RayE-Sub: Countering Subgraph Degradation via Perfect Reconstruction

Kuo Yang, Zhengyang Zhou*, *Member, IEEE,* Xu Wang, Pengkun Wang, *Member, IEEE,* Limin Li, Yang Wang*, *Senior Member, IEEE.*

***Abstract*—Subgraph learning has dominated most practices of improving the expressive power of Message Passing Neural Networks (MPNNs). Existing subgraph discovery policies can be classified into node-based and partition-based, which both achieve impressive performance in most scenarios. However, both mainstream solutions still face a subgraph degradation trap. Subgraph degradation is reflected in the phenomenon that the subgraph-level methods fails to offer any benefits over node-level MPNNs. In this work, we empirically investigate the existence of the subgraph degradation issue and introduce a unified perspective, perfect reconstruction, to provide insights for improving two lines of methods. We further propose a subgraph learning strategy guided by the principle of perfect reconstruction. To achieve this, two major issues should be well-addressed, i.e., *(i) how to ensure the subgraphs to possess with 'perfect' information? (ii) how to guarantee the 'reconstruction' power of obtained subgraphs?* Firstly, we propose a subgraph partition strategy *Rayleigh-resistance* to extract non-overlap subgraphs by leveraging the graph spectral theory. Secondly, we put forward a *Query* mechanism to achieve subgraph-level equivariant learning, which guarantees subgraph reconstruction ability. These two parts, *perfect subgraph partition* and *equivariant subgraph learning* are seamlessly unified as a novel *Rayleigh-resistance Equivariant Subgraph learning* architecture (*RayE-Sub*). Comprehensive experiments on both synthetic and real datasets demonstrate that our approach can consistently outperform previous subgraph learning architectures. Code is available at https://anonymous.4open.science/r/RayE-63C5.**

***Index Terms*—Subgraph learning, subgraph degradation, Message Passing Neural Networks, expressive ability, spectral theory.**

I. Introduction

GRAPH, a widely prevalent non-Euclidean data structure in the real world, play a crucial role in areas such as social networks, transportation networks, and biological networks [1]–[3]. A graph consists of nodes and edges, which represent the complex relationships between entities. In recent years, the rise and development of Graph Neural Networks (GNNs) have dominated the graph learning research, making a profound impact in areas such as recommendation systems [4], [5], smart cities [6], [7], and AI for science [8], [9]. GNNs are implemented through Message Passing Neural Networks (MPNNs), which embed nodes, edges, or subgraphs into a low-dimensional vector space while retaining the topological structure and attribute information.

However, the inherent limitations of the message-passing mechanism restrict the expressive power of GNNs on complex graph structures. It has been proved that MPNNs are with limited expressiveness, which are at most as expressive as 1-dimensional Weisfeiler-Lehman (1-WL) test [10], [11]. As a result, many studies have shifted their focus toward designing graph learning methods with more powerful expressive capabilities [12]–[18]. Among these, subgraph learning is one of the most promising solutions. Subgraph learning focuses on modeling the relationship between local structures and global features, thereby extracting the most critical topological information and the most label-relevant feature details. This graph learning paradigm effectively improves the expressive ability of network to learn complex structures while also enhancing model interpretability.

Subgraph learning aims to extract a bag of subgraphs from an original graph, and explore more powerful expressive frameworks based on subgraph-level encoding [15]. The practices of subgraph learning can be classified into two main research lines. (i) Node-based subgraph discovery employs predefined structure to extract subgraphs, wherein each subgraph is centered by a unique node in the graph [14], [15], [19]. The implementations of this category include node-deletion, node-marking, and ego-network subgraph extraction [16], [17], [20]. (ii) Partition-based subgraph discovery extracts a bag of non-overlapping subgraphs from original graphs. This category includes high-frequency substructures extraction, node clustering and edge dropping [1], [13], [21], [22]. This approach offers significant interpretability in addressing real-world tasks, such as finding functional groups in molecules, decoupling subnetworks in social networks, and discovering urban functional patterns in cities [2], [3], [23]. Both of them have revealed effectiveness in practices, where the former focuses more on studying the expressive capability based on WL-test, while the latter emphasizes providing better interpretability from a causal perspective. Although they all have achieved impressive performance in most practical applications, they fail to outperform traditional MPNNs in all scenarios.

We reflect on these subgraph learning strategies and observe they both suffer from the limitation that could potentially lead to the degradation of distinguishability. (i) As shown in the top panel of Figure 1, G_1 and G_2 are a pair of graphs with different structures, which cannot be distinguished

K. Yang, Z. Zhou, X. Wang, P. Wang, L. Li, Y. Wang are with University of Science and Technology of China, China. Z. Zhou, X. Wang, P. Wang, and Y. Wang are also with Suzhou Institute for Advanced Research, USTC, Suzhou, China.
E-mail: {yangkuo, lilimin}@mail.ustc.edu.cn, {zzy0929*, wx309, pengkun, angyan*}@ustc.edu.cn

*Prof. Yang Wang and Zhengyang Zhou are corresponding authors.

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 30, 2024; revised August 30, 2024.

by 1-WL algorithm. However, the subgraph sets obtained through the node-based subgraph extraction method remain indistinguishable, as the subgraph sets G_S are identical. This means that the subgraph learning method does not enhance the model's distinguishability. (ii) For partition-based subgraph discovery policy, researchers typically concentrate on implementing effective and interpretable subgraph partition strategies but often overlook subgraph-level permutation equivariance. As shown in the bottom panel of Figure 1, 2-Butanol (G_1) and 2-Methyl-1-Propanol (G_2), which share similar features and structures, cannot be distinguished by the 1-WL algorithm. However, we observe that the two molecules still share a same set of subgraphs (functional groups) through using a partition-based subgraph extraction approach, i.e., the subgraph sets G_S of the two graphs remain indistinguishable. This means that the subgraph learning method still fails to enable the network to distinguish between them.

The causes of the subgraph degradation phenomenon in the two lines of graph learning methods differ significantly. Node-based methods often focus excessively on enhancing the effective receptive field of the message passing mechanism, while subgraph-based methods lack proper modeling of interactions between subgraphs. This discrepancy limits existing methods from studying the two lines of subgraph learning approaches from a unified perspective. Actually, this is a primary challenge we encounter in this work.

