguish with high probability between the above two cases. Such algorithms are called *testers* for the property $\mathcal{P}$.

Preferably, a tester should look at a portion of the input whose size is a function of ε only. The general notion of property testing was first formulated by Rubinfeld and Sudan (1996), who were motivated by the study of various algebraic properties such as linearity of functions. Property testing is also motivated by questions in program checking, computational learning, approximation algorithms, and probabilistically checkable proofs, as well as by the need to access large datasets, like the graph of the Internet. The investigation of the subject relies heavily on probabilistic methods.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

The main focus of this chapter is in testing properties of graphs. In this case, a graph G on n vertices is said to be ϵ-far from satisfying a property $\mathcal{P}$ if one needs to add to or delete from G at least ϵn^2 edges in order to turn it into a graph satisfying $\mathcal{P}$. Here we assume that the tester can query an oracle whether a pair of vertices, i and j, are adjacent in the input graph G. If the graph satisfies the property, then the tester has to accept with probability at least, say, $2/3$, and if it is ϵ-far from satisfying it, then the algorithm has to reject with probability at least $2/3$.

The study of the notion of testability for combinatorial structures, and mainly for labeled graphs, was introduced by Goldreich, Goldwasser and Ron (1998). They showed that many natural graph properties such as k-colorability, having a large clique or having a large cut, admit a tester, whose *query complexity* (i.e., the number of oracle queries of type "does (i,j) belong to $E(G)$") as well as their total running time can be upper-bounded by a function of ϵ that is independent of the size of the input. We call properties having such efficient testers, that is, testers whose query complexity is a function of ϵ only, *testable*. In general, a property tester may have a small probability of accepting graphs that are ϵ-far from satisfying the tested property, as well as a small probability of rejecting graphs satisfying the property. In this case, the tester is said to have *two-sided* error. If the tester accepts graphs satisfying the property with probability 1, then the tester is said to have *one-sided* error.

It is worth noting that the model of graph property testing described here is often referred to as the *dense graph model*. Other models of graph property testing have also been investigated, see, for example, Goldreich and Ron (2002). For further reading and pointers on testing properties of graphs and other combinatorial structures, the reader is referred to the surveys Goldreich (1999), Fischer (2001), Ron (2001), Alon and Shapira (2006), and references therein.

17.2 TESTING COLORABILITY

Although the computational problem of deciding whether a given graph is k-colorable is NP-complete for every fixed $k \geq 3$, it turns out that, somewhat surprisingly, for every fixed $\epsilon > 0$ there is an efficient algorithm for distinguishing between graphs on n vertices that are k-colorable and graphs from which one has to delete at least ϵn^2 edges to make them k-colorable. This result, mentioned already in Alon et al. (1994), follows from the fact that the property of being k-colorable is testable, as proved implicitly in Rödl and Duke (1985) and explicitly (with a far better dependence on the parameter ϵ) in Goldreich et al. (1998). Indeed, as we show in this subsection, if a graph $G = (V, E)$ is ϵ-far from being k-colorable, then an induced subgraph of it on a randomly chosen set of $c(k)/\epsilon^2$ vertices is not k-colorable with high probability. This is proved in Alon and Krivelevich (2002), with $c(k) = 36k \ln k$, building on the work of Goldreich et al. (1998) who showed that a random set of $O(k^2 \ln k/\epsilon^3)$ vertices suffices. Note that the above supplies a very simple tester with one sided error for testing k-colorability; consider the induced subgraph on a randomly chosen set of $36k \ln k/\epsilon^2$ vertices, and accept iff this subgraph is k-colorable. Obviously, every k-colorable graph is accepted by this procedure, and graphs that

are ε-far from being k-colorable are likely to be rejected. Note also that the validity of this statement implies the nontrivial fact that every graph that is ε-far from being k-colorable contains a small witness (for being non-k-colorable), that is, a subgraph on only $c(\varepsilon, k) \leq O(k \ln k/\varepsilon^2)$ vertices which is not k-colorable. The existence of some such function $c(\varepsilon, k)$ has been conjectured by Erdős and first proved by Rödl and Duke (for some extremely fast growing function $c(\varepsilon, k)$ of ε and k – see Rödl and Duke (1985)). In this section we describe the improved $c(k)/\varepsilon^2$ bound. For simplicity, we present the proof only for $k = 3$, the proof for the general case being essentially identical. Throughout the proof we omit all floor and ceiling signs whenever these are not crucial.

Theorem 17.2.1 *Suppose* $0 < \varepsilon < 0.1$, *let* $G = (V, E)$ *be a graph on* $n > \frac{400}{\varepsilon^3}$ *vertices, and suppose that one has to delete from* G *at least* εn^2 *edges to make it* 3*-colorable. Then the probability that an induced subgraph of* G *on a randomly chosen set of* $s = 40/\varepsilon^2$ *vertices is* 3*-colorable does not exceed* 0.1.

Proof. We start with an outline of the proof. Given $G = (V, E)$ as in the theorem, pick a random subset $R \subset V$ of size $|R| = s = 40/\varepsilon^2$ in s rounds, each time choosing uniformly at random a single vertex r_j among the vertices not selected so far.

Suppose that some subset $S \subset R$ has already been 3-colored by $\phi : S \to C$, where $C = \{1, 2, 3\}$. The objective is to show that with high probability there is a witness showing that this partial coloring cannot be extended to a proper coloring of the induced subgraph on R. If a proper 3-coloring $c : V \to C$ of G is to coincide with ϕ on S, then for every vertex $v \in V \setminus S$, the colors of the neighbors of v in S under ϕ are forbidden for v in c. The rest of the colors are still feasible for v. It could be that v has no feasible colors left at all. Such a vertex will be called colorless with respect to S and ϕ. If the number of colorless vertices is large, then there is a decent chance that among the next few randomly chosen vertices of R there will be one such colorless vertex v^*. Obviously, adding v^* to S provides the desired witness for non-extendibility of ϕ.

If the set of colorless vertices is small, then one can show that, as G is ε-far from being 3-colorable, there is a relatively large subset W of vertices (which will be called restricting) such that adding any vertex $v \in W$ to S and coloring it by any feasible color excludes this color from the lists of feasible colors of at least εn neighbors of v. If such a vertex v is found among the next few vertices of the random sample R, then adding v to S and coloring it by any of its feasible colors reduces substantially the total size of the lists of feasible colors for the remaining vertices of V, which helps to approach the first situation, that is, the case when there are many colorless vertices. This process can be represented by a tree in which every internal node corresponds to a restricting vertex v, and every edge from v to a child corresponds to a feasible color for v. The tree will not be very large. Indeed, each of its internal vertices has at most three children, and its depth cannot exceed $3/\varepsilon$, as the total size of the lists of feasible colors at the beginning is $3n$, and this size is reduced by at least εn in each step. It thus suffices to show that with high probability the construction of the whole tree (until no feasible colors are left to any of its leaves) can be completed using the vertices in our random set R.

We proceed with the formal proof. For a subset $S \subseteq V$, a 3-coloring of it $\phi : S \to C$, and a vertex $v \in V \setminus S$, let $L_\phi(v)$ be the set of all colors in C besides those that appear already on some neighbor of v. This is the set of feasible colors for v. Clearly, for $S = \emptyset$, $L_\phi(v) = C$ for every $v \in V$. A vertex $v \in V \setminus S$ is called *colorless* if $L_\phi(v) = \emptyset$. Let U denote the set of all colorless vertices under (S, ϕ).

For every vertex $v \in V \setminus (S \cup U)$, define

$$\delta_\phi(v) = \min_{i \in L_\phi(v)} |\{u \in N(v) \setminus (S \cup U) : i \in L_\phi(u)\}|.$$

Therefore, coloring v by any one of the colors from $L_\phi(v)$ and then adding it to S will result in deleting this color and thus shortening the lists of feasible colors of at least $\delta_\phi(v)$ neighbors of v outside S.

Claim 17.2.2 *For every set $S \subset V$ and every 3-coloring ϕ of S, the graph G is at most $(n-1)|S \cup U| + \frac{1}{2}\sum_{v \in V \setminus (S \cup U)} \delta_\phi(v)$ edges far from being k-colorable.*

Proof. Consider the following coloring of G: every $v \in S$ is colored by $\phi(v)$, every $v \in U$ is colored by an arbitrary color, and every $v \in V \setminus (S \cup U)$ is colored by a color $i \in L_\phi(v)$ for which $\delta_\phi(v) = |\{u \in N(v) \setminus (S \cup U) : i \in L_\phi(u)\}|$. The number of monochromatic edges incident with $S \cup U$ is at most $(n-1)|S \cup U|$. Every vertex $v \in V \setminus (S \cap U)$ has exactly $\delta_\phi(v)$ neighbors $u \in V \setminus (S \cup U)$, whose color list $L_\phi(v)$ contains the color chosen for v. Therefore, v will have at most $\delta_\phi(v)$ neighbors in $V \setminus (S \cup U)$ colored in the same color as v itself. Hence the total number of monochromatic edges is at most $(n-1)|S \cup U| + \frac{1}{2}\sum_{v \in V \setminus (S \cup U)} \delta_\phi(v)$, as claimed. ■

Given a pair (S, ϕ), a vertex $v \in V \setminus (S \cup U)$ is called *restricting* if $\delta_\phi(v) \geq \varepsilon n$. We denote by W the set of all restricting vertices.

Claim 17.2.3 *For every pair (S, ϕ), where $S \subset V$ and $\phi : S \to C$, $|U \cup S \cup W| \geq \frac{\varepsilon n}{2}$.*

Proof. By the previous claim, and since G is ε-far from being 3-colorable,

$$\varepsilon n^2 < n(|S| + |U|) + \frac{1}{2} \sum_{v \in V \setminus (S \cup U)} \delta_\phi(v)$$

$$\leq n(|S| + |U|) + \frac{1}{2}|W|(n-1) + \frac{1}{2} \sum_{v \in V \setminus (S \cup U \cup W)} \delta_\phi(v) < n(|S| + |U| + |W|) + \frac{1}{2}\varepsilon n^2.$$

■

Returning to our randomly chosen vertices $r_1, \ldots, r_s$ of R, construct an auxiliary ternary tree T. To distinguish between the vertices of G and those of T, we call the latter nodes. Each node of T is labeled either by a vertex of G or by the special symbol #, whose meaning will be explained in what follows. If a node t of T is labeled by #, then t is called a *terminal node*. The edges of T are labeled by integers from C.

Let t be a node of T. Consider the path from the root of T to t, not including t itself. The labels of the nodes along this path form a subset $S(t)$ of V. The labels of the edges along the path define a 3-coloring $\phi(t)$ of $S(t)$ in a natural way: the label of the edge following a node t' in the path determines the color of its label $v(t')$. The labeling of the nodes and edges of T will have the following property: if t is labeled by v and v has a neighbor in $S(t)$ whose color in $\phi(t)$ is i, then the son of v along the edge labeled by i is labeled by #. This label indicates the fact that in this case color i is infeasible for v, given $(S(t), \phi(t))$.

At each step of the construction of T, we will maintain the following: all leafs of T are either unlabeled or are labeled by #. Also, only leafs of T can be labeled by #. We start the construction of T from an unlabeled single node, the root of T.

Suppose that $j-1$ vertices of T have already been chosen, and we are about to choose vertex r_j of R. Consider a leaf t of T. If t is labeled by #, we do nothing for this leaf. (That is the reason why such a t is called a terminal node; nothing will ever grow out of it.) Assume now that t is unlabeled. Define the pair $(S(t), \phi(t))$ as described above. Now, for the pair $(S(t), \phi(t))$, we define the set $U(t)$ of colorless vertices and the set $W(t)$ of restricting vertices as described before. Round j is called *successful* for the node t if the random vertex r_j satisfies $r_j \in U(t) \cup W(t)$. If round j is indeed successful for t, then we label t by r_j, create 3 sons of t, and label the corresponding edges by $1, 2, 3$. Now, if color i is infeasible for r_j, given $(S(t), \phi(t))$, we label the son of t along the edge with label i by #; otherwise we leave this son unlabeled. Note that if $r_j \in U(t)$, then none of the colors from C is feasible for r_j, and thus all the sons of t will be labeled by #. This completes the description of the process of constructing T. As each edge along a path from a root to a leaf of the tree corresponds to a restricting vertex, and the total size of all lists starts with $3n$ and is reduced by at least εn with each coloring of a restricting vertex, we have

Claim 17.2.4 *The depth of T is at most $\frac{3}{\varepsilon}$.* ■

Our construction also implies that, if a leaf t^* of T is labeled by #, then $\phi(t^*)$ is not a proper 3-coloring of $S(t^*)$. We thus have:

Claim 17.2.5 *If after round j all leafs of the tree T are terminal nodes, then the induced subgraph of G on $\{r_1, \ldots, r_j\}$ is not 3-colorable.* ■

To complete the proof, it suffices to show that

Claim 17.2.6 *After $s = 40/\varepsilon^2$ rounds, with probability at least 0.9, all leaves of T are terminal nodes.*

Proof. As every non-leaf node of T has at most three sons and by Claim 17.2.4 the depth of T is at most $3/\varepsilon$, it can be embedded naturally in the ternary tree $T_{3,\frac{3}{\varepsilon}}$ of depth $3/\varepsilon$. Moreover, this embedding can be prefixed even before exposing R and T. Note that the number of vertices of $T_{3,\frac{3}{\varepsilon}}$ is $1 + 3 + \cdots + 3^{\frac{3}{\varepsilon}} < 3^{\frac{3}{\varepsilon}+1}$.

Recall that, during the construction of the random sample R and the tree T, a successful round for a leaf t of T resulted in creating three sons of t. Fix a node t of $T_{3,\frac{3}{\varepsilon}}$. If after $40/\varepsilon^2$ rounds t is a leaf of T, then the total number of successful rounds for the path from the root of T to t is equal to the depth of t. As $S(t) \subseteq R$ and thus $|S(t)| \leq \frac{40}{\varepsilon^2} \leq \frac{\varepsilon n}{10}$, by Claim 17.2.3 each round has probability of success at least 0.4ε. Therefore, the probability that t is a nonterminal leaf of T after $40/\varepsilon^2$ steps can be bounded from above by the probability that the binomial random variable $B(40/\varepsilon^2, 0.4\varepsilon)$ is at most $3/\varepsilon$. The latter probability is at most

$$e^{-\frac{(16/\varepsilon - 3/\varepsilon)^2}{2 \cdot 16/\varepsilon}} = e^{-\frac{169}{32\varepsilon}}.$$

Thus by the union bound we conclude that the probability that some node of $T_{s,3/\varepsilon}$ is a leaf of T, not labeled by '#', is at most

$$|V(T_{3,\frac{3}{\varepsilon}})| e^{-\frac{169}{32\varepsilon}} < \frac{1}{10}.$$

■

The assertion of the theorem follows from Claims 17.2.5 and 17.2.6. ■

17.3 TESTING TRIANGLE-FREENESS

The relevance of the Regularity Lemma discussed in Chapter 9 to property testing is nicely illustrated in the proof that the property of containing no triangle is testable with one-sided error. The required combinatorial lemma here is the (intuitive, yet nontrivial) fact that, if one has to delete at least εn^2 edges of an n-vertex graph to destroy all triangles in it, then the graph must contain at least δn^3 triangles, where $\delta = \delta(\varepsilon) > 0$. As shown in Exercises, following Ruzsa and Szemerédi (1978), this fact implies that any set of integers with positive upper density contains a three-term arithmetic progression.

Lemma 17.3.1 *For any positive $\varepsilon < 1$, there is a $\delta = \delta(\varepsilon) > 0$ so that, if $G = (V, E)$ is a graph on $|V| = n$ vertices which is ε-far from being triangle-free, then it contains at least δn^3 triangles.*

Proof. We prove the lemma with

$$\delta = \frac{\varepsilon^3}{2^9 T^3(\varepsilon/4, \lceil 4/\varepsilon \rceil)},$$

where T is as in Theorem 9.4.1. Let $G = (V, E)$ satisfy the assumption. Note, first, that if $n < T(\varepsilon/4, \lceil 4/\varepsilon \rceil)$, then the assertion is trivial, as in this case δn^3 is less than 1, and it is trivial that if G is ε-far from being triangle-free then it contains a triangle. We thus assume that n is at least $T(\varepsilon, t)$, where $t = \lceil 4/\varepsilon \rceil$. By Theorem 9.4.1,

there is an $\varepsilon/4$-regular partition $(V_0, V_1, \ldots, V_k)$ of G, where $t \leq k \leq T = T(\varepsilon, t)$. Put $c = |V_1| = |V_2| = \cdots = |V_k|$. Let G' be the graph obtained from G by deleting the following edges:

- All edges of G that are incident with a vertex of the exceptional set V_0 (there are less than $\varepsilon n^2/4$ such edges);
- All edges of G that lie inside some set V_i (there are less than $\varepsilon n^2/8$ such edges);
- All edges of G that lie in irregular pairs (the number of such edges is at most $\frac{\varepsilon}{4}k^2c^2 \leq \varepsilon n^2/4$);
- All edges of G that lie in regular pairs (V_i, V_j), where the density $d(V_i, V_j)$ is smaller than $\varepsilon/2$. (There are less than $\binom{k}{2}\frac{\varepsilon}{2}c^2 < \varepsilon n^2/4$ such edges.)

