

# Google matrix

 +1 Recommend this on Google

Leonardo Ermann et al. (2016), Scholarpedia, 11(11):30944.

doi:10.4249/scholarpedia.30944

revision #169983 [link to/cite this article]

- **Dr. Leonardo Ermann**, Laboratorio Tandar, Comision Nacional de Energia Atomica, Buenos Aires, Buenos Aires, Argentina
- **Prof. Klaus Frahm**, Laboratoire de Physique Theorique du CNRS, Universite Paul Sabatier, Toulouse
- **Dima Shepelyansky**, Laboratoire de Physique Théorique, CNRS, Université Paul Sabatier, Toulouse

The **Google matrix**  $G$  of a directed network is a stochastic square matrix with nonnegative matrix elements and the sum of elements in each column being equal to unity. This matrix describes a Markov chain (Markov, 1906-a) of transitions of a random surfer performing jumps on a network of nodes connected by directed links. The network is characterized by an adjacency matrix  $A_{ij}$  with elements  $A_{ij} = 1$  if node  $j$  points to node  $i$  and zero otherwise. The matrix of Markov transitions  $S_{ij}$  is constructed from the adjacency matrix  $A_{ij}$  by normalization of the sum of column elements to unity and replacing columns with only zero elements (dangling nodes) with equal elements  $1/N$  where  $N$  is the matrix size (number of nodes). Then the elements of the Google matrix are defined as

$$G_{ij} = \alpha S_{ij} + (1 - \alpha)/N, \quad (1)$$

where the damping factor  $0 < \alpha < 1$  is the probability that a random surfer follows a link according to the stochastic matrix  $S$  while with probability  $1 - \alpha$  he may jump to any network node. In this form the Google matrix was introduced by Brin and Page in 1998 (Brin, 1998-a) for the description of the World Wide Web (WWW). The right eigenvector of  $G$  with the largest (by modulus) unit eigenvalue is the PageRank vector whose non-negative elements correspond to the stationary probability to find a random surfer on a given node. The product of two Google matrices is also a Google matrix. The above construction of  $S$  can be directly generalized to the case of weighted transitions with the sum of elements in each column of  $S$  equal to unity. The general spectral properties of  $G$  matrix are described below with concrete examples of various real networks. An example image of  $G$  is shown in Figure 1 for the Wikipedia network. The Google matrix belongs to the class of Perron-Frobenius operators which appear in the description of dynamical chaotic systems (Brin, 2002-b) and related Ulam networks (Ulam, 1960-b, Ermann, 2010-a).

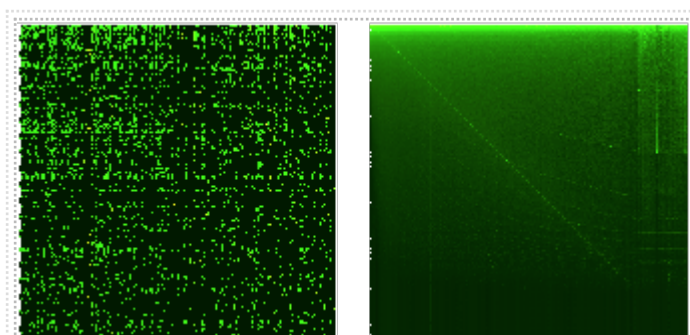


Figure 1: Google matrix  $G$  of the network Wikipedia English articles for Aug 2009 in the basis of nodes ordered by PageRank index  $K$ ; matrix indexes are  $K, K'$  in  $x, y$  axes with top values for  $K = K' = 1$  in the top left corner (see text for definition of indexes  $K, K'$ ). Left panel shows first  $200 \times 200$  matrix elements, right panel shows density of all matrix elements coarse-grained on  $500 \times 500$  cells. Color shows the density of matrix elements changing from black for minimum value  $((1 - \alpha)/N)$  to white for maximum value via green and yellow; here the damping factor is  $\alpha = 0.85$ , the matrix size is  $N = 3282257$ . (from Ermann, 2015-b)

Contents

- 1 Simple network example
- 2 PageRank and CheiRank eigenvectors
- 3 Numerical methods for  $G$  matrix
- 4 Spectrum of  $G$  matrix
- 5 Fractal Weyl law
- 6 Ulam networks
- 7 Linux Kernel networks
- 8 WWW networks of UK universities
- 9 Wikipedia networks
  - 9.1 Top 100 historical figures of Wikipedia
  - 9.2 Wikipedia ranking of world universities
- 10 Multiproduct world trade network
- 11 Other networks
- 12 Outlook
  - 12.1 Random matrix theory for  $G$  ?
  - 12.2 Anderson localization for Google matrix eigenstates
  - 12.3 Reduced Google matrix
- 13 Historical notes
- 14 Lectures about Google matrix
- 15 Research articles
- 16 Books and reviews
- 17 External links to data sets
- 18 Video lectures

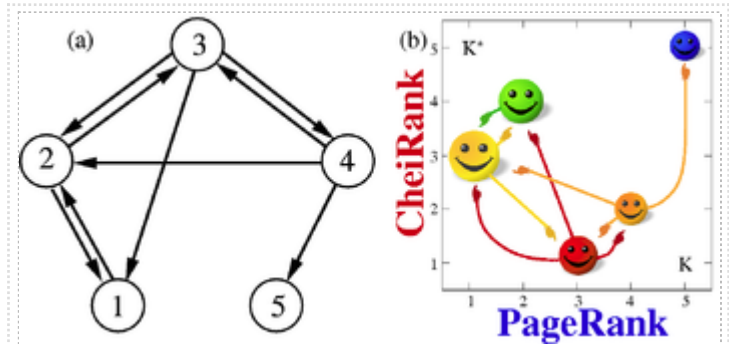


Figure 2: (a) Example of simple network with directed links between 5 nodes. (b) Distribution of 5 nodes from (a) on the PageRank-CheiRank plane ( $K, K^*$ ), where the size of node is proportional to PageRank probability  $P(K)$  and color of node is proportional to CheiRank probability  $P^*(K^*)$ , with maximum at red and minimum at blue; the location of nodes of panel (a) on PageRank-CheiRank plane is: (2,4), (1,3), (3,1), (4,2), (5,5) for original nodes  $i = 1, 2, 3, 4, 5$  respectively; PageRank and CheiRank vectors are computed from the Google matrices  $G$  and  $G^*$  shown in Figure 3 at a damping factor  $\alpha = 0.85$ . (from Ermann, 2015-b)

## Simple network example

An example of simple directed network with 5 nodes is shown in Figure 2(a), here nodes are numbered from 1 to 5. The distribution of nodes on the PageRank-CheiRank plane of indexes ( $K, K^*$ ) is shown in Figure 2(b) (see definition of ( $K, K^*$ ) in next Section). The corresponding adjacency matrix  $A$  and matrices  $S, G$  are given in Figure 3(a,c,e). In addition it is useful to consider the network with inverted link directions. The corresponding adjacency matrix  $A^*$  and related matrices  $S^*, G^*$  are shown for this case in Figure 3(b,d,f).

## PageRank and CheiRank eigenvectors

According to the Perron-Frobenius theorem all eigenvalues  $\lambda_i$  of  $G$  are distributed inside the unitary circle  $|\lambda_i| \leq 1$ . The right eigenvectors  $\psi_i(j)$  are defined by the equation  $\sum_{j'} G_{jj'} \psi_i(j') = \lambda_i \psi_i(j)$ . In the following we will also use the notation *eigenstates* for such eigenvectors in analogy to eigenstates in quantum Hamiltonian systems. It can be shown that for  $\alpha < 1$  the eigenvalue  $\lambda_0 = 1$  is not degenerate with only one right eigenvector called the PageRank vector  $P$ . The positive elements  $P(j)$  of the PageRank vector, when the sum of them is normalized to unity, give the probability to find a random surfer on a node  $j$  in the stationary limit of long times. Only the eigenvectors of  $S$  for  $\lambda_0 = 1$  (which may be degenerate) are affected by the damping factor while other eigenvectors of  $S$  (with eigenvalues  $\lambda_i \neq 1$ ) are also eigenvectors of  $G$  independent of  $\alpha$  due to their orthogonality to the left eigenvector (with identical unit entries) at  $\lambda_0 = 1$  but with rescaled eigenvalues  $\alpha \lambda_i$  for  $G$  (Langville, 2006-b, Gantmacher, 2000-b). The variation of  $\alpha$  in the range  $0.5 \leq \alpha \leq 0.95$  does not

significantly affect the PageRank probabilities so that the results are usually presented for a typical value  $\alpha = 0.85$  (Langville, 2006-b, Ermann, 2015-b).

The network with inverted link directions is described by the matrix  $G^*$ , the PageRank eigenvector of  $G^*$  is called the CheiRank vector. The statistical properties of the CheiRank vector  $P^*$  have been first studied in (Chepelianskii, 2010-a) for the Linux Kernel network and later extended to the Wikipedia network (Zhirov, 2010-a).

All network nodes can be ordered by monotonically decreasing probabilities of PageRank or CheiRank vectors providing indexes  $K$  and  $K^*$  with the maximal probability at  $K = K^* = 1$ , and minimum probability at  $K = K^* = N$ . The PageRank index  $K$  is used for the presentation of  $G$  in Figure 1: here all nodes are ordered by the PageRank index  $K$  and the strength of matrix elements  $G_{KK'}$  is shown by color on a small scale (left panel) and on the whole matrix scale with coarse-graining (right panel).

The distribution of nodes on the PageRank-CheiRank plane for the simple network example is shown in Figure 2(b).