As a result, we establish a unified perspective to understand the two lines of graph learning methods through a profound analysis. We argue that both subgraph learning methods share a common issue, i.e., the obtained subgraphs fail to *reconstruct* the original graph *perfectly*. Specifically, (i) the subgraphs generated from node-based strategies tend to contain redundant information due to the overlap among subgraphs, making them imperfect for reconstruction; (ii) the subgraphs extracted from partition-based methods often lack the ability to be inversely reconstructed. To verify this insight, we construct a theoretical bridge investigating the expressive power of subgraph learning models through the lens of reconstruction ability. We attribute the limitations of existing efforts to the limited capacity for reconstruction. We theoretically demonstrate that when the extracted subgraph has *perfect reconstruction* ability, the existing limitations of subgraph learning will be greatly addressed. Therefore, this naturally raises the specific challenge that *how to obtain subgraphs with perfect reconstruction ability?*

Present work. We address the aforementioned issue by two sub-solutions: (i) designing non-overlapped subgraph learning approaches to ensure the *perfect* property, (ii) devising subgraph-level equivariance learning architecture to guarantee *reconstruction* property. To achieve it, we propose a novel subgraphs partition strategy guided by spectral graph theory, and design a *Query* mechanism to achieve subgraph-level equivariant learning.

Firstly, we prioritize ensuring that the extracted subgraphs contain perfect and non-redundant information. Partition-based subgraph learning, which enjoys both non-overlapped and interpretable properties, is considered to be closer to guarantee the perfectness. Subgraph partition aims to identify optimally meaningful boundaries among complex graph connections. Fortunately, spectral theory possesses the powerful capability to draw graphs and identify potential boundaries from the spectral domain [18], [24]. Given its potent capabilities and numerous successful applications, we propose a novel spectrum-based subgraph partition strategy *Rayleigh-resistance Extractor*. Specifically, we quantify the structural stability using *Rayleigh entropy* and achieve subgraph partition by the edge-wise *Resistance distance*. The validity and feasibility of this design are verified through theoretical analysis.

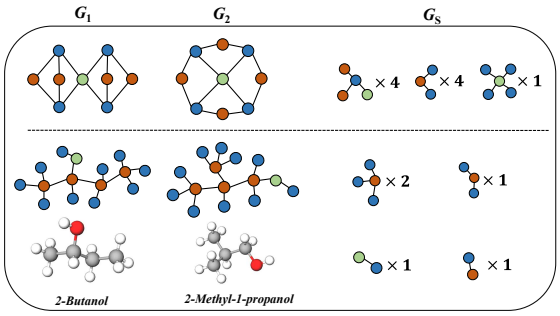


Fig. 1. Illustration of subgraph failures. **Top panel:** $m = 2, k = 2$ in Example 1, they will generate the same EGO-based subgraphs. **Bottom panel:** two isomers cannot be distinguished by partition-based methods, 2-Butanol and 2-Methyl-1-Propanol.

Secondly, although a collection of subgraphs obtained through partitioning contain suitable and non-redundant information, they still lack reconstruction ability. To significantly enhance the reconstructability of discovered subgraphs, we employ the equivariance principle to examine the equivalent relationships among subgraphs, thereby ensuring robust reconstruction power [16]. Specifically, we propose a *Siamese-Query* scheme to implement our equivariant architecture, where a Siamese network processes each subgraph independently with same parameters, and the *Query* mechanism aggregates all subgraphs with their equivariance information. Altogether, above two parts are composed of our ***Ray****leigh-resistance* ***E****quivariant* ***Sub****graph learning architecture* (***RayE-Sub***), which achieves the perfect reconstruction of extracted subgraphs. Contributions of this paper are summarized as follows:

- Through a theoretical and intuitive analysis of existing subgraph learning models, we observe that they exhibit a significant limitation known as subgraph degradation and attribute such limitation to the failure of perfect reconstruction.
- We propose a subgraph learning architecture RayE-Sub, including a novel subgraph partition strategy Rayleigh-resistance Extractor and a subgraph-level equivariant encoding framework Siamese-Query Network to achieve perfect reconstruction ability.
- Comprehensive empirical results validate the competitive performance of our approach on both synthetic and realworld datasets, underscoring its effectiveness in diverse scenarios.

II. Related Work

a) Subgraph Learning: According to different subgraph discovery strategies, subgraph learning can be classified into node-based [14]–[17] and partition-based [1], [13], [22], [23] approaches.

Node-based subgraph learning primarily concentrates on utilizing the Weisfeiler-Lehman (WL) test as a standard to enhance the expressivity of graph learning architectures. GNN-AK [17] observes that MPNNs' local neighbor aggregation follows a star pattern, and propose a design which naturally generalizes from encoding the star to encoding a more flexibly defined subgraph. ESAN [16] implements an subgraph equivariant learning architecture and achieves better expressiveness by per-layer aggregation across subgraphs. SUN [15] profoundly studies the characteristics of node-based subgraph learning. Further, SUN aligns the permutation group of nodes and subgraphs, and models the symmetry with a smaller single permutation group. These efforts are aimed at enhancing the efficiency of messaging within the models but often do not prioritize interpretability. In addition, the ineffectiveness of node-based subgraph learning in all scenarios has been reflected in some studies. For example, ESAN [16] involves performing stochastic sampling of extracted subgraphs and subsequently feeding the sampled subgraphs into networks. This design can be seen as a way to prevent performance degradation by minimizing overlap between subgraphs, which has been widely followed in subsequent works [12]. Their practices are uniformly analyzed and theoretically proved in our work.

Different from node-based subgraph learning methods, partition-based approaches place a greater emphasis on interpretability. Leveraging information theory to achieve subgraph partition is a notable research line [13], [25], where GSAT [13] is a representative method. Guided by the information bottleneck theory, GSAT designs a subgraph extraction strategy with edge deletions based on stochastic attention mechanism. Numerous partition based subgraph learning practices aim to methods seek invariant subgraphs from a causal perspective [26]–[29]. DIR [26] extracts causal invariant subgraph by conducting interventions on graphs to create interventional distributions. To tackle the problem of learning invariant graph representations under distribution shifts, GIL [27] designs three tailored modules to encourage the graph representations to capture the invariant relationships between predictive graph structural information and labels. Such methods prioritize achieving interpretable learning and often overlook the theoretical investigation of expressive capacity, where the neglect of subgraph-level permutation equivariance presents a significant challenge.