Since G' is obtained from G by deleting less than εn^2 edges, it contains a triangle, as G is ε-far from being triangle-free. By the definition of G', the vertices of this triangle must lie in three distinct sets V_i, any two of which form a regular pair of density at least $\varepsilon/2$. Without loss of generality, assume that these sets are V_1, V_2, V_3. Call a vertex $v_1 \in V_1$ *typical* if it has at least $\varepsilon c/4$ neighbors in V_2 and at least $\varepsilon c/4$ neighbors in V_3. We claim that all vertices of V_1 but at most $2\frac{\varepsilon}{4}c < c/2$ are typical. Indeed, if X_1 is the set of all vertices of V_1 that have less than $\varepsilon c/4$ neighbors in V_2, then its cardinality must be smaller than $\varepsilon c/4$, since otherwise the pair X_1 and $X_2 = V_2$, together with the fact that $d(V_1, V_2) \geq \varepsilon/2$ would violate the $\varepsilon/4$-regularity of this pair. Similarly, there are less than $\varepsilon c/4$ vertices of V_1 that have less than $\varepsilon c/4$ neighbors in V_3, proving the claim.

Fix a typical vertex $v_1 \in V_1$, and let N_2, N_3 denote the sets of all its neighbors in V_2 and V_3, respectively. Thus, $|N_2|, |N_3| \geq \varepsilon c/4$, and hence, by the $\frac{\varepsilon}{4}$-regularity of the pair (V_2, V_3) and the fact that its density is at least $\varepsilon/2$, there are at least $\frac{\varepsilon}{4}|N_2||N_3| \geq (\varepsilon/4)^3c^2$ edges between N_2 and N_3. We conclude that v_1 lies in at least $(\varepsilon/4)^3c^2$ triangles. As there are at least $c/2$ typical vertices in V_1, and since

$$c^3 \geq (1 - \varepsilon/4)^3 n^3/T^3 > \frac{n^3}{4T^3},$$

the desired result follows. ∎

Corollary 17.3.2 *The property of being triangle-free is testable with one-sided error.*

Proof. For $\varepsilon > 0$, let $\delta = \delta(\varepsilon)$ be as in Lemma 17.3.1. Given a graph $G = (V, E)$ on n vertices, consider the following randomized algorithm for testing whether G is triangle-free. Let s be a confidence parameter. Pick randomly and independently s/δ triples of vertices of the graph, and check whether at least one of them forms a triangle. If so, then report that the graph is not triangle-free, otherwise, report that the graph is triangle-free. Clearly, if G is triangle-free, the algorithm will decide so. If it is ε-far from being triangle-free, then, by Lemma 17.3.1, the probability that the algorithm will err and report that G is triangle-free does not exceed $(1 - 6\delta)^{s/\delta} \leq e^{-6s}$. This completes the proof. ∎

17.4 CHARACTERIZING THE TESTABLE GRAPH PROPERTIES

In this section we describe several more recent results on graph property testing. The proofs of these results apply a strong variant of the Regularity Lemma, proved in Alon et al. (2000). The detailed proofs are somewhat technical, and will not be given here.

A graph property is *monotone* if it closed under removing vertices and edges. Thus, being k-colorable or triangle-free is a monotone property. A property is *hereditary* if it is closed under removal of vertices (and not necessarily under removal of edges). Clearly, every monotone graph property is also hereditary, but there are also many well-studied hereditary properties that are not monotone. Examples are being a perfect graph, a chordal graph, an interval graph, and many more. The results discussed in the previous subsections deal with two special cases of hereditary properties that are also monotone, namely being triangle-free and being k-colorable. Handling hereditary non-monotone graph properties, such as being perfect or not containing an induced cycle of length 4, is more involved than handling monotone properties.

For a (possibly infinite) family of graphs $\mathcal{F}$, a graph G is said to be *induced* $\mathcal{F}$-free if it contains no $F \in \mathcal{F}$ as an induced subgraph. The following lemma is not difficult.

Lemma 17.4.1 *Let $\mathcal{F}$ be a (possibly infinite) family of graphs, and suppose there are functions $f_{\mathcal{F}}(\varepsilon)$ and $\delta_{\mathcal{F}}(\varepsilon)$ such that the following holds for every $\varepsilon > 0$: every graph G on n vertices, which is ε-far from being **induced** $\mathcal{F}$-free, contains at least $\delta_{\mathcal{F}}(\varepsilon)n^f$ **induced** copies of a graph $F \in \mathcal{F}$ of size $f \leq f_{\mathcal{F}}(\varepsilon)$. Then, being induced $\mathcal{F}$-free is testable with one-sided error.*

The following general result is proved in Alon and Shapira (2005). A subsequent different, elegant but noneffective, proof can be found in Lovász and Szegedy (2010).

Theorem 17.4.2 Alon and Shapira (2005) *For any (possibly infinite) family of graph $\mathcal{F}$, there are functions $f_{\mathcal{F}}(\varepsilon)$ and $\delta_{\mathcal{F}}(\varepsilon)$ satisfying the conditions of Lemma 17.4.1.*

It is easy to see that one can define for any hereditary property $\mathcal{P}$ a (possibly infinite) family of graphs $\mathcal{F}_{\mathcal{P}}$ such that satisfying $\mathcal{P}$ is equivalent to being *induced* $\mathcal{F}_{\mathcal{P}}$-free. Indeed, we simply put a graph F in $\mathcal{F}_{\mathcal{P}}$ if and only if F does not satisfy $\mathcal{P}$. It thus follows that Theorem 17.4.2, combined with Lemma 17.4.1, implies the following.

Theorem 17.4.3 *Every hereditary graph property is testable with one-sided error.*

An easy consequence of Theorem 17.4.2 is the following:

Corollary 17.4.4 *For every hereditary graph property $\mathcal{P}$, there is a function $W_{\mathcal{P}}(\varepsilon)$ with the following property: If G is ε-far from satisfying $\mathcal{P}$, then G contains an induced subgraph of size at most $W_{\mathcal{P}}(\varepsilon)$, which does not satisfy $\mathcal{P}$.*

Using Theorem 17.4.3, one can obtain a characterization of the "natural" graph properties, which are testable with one-sided error.

Definition (Oblivious Tester)

A tester (one-sided or two-sided) is said to be *oblivious* if it works as follows: given ε, the tester computes an integer $Q = Q(\varepsilon)$ and asks an oracle for a subgraph induced by a set of vertices S of size Q, where the oracle chooses S randomly and uniformly from the vertices of the input graph. If Q is larger than the size of the input graph, then the oracle returns the entire graph. The tester then accepts or rejects according to the graph induced by S.

In some sense, oblivious testers capture the essence of property testing, as essentially all the testers that have been analyzed in the literature are in fact oblivious, or could easily be turned into oblivious testers. Clearly, some properties cannot have oblivious testers, but these properties are not natural. An example is the property of not containing an induced cycle of length 4 if the number of vertices is even, and not containing an induced cycle of length 5 if the number of vertices is odd.

Using Theorem 17.4.3, it can be shown that, if one considers only oblivious testers, then it is possible to precisely characterize the graph properties that are testable with one-sided error. To state this characterization, we need the following definition:

Definition (Semi-Hereditary)

A graph property $\mathcal{P}$ is semi-hereditary if there exists a hereditary graph property $\mathcal{H}$ such that the following holds:

1. Any graph satisfying $\mathcal{P}$ also satisfies $\mathcal{H}$.
2. For any $\varepsilon > 0$, there is an $M(\varepsilon)$, such that any graph of size at least $M(\varepsilon)$, which is ε-far from satisfying $\mathcal{P}$, does not satisfy $\mathcal{H}$.

Clearly, any hereditary graph property $\mathcal{P}$ is also semi-hereditary because we can take $\mathcal{H}$ in the above definition to be $\mathcal{P}$ itself. In simple words, a semi-hereditary $\mathcal{P}$ is obtained by taking a hereditary graph property $\mathcal{H}$ and removing from it a (possibly infinite, carefully chosen) set of graphs. This means that the first item in the definition above is satisfied. The only restriction, which is needed in order to get item 2 in the definition, is that $\mathcal{P}$ will be such that for any $\varepsilon > 0$ there will be only finitely many graphs that are ε-far from satisfying it, and yet satisfy $\mathcal{H}$. We are now ready to state the characterization.

Theorem 17.4.5 *A graph property $\mathcal{P}$ has an oblivious one-sided tester if and only if $\mathcal{P}$ is semi-hereditary.*

The proof can be found in Alon and Shapira (2005). The Regularity Lemma and its strong variant mentioned in the beginning of this subsection play a crucial role in this proof. This is not a coincidence. In Alon et al. (2006), it is shown that the property defined by having any given Szemerédi-partition is testable with a constant number of queries. This leads to a combinatorial characterization of the graph properties that are testable with a constant number of queries. This characterization (roughly) says

that a graph property $\mathcal{P}$ can be tested by a two-sided error tester with a constant number of queries if and only if testing $\mathcal{P}$ can be reduced to testing the property of satisfying one of finitely many Szemerédi-partitions. See Alon et al. (2006) for the precise formulation and detailed proof.

17.5 EXERCISES

1. (The removal lemma) Show that, for any fixed graph H on h vertices and for any $\varepsilon > 0$, there is a $\delta = \delta(\varepsilon, H) > 0$ so that, if one has to delete at least εn^2 edges from an n-vertex graph G to destroy all copies of H, then G contains at least δn^h copies of H.
2. ($*$) Using Lemma 17.3.1 prove that for any $\varepsilon > 0$ there is an n_0 so that, if $n > n_0$, then every subset $A \subset \{1, 2, \ldots, n\}$ of size $|A| \geq \varepsilon n$ contains a three-term arithmetic progression.
3. A graph is chordal if any cycle of length at least 4 in it has a chord. Apply Corollary 17.4.4 to show that, for every $\varepsilon > 0$, there is a $k = k(\varepsilon)$ so that every graph on n vertices in which every cycle of length at least 4 and at most k has a chord can be transformed into a chordal graph by adding and/or deleting at most εn^2 edges.
4. ($*$) A construction of Behrend (1946) gives a subset A of $\{1, 2, \ldots, m\}$ of size $|A| \geq m/e^{c\sqrt{\log m}}$ with no three-term arithmetic progression. Show how to construct from such an A a graph on n vertices, which is ε-far from being triangle-free and yet contains only $\varepsilon^{b \log(1/\varepsilon)} n^3$ triangles.
5. Prove that the property of being triangle-free is not testable with a one-sided error tester whose query complexity is polynomial in $1/\varepsilon$.
6. A graph G is H-free if it contains no copy of H. Prove that, for every bipartite graph H with h vertices, there is a $c = c(h) > 0$ so that any graph G on n vertices that is ε-far from being H-free contains at least $\varepsilon^c n^h$ copies of H.

THE PROBABILISTIC LENS: Turán Numbers and Dependent Random Choice

For a graph H and an integer n, the Turán number $ex(n, H)$ is the maximum possible number of edges in a simple graph on n vertices that contains no copy of H. The asymptotic behavior of these numbers for graphs of chromatic number at least 3 is well known, see, for example, Exercise 6 in Chapter 9. For bipartite graphs H, however, the situation is considerably more complicated, and there are relatively few nontrivial bipartite graphs H for which the order of magnitude of $ex(n, H)$ is known. Here we prove that, for every fixed bipartite graph H in which the degrees of all vertices in one color class are at most r, there is a constant $c = c(H)$ so that $ex(n, H) \leq cn^{2-1/r}$. This is tight for all values of r, as it is known that for every r and $t > (r-1)!$ there is a simple graph with n vertices and at least $c_{r,t}n^{2-1/r}$ edges, containing no copy of the complete bipartite graph $K_{r,t}$.

The basic tool in the proof is a simple and yet surprisingly powerful method, whose probabilistic proof may be called "dependent random choice", as it involves a random selection of a set of vertices, where the choices are dependent in a way that increases the probability that r-tuples of selected vertices will have many common neighbors. An early variant of this lemma was first proved in Kostochka and Rödl (2004) and Gowers (1998). The proof given here is from Alon, Krivelevich and Sudakov (2003).

Lemma 1 *Let a, b, n, r be positive integers. Let $G = (V, E)$ be a graph on $|V| = n$ vertices with average degree $d = 2|E|/n$. If*

$$\frac{d^r}{n^{r-1}} - \binom{n}{r}\left(\frac{b-1}{n}\right)^r > a - 1\,, \tag{1}$$

then G contains a subset A_0 of at least a vertices so that every r vertices of A_0 have at least b common neighbors.

Proof. Let T be a (multi)-set of r random vertices of G, chosen uniformly with repetitions. Set

$$A = \{v \in V : T \subseteq N(v)\},$$

where $N(v)$ denotes the set of all neighbors of v. Denote by X the cardinality of A. By Linearity of Expectation

$$\begin{aligned} \mathrm{E}[X] &= \sum_{v \in V} \left(\frac{|N(v)|}{n} \right)^r = \frac{1}{n^r} \sum_{v \in V} |N(v)|^r \\ &\geq \frac{1}{n^r} n \left(\frac{\sum_{v \in V} |N(v)|}{n} \right)^r = \frac{1}{n^{r-1}} \left(\frac{2|E|}{n} \right)^r = \frac{d^r}{n^{r-1}}, \end{aligned}$$

where the inequality follows from the convexity of $f(x) = x^r$.

Let Y denote the random variable counting the number of r-tuples in A with fewer than b common neighbors. For a given r-tuple $R \subseteq V$, the probability that R will be a subset of A is precisely $\left(\frac{|N^*(R)|}{n} \right)^r$, where $N^*(R)$ denotes the set of all common neighbors of the vertices in R. As there are at most $\binom{n}{r}$ subsets R of cardinality $|R| = r$ for which $|N^*(R)| \leq b - 1$, it follows that

$$\mathrm{E}[Y] \leq \binom{n}{r} \left(\frac{b-1}{n} \right)^r.$$

Applying Linearity of Expectation once again, we conclude, by (1), that

$$\mathrm{E}[X - Y] \geq \frac{d^r}{n^{r-1}} - \binom{n}{r} \left(\frac{b-1}{n} \right)^r > a - 1.$$

Hence there exists a choice for T so that for the corresponding set A we get $X - Y \geq a$. Pick such a set, and omit a point from every r-tuple in it with fewer than b common neighbors. This gives a set A_0 of at least a vertices so that every r vertices in it have at least b common neighbors. ■

Theorem 2 *Let H be a bipartite graph with maximum degree r on one side. Then there exists a constant $c = c(H) > 0$ such that*

$$ex(n, H) < cn^{2-\frac{1}{r}}.$$

Proof. Let A and B be the vertex classes of H, and suppose $|A| = a$, $|A| + |B| = b$, where the degree of every vertex $b \in B$ in H does not exceed r. Let $G = (V, E)$ be a graph on $|V| = n$ vertices with average degree $d = 2|E|/n \geq cn^{1-1/r}$, where c satisfies $c^r > \frac{(b-1)^r}{r!} + a - 1$. It is easy to check that (1) holds. To complete the proof,

it suffices to show that G must contain a copy of H. By Lemma 1, there is a subset $A_0 \subset V$ of cardinality $|A_0| = a$ so that every r-subset of A_0 has at least b common neighbors in G. It is now an easy matter to embed H in G. To do so, start by embedding the vertices of A by an arbitrary injective function from A to A_0. Proceed by embedding the vertices of B one by one in an arbitrary order, making sure that in each step the image of the new embedded vertex is connected to the images of its neighbors in H and is different from the images of all previously embedded vertices. Since every set of (at most) r vertices of A_0 has at least b common neighbors in G, this process can be performed until the images of all b vertices of H are found. This completes the proof. ■

Appendix A
Bounding of Large Deviations

A.1 CHERNOFF BOUNDS

We give here some basic bounds on large deviations that are useful when employing the probabilistic method. Our treatment is self-contained. Most of the results may be found in, or immediately derived from, the seminal paper of Chernoff (1952). While we are guided by asymptotic considerations, the inequalities are proven for all values of the parameters in the specified region. The first result, while specialized, contains basic ideas found throughout the Appendix.

Theorem A.1.1 *Let $X_i, 1 \leq i \leq n$, be mutually independent random variables with*

$$\Pr\left[X_i = +1\right] = \Pr\left[X_i = -1\right] = \frac{1}{2}$$

and set, following the usual convention,

$$S_n = X_1 + \cdots + X_n.$$

Let $a > 0$. Then

$$\Pr\left[S_n > a\right] < e^{-a^2/2n}.$$

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

We require Markov's inequality, which is as follows: Suppose that Y is an arbitrary nonnegative random variable, $\alpha > 0$. Then

$$\Pr[Y > \alpha E[Y]] < 1/\alpha.$$

Proof [A.1.1] Fix n, a, and let, for the moment, $\lambda > 0$ be arbitrary. For $1 \leq i \leq n$,

$$\mathrm{E}\left[e^{\lambda X_i}\right] = (e^{\lambda} + e^{-\lambda})/2 = \cosh(\lambda).$$

We require the inequality

$$\cosh(\lambda) \leq e^{\lambda^2/2},$$

valid for all $\lambda > 0$, the special case $\alpha = 0$ of Lemma A.1.5 below. (The inequality may be more easily shown by comparing the Taylor series of the two functions termwise.)

$$e^{\lambda S_n} = \prod_{i=1}^{n} e^{\lambda X_i}.$$

Since the X_i are mutually independent, so are the $e^{\lambda X_i}$, expectations multiply and

$$\mathrm{E}\left[e^{\lambda S_n}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{\lambda X_i}\right] = [\cosh(\lambda)]^n < e^{\lambda^2 n/2}.$$

We note that $S_n > a$ if and only if $e^{\lambda S_n} > e^{\lambda a}$, and apply Markov's inequality so that

$$\Pr\left[S_n > a\right] = \Pr\left[e^{\lambda S_n} > e^{\lambda a}\right] < \mathrm{E}\left[e^{\lambda S_n}\right] / e^{\lambda a} \leq e^{\lambda^2 n/2 - \lambda a}.$$

We set $\lambda = a/n$ to optimize the inequality $\Pr\left[S_n > a\right] < e^{-a^2/2n}$, as claimed. ∎

By symmetry, we immediately have:

Corollary A.1.2 *Under the assumptions of Theorem A.1.1,*

$$\Pr\left[\left|S_n\right| > a\right] < 2e^{-a^2/2n}.$$

The proof of Theorem A.1.1 illustrates the basic idea of the Chernoff bounds. We wish to bound $\Pr[X > a]$ for some random variable X. For any positive λ, we bound

$$\Pr[X > a] = \Pr\left[e^{\lambda X} > e^{\lambda a}\right] \leq \mathrm{E}\left[e^{\lambda X}\right] e^{-\lambda a}.$$

The core idea of the Chernoff bounds is to select the λ that minimizes $\mathrm{E}\left[e^{\lambda X}\right] e^{-\lambda a}$. The art to the Chernoff bounds is to select a λ that is reasonably close to optimal and easy to work with, yielding upper bounds on $\Pr[X > a]$, which are, one hopes,

good enough for our purposes. Bounds on $\Pr[X < a]$ are similar. For any positive λ, we bound

$$\Pr[X < a] = \Pr\left[e^{-\lambda X} > e^{-\lambda a}\right] \leq \mathrm{E}\left[e^{-\lambda X}\right] e^{\lambda a}.$$

Chernoff-bound arguments tend to be cleaner when $\mathrm{E}[X] = 0$. A simple translation, replacing X by $X - \mu$ where $\mu = \mathrm{E}[X]$, is often quite helpful.