It is known that on average the PageRank probability is proportional to the number of ingoing links, characterizing how popular or known a given node is (Langville, 2006-b). Real networks are often characterized by power law distributions for the number of ingoing and outgoing links per node  $w_{in,out} \propto 1/k^{\mu_{in,out}}$  with typical exponents  $\mu_{in} \approx 2.1$  and  $\mu_{out} \approx 2.7$  for the WWW (Donato, 2004-a, Dorogovtsev, 2010-b, Newman, 2010-b). Assuming that the PageRank probability decays algebraically as  $P(i) \sim 1/K(i)^\beta$  we obtain that the number of nodes  $N_P$  with PageRank probability  $P$  scales as  $N_P \sim 1/P^{\mu_{in}}$  with  $\mu_{in} = 1 + 1/\beta$ . Thus for the typical above values of  $\mu_{in}$  we have  $\beta = 1/(\mu_{in,out} - 1) \approx 0.9$  for PageRank  $P$  and  $\beta \approx 0.6$  for CheiRank  $P^*$  which is proportional to the number of outgoing links due to the inversion of direction. Examples of the probability decay of  $P, P^*$  are shown in Figure 4 for networks of Wikipedia and University of Cambridge. It should be noted that the decay is only approximately described by a power law. WWW networks of larger sizes (about 3.5 billions) also only approximately described by an algebraic decay (Meusel, 2015-a).

For the case of the simple network visible in (a) the distribution of nodes on the PageRank-CheiRank plane is shown in (b). The distributions for Wikipedia and Linux Kernel networks are shown in Figure 5. It is convenient to characterize the network by the PageRank-CheiRank correlator  $\kappa = N \sum_{i=1}^N P(K(i))P^*(K^*(i)) - 1$  (Chepelianskii, 2010-a) which takes different values depending on internal network properties even if the decay of PageRank and CheiRank probabilities is approximately the same in these networks. Thus we have  $\kappa = 4.08; -0.034$  for panels (a;b) of Figure 5. At small correlators the density is homogeneous along the line  $\ln K \approx -\ln K^*$  while for large positive values it is more concentrated along the line  $\ln K \approx \ln K^*$ . More correlator values for different networks are given in (Ermann, 2015-b)].

It is also useful to rank network nodes by a 2DRank using a combination of PageRank and CheiRank: for this one considers a sequence of squares on the PageRank-CheiRank plane with the left bottom corner at  $K = K^* = 1$  and increasing size placing nodes in 2DRank  $K_2$  in order of their appearance on square sides (see

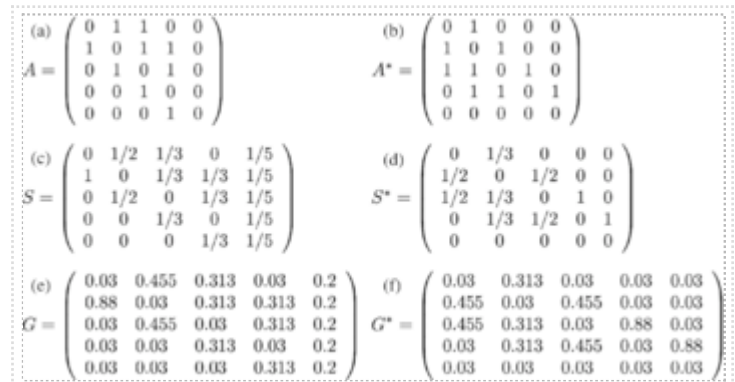


Figure 3: (a) Adjacency matrix  $A$  of network of (a) with indexes used there, (b) adjacency matrix  $A^*$  for the network with inverted links; matrices  $S$  (c) and  $S^*$  (d) corresponding to the matrices  $A, A^*$ ; the Google matrices  $G$  (e) and  $G^*$  (f) corresponding to matrices  $S$  and  $S^*$  for  $\alpha = 0.85$  (only 3 digits of matrix elements are shown. (from Ermann, 2015-b)

more detail at (Zhirov, 2010-a)).

The characterization of a directed network by both PageRank and CheiRank probabilities allows to be characterized in a better way the information flow on the network taking into account ingoing flows, related to PageRank, and outgoing flows, related to CheiRank (see more detail in Ermann, 2015-b).

The density distribution of nodes on the PageRank-CheiRank plane is shown in Figure 5 for Wikipedia (a) and Linux (b) networks. The density

$W(K, K^*) = d^2 N_i / dK dK^*$  is computed on logarithmic-equidistant grid (cells) so that  $W(K, K^*)$  is given by the number  $N_i$  of network nodes appearing in a given cell divided by the cell area on  $(K, K^*)$  plane.

## Numerical methods for $G$ matrix

Usually scale-free networks have algebraic distributions of ingoing and outgoing links with a relatively small average number of links  $N_\ell/N$  per node (see e.g. Dorogovtsev, 2010-b, Newman, 2010-b) corresponding to a very sparse adjacency matrix. For example for the networks of Figure 1, Figure 4, Figure 5 we have  $N_\ell/N \approx 22, 10, 2.3$ . Therefore the PageRank vector can be efficiently computed by the power method which consists of multiplying repeatedly the matrix  $G$  to a random initial (sum normalized) vector. Each such matrix vector multiplication can be implemented by a loop over the link index and has therefore a complexity  $N_\ell$  which is much smaller than the matrix size  $N^2$ . The particular contributions due to the dangling nodes or the damping factor in the Google matrix correspond to a complexity  $N$  and do not increase the overall complexity. Due to the presence of a gap between  $\lambda = 1$  and the next eigenvalue with  $|\lambda| \leq 1 - \alpha$  the convergence of the PageRank vector is exponential (e.g. after about 150 iterations the variation

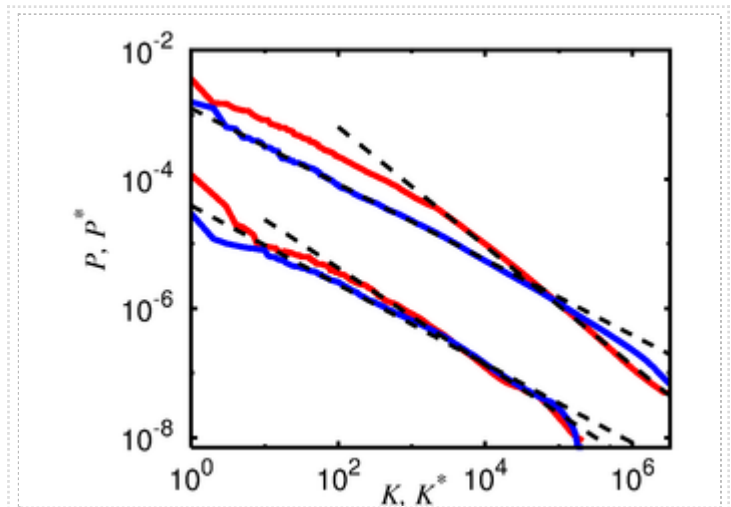


Figure 4: Dependence of probabilities of PageRank  $P$  (red curve) and CheiRank  $P^*$  (blue curve) vectors on the corresponding rank indexes  $K$  and  $K^*$  for networks of Wikipedia Aug 2009 (top curves) and University of Cambridge (bottom curves, moved down by a factor 100). The straight dashed lines show the power law fits for PageRank and CheiRank with the slopes  $\beta = 0.92; 0.58$  respectively, corresponding to  $\beta = 1/(\mu_{in,out} - 1)$  for Wikipedia, and  $\beta = 0.75, 0.61$  for Cambridge. (from Ermann, 2015-b)

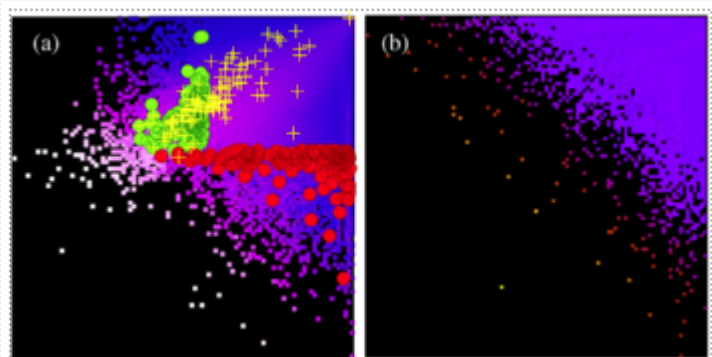


Figure 5: Density distribution of network nodes  $W(K, K^*) = d^2 N_i / dK dK^*$  shown on the plane of PageRank and CheiRank indexes in log-scale ( $\log_N K, \log_N K^*$ ) for all  $1 \leq K, K^* \leq N$ , density is computed over equidistant grid in plane ( $\log_N K, \log_N K^*$ ) with  $100 \times 100$  cells; color shows average value of  $W$  in each cell for the unit normalization condition for all nodes. Density  $W(K, K^*)$  is shown by color with blue for minimum in (a),(b) and white (a) and yellow (b) for maximum (black for zero). Panel (a): data for Wikipedia Aug (2009),  $N = 3282257$ , green/red points show top 100 persons from PageRank/CheiRank, yellow pluses show top 100 persons from (Hart, 1992-b). Panel (b): density distribution for Linux Kernel V2.4 network with  $N = 85757$ . (from Ermann, 2015-b)

of the vector norm becomes less than  $10^{-12}$  for the Wikipedia network of Figure 1).

For typical networks the whole set of nodes can be decomposed in invariant subspace nodes and fully connected core space nodes leading to a block structure of the matrix  $S$  (Frahm, 2011-a):

$$S = \begin{pmatrix} S_{ss} & S_{sc} \\ 0 & S_{cc} \end{pmatrix}.$$

The core space block  $S_{cc}$  contains links between core space nodes and the coupling block  $S_{sc}$  may contain links from certain core space nodes to certain invariant subspace nodes. In contrast there are no links from subspace nodes to the nodes of core space (block with zero elements). By construction there are no links from nodes of invariant subspaces to the nodes of the core space. The subspace-subspace block  $S_{ss}$  is actually composed of many diagonal blocks for different invariant subspaces whose number can generally be rather large. Each of these blocks corresponds to a column sum normalized matrix with positive elements of the same type as  $G$  and has therefore at least one unit eigenvalue. This leads to a high degeneracy  $N_1$  of the eigenvalue  $\lambda_0 = 1$  of  $S$ , for example  $N_1 \sim 10^3$  for the case of UK universities (see below). For each initial node one can iteratively determine a limit set of nodes that can be reached by a chain of non-zero matrix elements of  $S$  from the initial node. This set extends either over (nearly) the full network or it is limited, e.g. less than 10% of all network nodes. In the first case the initial node is attributed to the core space and in the second case the limit set defines an invariant subspace. For example for the WWW networks of UK universities, all invariant subspaces typically represent about 20 – 30% of the whole network.