Given the various limitations of existing works, there is currently no universally accepted solution to achieve the unified advancement of partition-based and node-based subgraph discovery. In this work, we propose a novel perspective to unify the underlying reasons for the limitation observed in existing subgraph learning practices, i.e., *perfect reconstruction*.

b) Expressive power of MPNNs: Exploring more expressive learning architectures is the primary goal in graph representation learning. Current researches tend to be divided into three lines, i.e., the MPNN-based methods aligned with WL-Test, transformer-based methods and the solutions derived from novel representation power measures. First, MPNN-based methods improves expressiveness on WL-Test by devising higher-order message-passing [11], [30], [31], position and structure encoding [32]–[34]. However, the computational cost for $k > 3$ expressive power in WL-Test should be unacceptable. Second, instead of conventional message passing, transformer like Graphormer computes the soft attention scores for aggregation [35]. Third, Zhang et al. [12] take a novel perspective, the graph bi-connectivity, as measure of expressiveness and promotes the representation on bi-connectivity aspect. In this work, we inherit the third line and further exploits elegant theoretical paradigms, to construct a novel *perfect reconstruction* subgraph learning scheme towards more interpretable and powerful graph representations.

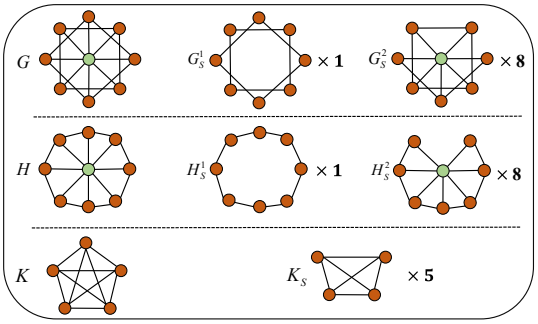


Fig. 2. The support graphs ($\boldsymbol{G}$ and $\boldsymbol{H}$) of ND-based subgraph learning failure by setting $m = 1, k = 4$ in Example 1. And a complete graph $\boldsymbol{K}$ with 5 nodes to support Theorem 1.

III. Subgraph Learning from Reconstruction Perspective

In this section, we first highlight the degradation of extracted subgraphs in existing methods and then attribute it to limited reconstruction ability rationally.

Notation: Let $\boldsymbol{G} = (\boldsymbol{A}, \boldsymbol{X})$ be an undirected graph with n nodes. $\boldsymbol{V}$ and $\boldsymbol{E}$ represent the node set and edge set, respectively. The adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ denotes the connectivity of $\boldsymbol{G}$. The feature matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ represents the features of all nodes, where $x(u) \in \mathbb{R}^{1\times d}$ is the feature of u. Let $[n] = 1, ..., n$. $\boldsymbol{G}_S = \{\boldsymbol{G}_S^1, \cdots, \boldsymbol{G}_S^k\}$ represents the subgraph set generated by subgraph discovery policy $\pi(G)$. Each subgraph is $\boldsymbol{G}_S^i = (\boldsymbol{A}_S^i, \boldsymbol{X}_S^i)$ with $\boldsymbol{V}_S^i \subseteq \boldsymbol{V}$, $\boldsymbol{E}_S^i \subseteq \boldsymbol{E}$, where $1 \le i \le k$. We denote $x_i \in \mathbb{R}^{|\boldsymbol{V}_i|\times d}$ as the feature of all nodes in subgraph $\boldsymbol{G}_S^i$, which is different from $x(u)$.

A. Node-based subgraph learning

The node-based subgraph discovery approach has emerged as the most popular strategy in subgraph learning, owing to its simplicity and effectiveness [15]–[17]. The specific implementations consist of node-deletion (ND), node-marking (NM), and ego-networks (EGO) policies. While such methods have achieved remarkable success, there is still a significant degradation of learning ability. We can easily observe two family graphs can be indistinguishable for existing node-based

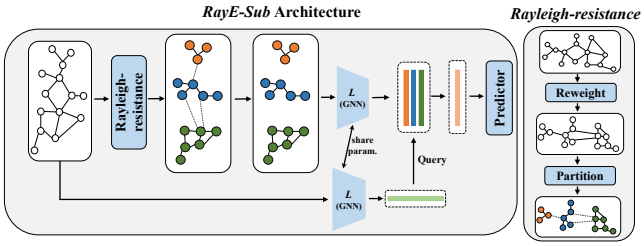


Fig. 3. The architecture of ***RayE-Sub***. **Left panel:** our ***RayE-Sub*** is composed of two stages: a partition block *Rayleigh-resistance Extractor*, a subgraph-level equivariant module *Siamese-Query Network*. **Right panel:** the detailed process of *Rayleigh-resistance*.

subgraph learning practices, as illustrated in Example 1. We sample some pairs of graphs to illustrate this phenomenon as shown in Figure 1.

Example 1. ***(The degeneration of Node-based subgraphs.)*** *Let $\boldsymbol{G}_1 = \{\boldsymbol{V}_1, \boldsymbol{E}_1\}$ and $\boldsymbol{G}_2 = \{\boldsymbol{V}_2, \boldsymbol{E}_2\}$ be a pair of graphs with $n = 2kl + 1$, where k, l are two positive integers satisfying $kl > 3$. Note that $\boldsymbol{V}_1 = \boldsymbol{V}_2 = [n]$, $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ satisfy the following conditions,*

$$\begin{aligned}\boldsymbol{E}_1 =& \{\{i, (i \bmod kl) + 1\} : i \in [kl]\} \\ &\cup \{\{i + kl, (i \bmod kl) + kl + 1\} : i \in [kl]\} \\ &\cup \{\{n, i\} : i \in [2kl], i \bmod l = 0\}, \\ \boldsymbol{E}_2 =& \{\{i, (i \bmod 2kl) + 1\} : i \in [2kl]\} \\ &\cup \{\{n, i\} : i \in [2kl], i \bmod l = 0\}.\end{aligned} \tag{1}$$

Given that ego-network is the simplest implementation of a node-based approach, we visualize a pair of graphs that cannot be distinguished by the EGO-based methods in the top panel of Figure 1. By setting $m = 2$ and $k = 2$, we can obtain a pair of graphs with identical EGO subgraphs set. Therefore, EGO-based methods are powerless to distinguish this pair of graphs. Moreover, we also provide more observations of subgraphs failure on ND-based and NM-based strategies as shown in Figure 2. We can obtain a pair of graphs $\boldsymbol{G}$ and $\boldsymbol{H}$, by setting $m = 1, k = 4$. For original $\boldsymbol{G}$ and $\boldsymbol{H}$, they cannot be distinguished by 1-WL [12]. Thus, we tend to explore whether we can distinguish them by ND-based and NM-based strategies. We first extract their subgraph set from $\boldsymbol{G}_S$ and $\boldsymbol{H}_S$ based on ND-based policy, respectively. Unfortunately, they also can't be distinguished by ND-based learning strategy. We provide detailed explanations for these reasons in the subsequent analysis.