It is instructive to examine the case when N is the standard normal distribution and a is positive. In this instance, $\mathrm{E}\left[e^{\lambda N}\right] = e^{\lambda^2/2}$ and so

$$\Pr[N > a] = \Pr\left[e^{\lambda N} > e^{\lambda a}\right] \leq \mathrm{E}\left[e^{\lambda X}\right] e^{-\lambda a} = e^{\lambda^2/2 - \lambda a}.$$

Elementary calculus leads to the optimal choice $\lambda = a$, so that

$$\Pr[N > a] < e^{-a^2/2}.$$

This compares well with the actual asymptotics

$$\Pr[N > a] = (2\pi)^{-1/2} \int_a^\infty e^{-t^2/2} dt \sim (2\pi a)^{-1/2} e^{-a^2/2}$$

as $a \to \infty$. Results with N being normal with mean μ and variance σ^2 are similarly good. This explains, to some extent, the efficacy of the Chernoff bounds. When a random variable X is "roughly" normal, the Chernoff bounds on $\Pr[X > a]$ should be quite close to the actual values for a large. In practice, however, precise calculations of $\mathrm{E}\left[e^{\lambda X}\right]$ can be difficult or impossible to achieve, and there can be considerable art in finding approximations for $\mathrm{E}\left[e^{\lambda X}\right]$ that will allow good bounds on $\Pr[X > a]$.

Many of our remaining results will deal with distributions X of the following prescribed type:

Assumptions A.1.3

$$p_1, \ldots, p_n \in [0, 1]$$
$$p = (p_1 + \cdots + p_n)/n$$
$$X_1, \ldots, X_n \text{ mutually independent with}$$
$$\Pr\left[X_i = 1 - p_i\right] = p_i$$
$$\Pr\left[X_i = -p_i\right] = 1 - p_i$$
$$X = X_1 + \cdots + X_n.$$

Remark. Clearly, $\mathrm{E}[X] = \mathrm{E}\left[X_i\right] = 0$. When all $p_i = 1/2$, X has distribution $S_n/2$. When all $p_i = p$, X has distribution $B(n, p) - np$, where $B(n, p)$ is the usual binomial distribution.

Theorem A.1.4 *Under assumptions A.1.3 and with $a > 0$,*

$$\Pr[X > a] < e^{-2a^2/n}.$$

Lemma A.1.5 *For all reals α, β with $|\alpha| \leq 1$,*

$$\cosh(\beta) + \alpha \sinh(\beta) \leq e^{\beta^2/2 + \alpha\beta}.$$

Proof. This is immediate if $\alpha = +1$ or $\alpha = -1$ or $|\beta| \geq 100$. If the Lemma were false, the function

$$f(\alpha, \beta) = \cosh(\beta) + \alpha \sinh(\beta) - e^{\beta^2/2 + \alpha\beta}$$

would assume a positive global maximum in the interior of the rectangle

$$R = \{(\alpha, \beta) : |\alpha| \leq 1, |\beta| \leq 100\}.$$

Setting partial derivatives equal to zero, we find

$$\sinh(\beta) + \alpha \cosh(\beta) = (\alpha + \beta) e^{\beta^2/2 + \alpha\beta},$$

$$\sinh(\beta) = \beta e^{\beta^2/2 + \alpha\beta},$$

and thus $\tanh(\beta) = \beta$ which implies $\beta = 0$. But $f(\alpha, 0) = 0$ for all α, a contradiction. ■

Lemma A.1.6 *For all $\theta \in [0, 1]$ and all λ,*

$$\theta e^{\lambda(1-\theta)} + (1 - \theta) e^{-\lambda\theta} \leq e^{\lambda^2/8}.$$

Proof. Setting $\theta = (1 + \alpha)/2$ and $\lambda = 2\beta$, Lemma A.1.6 reduces to Lemma A.5. ■

Proof [Theorem A.1.4] Let, for the moment, $\lambda > 0$ be arbitrary.

$$\mathrm{E}\left[e^{\lambda X_i}\right] = p_i e^{\lambda(1-p_i)} + (1 - p_i) e^{-\lambda p_i} \leq e^{\lambda^2/8}$$

by Lemma A.1.6. Then

$$\mathrm{E}\left[e^{\lambda X}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{\lambda X_i}\right] \leq e^{\lambda^2 n/8}.$$

Applying Markov's inequality,

$$\Pr[X > a] = \Pr\left[e^{\lambda X} > e^{\lambda a}\right] < \mathrm{E}\left[e^{\lambda X}\right] / e^{\lambda a} \leq e^{\lambda^2 n/8 - \lambda a}.$$

We set $\lambda = 4a/n$ to optimize the inequality $\Pr[X > a] < e^{-2a^2/n}$, as claimed. ■

Again, by symmetry we immediately have the following:

Corollary A.1.7 *Under assumptions A.1.3 and with $a > 0$,*

$$\Pr[|X| > a] < 2e^{-2a^2/n}.$$

Under assumptions A.1.3 with λ arbitrary,

$$\begin{aligned}\mathrm{E}\left[e^{\lambda X}\right] &= \prod_{i=1}^{n} \mathrm{E}\left[e^{\lambda X_i}\right] = \prod_{i=1}^{n} [p_i e^{\lambda(1-p_i)} + (1-p_i)e^{-\lambda p_i}] \\ &= e^{-\lambda pn} \prod_{i=1}^{n} [p_i e^{\lambda} + (1-p_i)] \ .\end{aligned}$$

With λ fixed, the function

$$f(x) = \ln[xe^{\lambda} + 1 - x] = \ln[Bx + 1] \ \text{ with } B = e^{\lambda} - 1$$

is concave and hence (Jensen's Inequality)

$$\sum_{i=1}^{n} f(p_i) \leq nf(p).$$

Exponentiating both sides,

$$\prod_{i=1}^{n} [p_i e^{\lambda} + (1-p_i)] \leq [pe^{\lambda} + (1-p)]^n \ ,$$

so that we have:

Lemma A.1.8 *Under the assumptions A.1.3,*

$$\mathrm{E}\left[e^{\lambda X}\right] < e^{-\lambda pn}[pe^{\lambda} + (1-p)]^n.$$

Theorem A.1.9 *Under the assumptions A.1.3 and with $a > 0$,*

$$\Pr[X \geq a] < e^{-\lambda pn}[pe^{\lambda} + (1-p)]^n e^{-\lambda a}$$

for all $\lambda > 0$.

Proof. $\Pr[X > a] = \Pr\left[e^{\lambda X} > e^{\lambda a}\right] < \mathrm{E}\left[e^{\lambda X}\right] / e^{\lambda a}$. Now apply Lemma A.1.8. ■

Remark. For given p, n, a, an optimal assignment of λ in Theorem A.1.9 is found by elementary calculus to be

$$\lambda = \ln\left[\left(\frac{1-p}{p}\right)\left(\frac{a+np}{n-(a+np)}\right)\right].$$

This value is oftentimes too cumbersome to be useful. We employ suboptimal λ to achieve more convenient results.

Setting $\lambda = \ln[1 + a/pn]$ and using the fact that $(1 + a/n)^n \leq e^a$, Theorem A.1.9 implies:

Corollary A.1.10

$$\Pr[X \geq a] < e^{a-pn\ln(1+a/pn)-a\ln(1+a/pn)}.$$

Theorem A.1.11

$$\Pr[X \geq a] < e^{-a^2/2pn+a^3/2(pn)^2}.$$

Proof. With $u = a/pn$, apply the inequality

$$\ln(1+u) \geq u - u^2/2,$$

valid for all $u \geq 0$, to Corollary A.1.10. ∎

When all $p_i = p$, X has variance $np(1-p)$. With $p = o(1)$ and $a = o(pn)$, this bound reflects the approximation of X by a normal distribution with variance $\sim np$. The bound of Theorem A.1.11 hits a minimum at $a = 2pn/3$. For $a > 2pn/3$, we have the simple bound

$$\Pr[X > a] \leq \Pr\left[X > 2pn/3\right] < e^{-2pn/27}.$$

This is improved by the following.

Theorem A.1.12 *For $\beta > 1$,*

$$\Pr\left[X \geq (\beta-1)pn\right] < [e^{\beta-1}\beta^{-\beta}]^{pn}.$$

Proof. Direct "plug in" to Corollary A.1.10. ∎

$X + pn$ may be interpreted as the number of successes in n independent trials when the probability of success in the ith trial is p_i.

Theorem A.1.13 *Under assumptions A.1.3 and with $a > 0$,*

$$\Pr[X < -a] < e^{-a^2/2pn}.$$

Note that one cannot simply employ "symmetry", because then the roles of p and $1 - p$ are interchanged.

Proof. Let $\lambda > 0$ be, for the moment, arbitrary. Then by the argument preceding Lemma A.1.8,

$$\mathrm{E}\left[e^{-\lambda X}\right] \leq e^{\lambda pn}[pe^{-\lambda} + (1-p)]^n \ .$$

Thus

$$\Pr[X < -a] = \Pr\left[e^{-\lambda X} > e^{\lambda a}\right] < e^{\lambda pn}[pe^{-\lambda} + (1-p)]^n e^{-\lambda a},$$

analogous to Theorem A.1.9. We employ the inequality

$$1 + u \leq e^u,$$

valid for all u, so that

$$pe^{-\lambda} + (1-p) = 1 + (e^{-\lambda} - 1)p < e^{p(e^{-\lambda}-1)}$$

and

$$\Pr[X < -a] \leq e^{\lambda pn + np(e^{-\lambda}-1) - \lambda a} = e^{np(e^{-\lambda}-1+\lambda) - \lambda a}.$$

We employ the inequality

$$e^{-\lambda} \leq 1 - \lambda + \lambda^2/2 \ ,$$

valid for all $\lambda > 0$. (Note: The analogous inequality $e^{\lambda} \leq 1 + \lambda + \lambda^2/2$ is *not* valid for $\lambda > 0$ and so this method, when applied to $\Pr[X > a]$, requires an "error" term as the one found in Theorem A.1.11.) Now

$$\Pr[X < -a] \leq e^{np\lambda^2/2 - \lambda a} \ .$$

We set $\lambda = a/np$ to optimize the inequality $\Pr[X < -a] < e^{-a^2/2pn}$, as claimed. ∎

For clarity, the following result is often useful.

Corollary A.1.14 *Let Y be the sum of mutually independent indicator random variables, $\mu = \mathrm{E}[Y]$. For all $\epsilon > 0$,*

$$\Pr[|Y - \mu| > \epsilon\mu] < 2e^{-c_\epsilon \mu},$$

where $c_\epsilon > 0$ depends only on ϵ.

Proof. Apply Theorems A.1.12 and A.1.13 with $Y = X + pn$ and

$$c_\epsilon = \min\{-\ln(e^\epsilon(1+\epsilon)^{-(1+\epsilon)}), \epsilon^2/2\}.$$

∎

The asymmetry between $\Pr[X < a]$ and $\Pr[X > a]$ given by Theorems A.1.12 and A.1.13 is real. The estimation of X by a normal distribution with mean zero and variance np is roughly valid for estimating $\Pr[X < a]$ for any a and for estimating

$\Pr[X > a]$ while $a = o(np)$. But when a and np are comparable, or when $a \gg np$, the Poisson behavior "takes over" and $\Pr[X > a]$ *cannot* be accurately estimated by using the normal distribution.

We conclude with several large deviation results involving distributions other than sums of indicator random variables.

Theorem A.1.15 *Let P have Poisson distribution with mean μ. For $\epsilon > 0$,*

$$\Pr[P \le \mu(1-\epsilon)] \le e^{-\epsilon^2\mu/2},$$
$$\Pr[P \ge \mu(1+\epsilon)] \le \left[e^{\epsilon}(1+\epsilon)^{-(1+\epsilon)}\right]^{\mu}.$$

Proof. For any s

$$\Pr[P = s] = \lim_{n\to\infty} \Pr\left[B\left(n, \frac{\mu}{n}\right) = s\right].$$

Apply Theorems A.1.12 and A.1.13. ∎

Theorem A.1.16 *Let X_i, $1 \le i \le n$, be mutually independent with all $\mathrm{E}\left[X_i\right] = 0$ and all $|X_i| \le 1$. Set $S = X_1 + \cdots + X_n$. Then*

$$\Pr[S > a] < e^{-a^2/2n}.$$

Proof. Set, as in the proof of Theorem A.1.1, $\lambda = a/n$. Set

$$h(x) = \frac{e^{\lambda} + e^{-\lambda}}{2} + \frac{e^{\lambda} - e^{-\lambda}}{2}x\ .$$

For $x \in [-1, 1]$, $e^{\lambda x} \le h(x)$. ($y = h(x)$ is the chord through the points $x = \pm 1$ of the convex curve $y = e^{\lambda x}$.) Thus

$$\mathrm{E}\left[e^{\lambda X_i}\right] \le \mathrm{E}\left[h(X_i)\right] = h(\mathrm{E}\left[X_i\right]) = h(0) = \cosh\lambda.$$

The remainder of the proof follows as in Theorem A.1.1. ∎

Theorem A.1.17 *Suppose $\mathrm{E}[X] = 0$ and no two values of X are ever more than 1 apart. Then for all $\lambda \ge 0$,*

$$\mathrm{E}\left[e^{\lambda X}\right] \le e^{\lambda^2/8}.$$

Proof. Fix $b \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ with $X \in \left[\frac{-1+b}{2}, \frac{+1+b}{2}\right]$. Let $y = h(x)$ be the straight line intersecting the curve $y = e^{\lambda x}$ at the points $(\pm 1 + b)/2$. As $e^{\lambda x}$ is a convex function, $e^{\lambda x} \le h(x)$ for all $x \in [\frac{-1+b}{2}, \frac{+1+b}{2}]$. Thus

$$\mathrm{E}\left[e^{\lambda X}\right] \le \mathrm{E}\left[h(X)\right] = h\left(\mathrm{E}\left[X\right]\right) = h(0).$$

We calculate $h(0) = e^{\lambda b/2}[\cosh(\lambda/2) - b\sinh(\lambda/2)]$, which is at most $e^{\lambda^2/8}$ by Lemma A.1.5. ∎

Theorem A.1.18 *Let X_i, $1 \le i \le n$, be independent random variables with each* $\mathrm{E}\left[X_i\right] = 0$ *and no two values of any X_i ever more than* 1 *apart. (We allow, however, values of different X_i, X_j to be further apart.) Set $S = X_1 + \cdots + X_n$. Then*

$$\Pr\left[S > a\right] < e^{-2a^2/n}.$$

Proof. $\mathrm{E}\left[e^{\lambda S}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{\lambda X_i}\right] \le e^{n\lambda^2/8}$ by Theorem A.1.17. Then for $\lambda \ge 0$,

$$\Pr\left[S > a\right] = \Pr\left[e^{\lambda S} \ge e^{\lambda a}\right] \le \exp\left[\frac{n\lambda^2}{8} - \lambda a\right]$$

and we set $\lambda = 4a/n$. ∎

We have been roughly guided by the notion that, if X has mean zero and variance σ^2, then $\Pr\left[X \ge a\sigma\right]$ should go like $e^{-a^2/2}$. There are times when this idea is badly wrong. Consider Assumptions A.3 with all $p_i = 1/n$, so that $X = P_n - 1$ where P_n has the binomial distribution $B(n, 1/n)$ which is asymptotically P, the Poisson distribution with mean 1. Then $\mathrm{E}\left[X\right] = 0$ and $\mathrm{Var}\left[X\right] \sim 1$. For a fixed, $\Pr\left[X = a\right] \to \frac{1}{e(a+1)!}$, which is far bigger than $e^{-a^2/2}$. With this cautionary preamble, we give a general situation for which the notion is asymptotically correct when a is not too large.