The largest eigenvalues of  $S_{cc}$  (taken by their modulus) can be efficiently obtained by the powerful Arnoldi method [1] ([https://en.wikipedia.org/wiki/Arnoldi\\_iteration](https://en.wikipedia.org/wiki/Arnoldi_iteration)) (see also Stewart, 2001-b, Ermann, 2015-b). The main idea of this method is to construct, by an iterative scheme of matrix vector multiplication and orthogonalization, an orthonormal basis on a subspace of modest dimension  $n_A$ , called *Krylov space*, and to diagonalize the representation matrix of  $G$  on this subspace which provides typically good approximations for the largest eigenvalues of  $G$  (taken by their modulus). Also the corresponding eigenvectors are available by this method.

For the particular case of networks with a nearly triangular adjacency matrix the effects of numerical and round-off errors on the precision of eigenvalues may become very important and require high precision computations for the Arnoldi method or other particular special methods (Frahm, 2014-a, Ermann, 2015-b).

## Spectrum of $G$ matrix

Typical complex eigenvalue spectra of  $G$ ,  $G^*$  are shown in Figs.6,7 for examples of UK universities and Wikipedia networks.

The spectra of  $S$ ,  $S^*$  of universities of Cambridge and Oxford in 2006 are shown in Figure 6. These networks have a size  $N \approx 2 \times 10^5$ . All subspace eigenvalues and  $n_A = 20000$  core eigenvalues with maximal  $|\lambda|$  are shown. There is a strong degeneracy of the unit eigenvalue (about 2% of all eigenvectors). The global spectral structure has visible similarities with the spectra of random orthostochastic matrices of small size  $N = 3, 4$  analyzed numerically and analytically in (Zyczkowski, 2003a). The spikes visible at certain angles  $2\pi/p$  for  $p = 2, 3, 4, 6$  correspond to approximate cycles of length  $p$  for the links between particular nodes ("close friends") that appear in top rank positions of the corresponding eigenstates of such eigenvalues.

The spectrum of the core space of  $S$  for the Wikipedia network (Aug 2009) is shown in Figure 7. The eigenstates with maximal values of  $|\lambda|$  correspond to certain quasi-isolated communities, they are marked by the most

frequent words appearing in largest amplitudes of the corresponding eigenvectors. The results show that the eigenvectors of  $G$  clearly identify interesting specific communities of the network.

## Fractal Weyl law

In quantum mechanics the Weyl law (1912) gives a fundamental relation between the number of states and the phase volume of a Hamiltonian closed system of dimension  $d$ . The generalization to operators of open quantum systems, appearing in the problems of quantum chaotic scattering with complex eigenenergies (Gaspard, 2014b), has been done relatively recently by (Sjostrand, 1990a). The spectrum of corresponding operators has a complex spectrum  $\lambda$ . The spread width  $\gamma = -2 \ln |\lambda|$  of eigenvalues  $\lambda$  determines the life time of a corresponding eigenstate.

According to the fractal Weyl law the number of eigenvalues  $N_\gamma$ , which have escape rates  $\gamma$  in a finite band width  $0 \leq \gamma \leq \gamma_b$ , scales as

$$N_\gamma \propto \hbar^{-d/2} \propto N^\nu, \quad \nu = d/2,$$

where  $d$  is a fractal dimension of a classical strange repeller formed by classical orbits nonescaping in future and past times,  $\hbar$  is the Planck constant. In the context of eigenvalues  $\lambda$  of the Google matrix we have  $\gamma = -2 \ln |\lambda|$ . As usual the Planck constant is inversely proportional to the number of states, which is determined by the matrix size, so that  $\hbar \propto 1/N$ .

The fractal Weyl law of open systems with a fractal dimension  $d < 2$  leads to a striking consequence: only a relatively small fraction of eigenvalues  $\mu_W \sim N_\gamma/N \propto \hbar^{(2-d)/2} \propto N^{(d-2)/2} \ll 1$  has finite values of  $|\lambda|$  while almost all eigenstates of the matrix operator of size  $N \propto 1/\hbar$  have  $\lambda \rightarrow 0$ . The eigenstates with finite  $|\lambda| > 0$  are related to the classical fractal sets of orbits non-escaping neither in the future neither in the past. The fractal Weyl law for the Ulam networks is discussed in next Section. This law has been shown to be valid for the Linux Kernel network with  $d \approx 1.3$  (see Figure 8 and related Section). For the Physical Review network it is found that  $d \approx 1$  (Frahm, 2014-a).

There is an expectation that the eigenstates with large  $|\lambda|$ , forming the fractal Weyl law, capture certain hidden interesting communities. It is qualitatively confirmed by the analysis of eigenvectors of Wikipedia matrix  $G$  (see

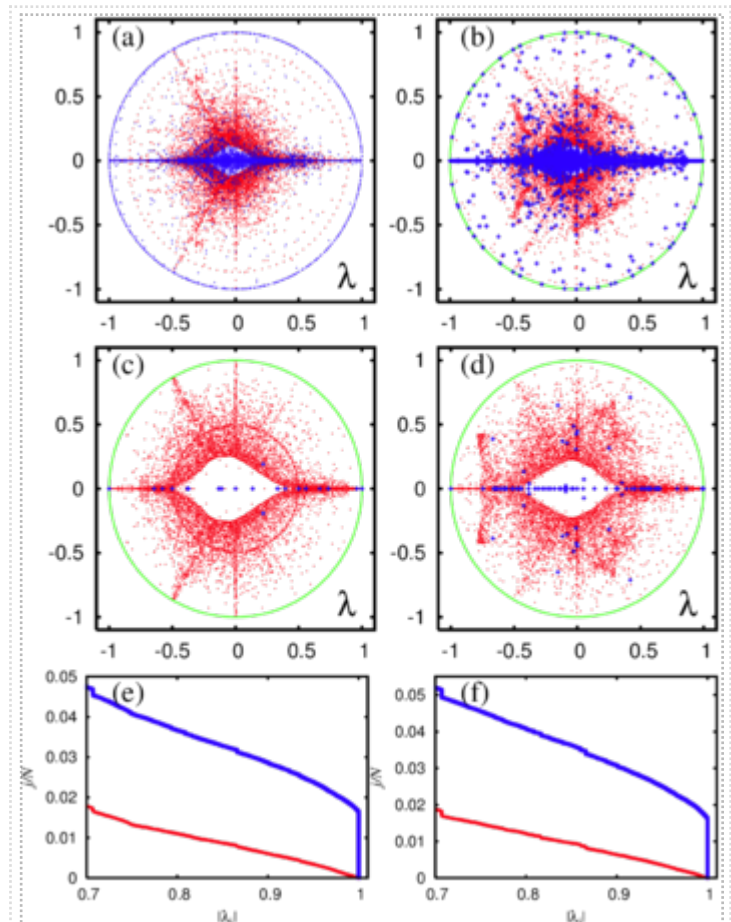


Figure 6: Panels (a) and (b) show the complex eigenvalue spectrum  $\lambda$  of the matrix  $S$  for the University of Cambridge 2006 and Oxford 2006 respectively. The spectrum  $\lambda$  of the matrix  $S^*$  for Cambridge 2006 and Oxford 2006 are shown in panels (c) and (d). Eigenvalues  $\lambda$  of the core space are shown by red points, eigenvalues of isolated subspaces are shown by blue points and the green curve (when shown) is the unit circle. Panels (e) and (f) show the fraction  $j/N$  of eigenvalues with  $|\lambda| > |\lambda_j|$  for the core space eigenvalues (red bottom curve) and all eigenvalues (blue top curve) from top row data for Cambridge 2006 and Oxford 2006. (from Ermann, 2015-b)

Figure 7 and Frahm, 2014-a).

Mathematical aspects of the fractal Weyl law are reviewed in (Nonnenmacher, 2014b).

## Ulam networks

By construction the Google matrix belongs to the class of Perron-Frobenius operators which naturally appear in ergodic theory and dynamical systems with Hamiltonian or dissipative dynamics (Brin, 2002-b). In 1960 Ulam (Ulam, 1960-b) proposed a method, now known as the Ulam method, for a construction of finite size approximants for the Perron-Frobenius operators of dynamical maps. The method is based on discretization of the phase space and construction of a Markov chain based on probability transitions between such discrete cells given by the dynamics. Using as an example a simple chaotic map Ulam made a conjecture that the finite size approximation converges to the continuous limit when the cell size goes to zero. Indeed, it has been proven that for hyperbolic maps in one and higher dimensions the Ulam method converges to the spectrum of the continuous system. The probability flows in dynamical systems have rich and nontrivial features of general importance, like simple and strange attractors with localized and delocalized dynamics governed by simple dynamical rules. Such objects are generic for nonlinear dissipative dynamics and therefore they can have relevance for actual WWW structure. The analysis of Ulam networks, generated by the Ulam method, allows to obtain a better intuition about the spectral properties of Google matrix.