The $\boldsymbol{G}_S^1$ extracted from $\boldsymbol{G}$ is two 4-cycle regular graphs, while the $\boldsymbol{H}_S^1$ extracted from $\boldsymbol{H}$ is a 8-cycle regular graph. Almost all node-based subgraph learning methods are upper bounded by 3-WL. Thus, they are unable to distinguish $\boldsymbol{G}_S^1$ and $\boldsymbol{H}_S^1$. Similar to $\boldsymbol{G}_S^1$ and $\boldsymbol{H}_S^1$, existing approaches with less than 3-WL expressive ability are unable to distinguish between $\boldsymbol{G}_S^2$ and $\boldsymbol{H}_S^2$. Intuitively, the subgraph generated from each node in $\boldsymbol{G}_S^2$ and $\boldsymbol{H}_S^2$ shares the same information. This homogenization will result in their indistinguishability. Theoretically, we can verify this fact with the simple aggregation practice of Algorithm 1. The failure of NM-based method is similar to ND-based strategy, so we don't repeat it.

Building upon the observed phenomenon of subgraph degradation, we aim to delve into the underlying theory of this phenomenon. As a result, we theoretically prove the existence of this degeneration from the perspective of message passing. Based on [36], we first derive the Proposition 1 regarding function composition.

Proposition 1. *MPNNs can repeatedly update each node's embedding by aggregating information from their neighbors. The graph-level embedding $\mathbf{h}_G$ can be obtained by,*

$$\mathbf{h}_G = \mathbf{MPNN}(\boldsymbol{A}, \boldsymbol{X}). \tag{2}$$

There exists a global correlation matrix $\mathbf{T}$ to indicate node-wise relevance learned by $\mathbf{MPNN}$, where each element $\mathbf{T}(i, j)$ is the aggregation coefficient from j to i.

This proposition stems from a profound understanding that the feature space tends to be linearly correlated due to the repeated aggregations of MPNNs [36]. We evaluate the expressive ability of different subgraph models by considering their power as learning over ground-truth correlation matrix $\boldsymbol{T}$. If extracted subgraphs fail to learning the correlation matrix closer to $\boldsymbol{T}$, these subgraph solutions will not contribute to better distingushment ability [37], [38].

Theorem 1. ***(The existence of subgraph degradation.)*** *Given two graphs G and H, the correlation matrices obtained by node-level MPNNs be $\mathbf{T}_G$ and $\mathbf{T}_H$ respectively. The correlation matrices implemented by node-based subgraph learning be $\mathbf{T}_G^S$ and $\mathbf{T}_H^S$. The subgraph degradation phenomenon occurs when one of the following two cases is existing:*

$$(i)\ \mathbf{T}_G^S = \mathbf{T}_G \ or\ \mathbf{T}_H^S = \mathbf{T}_H, \tag{3}$$

$$(ii)\ \mathbf{T}_G = \mathbf{T}_H \ and\ \mathbf{T}_G^S = \mathbf{T}_H^S. \tag{4}$$

Proof. Node-based subgraph learning first extract a subgraph set $\boldsymbol{G}_S = \{\boldsymbol{G}_S^1, \boldsymbol{G}_S^2, \cdots, \boldsymbol{G}_S^n\}$, wherein each subgraph is associated with a unique node. The aggregation process of each subgraph $\boldsymbol{G}_S^i$ generates a corresponding correlation matrix $\mathbf{T}_S^i$. For the whole graph, the correlation matrix of information passing among all nodes is $\mathbf{T}_S = \sum_{i \in [n]} \mathbf{T}_S^i$. Therefore, we turn the question into proving the existence of $\sum_{i \in [n]} \mathbf{T}_S^i = \mathbf{T}_G$. Since [15] has shown that the ND-based methods are the most expressive node-based subgraph learning strategies, we can only investigate ND-based subgraph discovery for demonstrating the existence of above issue. We then demonstrate that both cases leading to subgraph degradation are present.

Equation 3 indicates that the introduction of subgraphs does not enhance the 1-WL message passing mechanism. Consider a complete graph $\boldsymbol{K}$ with 5 nodes and its subgraph set $\boldsymbol{K}_S$ as shown in Figure 2. We observe that each subgraph is a complete graph with 4 nodes, there is no distinction among them. And we can easily deduce that $\mathbf{T_G}$ is a constant multiple of $\sum_{i \in [n]} \mathbf{T}_S^i$, which doesn't change the final result. Therefore, there exists graph $\boldsymbol{G}$ satisfying Equation 3, i.e., $\mathbf{T}_G^S = \mathbf{T}_G$.

Equation 4 indicates that although the introduction of subgraphs results in different correlation matrices with 1-WL,

it still cannot distinguish isomorphic graphs. Consider two graphs G and H as shown in Figure 2, they are isomorphic graphs and are indistinguishable under the 1-WL-based message passing mechanism [12]. Therefore, there exists $\mathbf{T}_G = \mathbf{T}_H$. The subgraphs extracted from G exhibit two distinct forms, namely G_S^1 and G_S^2. Similarly, the subgraphs extracted from H also exhibit two distinct forms, H_S^1 and H_S^2. Next, we will examine whether the introduction of subgraphs can alter the indistinguishability of G and H. G_S^1 consists of two 4-cycles, while H_S^1 forms a single 8-cycle. Actually, any node in these subgraphs shares the same local structure, which cannot be distinguished by the 1-WL-based message passing mechanism [39]. Therefore, there exists $\mathbf{T}^i_{G_S^1} = \mathbf{T}^j_{H_S^1}$ for $i \in G_S^1, j \in H_S^1$. Every node in G_S^2 has a corresponding node in H_S^2 with the same local structure. In other words, G_S^2 and H_S^2 are identical in shape and connectivity, i.e, they are isomorphic graphs. Thus, there exists $\mathbf{T}^i_{G_S^2} = \mathbf{T}^j_{H_S^2}$ for $i \in G_S^2, j \in H_S^2$. In conclusion, subgraph-level learning does not alter original distinguishability, meaning the following equation still holds on:

$$\sum_{i \in G_S^1} \mathbf{T}^i_{G_S^1} + 8 \sum_{j \in G_S^2} \mathbf{T}^j_{G_S^2} = \sum_{i \in H_S^1} \mathbf{T}^i_{H_S^1} + 8 \sum_{j \in H_S^2} \mathbf{T}^j_{H_S^2}. \quad (5)$$

Theoretical Insight. Our proof aims to confirm the existence of the samples leading to subgraph degeneration. Specifically, the process of the proof is divided into two steps. (i) We first demonstrate that the utilization of subgraph learning does not alter the original message passing correlation matrix by introducing a complete graph $\boldsymbol{K}$ with 5 nodes. (ii) Using two families of samples from Example 1, we then show that the use of subgraphs does not enhance the raw distinguishability of networks. In other words, a pair of graphs that are initially indistinguishable remain so even after applying subgraph learning.