Theorem A.1.19 *For every $C > 0$ and $\epsilon > 0$, there exists $\delta > 0$ so that the following holds: Let X_i, $1 \le i \le n$, n arbitrary, be independent random variables with* $\mathrm{E}[X_i] = 0$, $|X_i| \le C$, *and* $\mathrm{Var}[X_i] = \sigma_i^2$. *Set $X = \sum_{i=1}^{n} X_i$ and $\sigma^2 = \sum_{i=1}^{n} \sigma_i^2$ so that* $\mathrm{Var}[X] = \sigma^2$. *Then for $0 < a \le \delta\sigma$,*

$$\Pr\left[X > a\sigma\right] < e^{-\frac{a^2}{2}(1-\epsilon)}.$$

Proof. We set $\lambda = a/\sigma$ so that $0 \le \lambda \le \delta$. Then

$$\mathrm{E}\left[e^{\lambda X_i}\right] = \sum_{k=0}^{\infty} \mathrm{E}\left[\frac{\lambda^k}{k!} X_i^k\right] = 1 + \frac{\lambda^2}{2}\sigma_i^2 + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!} \mathrm{E}\left[X_i^k\right].$$

As $|X_i^k| \le C^{k-2} X_i^2$, we bound

$$\mathrm{E}\left[X_i^k\right] \le \mathrm{E}\left[|X_i^k|\right] \le C^{k-2}\mathrm{E}\left[X_i^2\right] = C^{k-2}\sigma_i^2.$$

For $k \ge 3$, we bound $\frac{2}{k!} \le \frac{1}{(k-2)!}$ so that

$$\mathrm{E}\left[e^{\lambda X_i}\right] \le 1 + \frac{\lambda^2}{2}\sigma_i^2 \left[1 + \sum_{k=3}^{\infty} \frac{(C\lambda)^{k-2}}{(k-2)!}\right] = 1 + \frac{\lambda^2}{2}\sigma_i^2 e^{\lambda C}.$$

We choose δ to satisfy $e^{C\delta} \leq 1 + \epsilon$. As $\lambda \leq \delta$,

$$\mathrm{E}\left[e^{\lambda X_i}\right] \leq 1 + \frac{\lambda^2}{2}\sigma_i^2(1+\epsilon) < \exp\left[\frac{\lambda^2}{2}\sigma_i^2(1+\epsilon)\right] .$$

This inequality has held for all X_i so

$$\mathrm{E}\left[e^{\lambda X}\right] = \prod_{i=1}^{n} \mathrm{E}\left[e^{\lambda X_i}\right] < \exp\left[\frac{\lambda^2}{2}\sigma^2(1+\epsilon)\right]$$

and

$$\Pr[X > a\sigma] \leq \mathrm{E}\left[e^{\lambda X}\right] e^{-\lambda a\sigma} < e^{-\frac{a^2}{2}(1-\epsilon)}.$$

■

A.2 LOWER BOUNDS

The Chernoff bounds of the previous section give upper bounds for $\Pr[X > a]$ by examining one value (albeit, the right one!) of the Laplace transform $\mathrm{E}\left[e^{\lambda X}\right]$. Here we use three values of the Laplace transform to give lower bounds for $\Pr[X > a]$. We shall set

$$f(\lambda) = \mathrm{E}\left[e^{\lambda X}\right]$$
$$g_a(\lambda) = f(\lambda)e^{-\lambda a}.$$

With this notation, $\Pr[X > a] \leq g_a(\lambda)$ and the Chernoff bound is achieved by taking that λ minimizing $g_a(\lambda)$. For any positive u and ϵ,

$$X \geq a + u \Rightarrow \lambda X \leq (\lambda + \epsilon)X - \epsilon a - \epsilon u$$
$$X \leq a - u \Rightarrow \lambda X \leq (\lambda - \epsilon)X + \epsilon a - \epsilon u,$$

so that

$$\mathrm{E}\left[e^{\lambda X}\chi(X \geq a + u)\right] \leq f(\lambda + \epsilon)e^{-\epsilon a}e^{-\epsilon u}$$
$$\mathrm{E}\left[e^{\lambda X}\chi(X \leq a - u)\right] \leq f(\lambda - \epsilon)e^{+\epsilon a}e^{-\epsilon u}.$$

Subtracting these from $\mathrm{E}\left[e^{\lambda X}\right]$,

$$\mathrm{E}\left[e^{\lambda X}\chi(|X - a| < u)\right] \geq f(\lambda) - e^{-\epsilon u}[f(\lambda + \epsilon)e^{-\epsilon a} + f(\lambda - \epsilon)e^{+\epsilon a}].$$

When $|X - a| < u$, $e^{\lambda X} \leq e^{\lambda u}e^{\lambda a}$, so

$$\Pr[|X - a| < u] \geq e^{-\lambda u}e^{-\lambda a}\mathrm{E}\left[e^{\lambda X}\chi(|X - a| < u)\right]$$

but $\Pr[X > a - u] \geq \Pr[|X - a| < u]$, giving our general result:

Theorem A.2.1 *For any a, u, λ, ϵ with $u, \lambda, \epsilon, \lambda - \epsilon$ all positive*

$$\Pr[X > a - u] \geq e^{-\lambda u}\left[g_a(\lambda) - e^{-\epsilon u}[g_a(\lambda + \epsilon) + g_a(\lambda - \epsilon)]\right].$$

We note that this bound has used only three values of the Laplace transform, $f(\lambda)$, $f(\lambda - \epsilon)$, and $f(\lambda + \epsilon)$.

It is instructive to examine the case when N is the standard normal distribution. We assume a is positive and are interested in the asymptotics as $a \to +\infty$. We set $\lambda = a$, so that $g_a(\lambda) = e^{-a^2/2}$. Now

$$g_a(\lambda \pm \epsilon) = e^{(\lambda \pm \epsilon)^2/2 - a(\lambda \pm \epsilon)} = g_a(\lambda)e^{\epsilon^2/2}.$$

The cancellation of the linear (in ϵ) terms was not serendipitous, but rather reflected the critical choice of λ to minimize $\ln\ (g_a(\lambda))$. Now

$$\Pr[N > a - u] \geq g_a(a)e^{-au}\left[1 - 2e^{-\epsilon u}e^{\epsilon^2/2}\right].$$

Suppose we take $\epsilon = u = 2$. This gives

$$\Pr[N > a - 2] \geq e^{-a^2/2}e^{-2a}\left[1 - 2e^{-2}\right].$$

Recall $\Pr[N > a] = \Omega(e^{-a^2/2}e^{-4a})$. In contrast, we have the upper bound $\Pr[N > a] \leq e^{-a^2/2}$.

In many applications, one does not have the precise values of the Laplace transform $f(\lambda)$. Suppose, however, that we have reasonably good estimates *in both directions* on $f(\lambda)$. Then Theorem A.2.1 will give a lower bound for $\Pr[X > a - u]$ by using a lower bound for $g_a(\lambda)$ and upper bounds for $g_a(\lambda \pm \epsilon)$. Our goal will be less ambitious than the estimate achieved for the standard normal N. We shall be content to find the asymptotics of the logarithm of $\Pr[X > a]$. In the next result, X_n may be imagined to be near the normal distribution. The interval for λ could easily be replaced by $[(1 - \gamma)a_n, (1 + \gamma)a_n]$ for any fixed positive γ.

Theorem A.2.2 *Let X_n be a sequence of random variables, and let a_n be a sequence of positive reals with $\lim_{n\to\infty} a_n = \infty$. Assume*

$$\mathrm{E}\left[e^{\lambda X_n}\right] = e^{\frac{\lambda^2}{2}(1+o(1))}$$

uniformly for $\frac{1}{2}a_n \leq \lambda \leq \frac{3}{2}a_n$. Then

$$\ln\left[\Pr\left[X_n > a_n\right]\right] \sim -\frac{a_n^2}{2}.$$

Remark. For $X_n = S_n n^{-1/2}$, $\mathrm{E}\left[e^{\lambda X_n}\right] = \cosh^n(\lambda n^{-1/2})$. As $u \to 0$, $\ln\cosh(u) \sim \frac{1}{2}u^2$. The conditions of Theorem A.2.2 therefore hold when $a_n = o(\sqrt{n})$ and $a_n \to +\infty$. That is, $\ln\left[\Pr\left[S_n > b_n\right]\right] \sim -b_n^2/2n$ when $\sqrt{n} \ll b_n \ll n$.

Proof. The upper bound is the Chernoff bound with $\lambda = a_n$.

$$\Pr\left[X_n > a_n\right] \leq \mathrm{E}\left[e^{\lambda X_n}\right] e^{-a_n\lambda} = e^{-\frac{a_n^2}{2}(1+o(1))}.$$

For the lower bound we first let $\delta \in (0, 0.01)$ be fixed. We set $\lambda = a = a_n(1+\delta)$, $u = a_n\delta$, $\epsilon = \lambda\delta/10$. Applying Theorem A.2.1

$$\Pr\left[X > a_n\right] \geq e^{-\lambda u}B$$

with

$$B = g_a(a) - e^{-\epsilon u}[g_a(a+\epsilon) + g_a(a-\epsilon)]$$

but

$$\ln[g_a(a)] \sim -\frac{a^2}{2}$$

and, analogous to our result for the standard normal,

$$\ln[g_a(a \pm \epsilon)] \sim \frac{a^2}{2}\left(1 \pm \frac{\delta}{10}\right)^2 - a^2\left(1 \pm \frac{\delta}{10}\right) = \frac{a^2}{2}\left(-1 + \frac{\delta^2}{100}\right).$$

As $\epsilon u = a^2\delta^2/10(1+\delta)$, we have $e^{-\epsilon u}g_a(a \pm \epsilon) \ll g_a(a)$. Now B is dominated by its initial term and

$$\Pr\left[X > a_n\right] \geq e^{-\lambda u}g_a(a)(1 - o(1)).$$

Taking logarithms

$$\ln\left[\Pr\left[X > a_n\right]\right] \geq -a_n^2\delta(1+\delta) - \frac{a_n^2}{2}(1+\delta)^2(1+o(1)) - o(1).$$

As this holds for any fixed $\delta \in (0, 0.01)$,

$$\ln\left[\Pr\left[X > a_n\right]\right] \geq -\frac{a_n^2}{2}(1+o(1)).$$ ∎

We have seen that $\Pr[S_n > b_n]$ can be well approximated by $\Pr[\sqrt{n}N > b_n]$ as long as $\sqrt{n} \ll b_n \ll n$. For $b_n = \Theta(n)$, this approximation by the normal distribution is no longer valid. Still, we shall see that the Chernoff bounds continue to give the right asymptotic value for $\ln[\Pr[S_n > b_n]]$. We place this in a somewhat wider context. Ellis (1984) has given far more general results.

Theorem A.2.3 *Let Z_n be a sequence of random variables. Let a be a fixed positive real. Set*

$$F(\lambda) = \lim_{n\to\infty} \frac{1}{n} \ln \mathrm{E}\left[e^{\lambda Z_n}\right].$$

Suppose that there exists $\lambda > 0$ and an open interval I containing λ such that

1. *$F(s)$ exists and has a first and a second derivative for all $s \in I$;*
2. *$F'(\lambda) = a$;*

3. *F' is a strictly increasing function in I;*
4. *There is a K so that $|F''(s)| \le K$ for all $s \in I$.*

Then

$$\lim_{n\to\infty} \frac{1}{n} \Pr\left[Z_n > an\right] = F(\lambda) - a\lambda \ .$$

Remark. Let X be a random variable whose Laplace transform is well defined. Let Z_n denote the sum of n independent copies of X. Then $F(\lambda) = \ln \mathrm{E}\left[e^{\lambda X}\right]$. In particular, suppose $\Pr[X = 1] = \Pr[X = -1] = 1/2$ so that $Z_n = S_n$. Then $F(\lambda) = \ln\cosh(\lambda)$. For any $a \in (0, 1)$, there is a positive λ for which $a = F'(\lambda) = \tanh(\lambda)$. The conditions of Theorem A.2.3 hold and give the asymptotics of $\ln[S_n > an]$.

Proof. The upper bound is the Chernoff bound, as

$$\Pr\left[Z_n > an\right] \le \mathrm{E}\left[e^{\lambda Z_n}\right] e^{-a\lambda n} = e^{n(F(\lambda)-a\lambda+o(1))} \ .$$

For the lower bound we will apply Theorem A.2.1. First, note that since F' is continuous and monotone over I, it has a continuous inverse H defined over some interval J containing a. Note $H(a) = \lambda$. Let u be positive and sufficiently small so that $H(a + u) \pm \frac{u}{K} \in I$. As $\lim_{u\to 0}, H(a + u) \pm \frac{u}{K} = H(a) = \lambda$, all sufficiently small u satisfy this criterion.

Set $a^* = a + u$ and $\lambda^* = H(a^*)$ so that $F'(\lambda^*) = a^*$. We define

$$g_n(s) = \mathrm{E}\left[e^{sZ_n}\right] e^{-sa^*}.$$

Theorem A.2.1 (noting that $an = a^*n - un$) states

$$\Pr\left[Z_n > an\right] \ge e^{-\lambda^* a^* n}[g_n(\lambda^*) - e^{-\epsilon un}[g_n(\lambda^* + \epsilon) + g_n(\lambda^* - \epsilon)].$$

We select $\epsilon = \frac{u}{K}$. Our selection of u assures us that $\lambda^* \pm \epsilon$ belong to I. We have

$$\lim_{n\to\infty} \frac{1}{n} \ln \left[\frac{e^{-\epsilon un} g_n(\lambda^* + \epsilon)}{g_n(\lambda^*)}\right] = -\epsilon u + F(\lambda^* + \epsilon) - F(\lambda^*) - \epsilon a^*.$$

We have selected λ^* so that $F'(\lambda^*) = a^*$. Since $|F''(s)| \le K$ in the interval I, Taylor series bounds

$$|F(\lambda^* + \epsilon) - F(\lambda^*) - \epsilon a^*| \le \frac{K}{2}\epsilon^2.$$

Our choice of ϵ (chosen to minimize the quadratic though any sufficiently small ϵ would do) gives that

$$-\epsilon u + F(\lambda^* + \epsilon) - F(\lambda^*) - \epsilon a^* \le -\frac{u^2}{2K}.$$

Thus $e^{-\epsilon n} g_n(\lambda^* + \epsilon)/g_n(\lambda^*)$ drops exponentially quickly. We only use that, for n sufficiently large, the ratio is less than 0.25. The same argument shows that for n

sufficiently large, $e^{-\epsilon n} g_n(\lambda^* - \epsilon)/g_n(\lambda^*) < 0.25$. For such n, we then have

$$\Pr\left[Z_n > an\right] \geq \frac{1}{2} e^{-\lambda^* a^* n} g_n(\lambda^*).$$

This lower bound is $\exp[n(F(\lambda^*) - \lambda^* a^* + o(1))]$. Now consider $F(\lambda^*) - \lambda^* a^*$ as a function of u. As $u \to 0$, $\lambda^* = H(a+u) \to H(a) = \lambda$. As F is continuous, $F(\lambda^*) \to F(\lambda)$. Clearly, $a^* = a + u \to a$ and therefore $\lambda^* a^* \to \lambda a$. Thus

$$F(\lambda^*) - \lambda^* a^* \to F(\lambda) - \lambda a$$

so

$$\Pr\left[Z_n > an\right] \geq \exp[n(F(\lambda) - \lambda a + o(1))]. \qquad \blacksquare$$

Remark. Let Z_n be a sequence of random variables with mean and variance μ_n, σ_n^2, respectively. The analysis of $\Pr[Z_n > \mu_n + \lambda_n \sigma_n]$ frequently (S_n being the premier example) splits into three parts:

1. Small Deviations. $\lambda_n \to \lambda$, a positive constant. One hopes to prove asymptotic normality so that $\Pr[Z_n > \mu_n + \lambda_n \sigma_n] \to \Pr[N > \lambda]$. There is a huge literature on asymptotic normality, but, for the most part, asymptotic normality is not covered in this work.
2. Large Deviations. $\lambda_n \to +\infty$ and $\lambda_n = o(\sigma_n)$. One hopes to show that Z_n is approximately normal in the sense that $\ln[\Pr[Z_n > \mu_n + \lambda_n \sigma_n]] \sim -\lambda_n^2/2$.
3. Very Large Deviations. $\lambda_n \to +\infty$ and $\lambda_n = \Omega(\sigma_n)$. Here the approximation of Z_n by the normal distribution generally fails, but one hopes that the asymptotics of $\ln[\Pr[Z_n > \mu_n + \lambda_n \sigma_n]]$ may still be found by the methods we have given.

A.3 EXERCISES

1. The *Hajós number* of a graph G is the maximum number k such that there are k vertices in G with a path between each pair so that all the $\binom{k}{2}$ paths are internally pairwise vertex disjoint (and no vertex is an internal vertex of a path and an endpoint of another). Is there a graph whose chromatic number exceeds twice its Hajós number ?
2. For two subsets A and B of the set Z_m of integers modulo m, and for a $g \in Z_m$, denote
$$s(A, B, g) = |\{(a, b) : a \in A, b \in B, a + b = g\}|.$$
For a partition of Z_m into two disjoint sets $Z_m = A \cup B, A \cap B = \emptyset$, denote
$$c(A, B) = \max_{x \in Z_m} |s(A, A, x) + s(B, B, x) - 2s(A, B, x)|.$$
Prove that for every odd m there is a partition of Z_m into two disjoint sets A and B such that $c(A, B) = O(\sqrt{m \log m})$.

3. For $a \in (0,1)$, apply Theorem A.2.3 to find $\lim_n \frac{1}{n} \ln \Pr[S_n > an]$ explicitly. Express $\Pr[S_n > an]$ combinatorially as 2^{-n} times the sum of binomial coefficients. Use Stirling's formula to asymptotically evaluate this sum and show that you get the same result for $\lim_n \frac{1}{n} \ln \Pr[S_n > an]$.
4. More generally, for $p \in (0,1)$ fixed, apply Theorem A.2.3 to find the asymptotics of $\ln \Pr[\text{ BIN}(n,p) > an]$ for $p < a < 1$ and of $\ln \Pr[\text{ BIN}(n,p) < an]$ for $0 < a < p$. Show that an application of Stirling's formula gives the same answer.
5. Let X_i, $1 \le i \le n$, be independent, each taking the values $+1, +2, -3$ with probability $1/3$. Set $Y_n = \sum_{i=1}^n X_i$. Let $f(n)$ be the minimal value so that $\Pr[Y_n > f(n)] < n^{-1}$. Find the asymptotics of $f(n)$. Redo with n^{-1} replaced by n^{-50}. (Note that it does not change the answer much!)