The Ulam method works as following: the phase space of a dynamical map is divided in equal cells and a number of trajectories  $N_c$  is propagated by a map iteration. Thus a number of trajectories  $N_{ij}$  arriving from cell  $j$  to cell  $i$  is determined. Then the matrix of Markov

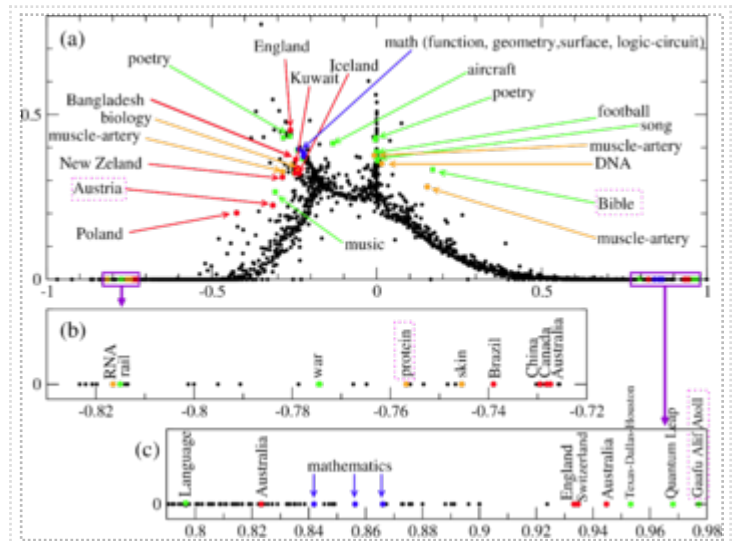


Figure 7: Complex eigenvalue spectrum of the matrices  $S$  for English Wikipedia Aug 2009. Highlighted eigenvalues represent different communities of Wikipedia and are labeled by the most repeated and important words following word counting of first 1000 nodes. Panel (a) shows complex plane for positive imaginary part of eigenvalues, while panels (b) and (c) zoom in the negative and positive real parts. (from Ermann, 2015-b)

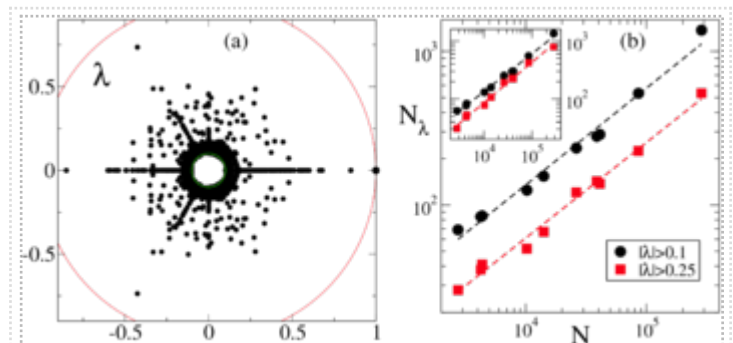


Figure 8: Panel (a) shows distribution of eigenvalues  $\lambda$  in the complex plane for the Google matrix  $G$  of the Linux Kernel version 2.6.32 with  $N = 285509$  and  $\alpha = 0.85$ ; the solid curves represent the unit circle and the lowest limit of computed eigenvalues. Panel (b) shows dependence of the integrated number of eigenvalues  $N_\lambda$  with  $|\lambda| > 0.25$  (red squares) and  $|\lambda| > 0.1$  (black circles) as a function of the total number of processes  $N$  for versions of Linux Kernels. The values of  $N$  correspond (in increasing order) to Linux Kernel versions 1.0, 1.1, 1.2, 1.3, 2.0, 2.1, 2.2, 2.3, 2.4, 2.6. The power law  $N_\lambda \propto N^\nu$  has fitted values  $\nu_{|\lambda|>0.25} = 0.622 \pm 0.010$  and  $\nu_{|\lambda|>0.1} = 0.630 \pm 0.015$ . Inset shows data for the Google matrix  $G^*$  with inverse link directions, the corresponding exponents are  $\nu_{|\lambda|>0.25}^* = 0.696 \pm 0.010$  and  $\nu_{|\lambda|>0.1}^* = 0.652 \pm 0.007$ . (from Ermann, 2015-b)

transition is defined as  $S_{ij} = N_{ij}/N_c$ . By construction this matrix belongs to the class of Perron-Frobenius operators which includes the Google matrix. The physical meaning of the coarse grain description by a finite number of cells is that it introduces in the system a noise of cell size amplitude. More details can be found at (Ermann, 2015-b).

Examples of eigenstates of the Ulam approximate of Perron-Frobenius operators (UPFO) of two Ulam networks are shown in Figure 9. The networks are generated by the Ulam method applied to the dynamical map

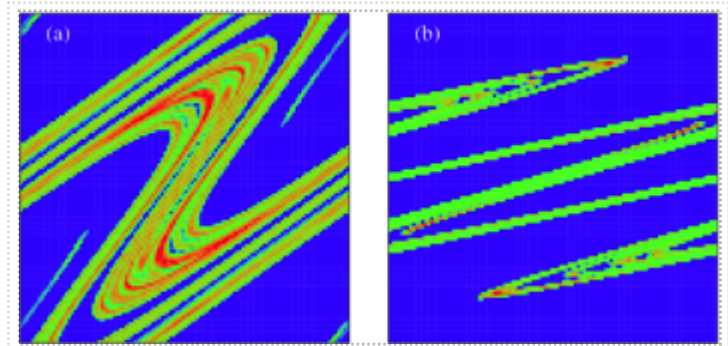


Figure 9: Phase space representation of eigenstates of the Ulam approximate of the Perron-Frobenius operator (UPFO)  $S$  for  $N = 110 \times 110$  cells (color is proportional to absolute value  $|\psi_i|$  with red for maximum and blue for zero). Panel (a) shows an eigenstate with maximum eigenvalue  $\lambda_1 = 0.756$  for the Chirikov standard map with absorption at  $K_s = 7$ ,  $a = 2$ ,  $\eta = 1$ , the space region is  $(-aK_s/4\pi \leq y \leq aK_s/4\pi, 0 \leq x \leq 1)$ , the fractal dimension of the strange repeller set nonescaping in future is  $d_e = 1 + d/2 = 1.769$ . Panel (b) shows an eigenstate at  $\lambda = 1$  of the UPFO of the map without absorption at  $K_s = 7$ ,  $\eta = 0.3$ , the shown space region is  $(-1/\pi \leq y \leq 1/\pi, 0 \leq x \leq 1)$  and the fractal dimension of the strange attractor is  $d = 1.532$ . (from Ermann, 2015-b)

$$\bar{y} = \eta y + \frac{K_s}{2\pi} \sin(2\pi x), \quad \bar{x} = x + \bar{y} \pmod{1}.$$

Here bars mark the variables after one map iteration and we consider the dynamics to be periodic on a torus so that  $0 \leq x \leq 1$ ,  $-1/2 \leq y \leq 1/2$ ;  $K_s$  is a dimensionless parameter of chaos. At  $\eta = 1$  we have the area-preserving symplectic map, known as the Chirikov standard map (Chirikov, 2008-b), for  $0 < \eta < 1$  we have a dissipative dynamics with a strange attractor. At  $\eta = 1$  the absorption is introduced so that all orbits leaving the interval  $-aK_s/4\pi \leq y \leq aK_s/4\pi$  are absorbed after one iteration. Thus the UPFO has the maximal eigenvalue  $\lambda < 1$  with a strange repeller of orbits remaining in the system after many map iterations. For the dissipative case at  $\eta < 1$  the orbits drop on a strange attractor (see Figure 9). The fractal dimension  $d$  of these strange sets depends on the system parameters that allows to vary it in a large range  $0 < d < 2$ . The spectral analysis of UPFO in these systems confirms the validity of the fractal Weyl law for variation of the exponent  $\nu$  in the interval  $0 < \nu < 1$  (Ermann, 2010-a).

## Linux Kernel networks

Modern software codes represent now complex large scale structures and analysis and optimization of their architecture become a challenge. An interesting approach to this problem was proposed in (Chepelienskii, 2010-a) on the basis of directed network analysis. Thus the Procedure Call Networks (PCN) are constructed for the open source programs of Linux Kernel written in the C programming language. In this language the code is structured as a sequence of procedures calling each other. Due to that feature the organization of a code can be naturally represented as a PCN, where each node represents a procedure and each directed link corresponds to a procedure call. For the Linux source code such a directed network is built by its lexical scanning with the identification of all the defined procedures. For each of them a list keeps track of the procedures calls inside their definition.



It is found that the PageRank and CheiRank probabilities in this network decay as a power law with the approximate exponent values  $\beta \approx 1; 0.5$  respectively. For V2.6.32 the top three procedures of PageRank are *printk*, *memset*, *kfree*, while at the top of CheiRank we have *start\_kernel*, *btrfs\_ioctl*, *menu\_finalize*. These procedures perform rather different tasks with *printk* reporting messages and *start\_kernel* initializing the Kernel and managing the repartition of tasks. This gives an idea that both PageRank and CheiRank order can be useful to highlight different aspects of directed and inverted flows on our network. Of course, in the context of WWW ingoing links related to PageRank are less vulnerable as compared to outgoing links related to CheiRank, which can be modified by a user rather easily.

For the Linux Kernel network the correlator  $\kappa$  between PageRank and CheiRank vectors is close to zero. This confirms the statistical independence of these two vectors. The density distribution of nodes of the Linux Kernel network, shown in Figure 5(b), has a homogeneous distribution along  $\ln K + \ln K^* = \text{const}$  lines demonstrating once more absence of correlations between  $P(K(i))$  and  $P^*(K^*(i))$ . Indeed, such homogeneous distributions appear if nodes are generated randomly with factorized probabilities  $P(i), P^*(i)$ .

The physical reasons for absence of correlations between  $P(K), P^*(K^*)$  have been explained (Chepelianskii, 2010-a) on the basis of the concept of *separation of concerns* in software architecture. It is argued that a good code should decrease the number of procedures that have high values of both PageRank and CheiRank since such procedures will play a critical role in error propagation since they are both popular and highly communicative at the same time. For example in the Linux Kernel, *do\_fork*, that creates new processes, belongs to this class. Such critical procedures may introduce subtle errors because they entangle otherwise independent segments of code. The above observations suggest that the independence between popular procedures, which have high  $P(K(i))$  and fulfill important but well defined tasks, and communicative procedures, which have high  $P^*(K^*(i))$  and organize and assign tasks in the code, is an important ingredient of well structured software.