B. Partition-based Subgraph Learning

Partition-based subgraph learning aims to extract a subgraph set $\boldsymbol{G}_S = \{\boldsymbol{G}_S^1, \boldsymbol{G}_S^2, \cdots, \boldsymbol{G}_S^k\}$ which are with no overlaps among any pair of elements, as described in following,

$$\boldsymbol{V} = \boldsymbol{V}_S^1 \cup \boldsymbol{V}_S^2 \cup ... \cup \boldsymbol{V}_S^k \ , \quad (6)$$

$$\boldsymbol{V}_S^i \cap \boldsymbol{V}_S^j = \emptyset \ (\forall i, j \in [k], i \neq j). \quad (7)$$

Given that existing practices aim to refine a minimal label-relevant subgraph set [13], [26], Equation 6 is thus not always a compulsory condition. Compared with node-based strategies, partition-based methods are more interpretable for real-world tasks, such as finding the functional groups in molecules [1].

However, most previous works [13], [21], [38], [40] pay more attention to explore effective and interpretable partition principle, but ignore subgraph-level permutation equivariance analysis. For example, this deficiency makes it challenging for the model to distinguish isomers and chiral molecules, leading to invalidated predictions. It is shown in the bottom panel of Figure 1. Therefore, this limitation results in failing to achieve the desired powerful representation ability, namely subgraph degradation.

We can denote these practices, which ignore subgraph-level permutation invariance, as $f : f(\tau \cdot \boldsymbol{G}_S) = f(\boldsymbol{G}_S)$, where τ denotes the permutation operation of subgraph. Obviously, this design cannot be suitable for real-world tasks of subgraph learning, where one of the most notable examples is the challenge of distinguishing isomers. In contrast, equivariant learning $g : g(\tau \cdot \boldsymbol{G}_S) = \tau \cdot g(\boldsymbol{G}_S)$ at the subgraph-level can achieve better distinguishability and interpretability [16].

C. Unifying Subgraph Learning via Reconstruction Ability

Based on this fresh review of previous practices, we can summarize the root view for subgraph degradation to two key limitations as follows.

a) Node-based subgraphs with redundant information: The most notable characteristic of the subgraphs $\boldsymbol{G}_S$ extracted by node-based methods is the presence of numerous pairwise subgraphs with overlapping information, namely $\boldsymbol{G}_S^i \cap \boldsymbol{G}_S^j \neq \emptyset$ for $i \neq j$. Moreover, we have demonstrated that the degradation of message passing space (neighbor information) is the main reason for node-based subgraphs failure. Therefore, overlapping subgraphs precisely give each node an independent and 1-WL-similar messaging passing space.

Actually, some empirical results from prior work have potentially supported this view. For example, ESAN [16] conducts a stochastic sampling of the extracted subgraphs $\boldsymbol{G}_S^m \subseteq \boldsymbol{G}_S^n$ and subsequently feeding the sampled subgraphs into networks. This design can be seen as a way to prevent the degradation by minimizing overlap between subgraphs, which has been widely followed in subsequent works [12]. Therefore, such redundancy in $\boldsymbol{G}_S$ poses a significant challenge for node-based strategies.

b) Partition-based subgraphs with inadequate information: From the perspective of information overlap, we can also summarize the characteristics of $\boldsymbol{G}_S$ obtained by partition-based subgraph selection. Specifically, (i) there is no information overlap between any pair of subgraphs, denoted as $\boldsymbol{G}_S^i \cap \boldsymbol{G}_S^j = \emptyset$ for $i \neq j$, (ii) $\boldsymbol{G}_S$ is disordered with no positional associations among subgraphs. The above observations indicate that the elements in $\boldsymbol{G}_S$ are independent of each other. Therefore, it is evident that the lack of subgraph-level interaction inevitably leads to the failure of distinguishment.

In summary, we attribute the limitations of the above two strategies to the redundancy and inadequacy of extracted subgraphs, respectively. The redundant and insufficient subgraphs result in failing to perfectly reconstruct the whole graph. Inspired by this observation, we take the perspective of perfect reconstruction to remedy above two limitations.

Definition 1. *Let subgraph set $\boldsymbol{G}_S$ be extracted by $\boldsymbol{G}$. The subgraph set $\boldsymbol{G}_S$ will be equipped with **reconstruction ability** if there exist a reconstruction function $r(\cdot)$ satisfying $r(\boldsymbol{G}_S) = \boldsymbol{G}$. Specifically, $\boldsymbol{G}_S$ with **redundant reconstruction ability** is defined by,*

$$r(\boldsymbol{G}_{S^*}) = \boldsymbol{G}, \ \exists \boldsymbol{G}_{S^*} \subset \boldsymbol{G}_S. \quad (8)$$

*$\boldsymbol{G}_S$ with **perfect reconstruction ability** is defined by,*

$$r(\boldsymbol{G}_{S^*}) \neq \boldsymbol{G}, \ \forall \boldsymbol{G}_{S^*} \subset \boldsymbol{G}_S. \quad (9)$$

If there does not existence a reconstruction function $r(\cdot)$ satisfying $r(\boldsymbol{G}_S) = \boldsymbol{G}$, $\boldsymbol{G}_S$ is with ***no reconstruction ability***.

In this section, we present a unified perspective that encapsulates the subgraph degeneration phenomena observed in two existing categories of subgraph learning methods. These failures stem from their inability to adhere to the principle of perfect reconstruction in subgraph learning. In next section, we thus propose a novel subgraph learning framework equipped with perfect reconstruction capabilities to effectively address the limitations of existing methods.

IV. RayE-Sub: Rayleigh-resistance Equivariant Subgraph Learning

In this section, we design a subgraph learning framework ***RayE-Sub*** with perfect reconstruction ability and further demonstrate its powerful expressive capacity theoretically.

A. Rayleigh-resistance Extractor for Subgraph Partition

Compared to node-based subgraph learning methods, partition-based methods often yield subgraphs closer to perfect reconstruction. The reason lies in that the excellent properties of the extracted subgraphs can only be guaranteed by incorporating the permutation invariant learning module. Therefore, we are pursuing the partition-based research line to tackle the challenge of subgraph degradation of existing approaches.