THE PROBABILISTIC LENS:
Triangle-Free Graphs Have Large Independence Numbers

Let $\alpha(G)$ denote the independence number of a graph G. It is easy to see and well known that for every graph G on n vertices with maximum degree d, $\alpha(G) \geq n/(d+1)$. Ajtai, Komlós and Szemerédi (1980) showed that, in case G is triangle-free, this can be improved by a logarithmic factor and in fact $\alpha(G) \geq cn \log d/d$, where c is an absolute positive constant. Shearer (1983) simplified the proof and improved the constant factor to its best possible value $c = 1 + o(1)$. Here is a very short proof, without any attempts to optimize c, which is based on a different technique of Shearer (1985) and its modification in Alon (1996).

Proposition 1 *Let $G = (V, E)$ be a triangle-free graph on n vertices with maximum degree at most $d \geq 1$. Then*

$$\alpha(G) \geq \frac{n \log d}{8d},$$

where the logarithm here and in what follows is to the base 2.

Proof. If, say, $d < 16$, the result follows from the trivial bound $\alpha(G) \geq n/(d+1)$ and hence we may and will assume that $d \geq 16$. Let W be a random, independent set of vertices in G, chosen uniformly among all independent sets in G. For each vertex $v \in V$, define a random variable $X_v = d|\{v\} \cap W| + |N(v) \cap W|$, where $N(v)$ denotes the set of all neighbors of v. We claim that the expectation of X_v satisfies $E[X_v] \geq \frac{\log d}{4}$.

To prove this claim, let H denote the induced subgraph of G on $V - (N(v) \cup \{v\})$, fix an independent set S in H, and let X denote the set of all non-neighbors of S in the set $N(v)$, $|X| = x$. It suffices to show that the conditional expectation

$$\mathrm{E}\left[X_v \,|\, W \cap V(H) = S\right] \geq \frac{\log d}{4} \tag{1}$$

for each possible S. Conditioning on the intersection $W \cap V(H) = S$, there are precisely $2^x + 1$ possibilities for W: one in which $W = S \cup \{v\}$ and 2^x in which $v \notin W$ and W is the union of S with a subset of X. It follows that the conditional expectation considered in (1) is precisely $\frac{d}{2^x+1} + x2^{x-1}/(2^x + 1)$. To check whether the last quantity is at least $\log d/4$, observe that the assumption that this is false implies that $x \geq 1$ and $2^x(\log d - 2x) > 4d - \log d$, showing that $\log d > 2x \geq 2$ and hence $4d - \log d < \sqrt{d}(\log d - 2)$, which is false for all $d \geq 16$. Therefore,

$$\mathrm{E}\left[X_v \,|\, W \cap V(H) = S\right] \geq \frac{\log d}{4},$$

establishing the claim.

By Linearity of Expectation, we conclude that the expected value of the sum $\sum_{v \in V} X_v$ is at least $\frac{n \log d}{4}$. On the other hand, this sum is clearly at most $2d|W|$, since each vertex $u \in W$ contributes d to the term X_u in this sum, and its degree in G, which is at most d, to the sum of all other terms X_v. It follows that the expected size of W is at least $\frac{n \log d}{8d}$, and hence there is an independent set of size at least this expectation, completing the proof. ■

The *Ramsey number* $R(3, k)$ is the minimum number r such that any graph with at least r vertices contains either a triangle or an independent set of size k. The asymptotic behavior of this function has been studied for over 50 years. It turns out that $R(3, k) = \Theta(k^2/\log k)$. The lower bound is a result of Kim (1995), based on a delicate probabilistic construction together with some 30 pages of computation. There is no known explicit construction of such a graph, and the largest known explicit triangle-free graph with no independent set of size k, described in Alon (1994), has only $\Theta(k^{3/2})$ vertices. The tight upper bound for $R(3, k)$, proved in Ajtai et al. (1980), is a very easy consequence of the above proposition.

Theorem 2 Ajtai et al. (1980) *There exists an absolute constant b such that $R(3, k) \leq bk^2/\log k$ for every $k > 1$.*

Proof. Let $G = (V, E)$ be a triangle-free graph on $8k^2/\log k$ vertices. If G has a vertex of degree at least k, then its neighborhood contains an independent set of size k. Otherwise, by Proposition 1 above, G contains an independent set of size at least $\frac{8k^2}{\log k} \frac{\log k}{8k} = k$. Therefore, in any case $\alpha(G) \geq k$, completing the proof. ■

Appendix B
Paul Erdős

Working with Paul Erdős was like taking a walk in the hills. Every time when I thought that we had achieved our goal and deserved a rest, Paul pointed to the top of another hill and off we would go.
–Fan Chung

B.1 PAPERS

Paul Erdős was the most prolific mathematician of the twentieth century, with over 1500 written papers and more than 490 collaborators. This highly subjective list gives only some of the papers that created and shaped the subject matter of this volume. **MR** and **Zbl.** refer to reviews in Math Reviews and Zentralblatt, respectively. Chapter and section reference are to the pertinent areas of this volume.

- A combinatorial problem in geometry, *Compositio Math* **2** (1935), 463–470 (with George Szekeres) **Zbl.** 12, 270.
  Written when Erdős was still a teenager, this gem contains a rediscovery of Ramsey's Theorem and the Monotone Subsequence theorem. Many authors have written that this paper played a key role in moving Erdős toward a more combinatorial view of mathematics.
- Some remarks on the theory of graphs, *Bull. Amer. Math. Soc*. **53** (1947), 292–294, **MR** 8,479d; **Zbl** 32,192.
  The three-page paper that "started" the probabilistic method gives an exponential lower bound on Ramsey $R(k, k)$. Section 1.1.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

- The Gaussian law of errors in the theory of additive number theoretic functions, *Amer. J. Math*. **62** (1940), 738–742 (with Mark Kac) **MR** 2,42c; **Zbl.** 24,102.
  Shows that the number of prime factors of x chosen uniformly from 1 to n has an asymptotically normal distribution. A connection between probability and number theory that was extraordinary for its time. Section 4.2.
- Problems and results in additive number theory, *Colloque sur la Théorie des Nombres, Bruxelles, 1955*, 127–137, George Thone, Liège; Masson and Cie, Paris, 1956; **MR** 18,18a; **Zbl.** 73,31.
  Uses random subsets to prove the existence of a set of integers such that every n is represented $n = x + y$ at least once but at most $c \ln n$ times. Resolves a problem Sidon posed to Erdős in the 1930s. This problem continued to fascinate Erdős: see, for example, Erdős and Tetali (1990). Section 8.6.
- On a combinatorial problem, *Nordisk. Mat. Tidskr*. **11** (1963), 220–223 **MR** 28# 4068; **Zbl.** 122,248.
  On a combinatorial problem II, *Acta. Math. Acad. Sci. Hungar*. **15** (1964), 445–447; **MR** 29# 4700; **Zbl.** 201,337.
  Property B. Probabilistic proofs that any $m < 2^{n-1}$ n-sets can be two-colored with no set monochromatic yet there exist $cn^2 2^n$ n-sets that cannot be so colored. Section 1.3.
- On the evolution of random graphs, *Magyar. Tud. Akad. Mat. Kutató Int. Közl*. **5** (1960), 17–61 (with Alfred Renyi); **MR** 23# A2338; **Zbl.** 103,163.
  Rarely in mathematics can an entire subject be traced to one paper. For random graphs, this is the paper. Chapter 10.
- Graph theory and probability, *Canad. J. Math*. **11** (1959), 34–38; **MR** 21# 876; **Zbl.** 84,396.
  Proves by probabilistic methods the existence of graphs with arbitrarily high girth and chromatic number. This paper convinced many of the power of the methodology, as the problem had received much attention but no construction had been found. Lens, following Chapter 3.
- Graph theory and probability II, *Canad. J. Math*. **13** (1961), 346–352 **MR** 22# 10925; **Zbl.** 97,391.
  Shows the existence of a triangle-free graph on n vertices with no independent set of size $cn^{1/2} \ln n$ vertices, and hence that the Ramsey $R(3, k) = \Omega(k^2 \ln^{-2} k)$. A technical *tour de force* that uses probabilistic methods in a very subtle way, particularly considering the early date of publication.
- On circuits and subgraphs of chromatic graphs, *Mathematika* **9** (1962), 170–175; **MR** 25 # 3035; **Zbl.** 109,165.
  Destroying the notion that chromatic number is necessarily a local property, Erdős proves the existence of a graph on n vertices that cannot be k-colored but for which every ϵn vertices can be three-colored. Lens, following Chapter 8.
- On a combinatorial game, *J. Combinatorial Theory Ser. A* **14** (1973), 298–301 (with John Selfridge) **MR** 48# 5655; **Zbl.** 293,05004.

Players alternate turns selecting vertices and the second player tries to stop the first from getting a winning set. The weight function method used was basically probabilistic and was an early use of derandomization. Section 16.1.

B.2 CONJECTURES

Conjectures were always an essential part of the mathematical life of Paul Erdős. Here are some of our favorites.

- Do sets of integers of positive density necessarily contain arithmetic progressions of arbitrary length? In finite form, is there for all k and all $\epsilon > 0$, an n_0 so that if $n \geq n_0$ and S is a subset of the first n integers of size at least ϵn then S necessarily contains an arithmetic progression of length k? This conjecture was first made by Paul Erdős and Paul Turan in the 1930s. It was solved (positively) by Szemeredi in the 1970s. Let $F(k, \epsilon)$ denote the minimal n_0 that suffices above. The growth rate of F remains an intriguing question, with results due to Gowers.
- Call distinct S, T, U a Δ-system if $S \cap T = S \cap U = T \cap U$. Let $F(n)$ be the minimal m such that, given any m n-sets, some three form a Δ-system. Erdős and Rado showed that $F(n)$ exists and gave the upper bound $F(n) \leq 2^n n!$. Erdős conjectured that $F(n) < C^n$ for some constant C.
- What are the asymptotics of the Ramsey function $R(k, k)$? In particular, what is the value c (if it exists) of $\lim_k R(k, k)^{1/k}$? The classic 1947 paper of Erdős gives $c \geq \sqrt{2}$, and $c \leq 4$ follows from the proof of Ramsey's theorem but seventy years of strenuous efforts have seen no further improvements in c, though there have been some results on lower order terms.
- Write $r_S(n)$ for the number of solutions to the equation $n = x + y$ with $x, y \in S$. Does there exist a set S of positive integers such that $r_S(n) > 0$ for all but finitely many n yet $r_S(n)$ is bounded by some constant K? The 1955 paper of Erdős referenced above gives S with $r_S(n) = \Theta(\ln n)$.
- Let $m(n)$, as defined in Section 1.3, denote the minimal size of a family of n-sets that cannot be two-colored without forming a monochromatic set. What are the asymptotics of $m(n)$? In 1963 and 1964, Erdős found the bounds $\Omega(2^n) \leq m(n) = O(2^n n^2)$, and the lower bound of Radhakrishnan and Srinivasan (2000) (with the argument of Cherkashin and Kozik 2015), shown in Section 3.5, is now $\Omega(2^n (n/\ln n)^{1/2})$.
- Given $2^{n-2} + 1$ points in the plane, no three on a line, must some n of them form a convex set? This conjecture dates back to the 1935 paper of Erdős and Szekeres referenced above.
- Let $m(n, k, l)$ denote the size of the largest family of k-element subsets of an n-set such that no l-set is contained in more than one of them. Simple counting

gives $m(n,k,l) \leq \binom{n}{l} / \binom{k}{l}$. Erdős and Hanani conjectured in 1963 that for fixed $l < k$ this bound is asymptotically correct – that is, the ratio of $m(n,k,l)$ to $\binom{n}{l} / \binom{k}{l}$ goes to 1 as $n \to \infty$. Erdős had a remarkable ability to select problems that were very difficult but not impossible. This conjecture was settled affirmatively by Vojtech Rödl (1985), as discussed in Section 4.7.

B.3 ON ERDŐS

There have been numerous books and papers written about the life and mathematics of Paul Erdős. Three deserving particular mention are the following:

- *The Mathematics of Paul Erdős* (Ron Graham and Jarik Nešetřil, eds.), Berlin: Springer-Verlag, 1996. (Vols. I and II)
- *Combinatorics, Paul Erdős is Eighty* (D. Miklós, V. T. Sós, T. Szönyi, eds.), Bolyai Soc. Math. Studies, Vol. I (1990) and Vol. II (1993).
- *Erdős on Graphs – His Legacy of Unsolved Problems*, Fan Chung and Ron Graham, A.K. Peters, 1998.

Of the many papers by mathematicians, we note

- László Babai, In and out of Hungary: Paul Erdős, his friends, and times. In *Combinatorics, Paul Erdős is Eighty* (listed above), Vol. II, 7–93.
- Béla Bollobás, Paul Erdős – Life and work, in *The Mathematics of Paul Erdős* (listed above), Vol. II, 1–42.
- A. Hajnal, Paul Erdős' Set theory, in *The Mathematics of Paul Erdős* (listed above), Vol. II, 352–393.
- János Pach, Two places at once: a remembrance of Paul Erdős, *Math Intelligencer*, Vol. 19 (1997), no. 2, 38–48.

Two popular biographies of Erdős have appeared:

- *The Man Who Loved Only Numbers*, Paul Hoffman, Hyperion (New York), 1998.
- *My Brain is Open - The Mathematical Journeys of Paul Erdős*, Bruce Schechter, Simon & Schuster (New York), 1998.

Shortly before Paul Erdős death, George Csicsery created a documentary film *N is a Number, A Portrait of Paul Erdős*. Our younger readers, naturally, never knew Erdős personally. In this film one can see and hear Erdős in lecture and amongst friends, proving and conjecturing. The film is widely available on the web, including You Tube, for downloading.

B.4 UNCLE PAUL

Paul Erdős died in September 1996 at the age of 83. His theorems and conjectures permeate this volume. This tribute[1]*, given by Joel Spencer at the National Meeting of the American Mathematical Society in January 1997, attempts to convey some of the special spirit that we and countless others took from this extraordinary man.*

Paul Erdős was a searcher, a searcher for mathematical truth.

Paul's place in the mathematical pantheon will be a matter of strong debate for in that rarefied atmosphere he had a unique style. The late Ernst Straus said it best, in a commemoration of Erdős' seventieth birthday.

> In our century, in which mathematics is so strongly dominated by "theory constructors" he has remained the prince of problem solvers and the absolute monarch of problem posers. One of my friends – a great mathematician in his own right – complained to me that "Erdős only gives us corollaries of the great metatheorems which remain unformulated in the back of his mind." I think there is much truth to that observation but I don't agree that it would have been either feasible or desirable for Erdős to stop producing corollaries and concentrate on the formulation of his metatheorems. In many ways Paul Erdős is the Euler of our times. Just as the "special" problems that Euler solved pointed the way to analytic and algebraic number theory, topology, combinatorics, function spaces, etc.; so the methods and results of Erdős' work already let us see the outline of great new disciplines, such as combinatorial and probabilistic number theory, combinatorial geometry, probabilistic and transfinite combinatorics and graph theory, as well as many more yet to arise from his ideas.

Straus, who worked as an assistant to Albert Einstein, noted that Einstein chose physics over mathematics because he feared that one would waste one's powers in pursuing the many beautiful and attractive questions of mathematics without finding the central questions. Straus goes on,

> Erdős has consistently and successfully violated every one of Einstein's prescriptions. He has succumbed to the seduction of every beautiful problem he has encountered – and a great many have succumbed to him. This just proves to me that in the search for truth there is room for Don Juans like Erdős and Sir Galahads like Einstein.

I believe, and I'm certainly most prejudiced on this score, that Paul's legacy will be strongest in Discrete Math. Paul's interest in this area dates back to a marvellous paper with George Szekeres in 1935 but it was after World War II that it really flourished. The rise of the Discrete over the past half century has, I feel, two main causes. The first was The Computer, how wonderful that this physical object has led to such intriguing mathematical questions. The second, with due respect to the many others, was the constant attention of Paul Erdős with his famous admonition "Prove and Conjecture!" Ramsey Theory, Extremal Graph Theory, Random Graphs, how many turrets in our mathematical castle were built one brick at a time with Paul's theorems and, equally important, his frequent and always penetrating conjectures.

[1]Reprinted with permission from the Bulletin of the American Mathematical Society.

My own research specialty, The Probabilistic Method, could surely be called The Erdős Method. It was begun in 1947 with a three-page paper in the Bulletin of the American Mathematical Society. Paul proved the existence of a graph having certain Ramsey property without actually constructing it. In modern language, he showed that an appropriately defined random graph would have the property with positive probability and hence there must exist a graph with the property. For the next 20 years Paul was a "voice in the wilderness." His colleagues admired his amazing results but adaption of the methodology was slow. But Paul persevered – he was always driven by his personal sense of mathematical aesthetics in which he had supreme confidence – and today the method is widely used in both Discrete Math and in Theoretical Computer Science.

There is no dispute over Paul's contribution to the spirit of mathematics. Paul Erdős was the most inspirational man I have ever met. I began working with Paul in the late 1960s, a tumultuous time when "do your own thing" was the admonition that resonated so powerfully. But while others spoke of it, this was Paul's modus operandi. He had no job; he worked constantly. He had no home; the world was his home. Possessions were a nuisance, money a bore. He lived on a web of trust, traveling ceaselessly from Center to Center, spreading his mathematical pollen.