The different Linux versions from V1.0 to V2.6 provide a network size variation in a range  $2752 \leq N \leq 285510$  allowing to demonstrate the validity of the fractal Weyl law with the fractal dimension  $d \approx 1.3$  (see Figure 8). Linux network data sets are available at (FETNADINE, 2015-e).

## WWW networks of UK universities

The WWW networks of certain UK universities for the years between 2002 and 2006 are publicly available at (UK universities, 2006-e; selected networks are given at EU-FET-NADINE site FETNADINE, 2015-e). The universal emergence of PageRank, properties of PageRank and CheiRank vectors and the spectral properties of  $G, G^*$  are analyzed in detail at (Frahm, 2011-a, see also Figs.4,6). It is established that the rescaled distribution of sizes  $d_i$  of invariant subspaces of university networks is described by a universal function  $F(x) = 1/(1 + 2x)^{3/2}$  with  $x = d_i / \langle d \rangle$ , where  $\langle d \rangle$  is an average subspace dimension computed for a WWW of a given university. This is related with a universal power law decay of PageRank probability  $P \propto 1/K^{2/3}$  emerging at  $\alpha \rightarrow 0$ . It is shown that for certain universities the maximal eigenvalue  $\lambda_c$  of the core space is enormously close to unity (e.g  $1 - \lambda_c < 10^{-16}$ ); the corresponding eigenstates are localized on a small node subset. More results are available at (Frahm, 2011-a, Ermann, 2015-b).

## Wikipedia networks

The free online encyclopedia Wikipedia is a huge repository of human knowledge. Its size is growing permanently accumulating a enormous amount of information. The hyperlink citations between Wikipedia articles provide an important example of directed networks evolving in time for many different languages.

The decay of probabilities of PageRank and CheiRank are shown in Figure 4 for English Wikipedia edition of August 2009 (Zhirov, 2010-a). They are satisfactorily described by a power law decay with exponents  $\beta_{PR,CR} = 1/(\mu_{in,out} - 1) = 0.92; 0.58$ .

The density distribution of articles over the PageRank-CheiRank plane ( $\log_N K, \log_N K^*$ ) is shown in Figure 5(a). The density is very different from those generated by the product of independent probabilities of  $P$  and  $P^*$  which gives the distribution similar to the case of the Linux Kernel network shown in Figure 5(b) where the correlator  $\kappa$  between PageRank and CheiRank vectors is almost zero (while for Wikipedia  $\kappa = 4.08$ ).

The difference between PageRank and CheiRank is clearly seen from the names of articles with highest ranks. At the top of PageRank there are 1. *United States*, 2. *United Kingdom*, 3. *France* while for CheiRank one finds 1. *Portal:Contents/Outline of knowledge/Geography and places*, 2. *List of state leaders by year*, 3. *Portal:Contents/Index/Geography and places*. Clearly the PageRank selects first articles on a broadly known subject with a large number of ingoing links while the CheiRank selects first highly communicative articles with many outgoing links. The 2DRank combines these two characteristics of information flow on directed network. At the top of 2DRank  $K_2$  one has 1. *India*, 2. *Singapore*, 3. *Pakistan*. Thus, these articles are most known/popular and most communicative at the same time. Results of ranking of the Wikipedia Aug 2009 edition for various categories are available at (Wiki2009, 2010-e).

The complex spectrum of eigenvalues of  $G$  for this Wikipedia network is shown in Figure 7 (due to symmetry of eigenvalues  $\lambda = \lambda^*$  only the upper plane of  $\lambda$  is shown). As for university networks, the spectrum also has some invariant subspaces resulting in degeneracies of the leading eigenvalue  $\lambda = 1$  of  $S, S^*$ . However, due to the stronger connectivity of the Wikipedia network these subspaces are significantly smaller compared to university networks.

It is expected that the eigenstates with large values of  $|\lambda|$  select certain specific communities. If  $|\lambda|$  is close to unity then the relaxation of probability from such nodes is rather slow and we can expect that such eigenstates highlight some new interesting information even if these nodes are located in the tail of the PageRank. The important advantage of the Wikipedia network is that its nodes are Wikipedia articles with a relatively clear meaning allowing to understand the origins of appearance of certain nodes in one community. The frequency analysis of words appearing at the largest amplitudes of eigenvectors with large modulus of  $|\lambda|$  confirms this expectation (see Figure 7 and Ermann, 2015-b).

## Top 100 historical figures of Wikipedia

There is always significant public interest to know who are the most significant historical figures, or persons, of humanity. The Hart list of the top 100 people who, according to him, most influenced human history is available at (Hart, 1992-b). Hart “ranked these 100 persons in order of importance: that is, according to the total amount of influence that each of them had on human history and on the everyday lives of other human beings.” Of course, a human ranking can always be objected arguing that an investigator has his or her own preferences. Also investigators from different cultures can have different viewpoints on the same historical figure. Thus it is important to perform a ranking of historical figures on purely mathematical and statistical grounds which exclude any cultural and personal preferences of investigators.

A detailed two-dimensional ranking of persons of the English Wikipedia August 2009 was done by (Zhirov, 2010-a). Earlier studies had been done in a non-systematic way without any comparison with established top 100 lists. The distribution of the top 100 PageRank, CheiRank, and Hart’s persons on PageRank-CheiRank plane is shown in Figure 5(a). For the PageRank top 100 list the overlap with the Hart list is at 35% (PageRank), 10% (2DRank), and almost zero for CheiRank. This is attributed to a very broad distribution of historical figures on

the 2D plane, as shown in Figure 5(a), and a large variety of human activities. The distribution of the top 100 persons of the Wikipedia August 2009 remains stable and compact for PageRank and 2DRank for the period 2007–2011 while for CheiRank the fluctuations of positions are large (Ermann, 2015-b). This is due to the fact that outgoing links are easily modified and fluctuating.

However, it is important to take into account not only the view point of English Wikipedia but also to consider viewpoints of other language editions of Wikipedia representing other cultures. Thus the ranking of world historical figures was done on the basis of 24 editions (Eom, 2015-a). In 2014 these 24 languages cover 59 percent of world population, and the corresponding 24 editions cover 68 percent of the total number of Wikipedia articles in all 287 available languages. Also the selection of people from the rank list of each edition is now done in an automatic computerized way. For this a list of about 1.1 million biographical articles about people with their English names is generated. From this list of persons, with their biographical article title in the English Wikipedia, the corresponding titles in other language editions are determined using the inter-language links provided by Wikipedia. The rank score of each persons is averaged over all 24 editions thus equally taking into account the opinions of these 24 cultures.

For PageRank the top global three historical figures are Carl Linnaeus, Jesus, and Aristotle. All other ranks are available at (TopWikiPeople, 2014-e). The overlap of top 100 PageRank and Hart's lists have 43 common persons. The fact that Carl Linnaeus is the top historical figure of the Wikipedia PageRank list came as a surprise for media and the broad public (see Refs. in Ermann, 2015-b). This ranking is due to the fact that Carl Linnaeus created a classification of world species including animals, insects, herbs, trees, etc. Thus all articles of these species point to the article Carl Linnaeus in various languages. As a result Carl Linnaeus appears on almost all top positions in all 24 languages. Hence, even if a politician, like Barak Obama, takes the second position in his country language EN (Napoleon is at the first position in EN) he is usually placed at a low ranking in other language editions. As a result Carl Linnaeus takes the first global PageRank position. More details, including the distribution of historical figures over world countries and 35 centuries of human history, can be found at (Eom, 2015-a, Ermann, 2015-b, TopWikiPeople, 2014-e). The results of other research groups for ranking of historical figures of Wikipedia are referenced in (PantheonMIT, 2015-e, StonyBrookranking, 2015-e, see more Refs. in Eom, 2015-a, Ermann, 2015-b).

## Wikipedia ranking of world universities

The ranking of universities for the English Wikipedia edition Aug 2009 was done in (Zhirov, 2010-a) giving at the top of PageRank list: University of Harvard, University of Oxford, University of Cambridge with the overlap of 70% for the top 100 list of Academic ranking of world universities of Shanghai in 2009 (Shanghai, 2015-e). All results of ranking of universities are available at (Wiki2009, 2010-e). However, it is important also to take into account the opinions of other cultures and not only of the English edition to determine the university ranking.

Thus, the above approach for ranking of historical figures is also used for the Wikipedia ranking of world universities, using the same datasets of 24 Wikipedia editions. The combined results (Lages, 2016-a) obtained from top 100 universities of each edition give total global lists of 1025, 1379, 1560 universities for PageRank, CheiRank, and 2DRank algorithms respectively. All these results are available at (TopWikiUniversities, 2015-e). The distribution of 1025 PageRank universities over the world countries is shown in Figure 10. For the global PageRank list the top three positions are taken by University of Cambridge, University of Oxford, Harvard University. The overlap of top 100 PageRank list with top 100 of Academic ranking of world universities of

Shanghai (Shanghai, 2015-e) is equal to 62 universities (English, French, German editions have overlaps of 65, 41, 35 universities respectively; the comparison is done for the year 2013).

The time evolution of the geographical distribution of leading world universities over 10 centuries is given in (TopWikiUniversities, 2015-e). Before the 19th century universities of Germany dominate this ranking (thus among the top universities of PageRank list with 139 universities, founded before year 1800, the main part of 25 universities is located in Germany, see Fig.10 in (Lages, 2016-a)). However, already for the universities founded before the 20th century (before year 1900) the lead is taken by the USA (see Fig.9 in (Lages, 2016-a)). The analysis of the university ranking evolution through ten centuries shows that Wikipedia highlights significantly stronger historically important universities whose role is reduced in the Shanghai ranking. Nowadays the PageRank algorithm gives the top 5 countries: USA, UK, Germany, Sweden, and France, while the Shanghai ranking gives USA, UK, Canada, Switzerland, and Japan.

The Wikipedia ranking provides a sound mathematical statistical evaluation of world universities which can be viewed as a new independent ranking being complementary to already existing approaches. A comparison of various web-based rankings of world universities is reported in (Pagell, 2016-a). In the view of the importance of university ranking for higher education (Hazelkorn, 2015-b) it is possible to expect that the Wikipedia ranking of world universities will also find a broad usage together with other rankings.