Subgraph partition methods aim to find the significant boundary, which composes of a series of edges connecting two irrelevant nodes. In implementations, topological characteristics and feature contents are vital factors affecting the effectiveness of partition principles. Inspired by the great superiority of spectral theory in drawing graphs [18], [24], we exploit the idea of spectrum to realize the subgraph partitions with topological and feature information.

The Laplacian operator $L_{\boldsymbol{G}}$ is the entry to spectral theory, where *Rayleigh quotient* of $L_{\boldsymbol{G}}$ elegantly depict the stability of graph $\boldsymbol{G}$ [24].

Definition 2. *The Rayleigh quotient $q(\boldsymbol{G})$ of the Laplacian matrix $L_{\boldsymbol{G}}$ is defined as*

$$q(\boldsymbol{G}) = x^T L_{\boldsymbol{G}} x = \sum_{(u,v)\in E} w_{uv}(x(u) - x(v))^2,\ x^T x = 1, \quad (10)$$

where $x \in \mathbb{R}^{n\times 1}$ is the feature matrix of nodes [1] .

The value of $q(\boldsymbol{G})$ resonates with both the structure information w_{uv} and the feature contents $(x(u) - x(v))^2$ for all edges (u, v). Thus, *Rayleigh quotient* quantifies the stability of the graph with structure and feature information. The smaller $q(\boldsymbol{G})$ refers to those nodes are closer with each other so that the graph is more stable. Otherwise, the graph is fragile. The quantization property aligns well with our subgraph extraction strategy, which involves breaking fragile edges $B_{\boldsymbol{G}} \subset \boldsymbol{V}$ to obtain the most stable substructures. This leads us to further investigate a subgraph learning method that can extract subgraphs with perfect reconstruction from the perspective of Rayleigh entropy.

[1]We let the feature dimension be 1 to simplify subsequent analysis.

There exists an equation between the Laplacian matrix $L_{\boldsymbol{G}_S^i} \in \mathbb{R}^{n\times n}$ of the subgraph $\boldsymbol{G}_S^i$, boundary matrix $B_{\boldsymbol{G}_S^i} \in \mathbb{R}^{n\times n}$ and extraction matrix $L_{\boldsymbol{G}}(\boldsymbol{V}_i, \boldsymbol{V}_i) \in \mathbb{R}^{n\times n}$ of $L_{\boldsymbol{G}}$,

$$L_{\boldsymbol{G}}(\boldsymbol{V}_i, \boldsymbol{V}_i) = L_{\boldsymbol{G}_S^i} + B_{\boldsymbol{G}_S^i}, \quad (11)$$

where $B_{\boldsymbol{G}_S^i}$ represents the boundary between $\boldsymbol{G}_S^i$ and the rest of the graph with $B_{\boldsymbol{G}_S^i}(a, a) = \sum_{b\notin \boldsymbol{G}_S^i} w_{ab}$. Let the boundary of global graph $B_{\boldsymbol{G}}$ be $\sum_{i\in[k]} B_{\boldsymbol{G}_S^i}$, we can rewrite the Rayleigh quotient (Equation 10) from the perspective of partitioning graph,

$$x^T L_{\boldsymbol{G}} x = \sum_{i\in[k]} x^T L_{\boldsymbol{G}_S^i} x + x^T B_{\boldsymbol{G}} x. \quad (12)$$

Given a specific x, $x^T L_{\boldsymbol{G}} x$ becomes fixed. Thus, the implementation of partitioning subgraphs by $\max_{B_{\boldsymbol{G}}} x^T B_{\boldsymbol{G}} x$ is equivalent to optimizing,

$$\min_{\boldsymbol{G}_S} q(\boldsymbol{G}_S) = \min_{\boldsymbol{G}_S} \sum_{i\in[k]} x^T L_{\boldsymbol{G}_S^i} x. \quad (13)$$

Unfortunately, this optimization objective become trapped, as it may lead to including all edges in $B_{\boldsymbol{G}}$ to achieve the optimal value, resulting in a bag of single-node subgraphs. A straightforward solution is to consider the number of nodes as one of the descriptive factors of subgraph stability. Follow this idea, instead of $L_{\boldsymbol{G}}$, we utilize the normalized Laplacian matrix $N_{\boldsymbol{G}} = D^{-1/2} L_{\boldsymbol{G}} D^{-1/2}$ to rewrite the formation of *Rayleigh quotient* as,

$$q^*(\boldsymbol{G}) = x^T N_{\boldsymbol{G}} x = \sum_{(a,b)\in E} w_{ab}\left(\frac{x(a)}{\sqrt{d_a}} - \frac{x(b)}{\sqrt{d_b}}\right)^2, \quad (14)$$

where d_j represents the degree of j. Equation 14 effectively avoids single-node subgraphs case by considering the degree of node in subgraph. However, we are still curious about whether this rephrasing results in different semantics theoretically. We make the following derivation:

$$x^T N_{\boldsymbol{G}} x = x^T D^{-1/2} L_{\boldsymbol{G}} D^{-1/2} x, \quad (15)$$

we can further obtain

$$N_{\boldsymbol{G}}(V_i, V_i) = D^{-1/2} L_{\boldsymbol{G}_S^i} D^{-1/2} + D^{-1/2} B_{\boldsymbol{G}_S^i} D^{-1/2}, \quad (16)$$

we then get

$$x^T N_{\boldsymbol{G}} x = \sum_{i\in[k]} x^T D^{-1/2} L_{\boldsymbol{G}_S^i} D^{-1/2} x + x^T D^{-1/2} B_{\boldsymbol{G}} D^{-1/2} x. \quad (17)$$

Given x, $x^T N_{\boldsymbol{G}} x$ is fixed. Thus, discovering the significant boundary $\max_{B_{\boldsymbol{G}}} x^T D^{-1/2} B_{\boldsymbol{G}} D^{-1/2} x$ is equivalent to optimizing $\min_{\boldsymbol{G}_S} x^T N_{\boldsymbol{G}} x = \min_{\boldsymbol{G}_S} \sum_{i\in[k]} x^T D^{-1/2} L_{\boldsymbol{G}_S^i} D^{-1/2} x = \min_{\boldsymbol{G}_S} \sum_{i\in[k]} \left(D^{-1/2} x\right)^T L_{\boldsymbol{G}_S^i} D^{-1/2} x$. In this case, we can consider $y = D^{-1/2} x$ as the feature of nodes. Different from the previous one, topological information, such as the degree distribution of nodes, is also an important metric for graph partition. Therefore, our solution can not only improve the subgraph partition guided by *Rayleigh quotient* but also avoid the defect of single-node subgraph case.