What drew so many of us into his circle? What explains the joy we have in speaking of this gentle man? Why do we love to tell Erdős stories? I've thought a great deal about this and I think it comes down to a matter of belief, or faith. We mathematicians know the beauties of our subject and we hold a belief in its transcendent quality. God created the integers, the rest is the work of Man. Mathematical truth is immutable, it lies outside physical reality. When we show, for example, that two nth powers never add to an nth power for $n \geq 3$, we have discovered a Truth. This is our belief, this is our core motivating force. Yet our attempts to describe this belief to our nonmathematical friends are akin to describing the Almighty to an atheist. Paul embodied this belief in mathematical truth. His enormous talents and energies were given entirely to the Temple of Mathematics. He harbored no doubts about the importance, the absoluteness, of his quest. To see his faith was to be given faith. The religious world might better have understood Paul's special personal qualities. We knew him as Uncle Paul.

I do hope that one cornerstone of Paul's, if you will, theology will long survive. I refer to The Book. The Book consists of all the theorems of mathematics. For each theorem, there is in The Book just one proof. It is the most aesthetic proof, the most insightful proof, what Paul called The Book Proof. And when one of Paul's myriad conjectures was resolved in an "ugly" way, Paul would be very happy in congratulating the prover but would add, "Now, let's look for The Book Proof." This platonic ideal spoke strongly to those of us in his circle. The mathematics was there, we had only to discover it.

The intensity and the selflessness of the search for truth were described by the writer Jorge Luis Borges in his story "The Library of Babel". The narrator is a worker in this library which contains on its infinite shelves all wisdom. He wanders its infinite corridors in search of what Paul Erdős might have called The Book. He cries out,

> To me, it does not seem unlikely that on some shelf of the universe there lies a total book. I pray the unknown gods that some man – even if only one man, and though it have been thousands of years ago! – may have examined and read it. If honor and wisdom and happiness are not for me, let them be for others. May heaven exist though my place be in hell. Let me be outraged and annihilated but may Thy enormous Library be justified, for one instant, in one being.

In the summer of 1985, I drove Paul to what many of us fondly remember as Yellow Pig Camp – a mathematics camp for talented high-school students at Hampshire College. It was a beautiful day – the students loved Uncle Paul, and Paul enjoyed nothing more than the company of eager young minds. In my introduction to his lecture, I discussed The Book but I made the mistake of describing it as being "held by God." Paul began his lecture with a gentle correction that I shall never forget. "You don't have to believe in God," he said, "but you should believe in The Book."

THE PROBABILISTIC LENS:
The Rich Get Richer

Consider two bins, each of which initially has one ball. At each time $u = 1, 2, \cdots$, we add one ball to one of the bins. The ball is placed randomly, in proportion to the *square* of the number of balls already in the bin. (e.g., if the bins have 5 balls and 3 balls, respectively, the next ball is placed in the bin with 5 balls with probability $\frac{25}{25+9}$.)

Theorem 1 *With probability* 1*, one of the bins will get all but a finite number of the balls.*

We move to a continuous time model. Let X_i be independent random variables, X_i having the exponential distribution with mean i^{-2}. (That is, X_i has density function $i^2 e^{-ti^2}$ for $t \geq 0$.) At time zero, the first bin has one ball. It receives its second ball at time X_1. In general, it receives its ith ball time X_i after receiving its $(i-1)$st ball. Let X_i' also be independent exponential distributions with mean i^{-2}, independently chosen from the X_i. The second bin receives its balls according to the X_i'. The process ends when an infinite number of balls have been placed. The fictitious continuation, of defining the X_i, X_i' for all $i \geq 1$, shall be helpful in the analysis.

We use two basic properties of exponential distributions. Both are easy calculus exercises.

- Let X be exponential with mean μ and let $a > 0$. Then $X - a$, conditional on $X \geq a$, is also exponential with mean μ. This is often called the *forgetfulness* property.
- Let X, X' be independent exponentials with means μ, ν respectively. Then $\Pr[\min(X, X') = X] = \frac{\mu^{-1}}{\mu^{-1}+\nu^{-1}}$.

The continuous time process mirrors the sequential process. Clearly, the first ball is equally likely to go into either of the two bins. Suppose at some time $t > 0$ the first (say) bin has just received its ith ball and the second bin last received its jth ball at

time $t' < t$. (When the second bin has not yet received its second ball, set $j = 1$ and $t' = 0$.) The waiting time for the first bin is then X_i. The waiting time for the second was X_j at time t'. By the forgetfulness property, its conditional waiting time at time t is X_j^*, exponential with mean j^{-2}. The next ball goes into the first bin if and only if $\min(X_i, X_j^*) = X_i$, which occurs with probability $i^2/(i^2 + j^2)$ as desired.

Let $T = \sum_{i=1}^{\infty} X_i$, and $T' = \sum_{i=1}^{\infty} X_i'$ be the total times for the bins to receive (under fictitious continuation) an infinite number of balls. As $E[X_i] = E[X_i'] = i^{-2}$ and (critically!), $\sum_{i=1}^{\infty} i^{-2}$ converges, both T, T' have finite means and so are finite with probability 1. As sums of independent continuous distributions, $\Pr[T = T'] = 0$. Suppose $T < T'$, the other case being identical. At time T, the first bin has received an infinite number of balls. The second bin has not. Therefore, the second bin has received only a finite number of balls!

Appendix C
Hints to Selected Exercises

If you want to have good ideas you must have many ideas. Most of them will be wrong, and what you have to learn is which ones to throw away.
–Linus Pauling

CHAPTER 1

3. Let $T = (x_1, x_2, \ldots, x_m)$ be a sequence of not necessarily distinct reals. For any positive b, define

$$T_b = \{(x_i, x_j) : 1 \le i, j \le m, |x_i - x_j| \le b\}.$$

Prove, by induction on $|T|$, that

$$|T_3| < 3|T_1|.$$

4. Let A be, initially, a random set of vertices obtained by picking each vertex of G, randomly and independently, to be a member of A with probability $p = \frac{\ln \delta}{\delta}$. As long as there is a vertex in $V - A$ with no neighbor in A, or no neighbor in $V - A$, add it to A, until this process terminates. Note that, if a vertex v is not in A in the initial random selection, and it has two neighbors x, y where x is in A , y is not in A and y has a neighbor in A, then v (and y) will stay outside A during the whole process.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

5. Apply Theorem 1.3.3.

6. Show that for every set W of at most $k-1$ vertices of T_k, there are at least $k+1$ vertices v so that (v,w) is a directed edge for all $w \in W$.

9. Let m be the maximum length of a string in F. Show that for every k $(\sum_i \frac{N_i}{2^i})^k$ does not exceed the probability that some prefix of length at most km of a random binary sequence $x_1, x_2, \ldots$ is a concatenation of members of F.

CHAPTER 2

6. Assume n is even. By summing over all bipartitions of the complete graph, conclude that it suffices to prove the corresponding result for two-colored complete bipartite graphs. For such graphs, the number of alternating Hamiltonian cycles is at most the number of perfect red matchings times that of perfect blue matchings, and each of these can be expressed as a permanent.

8. By Cauchy–Schwartz, the sum of squares of the projections of the vectors in X on each fixed coordinate is at most 1, hence the expected value of the square of the norm of the projection of a randomly chosen vector in the collection on the first k coordinates is at most $k/|X|$.
To show tightness, consider the normalized rows of a Hadamard matrix of dimension 2^r.

CHAPTER 3

2. Apply Theorem 3.1.3 with $l=4$. As the precise constant is not called, we may select p so that $n^4(1-p)^6 \sim n/2$.

4. Put $k = \lfloor \log_2 n - \frac{1}{10}\log_2\log_2 n \rfloor$, and let V be the set of vertices of a directed graph satisfying the assumption. Let H be a $k+1$-uniform hypergraph on the set of vertices V, where for each $v \in V$, H has an edge e_v consisting of v together with k arbitrarily chosen outneighbors of v. Show that H is 2-colorable.

CHAPTER 4

2. If no a_i exceeds, say, 0.1, split the numbers into two disjoint parts with nearly equal sums of squares, apply Chebyshev to lower bound the probability that each signed sum over a part is in absolute value at most 1, and observe that with probability $1/2$ the two sums have opposite signs. If there are some large numbers a_i, multiply by the probability that their contribution will also keep the absolute value of the sum small.

3. This is similar to the previous exercise. If all a_i are of small norm, partition to a bounded number of pieces with nearly equal sums of squares of norms, apply Chebyshev to each signed piece separately, and show that they can be combined. If there are some vectors with large norms, deal with them separately.

6. Partition Z_p arbitrarily into $2k$ intervals I_j of nearly equal size. Choose a and b randomly and independently in Z_p, and let $z_x = ax + b (\bmod p)$. Show that the random

variables z_x for $x \in X$ are pairwise independent, and apply Chebyshev to show that with positive probability at least one of them falls into each of the intervals I_j.

CHAPTER 5

1. By König's Theorem, the edges can be colored by d colors, so that every color class forms a matching. Now use the Local Lemma to show that one can pick an edge from each two-colored cycle so that no two edges picked are within distance 1 in the graph.

2. By compactness, it is enough to prove, for every length m, the existence of a finite sequence of length m with the desired properties. Consider a random sequence of that length and apply the Local Lemma (General Case).

5. Let S be a random set of half the elements in the interval of the first $2ck \ln k$ integers. Show that, with high probability, there is no subset of size at most $4c \ln k$ of the first $4ck \ln k$ integers that intersects every shift $x + S$ for $0 \leq x \leq 2ck \ln k$.

CHAPTER 6

2. Apply induction on k using the fact that, without loss of generality, we may assume that each family $\mathcal{F}_i$ is monotone increasing.

CHAPTER 7

2. Fix a proper coloring of G by 1000 color classes $V_1, V_2, \ldots, V_{1000}$, and consider the martingale $X_0, X_1, \ldots, X_{1000}$, where $X_i(H)$ is the expected value of the chromatic number of H conditioned on the value of $U \cap V_j$ for all $j \leq i$.

CHAPTER 8

8. Let $G = G(n, 0.5)$ be the random graph on n vertices. For every fixed graph H on $k \leq (2 - \epsilon)\log_2 n$ vertices, apply the Janson inequalities to bound the probability that H is not an induced subgraph of G.

CHAPTER 9

3. Let X be a set of vertices of maximum cardinality satisfying

(i) $|X| \leq 2a$, and

(ii) $|N(X) \cap Y| < 2|X|$.

Show that in fact $|X| \leq a$.

11. Use martingales to get the stated inequality and apply the Borel–Cantelli Lemma to deduce the conclusion.

CHAPTER 10

1. Show that the random graph $G(n, 3/4)$ has the required properties almost always (one can give explicit constructions as well).

2. Let $p = p(n)$ be at the threshold for containing a complete graph on four vertices. Show that for every four vertices v_1, v_2, v_3, v_4 there will be a fifth vertex v_5 adjacent to v_1.

CHAPTER 11

1. Set $x = n\alpha$ and $y = n - 1 - x \sim n\beta$ with $\alpha + \beta = 1$. Use the asymptotic equation (11.6) to estimate $\Pr[X = x] \Pr[Y = y]$. The denominator $\Pr[X + Y = n - 1]$ is dominated by $\min(X, Y)$ being small. Perhaps surprisingly, the conditional distribution of X is highly skewed to the corners.

CHAPTER 12

3. Consider a depth-4 formula consisting of a tree in which the root (the output vertex, at level zero) is the OR of n children, each node at level one is the AND of n^2 children, each node at level 2 is the OR of $n \log n$ children, and each node at level 3 is the AND of $\log n$ randomly chosen variables among the n variables x_i. Show that almost always this computes a function with the required properties.

CHAPTER 13

1. As $i, i + m/2$ cancel out, sets A effectively have at most size $m/2$. When the split is random, the effective size of A is even lower.

3. Truncate each Gaussian by redefining it to be b whenever its value exceeds b, where b is chosen to ensure that its variance still exceeds $1 - \epsilon/2$, and use Azuma's Inequality for the sum of the squares of the truncated variables.

4. The second inequality follows from the first by adjusting the martingale so that, when $X_i > a\sigma$, all future τ_j (those with $j > i$) are set to zero. (One may think of this as the "go to sleep" martingale, representing a gambler whose strategy is to go to sleep after winning a certain predetermined amount.)

CHAPTER 14

1. Show that n triangles partition the plane into at most $O(n^2)$ regions.

CHAPTER 15

3. Let S_i denote the star consisting of all edges of the complete graph on $\{1, 2, \ldots, 2t\}$ incident with i. Each two graphs in $\mathcal{F}$ must have at least one common edge in each S_i, implying that the cardinality of the set $\{F \cap S_i : F \in \mathcal{F}\}$ is at most $\frac{1}{2}2^{2t-1}$.

CHAPTER 16

2. Apply Corollary 9.2.8.

CHAPTER 17

2. Construct a 3-partite graph on the classes of vertices $X = [n] = \{1, 2, \ldots, n\}$, $Y = [2n]$ and $Z = [3n]$, where for each $i \in X$ and $a \in A$ the graph contains the triangle on

the vertices $i \in X$, $i + a \in Y$ and $i + 2a \in Z$. If A contains no three-term progression, then the graph contains no triangles besides the ones above.

4. Consider the graph constructed from A as in Exercise 2, and the graph obtained from it by replacing each vertex v by an independent set I_v and each edge uv by a complete bipartite graph on the classes of vertices I_u and I_v.

Appendix A

1. Consider the random graph $G(n, 1/2)$ for large n.

References

Ahlswede, R. and Daykin, D. E. (1978). An inequality for the weights of two families of sets, their unions and intersections, *Z. Wahrscheinl. Verw. Geb* **43**: 183–185.

Aho, A. V., Hopcroft, J. E. and Ullman, J. D. (1974). *The Design and Analysis of Computer Algorithms*, Addison Wesley, Reading, MA.

Aizenman, M. and Newman, C. M. (1984). Tree graph inequalities and critical behavior in percolation models, *J. Stat. Phys.* **36**(1-2): 107–143.

Aizenman, M., Kesten, H. and Newman, C. M. (1987). Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation, *Commun. Math. Phys.* **111**(4): 505–531.

Ajtai, M. (1983). Σ_1^1-formulae on finite structures, *Ann. Pure Appl. Logic* **24**: 1–48.

Ajtai, M., Chvátal, V., Newborn, M. M. and Szemerédi, E. (1982). Crossing-free subgraphs, *Theory and Practice of Combinatorics*, North Holland Mathematics Studies, Vol. **60**, North-Holland, Amsterdam, pp. 9–12.

Ajtai, M., Komlós, J. and Szemerédi, E. (1980). A note on Ramsey numbers, *J. Comb. Theory, Ser. A* **29**: 354–360.

Ajtai, M., Komlós, J. and Szemerédi, E. (1981). The longest path in a random graph, *Combinatorica* **1**(1): 1–12.

Ajtai, M., Komlós, J. and Szemerédi, E. (1983). Sorting in $c \log n$ parallel steps, *Combinatorica* **3**: 1–19.

Ajtai, M., Komlós, J. and Szemerédi, E. (1987). Deterministic simulation in LOGSPACE, *Proceedings of the 19th Annual ACM STOC*, New York, pp. 132–140.

Akiyama, J., Exoo, G. and Harary, F. (1981). Covering and packing in graphs IV: linear arboricity, *Networks* **11**: 69–72.

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

Alon, N. (1986a). Eigenvalues and expanders, *Combinatorica* **6**: 83–96.

Alon, N. (1986b). Eigenvalues, geometric expanders, sorting in rounds and Ramsey Theory, *Combinatorica* **6**: 207–219.

Alon, N. (1988). The linear arboricity of graphs, *Isr. J. Math* **62**: 311–325.

Alon, N. (1990a). The maximum number of Hamiltonian paths in tournaments, *Combinatorica* **10**: 319–324.

Alon, N. (1990b). Transversal numbers of uniform hypergraphs, *Graphs and Combinatorics* **6**: 1–4.

Alon, N. (1994). Explicit Ramsey graphs and orthonormal labelings, *Electron. J. Combin.* **1**(8): R12.

Alon, N. (1996). Independence numbers of locally sparse graphs and a Ramsey type problem, *Random Struct. Algorithm.* **9**: 271–278.

Alon, N. (2000). Degrees and choice numbers, *Random Struct. Algorithm.* **16**(4): 364–368.

Alon, N. and Boppana, R. B. (1987). The monotone circuit complexity of boolean functions, *Combinatorica* **7**: 1–22.

Alon, N. and Chung, F. R. K. (1988). Explicit construction of linear sized tolerant networks, *Discrete Math.* **72**: 15–19.

Alon, N. and Frankl, P. (1985). The maximum number of disjoint pairs in a family of subsets, *Graphs Comb.* **1**: 13–21.

Alon, N. and Kleitman, D. J. (1990). Sum-free subsets, in: *A Tribute to Paul Erdős* (A. Baker, B. Bollobás and A. Hajnál, eds.), Cambridge University Press, Cambridge, pp. 13–26.

Alon, N. and Krivelevich, M. (1997). The concentration of the chromatic number of random graphs, *Combinatorica* **17**: 303–313.

Alon, N. and Krivelevich, M. (2002). Testing k-colorability, *SIAM J. Discrete Math.* **15**(2): 211–227.

Alon, N. and Linial, N. (1989). Cycles of length 0 modulo k in directed graphs, *J. Comb. Theory, Ser. B* **47**: 114–119.

Alon, N. and Milman, V. D. (1984). Eigenvalues, expanders and superconcentrators, *Proceedings of the 25th Annual FOCS*, IEEE, New York, pp. 320–322; See also: Alon, N. and Milman, V. D. (1985). λ_1, isoperimetric inequalities for graphs and superconcentrators, *J. Comb. Theory, Ser. B*, **38**: 73–88.

Alon, N. and Shapira, A. (2005). A characterization of the (natural) graph properties testable with one-sided error, *FOCS'05: Proceedings of the 46th Annual Symposium on Foundations of Computer Science*, pp. 429–438.