## Multiproduct world trade network

The Google matrix of the world trade network was constructed in (WTN, 2011-e) on the basis of the United Nations Commodity Trade Statistics Database (UNCOMTRADE, 2015-e) for all UN countries and various trade commodities for all available years from 1962 to 2009. The trade flows on this network are classified with the help of the PageRank and CheiRank algorithms and the distribution of countries on the PageRank-CheiRank plane is shown in for the trade in all commodities (or all products). This ranking treats all countries on equal democratic grounds independent of country richness but this method still puts at the top a group of industrially developed countries, reproducing about 75% of G20 members. The matrix  $S$  is obtained by column normalization of the monetary trade flow matrix  $M_{cc'p}$  available for each year at (UNCOMTRADE, 2015-e) for countries  $c, c'$  and product  $p$  ( $c \neq c'$ ). Then the matrix  $G$  is obtained by the general rule (1).

The case, when the trade is considered for all commodities, gives a typical distribution visible in with concentration of countries in a vicinity of the diagonal  $K = K^*$ . This is due to the economic trade balance which each country tries to equilibrate roughly. In a certain sense the PageRank corresponds to country import

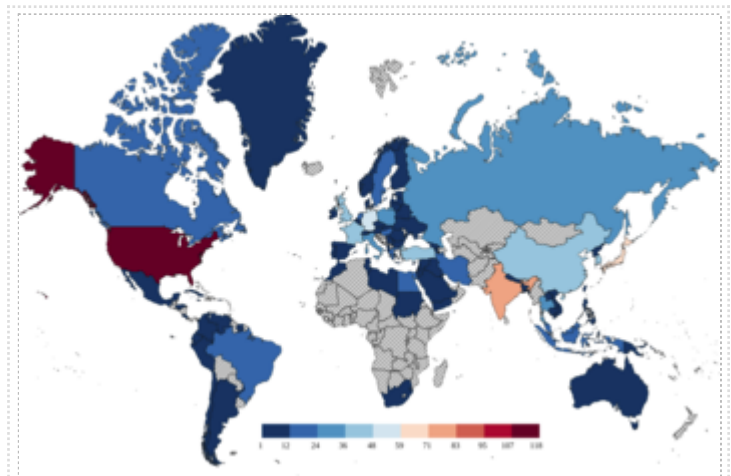


Figure 10: Geographical distribution of universities appearing in the top 100 universities of all 24 Wikipedia editions given by PageRank algorithm. The total number of universities is 1025. Colors range from dark blue (small number of universities) to dark red (maximum number of universities, here 118 for US). Countries filled by dashed lines pattern have no university in the top 100 lists of 24 editions. (from Lages, 2016-a)

(ingoing links) and CheiRank to export (outgoing links). However, the import and export take into account only one link trade between countries while the Google matrix analysis takes into account multiple links and significance of nodes. In general the country distribution on the PageRank-CheiRank plane is quite similar to the distribution on the Import-Export plane (see WTN, 2011-e). However, there are also some exceptions with noticeable differences such as Singapore (it improves its position from 15 in export rank to  $K^* = 11$  in CheiRank) showing the stability and broadness of its export trade in 2008. On the other hand Canada and Mexico have a lower ("better") position in export rank than in CheiRank due to a too strong orientation of their export to the USA.

The time evolution of PageRank and CheiRank indexes captures correctly known crises at certain years for certain countries (e.g Russia in 1998, Argentina in 2001) which typically lead to a strong increase of the country's PageRank index  $K$  related to the drop of its import during a crisis.

The approach developed in (WTN, 2011-e) allows to perform the Google matrix analysis for one specific product or for all commodities counted together. In this way the matrix size is always restricted to the number of countries  $N_c$  being significantly smaller than the total number of nodes  $N = N_c N_p$  for a trade with  $N_p$  products.

The Google matrix of the multiproduct world trade was constructed in (Ermann, 2015-a). This construction treats all countries on equal democratic grounds independently of their richness and at the same time it considers the contributions of trade products proportionally to their trade volume. This is achieved by the introduction of a personalized vector in the term of  $G$  with  $(1 - \alpha)$ , that makes the contribution of products being proportional to their trade volume, while all countries are treated on equal grounds. This analysis was done for  $M_p = 61$  products and up to  $N_c = 227$  countries. The obtained results show that the trade contribution of products is asymmetric: some of them are export oriented while others are import oriented even if the ranking by their trade volume is symmetric in respect to export and import after averaging over all world countries. The construction of the multiproduct Google

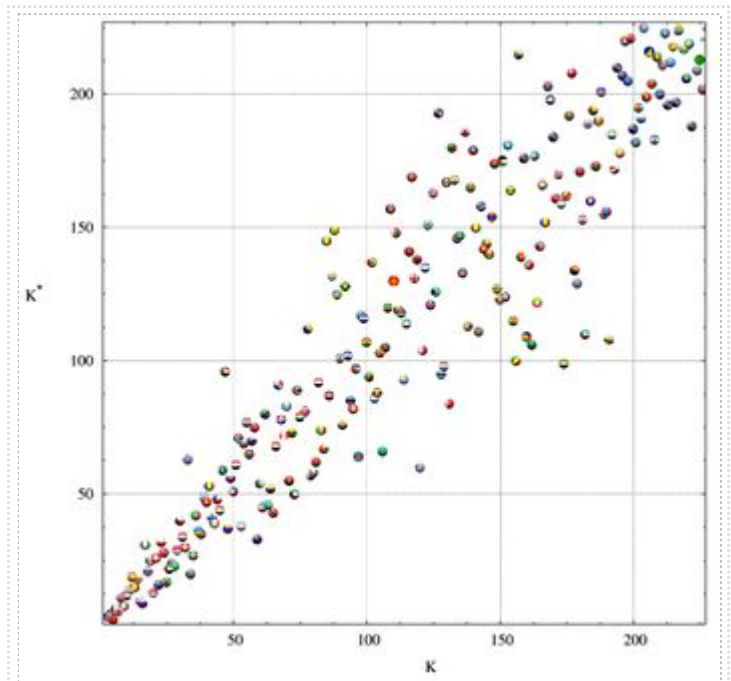


Figure 11: Country positions in PageRank-CheiRank plane ( $K, K^*$ ) for the world trade in all commodities in 2008. Each country is shown by circle with its own flag. (from Ermann, 2015-b)

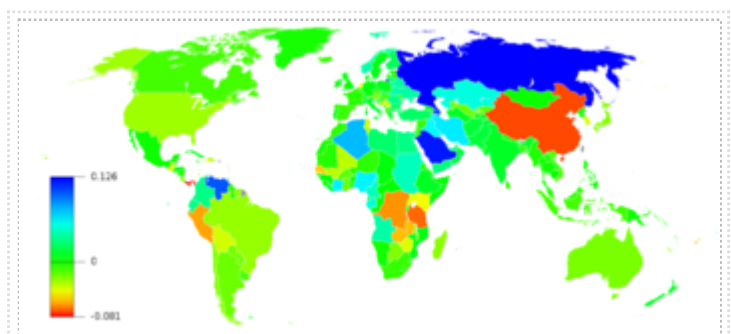


Figure 12: Geographical distribution of the derivative of probabilities balance  $dB_c/d\delta_{33}$  of world countries over petroleum price  $\delta_{33}$  for the year 2008. The country balance is determined from CheiRank and PageRank vectors as  $B_c = (P^*_c - P_c)/(P^*_c + P_c)$ . (from Ermann, 2015-a)

matrix allows to investigate the sensitivity of the trade balance with respect to price variations of products, e.g. petroleum and gas, taking into account the world connectivity of trade links. An example of the country sensitivity to the petroleum price increase  $\delta_{33}$  is shown in Figure 12. It shows that the dimensionless trade balance  $B_c = (P_c^* - P_c)/(P_c^* + P_c)$  is increased for petroleum producing countries like Russia and Saudi Arabia while the trade balance of China drops significantly ( $P_c, P_c^*$  are PageRank and CheiRank probabilities of a country  $c$  after summation over all products).

The Google matrix analysis of multiproduct world trade allows to establish hidden dependencies between various products and countries and opens new prospects for further studies of this interesting complex system of world importance.

This approach was successfully extended to the analysis of the world network of economic activities from the OECD-WTO TiVA database (Kandiah, 2015-a). This network describes the exchange of 37 activity sectors of 58 countries in years 1995 - 2008. In contrast to the UN COMTRADE these datasets contain also exchange between different sectors. The exchange balance  $B_c$  allows to determine economically rising countries with a robust network of economic relations. The sensitivity of  $B_c$  to price variations and labor cost in various countries determines the hidden relations between world economies being not visible for the usual export-import exchange analysis. The analysis of financial network transactions between various bank units can be also well suited for the Google matrix approach.

The Google matrix analysis can be considered as a further extension of the matrix analysis of Input-Output transactions broadly used in economy (Miller, 2009-b), starting from the fundamental works of Leontief (Leontief, 1953-a, Leontief, 1986-b).

## Other networks

The Google matrix approach allows to obtain interesting and useful results for a variety of directed networks: network of integers and citation network of Physical Review with nilpotent (triangular or nearly triangular) adjacency matrices, networks of game go (Kandiah, 2014-a), the entire Twitter network of 41 million size in 2009, network of business process management, neural network of a large-scale thalamocortical model (Izhikevich, 2008-a), neural network of *C.elegans*, networks of word transitions in DNA sequences, gene regulation networks (see Refs. in Ermann, 2015-b).