However, optimizing $\min_{\boldsymbol{G}_S} q^*(\boldsymbol{G}_S)$ is an NP (Non-deterministic Polynomial) problem, such optimization is an inaccessible target in practical implementation. To this end, we utilize resistance distance to implement the quantification defined by the *Rayleigh quotient*. Resistance distance (RD) is a basic metric in graph spectral theory [41]. It reflects both distance and the accessibility (number of pathways) between two nodes, which has the potential to characterize the global structural topology. This inspires us to exploit RD as an alternative to quantify the stability of subgraphs.

Definition 3. *The resistance distance between two vertices u and v in an electrical network, is measured by the resistance of the entire network when we treat it as an integrated complex resistor. It can be computed by*

$$RD_{uv} = (\delta_u - \delta_v)^T L^+ (\delta_u - \delta_v), \tag{18}$$

where δ_j is the elementary unit vector with $\mathbf{1}$ in coordinate j and L^+ is the pseudo-inverse of L.

Theorem 2. ***(The equivalence between Resistance distance and Rayleigh quotient.)*** *Let u and v be any two vertices connected by an edge. Under the accessible constraints, the resistance distance between u and v is approximately equivalent to the stability of pairwise vertices defined by* Rayleigh quotient.

Proof. We attempt to demonstrate that their equivalence can be achieved by integrating feature information x into the topology using the edge reweighting strategy $w_{uv} = ||x_u - x_v||^{-2}$. Given that we have $L_{\boldsymbol{G}} = U^T W U$, where $U \in \mathbb{R}^{m\times n}$ is the signed edge-vertex adjacency matrix and $W \in \mathbb{R}^{m\times m}$ is the diagonal matrix of edge weights. Thus, $N_{\boldsymbol{G}}$ can be rewritten as,

$$N_{\boldsymbol{G}} = D^{-1/2} U^T W U D^{-1/2}. \tag{19}$$

We can obtain the following two derivations, i.e., the derivation of objective of finding boudary of subgraph partition Equation 20 and Resistance distance Equation 21,

$$\begin{aligned} x^T N_{\boldsymbol{G}} x &= x^T D^{-1/2} U^T W U D^{-1/2} x \\ &= ||W^{1/2} U D^{-1/2} x||^2, \end{aligned} \tag{20}$$

$$\begin{aligned} RD_{uv} &= (\delta_u - \delta_v)^T L^+_{\boldsymbol{G}} L_{\boldsymbol{G}} L^+_{\boldsymbol{G}} (\delta_u - \delta_v) \\ &= ||W^{1/2} U L^+_{\boldsymbol{G}} (\delta_u - \delta_v)||^2. \end{aligned} \tag{21}$$

By comparing the two derivations, we find the only different between them is $D^{-1/2}x$ and $L^+_{\boldsymbol{G}}(\delta_u - \delta_v)$, which inspires us to utilize *Resistance distance* to quantify *Rayleigh quotient*. In other words, we will explore whether there exist the function $\phi(\cdot)$ satisfying $D^{-1/2}x = \phi(L^+_{\boldsymbol{G}}(\delta_u - \delta_v))$. We investigate this problem by integrating the feature information x into the topology, and prove this equivalence.

Since L_G is symmetric, we can diagonalize it and derive,

$$L_G = \sum_{i=1}^{n-1} \lambda_i \mu_i \mu_i^T, \tag{22}$$

where $\lambda_1, \lambda_2, ..., \lambda_{n-1}$ are the nonzero eigenvalues of $L_{\boldsymbol{G}}$ and $\mu_1, \mu_2, ..., \mu_{n-1}$ are a corresponding set of orthonormal eigenvectors. Thus, we can obtain the $L^+_{\boldsymbol{G}}$,

$$L^+_G = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \mu_i \mu_i^T. \tag{23}$$

For all nodes u and v connected by edges, we can easily obtain $(\delta_u - \delta_v) \in \mathbb{R}^{n\times|E|}$. We can derive that the i-th row element of $L^+_{\boldsymbol{G}}(\delta_u - \delta_v)$ is denoted as $\sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\mu_{ki}^2 - \mu_{ki}\mu_{kj})\kappa_i$, where $\kappa_i \in \mathbb{R}^{1\times n}$, and if j directly connected to i, then $\kappa_i[j]$ has a value of 1, otherwise it is 0. For $D^{-1/2}x$, its i-th row element is $\frac{x_i}{\sqrt{d_i}}$. Therefore, our proof is transformed into exploring whether there exists the function $\phi(\cdot)$ satisfying $\frac{x_i}{\sqrt{d_i}} = \phi(\sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\mu_{ki}^2 - \mu_{ki}\mu_{kj})\kappa_i)$. We define $\mathcal{N}(i) = \{j^1, ..., j^{d_i}\}$ and further simplify the latter term:

$$\begin{aligned} &\sum \sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\mu_{ki}^2 - \mu_{ki}\mu_{kj})\kappa_i \\ &= d_i \sum_{k=1}^{n-1} \frac{1}{\lambda_k}\mu_{ki}^2 - \sum_{j\in\mathcal{N}(i)} \sum_{k=1}^{n-1} \frac{1}{\lambda_k}\mu_{ki}\mu_{kj} \\ &= (\sum_{k=1}^{n-1} \frac{1}{\lambda_k}\mu_{ki}^2 - \sum_{k=1}^{n-1} \frac{1}{\lambda_k}\mu_{ki}\mu_{kj^1}) + ... \\ &+ (\sum_{k=1}^{n-1} \frac{1}{\lambda_k}\mu_{ki}^2 - \sum_{k=1}^{n-1} \frac{1}{\lambda_k}\mu_{ki}\mu_{kj^{d_i}}) \\ &= \sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\mu_{ki}^2 - \mu_{ki}\mu_{kj^1}) + ... + \sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\mu_{ki}^2 - \mu_{ki}\mu_{kj^{d_i}}). \end{aligned} \tag{24}$$

We borrow the concept of access time $H(i,j)$ in graph theory to find out the $\phi(\cdot)$. $H(i,j)$ is the expected number of steps before node j is visited, starting from node i,

$$H(i,j) = 2|E| \sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\frac{\mu_{ki}^2}{d_i} - \frac{\mu_{ki}\mu_{kj}}{\sqrt{d_i d_j}}). \tag{25}$$

Specifically, we define that $\phi(L^+_{\boldsymbol{G}}(\delta_u - \delta_v)) = \frac{\sum_{j\in N(i)} H(i,j)}{\sqrt{d_i}}$. Actually, the *access time* between neighbors is determined by the weight of the edges and the local topology. Therefore, integrating feature information x into the topology is the key to ensure this feasibility. This problem becomes into investigating the correlation between x_i and $\sum_{j\in N(i)} H(i,j)$. In practical implementation, however, the dimension of x is d (non-zero). We further explore the topological relationship between nodes from the embedded Euclidean space. Specifically, we reweight edges using $w_{uv} = ||x_u - x_v||^{-2}$. This design reveals that the isolation or centrality of node embedding in Euclidean space must lead to its topological isolation or centrality. For example, isolated node i (x_i) must be hard to reach, so the *access time* ($\sum_{j\in N(i)} H(i,j)$) to its neighbors must also be long, which satisfies our expectation.