Alon, N. and Shapira, A. (2006). Homomorphisms in graph property testing, *Topics in Discrete Mathematics*, Algorithms Combinatorics, Vol. **26**, Springer-Verlag, Berlin, pp. 281–313.

Alon, N., Babai, L. and Itai, A. (1986). A fast and simple randomized parallel algorithm for the maximal independent set problem, *J. Algorithms* **7**: 567–583.

Alon, N., Duke, R. A., Lefmann, H., Rödl, V. and Yuster, R. (1994). The algorithmic aspects of the regularity lemma, *J. Algorithms* **16**(1): 80–109.

Alon, N., Fischer, E., Krivelevich, M. and Szegedy, M. (2000). Efficient testing of large graphs, *Combinatorica* **20**(4): 451–476.

Alon, N., Fischer, E., Newman, I. and Shapira, A. (2006). A combinatorial characterization of the testable graph properties: it's all about regularity, *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, ACM, New York, pp. 251–260.

Alon, N., Frankl, P. and Rödl, V. (1985). Geometrical realization of set systems and probabilistic communication complexity, *Proceedings of the 26th FOCS*, IEEE, New York, pp. 277–280.

Alon, N., Goldreich, O., Håstad, J. and Peralta, R. (1990). Simple constructions of almost k-wise independent random variables, *Proceedings of the 31st FOCS, St. Louis*, IEEE, New York, pp. 544–553.

Alon, N., Grytczuk, J., Hałuszczak, M. and Riordan, O. (2002). Nonrepetitive colorings of graphs, *Random Struct. Algorithms* **21**(3-4): 336–346. Random structures and algorithms (Poznan, 2001).

Alon, N., Kim, J. H. and Spencer, J. H. (1997). Nearly perfect matchings in regular simple hypergraphs, *Isr. J. Math.* **100**: 171–187.

Alon, N., Krivelevich, M. and Sudakov, B. (2003). Turán numbers of bipartite graphs and related Ramsey-type questions, *Comb. Probab. Comput.* **12**(5-6): 477–494. Special issue on Ramsey theory.

Alon, N., Rónyai, L. and Szabó, T. (1999). Norm-graphs: variations and applications, *J. Comb. Theory, Ser. B* **76**: 280–290.

Andreev, A. E. (1985). On a method for obtaining lower bounds for the complexity of individual monotone functions, *Dokl. Akad. Nauk SSSR* **282**(5): 1033–1037. (in Russian). English translation in *Sov. Math. Dokl*. **31**(3): 530–534.

Andreev, A. E. (1987). On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes, *Vestn. Mosk. Univ. Ser. I Mat. Mekh* (1): 70–73. (in Russian).

Baik, J., Deift, P. and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations, *J. Am. Math. Soc.* **12**: 1119–1178.

Balogh, J., Morris, R. and Samotij, W. (2014). Independent sets in hypergraphs, *arXiv:1204.6530v2*.

Bansal, N. (2010). Constructive algorithms for discrepancy minimization, *2010 IEEE 51st Annual Symposium on Foundations of Computer Science FOCS 2010*, IEEE Computer Society, Los Alamitos, CA, pp. 3–10.

Barak, B., Rao, A., Shaltiel, R. and Wigderson, A. (2012) 2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl-Wilson construction, *Ann. of Math.* **176**(2): 1483–1543.

Bárány, I. and Füredi, Z. (1987). Empty simplices in Euclidean spaces, *Can. Math. Bull.* **30**: 436–445.

Beck, J. (1978). On 3-chromatic hypergraphs, *Discrete Math.* **24**: 127–137.

Beck, J. and Fiala, T. (1981). Integer-making theorems, *Discrete. Appl. Math.* **3**: 1–8.

Behrend, F. A. (1946). On sets of integers which contain no three terms in arithmetical progression, *Proc. Natl. Acad. Sci. U.S.A.* **32**: 331–332.

Bernstein, S. N. (1912). Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, *Commun. Soc. Math. Kharkov* **13**: 1–2.

Billingsley, P. (1995). *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, third edition, John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication.

Blum, N. (1984). A boolean function requiring 3n network size, *Theor. Comput. Sci.* **28**: 337–345.

Bollobás, B. (1965). On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16**: 447–452.

Bollobás, B. (1984). The evolution of random graphs, *Trans. Am. Math. Soc.* **286**(1): 257–274.

Bollobás, B. (1988). The chromatic number of random graphs, *Combinatorica* **8**: 49–55.

Bollobás, B. (2001). *Random Graphs*, Cambridge Studies in Advanced Mathematics, Vol. **73**, second edition, Cambridge University Press, Cambridge.

Bollobás, B. and Erdős, P. (1976). Cliques in random graphs, *Math. Proc. Cambridge Philos. Soc.* **80**: 419–427.

Boppana, R. B. and Spencer, J. H. (1989). A useful elementary correlation inequality, *J. Comb. Theory, Ser. A* **50**: 305–307.

Borgs, C., Chayes, J. T., van der Hofstad, R., Slade, G. and Spencer, J. H. (2005). Random subgraphs of finite graphs. I. The scaling window under the triangle condition, *Random Struct. Algorithms* **27**(2): 137–184.

Brégman, L. M. (1973). Some properties of nonnegative matrices and their permanents, *Sov. Math. Dokl.* **14**: 945–949.

Burton, R. M. and Keane, M. (1989). Density and uniqueness in percolation, *Commun. Math. Phys.* **121**(3): 501–505.

Chazelle, B. and Welzl, E. (1989). Quasi-optimal range searching in spaces of finite VC-dimension, *Discrete Comput. Geom.* **4**: 467–489.

Cherkashin, D. D. and Kozik, J. (2015). A note on random greedy coloring of uniform hypergraphs, *Random Struct. Algorithms*.

Chernoff, H. (1952). A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23**: 493–509.

Chung, F. R. K., Frankl, P., Graham, R. L. and Shearer, J. B. (1986). Some intersection theorems for ordered sets and graphs, *J. Comb. Theory, Ser. A* **43**: 23–37.

Chung, F. R. K., Graham, R. L. and Wilson, R. M. (1989). Quasi-random graphs, *Combinatorica* **9**: 345–362.

Cicalese, F. and Mundici, D. (2000). Optimal coding with one asymmetric error: below the sphere packing bound, *Computing and Combinatorics (Sydney, 2000)*, Lecture Notes in Computer Science, Vol. **1858**, Springer-Verlag, Berlin, pp. 159–169.

Cohen, A. and Wigderson, A. (1989). Dispersers, deterministic amplification, and weak random sources, *Proceedings of the 30th IEEE FOCS*, IEEE, New York, pp. 14–19.

Danzer, L. and Grünbaum, B. (1962). Űber zwei Probleme bezűglich konvexer Kőrper von P. Erdős und von V. L. Klee, *Math. Z.* **79**: 95–99.

de la Vega, W. F. (1983). On the maximal cardinality of a consistent set of arcs in a random tournament, *J. Comb. Theory, Ser. B* **35**: 328–332.

Dudley, R. M. (1978). Central limit theorems for empirical measures, *Ann. Probab.* **6**: 899–929.

Dumitriu, I. and Spencer, J. H. (2004). A Halfliar's game, *Theor. Comput. Sci.* **313**(3): 353–369. Algorithmic combinatorial game theory.

Eberhard, S., Green, B. and Manners, F. (2013) Sets of integers with no large sum-free subset, arXiv:1301.4579.

Elekes, G. (1997). On the number of sums and products, *Acta Arith.* **81**: 365–367.

Ellis, R. S. (1984). Large deviations for a general class of random vectors, *Ann. Probab.* **12**(1): 1–12.

Ellis, D., Filmus, Y. and Friedgut, E. (2012). Triangle-intersecting families of graphs, *J. Eur. Math. Soc. (JEMS)* **14**(3): 841–885.

Erdős, P. (1947). Some remarks on the theory of graphs, *Bull. Am. Math. Soc.* **53**: 292–294.

Erdős, P. (1956). Problems and results in additive number theory, *Colloque sur le Théorie des Nombres (CBRM, Bruselles)*, pp. 127–137.

Erdős, P. (1959). Graph theory and probability, *Can. J. Math.* **11**: 34–38.

Erdős, P. (1962). On circuits and subgraphs of chromatic graphs, *Mathematika* **9**: 170–175.

Erdős, P. (1963a). On a combinatorial problem, I, *Nordisk Mat. Tidskr.* **11**: 5–10.

Erdős, P. (1963b). On a problem of graph theory, *Math. Gaz.* **47**: 220–223.

Erdős, P. (1964). On a combinatorial problem II, *Acta Math. Acad. Sci. Hungar.* **15**: 445–447.

Erdős, P. (1965a). Extremal problems in number theory, *Proc. Symp. Pure Math. (AMS)* **VIII**: 181–189.

Erdős, P. (1965b). On extremal problems of graphs and generalized graphs, *Isr. J. Math.* **2**: 189–190.

Erdős, P. and Füredi, Z. (1983). The greatest angle among n points in the d-dimensional Euclidean space, *Ann. Discrete Math.* **17**: 275–283.

Erdős, P. and Hanani, H. (1963). On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10**: 10–13.

Erdős, P. and Kac, M. (1940). The Gaussian law of errors in the theory of additive number theoretic functions, *Am. J. Math.* **62**: 738–742.

Erdős, P. and Lovász, L. (1975). Problems and results on 3-chromatic hypergraphs and some related questions, in *Infinite and Finite Sets* (A. Hajnal et al., eds.), North-Holland, Amsterdam, pp. 609–628.

Erdős, P. and Moon, J. W. (1965). On sets of consistent arcs in a tournament, *Can. Math. Bull.* **8**: 269–271.

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**: 17–61.

Erdős, P. and Selfridge, J. L. (1973). On a combinatorial game, *J. Comb. Theory, Ser. A* **14**: 298–301.

Erdős, P. and Spencer, J. H. (1991). Lopsided Lovász local lemma and latin transversals, *Discrete Appl. Math.* **30**: 151–154.

Erdős, P. and Tetali, P. (1990). Representations of integers as the sum of k terms, *Random Struct. Algorithms* **1**(3): 245–261.

Erdős, P., Rubin, A. L. and Taylor, H. (1980). Choosability in graphs, *Proceedings of the West Coast Conference on Combinatorics, Graph Theory and Computing (Humboldt State Univ., Arcata, Calif., 1979), Congress. Numer., XXVI, Utilitas Math., Winnipeg, Man.*, pp. 125–157.

Fagin, R. (1976). Probabilities in finite models, *J. Symbolic Logic* **41**: 50–58.

Fischer, E. (2001). The art of uninformed decisions: a primer to property testing, *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS* (75): 97–126.

Fishburn, P. (1992). Correlation in partially ordered sets, *Discrete Appl. Math.* **39**: 173–191.

Fortuin, C. M., Kasteleyn, P. W. and Ginibre, J. (1971). Correlation inequalities on some partially ordered sets, *Commun. Math. Phys.* **22**: 89–103.

Füredi, Z. (1988). Matchings and covers in hypergraphs, *Graphs Comb.* **4**: 115–206.

Frankl, P. and Wilson, R. M. (1981). Intersection theorems with geometric consequences, *Combinatorica* **1**: 357–368.

Frankl, P., Rödl, V. and Wilson, R. M. (1988). The number of submatrices of given type in a Hadamard matrix and related results, *J. Comb. Theory, Ser. B* **44**: 317–328.

Furst, M., Saxe, J. and Sipser, M. (1984). Parity, circuits and the polynomial hierarchy, *Math. Syst. Theory* **17**: 13–27.

Glebskii, Y. V., Kogan, D. I., Liagonkii, M. I. and Talanov, V. A. (1969). Range and degree of realizability of formulas the restricted predicate calculus, *Cybernetics* **5**: 142–154. (Russian original: *Kibernetica* **5**, 17–27).

Goldreich, O. (1999). Combinatorial property testing (a survey), *Randomization Methods in Algorithm Design (Princeton, NJ, 1997)*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. **43**, American Mathematical Society, Providence, RI, pp. 45–59.

Goldreich, O. and Ron, D. (2002). Property testing in bounded degree graphs, *Algorithmica* **32**(2): 302–343.

Goldreich, O., Goldwasser, S. and Ron, D. (1998). Property testing and its connection to learning and approximation, *J. ACM* **45**(4): 653–750.

Gowers, W. T. (1997). Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.* **7**(2): 322–337.

Gowers, W. T. (1998). A new proof of Szemerédi's theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* **8**(3): 529–551.

Graham, R. L. and Spencer, J. H. (1971). A constructive solution to a tournament problem, *Can. Math. Bull.* **14**: 45–48.

Graham, R. L., Rothschild, B. L. and Spencer, J. H. (1990). *Ramsey Theory*, second edition, John Wiley & Sons, Inc., New York.

Grimmett, G. (1999). *Percolation*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Vol. **321**, second edition, Springer-Verlag, Berlin.

Grytczuk, J., Kozik, J. and Micek, P. (2013). New approach to nonrepetitive sequences, *Random Struct. Algorithms* **42**(2): 214–225.

Halberstam, H. and Roth, K. F. (1983). *Sequences*, second edition, Springer-Verlag, Berlin.

Hall, M. (1986). *Combinatorial Theory*, second edition, John Wiley & Sons, Inc., New York.

Hara, T. and Slade, G. (1990). Mean-field critical behaviour for percolation in high dimensions, *Commun. Math. Phys.* **128**(2): 333–391.

Hara, T. and Slade, G. (1994). Mean-field behaviour and the lace expansion, *Probability and Phase Transition (Cambridge, 1993)*, NATO Advanced Science Institutes Series C, Mathematical and Physical Sciences, Vol. **420**, Kluwer Academic Publishers, Dordrecht, pp. 87–122.

Hardy, G. H. and Ramanujan, S. (1917). The normal number of prime factors of a number n, *Quarterly Journal of Mathematics* **48**: 76–92.

Harper, L. (1966). Optimal numberings and isoperimetric problems on graphs, *J. Comb. Theory* **1**: 385–394.

Harris, T. E. (1960). Lower bound for the critical probability in a certain percolation process, *Math. Proc. Cambridge Philos. Soc.* **56**: 13–20.

Haussler, D. (1995). Sphere packing numbers for subsets of the boolean n-cube with bounded Vapnik-Chervonenkis dimension, *J. Comb. Theory, Ser. A* **69**: 217–232.

Haussler, D. and Welzl, E. (1987). ϵ-nets and simplex range queries, *Discrete Comput. Geom.* **2**: 127–151.

Håstad, J. (1988). Almost optimal lower bounds for small depth circuits, *in Advances in Computer Research* (S. Micali ed.), JAI Press, Chapter 5: Randomness and Computation, pp. 143–170.

Håstad, J. (1998). The shrinkage exponent of De Morgan formulas is 2, *SIAM J. Comput.* **27**: 48–64.

Janson, S. (1990). Poisson approximation for large deviations, *Random Struct. Algorithms* **1**: 221–230.

Janson, S. (1998). New versions of Suen's correlation inequality, *Random Struct. Algorithms* **13**: 467–483.

Janson, S., Knuth, D., Łuczak, T. and Pittel, B. (1993). The birth of the giant component, *Random Struct. Algorithms* **4**: 233–358.

Janson, S., Łuczak, T. and Ruciński, A. (1990). An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, *Random graphs '87 (Poznań, 1987)*, John Wiley & Sons, Ltd, Chichester, pp. 73–87.

Janson, S., Łuczak, T. and Rucinski, A. (2000). *Random Graphs*, John Wiley & Sons, Inc., New York.

Joffe, A. (1974). On a set of almost deterministic k-independent random variables, *Ann. Probab.* **2**: 161–162.

Kahn, J. (1996). Asymptotically good list colorings, *J. Comb. Theory, Ser. A* **73**: 1–59.

Karchmer, M. and Wigderson, A. (1990). Monotone circuits for connectivity require super-logarithmic depth, *SIAM J. Discrete Math.* **3**: 255–265.

Karp, R. M. and Ramachandran, V. (1990). Parallel algorithms for shared memory machines, in *Handbook of Theoretical Computer Science* (J. Van Leeuwen, ed.), Vol. A, Chapter 17, Elsevier, New York, pp. 871–941.

Katchalski, M. and Meir, A. (1988). On empty triangles determined by points in the plane, *Acta Math. Hungar.* **51**: 323–328.

Katona, G. O. H. (1972). A simple proof of the Erdős-Ko-Rado theorem, *J. Comb. Theory, Ser. B* **13**: 183–184.

Keevash, P. (2014). The existence of designs, *arXiv:1401.3665v1*.

Khrapchenko, V. M. (1971). A method of determining lower bounds for the complexity of Π-schemes, *Mat. Zametki* **10**(1): 83–92. (in Russian.) English translation in *Math. Notes Acad. Sci. USSR* **11**, 1972, 474–479.

Kim, J. H. (1995). The Ramsey number $R(3, t)$ has order of magnitude $t^2 / \log t$, *Random Struct. Algorithms* **7**: 173–207.

Kim, J. H. and Vu, V. (2000). Concentration of multivariate polynomials and its applications, *Combinatorica* **20**(3): 417–434.

Kleitman, D. J. (1966). Families of non-disjoint subsets, *J. Comb. Theory* **1**: 153–155.

Kleitman, D. J., Shearer, J. B. and Sturtevant, D. (1981). Intersection of k-element sets, *Combinatorica* **1**: 381–384.

Kolountzakis, M. N. (1999). An effective additive basis for the integers, *Discrete Math.* **145**: 307–313.

Komlós, J. and Simonovits, M. (1996). Szemerédi's regularity lemma and its applications in graph theory, *Combinatorics, Paul Erdős is eighty, (Keszthely, 1993)*, Bolyai Society Mathematical Studies, Vol. **2**, János Bolyai Mathematical Society, Budapest, pp. 295–352.