## Outlook

In physics, the Random matrix theory was introduced by Wigner (Wigner, 1967-b) to explain spectral properties of complex nuclei, atoms and molecules. This theory, developed for Hermitian and unitary matrices, captures universal spectral properties and find numerous applications in atomic, mesoscopic and nuclear systems (Guhr, 1998-b, Mehta, 2004-b, Fyodorov, 2011-b). This approach also describes the spectral properties of quantum chaotic systems which are characterized by matrices of a relatively simple structure (Haake, 2001-b). It is interesting to note that the quantum algorithm for computations with the Google matrix on a quantum computer has been also analyzed recently (Paparo, 2014-a). The development of a random matrix theory for Markov chains and Google matrix ensembles still remains a challenge. Some attempts in this direction are described below. It is

## Random matrix theory for $G$ ?

On a first glance there are various preferential attachment models generating complex scale-free networks (Dorogovtsev, 2010-b, Newman, 2010-b). A well known example is the Albert-Barabasi procedure (AB) which builds networks by an iterative process. Such a procedure has been generalized to generate directed networks with an expectation that such networks can generate spectra of Google matrices being close to real cases (see Refs. in Ermann, 2015-b). However, it has been found that the spectrum of  $G$  of the AB model has all  $|\lambda_i| < 0.4$  (except one unit eigenvalue). Thus, even if the decay of PageRank probability is well described by the relation  $P \sim 1/K$ , the spectrum of  $G$  for the AB model is drastically different from real cases of WWW and other networks described above.

A class of random matrix models of  $G$  has been analyzed in (Frahm, 2014-a). These models have  $Q$  positive random elements at random positions per column whose sum is normalized to unity. For this case it was shown that all eigenvalues (except the unit one) are concentrated inside a circle around zero with radius  $R \sim 1/\sqrt{Q}$ . Therefore these models are not suitable as well to reproduce spectral features of real networks.

The class of orthostochastic matrices of size  $N = 3; 4$  (Zyczkowski, 2003-a) approximately reproduces triplet and cross structures well visible for real networks (see Figs.6,7,8), but their size is too small to be used for real systems.

## Anderson localization for Google matrix eigenstates

The phenomenon of Anderson localization appears in a variety of quantum physical systems including electron transport in disordered solids and waves in random media (see Refs. in Guhr, 1998-b, Ermann, 2015-b, Zhirov, 2015-a). It is usually analyzed in the framework of Hermitian or unitary matrices. The possibilities of Anderson like localization and delocalization for matrices belonging to the class of Markov chains and Google matrices are considered in (Ermann, 2015-b, Zhirov, 2015-a). It was shown that certain matrix models, composed of blocks of orthostochastic matrices of size  $N = 3; 4$  (Zyczkowski, 2003-a), can have an algebraic decay of PageRank probability with the exponent  $\beta \sim 0.5$  (for the case  $\alpha = 1$ ) which is related to the existence of an Anderson transition of eigenstates and a mobility edge in the complex  $\lambda$  -plane. A further development of such models can allow to establish a closer link between the Anderson delocalization phenomenon in disordered solids and of delocalization of eigenstates for the Google matrix of directed networks.

## Reduced Google matrix

In many cases the real directed networks can be very large. However, in certain cases one may be interested in the particular interactions among a small reduced subset of  $N_r$  nodes with  $N_r \ll N$  instead of the interactions of the entire network. The interactions between these  $N_r$  nodes should be correctly determined taking into account that there are many indirect links between the  $N_r$  nodes via all other  $N_s = N - N_r$  nodes of the network. This leads to the problem of the reduced Google matrix  $G_R$  with  $N_r$  nodes which describes the interactions of a subset of  $N_r$  nodes. The matrix  $G_R$  has the form (Frahm, 2016-a):

$$G_R P_r = P_r \quad , \quad G_R = G_{rr} + G_{rs}(\mathbf{1} - G_{ss})^{-1} G_{sr} \quad ,$$

where  $G_{rr}$ ,  $G_{rs}$  and  $G_{sr}$ ,  $G_{ss}$  are sub blocks of the matrix  $G$  with respect to the decomposition of nodes in the reduced and the complementary subset of nodes:

$$G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix}.$$

The matrix  $G_R$  takes into account effective interactions between subset nodes by all their indirect links via the whole network. It belongs to the class of Google matrices and its PageRank vector has the same probabilities as the  $N_r$  nodes of matrix  $G$  (after rescaling due to normalization). The numerical methods of computation of  $G_R$  are described in (Frahm, 2016-a). This approach provides new possibilities to analyze effective interactions in a group of nodes embedded in large directed networks. An example of application of this approach to recovery of hidden links between political leaders is given in (Frahm, 2016b-a).

## Historical notes

Starting from the work of Markov (Markov, 1906-a) many scientists contributed to the development of spectral ranking of Markov chains. Research of Perron (1907) and Frobenius (1912) led to the Perron-Frobenius theorem for square matrices with positive entries (see e.g. Brin, 2002-b). A detailed historical description of spectral ranking research is reviewed by (Franceschet, 2011-a and Vigna, 2015-a). As described there, important steps have been done by researchers in psychology, sociology and mathematics including J.R. Seeley (1949), T.-H. Wei (1952), L. Katz (1953), C.H. Hubbell (1965). In the WWW context, the Google matrix in the form (1), with regularization of dangling nodes and damping factor  $\alpha$ , was introduced by (Brin, 1998-a).

The PageRank vector of a Google matrix  $G^*$  with inverted directions of links has been considered by (Fogaras, 2003-a, Hrisitidis, 2008-a), but no systematic statistical analysis of 2DRanking was presented there. An important step was done by (Chepelianskii, 2010-a) who analyzed  $\lambda = 1$  eigenvectors of  $G$  for directed network and of  $G^*$  for network with inverted links. The comparative analysis of the Linux Kernel network and WWW of the University of Cambridge demonstrated a significant difference in the correlator  $\kappa$  for these networks and different functions of top nodes in  $K$  and  $K^*$ . The term CheiRank was coined in (Zhirov, 2010-a) to have a clear distinction between eigenvectors of  $G$  and  $G^*$ . We note that top PageRank and CheiRank nodes have certain similarities with authorities and hubs appearing in the HITS algorithm (Kleinberg, 1999-a). However, the HITS is query dependent while the rank probabilities  $P(K(i))$  and  $P^*(K(i)^*)$  classify all nodes of the network.

## Lectures about Google matrix

Video lectures about Google matrix are available at (Frahm, 2014-v, Georgeot, 2014-v, Shepelyansky, 2014-v).

## Research articles

- (cited as "first\_author, year-a")
- Brin S., and Page L. (1998). *The anatomy of a large scale hypertextual web search engine*, *Computer Networks and ISDN Systems* **30**, 107 (<http://www.sciencedirect.com/science/article/pii/S016975529800110X>) .
- Chepelianskii A.D. (2010). *Towards physical laws for software architecture*, [arXiv:1003.5455](http://arxiv.org/abs/1003.5455) [cs.SE (<http://arxiv.org/abs/1003.5455>) ] .
- Donato D., Laura L., Leonardi S., and Milozzi S. (2004). *Large scale properties of the webgraph*, *Eur. Phys. J. B* **38**, 239 (<http://link.springer.com/article/10.1140%2Fepjb%2Fe2004-00056-6>) .



- Eom Y.-H., Aragon P., Laniano D., Kaltenbrunner A., Vigna S., and Shepelyansky D.L. (2015). *Interactions of cultures and top people of Wikipedia from ranking of 24 language editions*, *PLoS ONE* **10(3)**, e0114825. (<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0114825>)
- Ermann L., and Shepelyansky D.L. (2010). *Ulam method and fractal Weyl law for Perron-Frobenius operators*, *Eur. Phys. J. B* **75**, 299 (<http://link.springer.com/article/10.1140%2Fepjb%2Fe2010-00144-0>) .
- Ermann L., and Shepelyansky D.L. (2015). *Google matrix analysis of the mulitproduct world trade network*, *Eur. Phys. J. B* **88**, 84 (<http://epjb.epj.org/articles/epjb/abs/2015/04/b150047/b150047.html>) .
- Fogaras D. (2003). *Where to start browsing the Web?*, *Lect. Notes Comp. Sci.* (Springer) **2877**, 95.
- Frahm K.M., Eom Y.-H., and Shepelyansky D.L. (2014). *Google matrix of the citation network of Physical Review*, *Phys. Rev. E* **88**, 052814 (<http://journals.aps.org/pre/abstract/10.1103/PhysRevE.89.052814#fulltext#fulltext>) .
- Frahm K.M., Georgeot B., and Shepelyansky D.L. (2011). *Universal emergence of PageRank*, *J. Phys. A: Math. Theor.* **44**, 465101 (<http://iopscience.iop.org/article/10.1088/1751-8113/44/46/465101>) .
- Frahm K.M., and Shepelyansky D.L. (2016). *Reduced Google matrix*, *arXiv:1602.02394* [physics.soc-ph (<http://arxiv.org/abs/1602.02394>) ].
- Frahm K.M., Jaffres-Runser K., and Shepelyansky D.L. (2016)b. *Wikipedia mining of hidden links between political leaders*, *arXiv:1609.01948* [cs.SI (<http://arxiv.org/abs/1609.01948>) ].
- Franceschet M. (2011). *PageRank: standing on the shoulders of giants*, *Comm. ACM* **54(6)**, 92 (<http://dl.acm.org/citation.cfm?id=1953146&CFID=761690844&CFTOKEN=99512841>) .
- Hristidis V., Hwang H., and Papakonstantinou Y. (2008). *Authority-based keyword search in databases*, *ACM Trans. Database Systems* **33(1)**, 1 (<http://dl.acm.org/citation.cfm?id=1331905&CFID=761690844&CFTOKEN=99512841>) .
- Izhikevich E.M., and Edelman G.M. (2008). *Large-scale model of mammalian thalamocortical systems*, *Proc. Nat. Acad. Sci. USA* **105**, 3593 (<http://www.pnas.org/content/105/9/3593.full>) .
- Kandiah V., Georgeot B., and Giraud O. (2014). *Move ordering and communities in complex networks describing the game of go*, *Eur. Phys. J. B* **87**, 246 (<http://epjb.epj.org/articles/epjb/abs/2014/10/b140497/b140497.html>) .
- Kandiah V., Escaith H., and Shepelyansky D.L. (2015). *Google matrix of the world network of economic activities*, *Eur. Phys. J. B* **88**, 186 (<http://epjb.epj.org/articles/epjb/abs/2015/07/b150324/b150324.html>) .
- Kleinberg J.M. M. (1999). *Authoritative sources in a hyperlinked environment*, *Jour. ACM* **46(5)**, 604 (<http://dl.acm.org/citation.cfm?id=324140&CFID=761690844&CFTOKEN=99512841>) .
- Lages J., Patt A., and Shepelyansky D.L. (2016). *Wikipedia ranking of world universities*, *Eur. Phys. J. B* **89**, 69 (<http://epjb.epj.org/articles/epjb/abs/2016/03/b150922/b150922.html>) .
- Leontief W. (1953). *Domestic production and foreign trade; the American capital position re-examined*, *Proc. Am. Phios. Soc.* **97**, 332.
- Markov A.A. (1906). *Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga*, *Izvestiya Fiziko-matematicheskogo obschestva pri Kazanskom universitete*, 2-ya seriya, **15**, 135 (in Russia); English trans.: *Extension of the limit theorems of probability theory to a sum of variables connected in a chain*, reprinted in Appendix B of Howard B.A. *Dynamic Probabilistic Systems*, volume 1 *Markov models*, Dover Publ. (2007).
- Meusel R., Vigna S., Lehmborg O., and Bizer C. (2015). *The graph structure in the web – analyzed on different aggregation levels*, *J. Web Sci.* **1**, 33 ([http://www2014.kr/wp-content/uploads/2014/05/companion\\_p427.pdf](http://www2014.kr/wp-content/uploads/2014/05/companion_p427.pdf)) .
- Pagell R.A. (2016). *Wikipedia and Google Scholar as Sources for University Rankings – Influence and*