Komlós, J., Pach, J. and Woeginger, G. (1992). Almost tight bounds on epsilon-nets, *Discrete Comput. Geom.* **7**: 163–173.

Komlós, J., Pintz, J. and Szemerédi, E. (1982). A lower bound for Heilbronn's problem, *J. London Math. Soc.* **25**(2): 13–24.

Kostochka, A. V. and Rödl, V. (2004). On graphs with small Ramsey numbers. II, *Combinatorica* **24**(3): 389–401.

Krivelevich, M. and Sudakov, B. (2013). The phase transition in random graphs: a simple proof, *Random Struct. Algorithms* **43**(2): 131–138.

Loomis, L. H. and Whitney, H. (1949). An inequality related to the isoperimetric inequality, *Bull. Am. Math. Soc.* **55**: 961–962.

Lovász, L. (2012). *Large Networks and Graph Limits*, American Mathematical Society Colloquium Publications, Vol. **60**, American Mathematical Society, Providence, RI.

Lovász, L. and Szegedy, B. (2010). Testing properties of graphs and functions, *Isr. J. Math.* **178**: 113–156.

Lovász, L., Spencer, J. H. and Vesztergombi, K. (1986). Discrepancy of set systems and matrices, *Eur. J. Comb.* **7**: 151–160.

Lovett, S. and Meka, R. (2012). Constructive discrepancy minimization by walking on the edges, *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science—FOCS 2012*, IEEE Computer Society, Los Alamitos, CA, pp. 61–67.

Lubotzky, A., Phillips, R. and Sarnak, P. (1986). Explicit expanders and the Ramanujan conjectures, *Proceedings of the 18th ACM STOC*, pp. 240–246; See also: Lubotzky, A., Phillips, R. and Sarnak, P. (1988). Ramanujan graphs, *Combinatorica* **8**: 261–277.

Łuczak, T. (1990). Component behavior near the critical point of the random graph process, *Random Struct. Algorithms* **1**(3): 287–310.

Łuczak, T. (1991). A note on the sharp concentration of the chromatic number of random graphs, *Combinatorica* **11**: 295–297.

MacWilliams, F. J. and Sloane, N. J. A. (1977). *The Theory of Error Correcting Codes*, North-Holland, Amsterdam.

Mani-Levitska, P. and Pach, J. (1988). Decomposition problems for multiple coverings with unit balls, manuscript.

Margulis, G. A. (1973). Explicit constructions of concentrators, *Problemy Peredachi Informatsii* **9**: 71–80. (in Russian). English translation in *Prob. Inf. Transm.* **9**, 325–332.

Margulis, G. A. (1988). Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and superconcentrators, *Problemy Peredachi Informatsii* **24**: 51–60. (in Russian). English translation in *Prob. Inf. Transm.* **24**, 39–46.

Marica, J. and Schonheim, J. (1969). Differences of sets and a problem of Graham, *Can. Math. Bull.* **12**: 635–637.

Matoušek, J. (1997). On discrepancy bounds via dual shatter function, *Mathematika* **44**(1): 42–49.

Matoušek, J., Welzl, E. and Wernisch, L. (1993). Discrepancy and approximation for bounded VC dimension, *Combinatorica* **13**: 455–466.

Matula, D. W. (1976). *The largest clique size in a random graph*, Technical report CS 7608, Southern Methodist University, Dallas, TX.

Maurey, B. (1979). Construction de suites symétriques, *C.R. Acad. Sci. Paris* **288**: 679–681.

Men'shikov, M. V. (1986). Coincidence of critical points in percolation problems, *Dokl. Akad. Nauk SSSR* **288**(6): 1308–1311.

Milman, V. D. and Schechtman, G. (1986). *Asymptotic Theory of Finite Dimensional Normed Spaces*, Lecture Notes in Mathematics, Vol. **1200**, Springer-Verlag, Berlin and New York.

Molloy, M. and Reed, B., Graph colouring and the probabilistic method, Springer, 2002

Moon, J. W. (1968). *Topics on Tournaments*, Holt, Reinhart and Winston, New York.

Moser, R. A. and Tardos, G. (2010). A constructive proof of the general Lovász local lemma, *J. ACM* **57**(2): Article 11, 15.

Nakayama, A. and Peroche, B. (1987). Linear arboricity of digraphs, *Networks* **17**: 39–53.

Naor, J. and Naor, M. (1990). Small-bias probability spaces: efficient constructions and applications, *Proceedings of the 22nd Annual ACM STOC*, ACM Press, pp. 213–223.

Nilli, A. (1991). On the second eigenvalue of a graph, *Discrete Math.* **91**: 207–210.

Pach, J. and Agarwal, P. K. (1995). *Combinatorial Geometry*, John Wiley & Sons, Inc., New York.

Pach, J. and Woeginger, G. (1990). Some new bounds for epsilon-nets, *Proceedings of the 6th Annual Symposium on Computational Geometry*, ACM Press, New York, pp. 10–15.

Paturi, R. and Simon, J. (1984). Probabilistic communication complexity, *Proceedings of the 25th FOCS*, IEEE, New York, pp. 118–126.

Paul, W. J. (1977). A 2.5n lower bound on the combinational complexity of boolean functions, *SIAM J. Comput.* **6**: 427–443.

Pinsker, M. (1973). On the complexity of a concentrator, *7th International Teletraffic Conference*, Stockholm, pp. 318/1–318/4.

Pippenger, N. and Spencer, J. H. (1989). Asymptotic behaviour of the chromatic index for hypergraphs, *J. Comb. Theory, Ser. A* **51**: 24–42.

Pluhár, A. (2009). Greedy colorings of uniform hypergraphs, *Random Struct. Algorithms* **35**(2): 216–221.

Rabin, M. O. (1980). Probabilistic algorithms for testing primality, *J. Number Theory* **12**: 128–138.

Radhakrishnan, J. and Srinivasan, A. (2000). Improved bounds and algorithms for hypergraph two-coloring, *Random Struct. Algorithms* **16**: 4–32.

Raghavan, P. (1988). Probabilistic construction of deterministic algorithms: approximating packing integer programs, *J. Comput. Syst. Sci.* **37**: 130–143.

Ramsey, F. P. (1929). On a problem of formal logic, *Proc. London Math. Soc.* **30**(2): 264–286.

Razborov, A. A. (1985). Lower bounds on the monotone complexity of some boolean functions, *Dokl. Akad. Nauk SSSR* **281**(4): 798–801. (in Russian.) English translation in *Sov. Math. Dokl.* **31**, 354–357.

Razborov, A. A. (1987). Lower bounds on the size of bounded depth networks over a complete basis with logical addition, *Mat. Zametki* **41**(4): 598–607. (in Russian.) English translation in *Math. Notes Acad. Sci. USSR* **41**(4), 333–338.

Rödl, V. (1985). On a packing and covering problem, *Eur. J. Comb.* **6**: 69–78.

Rödl, V. and Duke, R. A. (1985). On graphs with small subgraphs of large chromatic number, *Graphs Comb.* **1**(1): 91–96.

Ron, D. (2001). Property testing, *Handbook of Randomized Computing, Vol. I, II*, Combinatorial Optimization, Vol. **9**, Kluwer Acadcademic Publishers, Dordrecht, pp. 597–649.

Rubinfeld, R. and Sudan, M. (1996). Robust characterizations of polynomials with applications to program testing, *SIAM J. Comput.* **25**(2): 252–271.

Ruzsa, I. Z. and Szemerédi, E. (1978). Triple systems with no six points carrying three triangles, *Combinatorics (Proceedings of the 5th Hungarian Colloquium, Keszthely, 1976), Vol. II*, Colloquium Mathematica Societatis János Bolyai, Vol. **18**, North-Holland, Amsterdam, pp. 939–945.

Shelah, S. (1972). A combinatorial problem: Stability and order for models and theories in infinitary language, *Pacific J. Math.*, **41**: 247–261.

Sapozhenko, A. A. (2001). On the number of independent sets in extenders, *Discrete Math.* **13**(1): 56–62.

Sauer, N. (1972). On the density of families of sets, *J. Comb. Theory, Ser. A* **13**: 145–147.

Saxton, D. and Thomason, A. (2012). Hypergraph containers, *arXiv:1204.0595v2*.

Schrijver, A. (1978). A short proof of Minc's conjecture, *J. Comb. Theory, Ser. A* **25**: 80–83.

Shamir, E. and Spencer, J. H. (1987). Sharp concentration of the chromatic number in random graphs $G_{n,p}$, *Combinatorica* **7**: 121–130.

Shearer, J. B. (1983). A note on the independence number of triangle-free graphs, *Discrete Math.* **46**: 83–87.

Shearer, J. B. (1985). On a problem of Spencer, *Combinatorica* **5**: 241–245.

Shearer, J. B. (1995). On the independence number of sparse graphs, *Random Struct. Algorithms* **7**: 269–271.

Shelah, S. and Spencer, J. H. (1988). Zero-one laws for sparse random graphs, *J. Am. Math. Soc.* **1**: 97–115.

Shepp, L. A. (1982). The *XYZ*-conjecture and the *FKG*-inequality, *Ann. Probab.* **10**: 824–827.

Slade, G. (2006). *The Lace Expansion and its Applications*, Lecture Notes in Mathematics, Vol. **1879**, Springer-Verlag, Berlin. Lectures from the 34th Summer School on Probability Theory held in Saint-Flour, July 6–24, 2004, Edited and with a foreword by Jean Picard.

Smolensky, R. (1987). Algebraic methods in the theory of lower bounds for boolean circuit complexity, *Proceedings of the 19th ACM STOC*, ACM Press, New York, pp. 77–82.

Spencer, J. H. (1977). Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20**: 69–76.

Spencer, J. H. (1985a). Six standard deviations suffice, *Trans. Am. Math. Soc.* **289**: 679–706.

Spencer, J. H. (1985b). Probabilistic methods, *Graphs Comb.* **1**: 357–382.

Spencer, J. H. (1987). *Ten Lectures on the Probabilistic Method*, SIAM, Philadelphia, PA.

Spencer, J. H. (1990a). Threshold functions for extension statements, *J. Comb. Theory, Ser. A* **53**: 286–305.

Spencer, J. H. (1990b). Counting extensions, *J. Comb. Theory, Ser. A* **55**: 247–255.

Spencer, J. H. (1995). Asymptotic packing via a branching process, *Random Struct. Algorithms* **7**: 167–172.

Subbotovskaya, B. A. (1961). Realizations of linear functions by formulas using $+, \cdot, -$, *Dokl. Akad. Nauk SSSR* **136**(3): 553–555. (in Russian.) English translation in *Sov. Math. Dokl.*, **2**, 110–112.

Suen, W. C. (1990). A correlation inequality and a Poisson limit theorem for nonoverlapping balanced subgraphs of a random graph, *Random Struct. Algorithms* **1**: 231–242.

Székely, L. (1997). Crossing numbers and hard Erdős problems in discrete geometry, *Comb. Probab. Comput.* **6**: 353–358.

Szele, T. (1943). Kombinatorikai vizsgálatok az irányitott teljes gráffal kapcsolatban, *Mat. Fiz. Lapok* **50**: 223–256; For a German translation see: Szele, T. (1966). *Publ. Math. Debrecen* **13**: 145–168.

Szemerédi, E. (1978). Regular partitions of graphs, *Problèmes combinatoires et théorie des graphes (Colloquium International CNRS, University Orsay, Orsay, 1976)*, Colloquium International CNRS, Vol. **260**, CNRS, Paris, pp. 399–401.

Talagrand, M. (1996). Concentration of measures and isopermetric inequalites in product spaces, *Publ. Math. de l'I.H.E.S.* **81**: 73–205.

Tanner, R. M. (1984). Explicit construction of concentrators from generalized *N*-gons, *SIAM J. Algebra. Discrete Meth.* **5**: 287–293.

Tarjan, R. E. (1983). *Data Structures and Network Algorithms*, SIAM, Philadelphia, PA.

Thomason, A. (1987). Pseudo-random graphs, *Ann. Discrete Math.* **33**: 307–331.

Turán, P. (1934). On a theorem of Hardy and Ramanujan, *J. London Math. Soc.* **9**: 274–276.

Turán, P. (1941). On an extremal problem in graph theory, *Mat. Fiz. Lapok* **48**: 436–452.

Valtr, P. (1995). On the minimum number of empty polygons in planar point sets, *Stud. Sci. Math. Hungar.* **30**: 155–163.

van der Hofstad, R. and Spencer, J. H. (2006). Counting connected graphs asymptotically, *Eur. J. Comb.* **27**(8): 1294–1320.

Vapnik, V. N. and Chervonenkis, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.* **16**: 264–280.

Vizing, V. G. (1976). Coloring the vertices of a graph in prescribed colors, *Diskret. Anal.* (29 Metody Diskret. Anal. v Teorii Kodov i Shem): 3–10, 101–.

Wegener, I. (1987). *The Complexity of Boolean Functions*, Wiley-Teubner, New York.

Weil, A. (1948). Sur les courbes algébriques et les variéstés qui sèn déduisent, *Actualités Sci. Ind.* **1041**. iv+85 pp.

Wendel, J. G. (1962). A problem in geometric probability, *Math. Scand.* **11**: 109–111.

Wright, E. M. (1977). The number of connected sparsely edged graphs, *J. Graph Theory* **1**: 317–330.

Yao, A. C. (1985). Separating the polynomial-time hierarchy by oracles, *Proceedings of the 26th Annual IEEE FOCS*, IEEE, New York, pp. 1–10.

Author Index

Note: Index

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

Subject Index

Note: Index

The Probabilistic Method, Fourth Edition. Noga Alon and Joel H. Spencer.

WILEY SERIES IN DISCRETE MATHEMATICS AND OPTIMIZATION

AARTS AND KORST • Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing
AARTS AND LENSTRA • Local Search in Combinatorial Optimization
ALON AND SPENCER • The Probabilistic Method, Fourth Edition
ANDERSON AND NASH • Linear Programming in Infinite-Dimensional Spaces: Theory and Application
ARLINGHAUS, ARLINGHAUS, AND HARARY • Graph Theory and Geography: An Interactive View E-Book
AZENCOTT • Simulated Annealing: Parallelization Techniques
BARTHÉLEMY AND GUÉNOCHE • Trees and Proximity Representations
BAZARRA, JARVIS, AND SHERALI • Linear Programming and Network Flows
BRUEN AND FORCINITO • Cryptography, Information Theory, and Error-Correction: A Handbook for the 21st Century
CHANDRU AND HOOKER • Optimization Methods for Logical Inference
CHONG AND ŻAK • An Introduction to Optimization, Third Edition
COFFMAN AND LUEKER • Probabilistic Analysis of Packing and Partitioning Algorithms
COOK, CUNNINGHAM, PULLEYBLANK, AND SCHRIJVER • Combinatorial Optimization
DASKIN • Network and Discrete Location: Modes, Algorithms and Applications
DINITZ AND STINSON • Contemporary Design Theory: A Collection of Surveys
DU AND KO • Theory of Computational Complexity
ERICKSON • Introduction to Combinatorics
GLOVER, KLINGHAM, AND PHILLIPS • Network Models in Optimization and Their Practical Problems
GOLSHTEIN AND TRETYAKOV • Modified Lagrangians and Monotone Maps in Optimization
GONDRAN AND MINOUX • Graphs and Algorithms *(Translated by S. Vajdā)*
GRAHAM, ROTHSCHILD, AND SPENCER • Ramsey Theory, Second Edition
GROSS AND TUCKER • Topological Graph Theory
HALL • Combinatorial Theory, Second Edition
HOOKER • Logic-Based Methods for Optimization: Combining Optimization and Constraint Satisfaction
IMRICH AND KLAVŽAR • Product Graphs: Structure and Recognition
JANSON, LUCZAK, AND RUCINSKI • Random Graphs
JENSEN AND TOFT • Graph Coloring Problems
KAPLAN • Maxima and Minima with Applications: Practical Optimization and Duality
LAWLER, LENSTRA, RINNOOY KAN, AND SHMOYS, Editors • The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization
LAYWINE AND MULLEN • Discrete Mathematics Using Latin Squares
LEVITIN • Perturbation Theory in Mathematical Programming Applications
MAHMOUD • Evolution of Random Search Trees
MAHMOUD • Sorting: A Distribution Theory
MARTELLI • Introduction to Discrete Dynamical Systems and Chaos
MARTELLO AND TOTH • Knapsack Problems: Algorithms and Computer Implementations
McALOON AND TRETKOFF • Optimization and Computational Logic
MERRIS • Combinatorics, Second Edition
MERRIS • Graph Theory
MINC • Nonnegative Matrices
MINOUX • Mathematical Programming: Theory and Algorithms *(Translated by S. Vajdā)*
MIRCHANDANI AND FRANCIS, Editors • Discrete Location Theory
NEMHAUSER AND WOLSEY • Integer and Combinatorial Optimization
NEMIROVSKY AND YUDIN • Problem Complexity and Method Efficiency in Optimization *(Translated bv E. R. Dawson)*

PACH AND AGARWAL • Combinatorial Geometry
PLESS • Introduction to the Theory of Error-Correcting Codes, Third Edition
ROOS AND VIAL • Ph. Theory and Algorithms for Linear Optimization: An Interior Point Approach
SCHEINERMAN AND ULLMAN • Fractional Graph Theory: A Rational Approach to the Theory of Graphs
SCHIFF • Cellular Automata: A Discrete View of the World
SCHRIJVER • Theory of Linear and Integer Programming
SPALL • Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control
SZPANKOWSKI • Average Case Analysis of Algorithms on Sequences
TOMESCU • Problems in Combinatorics and Graph Theory *(Translated by R. A. Melter)*
TUCKER • Applied Combinatorics, Second Edition
WOLSEY • Integer Programming
YE • Interior Point Algorithms: Theory and Analysis

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.