*popularity and open bibliometrics* (<http://librarylearningspace.com/ruths-rankings-17-wikipedia-google-scholar-sources-university-rankings-influence-popularity-open-bibliometrics/>) .

- Paparo G.D., Muller M., Comellas F., and Martin-Delgado M.A. (2014). *Quantum Google algorithm, construction and application to complex networks*, *Eur. Phys. J. Plus* **129**, 150 (<http://link.springer.com/article/10.1140/epjp/i2014-14150-y>) .
- Sjostrand J. (1990). *Geometry bounds on the density of resonances of quasi-bound states in open quantum systems*, *Duke Math. J.* **60**, 1 (<https://projecteuclid.org/euclid.dmj/1077297139>) .
- Vigna S. (2015). *Spectral ranking*, [arXiv:0912.0238v14](http://arxiv.org/pdf/0912.0238v14) (<http://arxiv.org/pdf/0912.0238v14.pdf>) .
- Zhironov A.O., Zhironov O.V., and Shepelyansky D.L. (2010). *Two-dimensional ranking of Wikipedia articles*, *Eur. Phys. J. B* **77**, 523 (<http://epjb.epj.org/articles/epjb/abs/2010/20/b100500/b100500.html>) .
- Zhironov O.V., and Shepelyansky D.L. (2015). *Anderson transition for Google matrix eigenstates*, *Annalen der Physik (Berlin)* **527**, 713.
- Zyczkowski K., Kus M., Slomczynski W., and Sommers H.-J. (2003). *Random unistochastic matrices*, *J. Phys. A: Math. Gen.* **36**, 3425 (<http://iopscience.iop.org/article/10.1088/0305-4470/36/12/333>) .

## Books and reviews

- (cited as "first\_author, year-b")
- Brin M., and Stuck G. (2002). *Introduction to dynamical systems*, Cambridge University Press, Cambridge UK.
- Chirikov B., and Shepelyansky D. (2008). *Chirikov standard map*, ([http://www.scholarpedia.org/article/Chirikov\\_standard\\_map](http://www.scholarpedia.org/article/Chirikov_standard_map)) Scholarpedia **3(3)**, 3550.
- Dorogovtsev S. (2010), "Lectures on complex networks", Oxford University Press, Oxford UK.
- Ermann L., Frahm K.M., and Shepelyansky D.L. (2015). *Google matrix analysis of directed networks*, *Rev. Mod. Phys.* **87**, 1261 (<http://journals.aps.org/rmp/abstract/10.1103/RevModPhys.87.1261>) .
- Fyodorov Y. (2011). *Random matrix theory*, *Scholarpedia* **6(3)**, 9886 ([http://www.scholarpedia.org/article/Random\\_matrix\\_theory](http://www.scholarpedia.org/article/Random_matrix_theory)) .
- Gantmacher F.R. (2000), "The theory of matrices, Vol.2", AMS Chelsea Publ., New York.
- Gaspard P. (2014). *Quantum chaotic scattering*, *Scholarpedia* **9(6)**, 9806 ([http://www.scholarpedia.org/article/Quantum\\_chaotic\\_scattering](http://www.scholarpedia.org/article/Quantum_chaotic_scattering)) .
- Guhr T., Muller-Groeling A., and Weidenmuller H.A. (1998). *Random matrix theories in quantum physics: common concepts*, *Phys. Reports* **299**, 189 (<http://www.sciencedirect.com/science/article/pii/S0370157397000884>) .
- Haake F. (2001), "Quantum signatures of chaos", Springer, Berlin.
- Hart M.H. (1992), "The 100: ranking the most influential persons in history", Citadel Press, New York.
- Hazelkorn E. (2015), "Rankings and the reshaping of higher education: the battle for world-class excellence", Palgrave Macmillan, New York.
- Langville A.M., and Meyer C.D. (2006), "Google's PageRank and beyond: the science of search engine rankings", Princeton University Press, Princeton NJ.
- Leontief W.W. (1986), "Input-Output economics", Oxford University Press, New York.
- Mehta M.L. (2004), "Random matrices (3rd edition)", Elsevier Academic Press, Amsterdam.
- Miller R.E., and Blair P.D. C.D. (2009), "Input-output analysis: foundations and extensions", Cambridge University Press, Cambridge UK.
- Newman M.E.J. (2010), "Networks: an introduction", Oxford University Press, Oxford UK.

- Nonnenmacher S., Sjostrand J., and Zworski M. (2014). *Fractal Weyl law for open chaotic maps*, *Ann. Mathematics* **179**, 179 (<http://annals.math.princeton.edu/2014/179-1/p03>) .
- Stewart G.W.. (2001). *Matrix algorithms v.II: Eigensystems*, SIAM, Philadelphia PA.
- Ulam S. (1960). *A collection of mathematical problems*, Interscience Tracts in Pure and Applied Mathematics, Interscience New York.
- Wigner E. (1967). *Random matrices in physics*, SIAM Reviews **9(1)**, 1.

## External links to data sets

- (cited as "name, year-e")
- Network data sets of EU FET NADINE project (2015) (<http://www.quantware.ups-tlse.fr/FETNADINE/datasets.htm>)
- Pantheon MIT Wikipedia project (2015) (<http://pantheon.media.mit.edu>)
- Shanghai Academic ranking of world universities (2015) (<http://www.shanghairanking.com/>)
- Stony-Brook Wikipedia ranking (2015) (<http://www.whoisbigger.com>)
- Top 100 historical figures of Wikipedia (2014) (<http://www.quantware.ups-tlse.fr/QWLIB/topwikipeople/>)
- Top 100 world universities of Wikipedia (2015) (<http://perso.utinam.cnrs.fr/~lages/datasets/WRWU/>)
- WWW of UK universities (2002-2006) (<http://cybermetrics.wlv.ac.uk/database/>)
- UN Comtrade Database (2015) (<http://comtrade.un.org/>)
- Wikipedia Aug 2009 ranking results (2010) (<http://www.quantware.ups-tlse.fr/QWLIB/2drankwikipedia/>)
- World trade network data (2011) (<http://www.quantware.ups-tlse.fr/QWLIB/tradecheirank>)

## Video lectures

- (cited as "name, year-v")
- Frahm K.M. (2014) *Spectral properties of Google matrix* Lectures at ecole de Luchon (slides+videos, 1+1+1 hours at 23,24,25 June) (<http://www.quantware.ups-tlse.fr/ecoledeluchon2014/program.html>)
- Georgeot B. (2014) *Game of GO as a complex network* Lectures at ecole de Luchon (slides+video, 1 hour at 3 July) (<http://www.quantware.ups-tlse.fr/ecoledeluchon2014/program.html>)
- Shepelyansky D.L. (2014) *Google matrix analysis of world trade network* Lectures at ecole de Luchon (slides+video, 1 hour at 4 July) (<http://www.quantware.ups-tlse.fr/ecoledeluchon2014/program.html>)

Sponsored by: Eugene M. Izhikevich, Editor-in-Chief of Scholarpedia, the peer-reviewed open-access encyclopedia

Reviewed by ([http://www.scholarpedia.org/w/index.php?title=Google\\_matrix&oldid=169983](http://www.scholarpedia.org/w/index.php?title=Google_matrix&oldid=169983)) : Prof. Thomas Guhr, Faculty of Physics, University of Duisburg-Essen, Germany

Reviewed by ([http://www.scholarpedia.org/w/index.php?title=Google\\_matrix&oldid=169983](http://www.scholarpedia.org/w/index.php?title=Google_matrix&oldid=169983)) : Anonymous (via Dr. Michelle L. Jones, Jones Biomediting, Gouda, Netherlands)

Accepted on: 2016-11-01 15:12:37 GMT ([http://www.scholarpedia.org/w/index.php?title=Google\\_matrix&oldid=169983](http://www.scholarpedia.org/w/index.php?title=Google_matrix&oldid=169983))

Category: Quantum Chaos

*This page was last modified on 1 November 2016, at 15:12.*

*This page has been accessed 11,857 times*



*"Google matrix" by Leonardo Ermann, Klaus Frahm and Dima Shepelyansky is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. Permissions beyond the scope of this license are described in the Terms of Use*