In conclusion, when the reweighting strategy of edge $w_{uv} = ||x_u - x_v||^{-2}$ is achieved, the resistance distance between u and v is approxi20mately equivalent to the stability of

pairwise vertices defined by *Rayleigh quotient*. The theoretical understanding of effective resistance distance supports this design: the edge weight and resistance distance have opposite tendency [24].

Theoretical Insight. We begin by aligning the formal expressions of Rayleigh entropy and resistance distance, as presented in Equation 20 and 21. Then, we analyze their differences through spectral decomposition, as shown in Equation 23. Based on the differences identified through re-comparison, our proof is transformed into verifying whether there exists the function $\phi(\cdot)$ satisfying $\frac{x_i}{\sqrt{d_i}} = \phi(\sum_{k=1}^{n-1} \frac{1}{\lambda_k}(\mu_{ki}^2 - \mu_{ki}\mu_{kj})\kappa_i)$. Ultimately, we develope a reweighting estimation method using access time as a bridge. In conclusion, the proof shows that the alignment of Rayleigh quotient and resistance distance can be achieved through the reweighting strategy.

B. Siamese-Query Network for Subgraph-level Equivariant Aggregation

The discussion above supports the extraction of a bag of non-overlapping subgraphs, which guarantees the perfection of extracted subgraphs. After that, we investigate methods for enabling interactions among subgraphs to achieve subgraphs reconstruction. Specifically, we propose a *Siamese-Query* scheme to realize equivariant subgraph learning, as shown in Figure 3. These layers map bags of subgraphs into representation $\boldsymbol{Z}$ as follows,

$$h_{\boldsymbol{G}} = L(A, X),\ h_S = \mathrm{CONCAT}[L(A_i, X_i)], \tag{26}$$

$$Q = h_{\boldsymbol{G}} W_Q,\ K = h_S W_K,\ V = h_S W_V, \tag{27}$$

where $W_Q, W_K, W_V \in \mathbb{R}^{d\times d}$ are learnable parameters. $h_{\boldsymbol{G}} \in \mathbb{R}^{1\times d}$ and $h_S \in \mathbb{R}^{m\times d}$ respectively denote the graph-level and subgraph-level embedding obtained by an MPNN encoder L. We can further get the representations $\boldsymbol{Z}$ for prediction,

$$\boldsymbol{Z} = \mathrm{softmax}(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d}}})\mathrm{V}. \tag{28}$$

Query mechanism introduces attention score to each subgraph by interacting with global information, which realizes subgraph-level equivariant learning. We define $g(h_{\boldsymbol{G}}, h_S) = \mathrm{softmax}(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d}}})$. For any permutation τ acting on subgraphs, $g(h_{\boldsymbol{G}}, \tau \cdot h_S) = \tau \cdot g(h_{\boldsymbol{G}}, h_S)$ is always hold on.

C. Learning Objective

RayE-Sub is a two-stage learning architecture, where *Rayleigh-resistance* realizes subgraph partition and *Siamese-Query* achieves subgraph-level equivariant learning.

Rayleigh-resistance Extractor aims to obtain boundary $B_{\boldsymbol{G}}$ to partition graph $\boldsymbol{G}$. Specifically, we employ the resistance distance between two connected nodes to quantify the stability of this edge $s(u, v)$. Similar to *Rayleigh quotient*, smaller resistance distance indicates a more stabler connection. Therefore, $B_{\boldsymbol{G}}$ is composed of the edges (u, v) belonging to top-β(S), which picks out the top β larger of S,

$$B_G := \bigcup_{(u,v)\in E} \{(u, v)\},\ \ s(u, v) \in \text{top-}\beta(\mathrm{S}), \tag{29}$$

where β is a hyper-parameter and its sensitivity analysis is provided in Section VI-H.

Siamese-Query Network is proposed to achieve subgraph-level equivariant learning in making predictions. For each graph with label Y_i and its prediction $\hat{Y}_i$, we impose the cross entropy loss on all N graphs as the learning objective,

$$\mathcal{L} := -\frac{1}{N}\sum_{i=1}^{N} Y_i \log(\hat{Y}_i). \tag{30}$$

D. The Expressive Power of RayE-WL

We introduce WL analogue (RayE-WL) for ***RayE-Sub*** to support our next study of the expressive ability. We initially introduce the central process of RayE-WL, the color refinement algorithm. A detailed outline of the algorithm is provided in Algorithm 3. On subgraph $\boldsymbol{G}_S^i \in \boldsymbol{G}_S$, the color (feature) of node $v \in \boldsymbol{G}_S^i$ is refined (updated) according to the rule,

$$c_{v,\boldsymbol{G}_S^i}^{t+1} := \mathrm{HASH}(c_{v,\boldsymbol{G}_S^i}^{t}, N_{v,\boldsymbol{G}_S^i}^{t}, c_{\boldsymbol{G}_S^i}^{t+1}), \tag{31}$$

$$c_{\boldsymbol{G}_S^i}^{t+1} := \mathrm{HASH}(c_{\boldsymbol{G}_S^i}^{t}, M_{\boldsymbol{G}_S^i,\boldsymbol{G}}^{t}, c_{\boldsymbol{G}}^{WL,t}), \tag{32}$$

where $N_{v,\boldsymbol{G}_S^i}^{t}$ denotes the multiset of colors in v's neighborhood over subgraph $\boldsymbol{G}_S^i$ after the t-th iteration [19], [42]. $c_{\boldsymbol{G}_S^i}^{t}$ represents the color of the subgraph $\boldsymbol{G}_S^i$ in which node v is located after the t-th iteration. $M_{\boldsymbol{G}_S^i,G}^{t}$ denotes the color multiset of all subgraphs of the graph $\boldsymbol{G}$ independently mapped by 1-WL after the t-th iteration, $M_{\boldsymbol{G}_S^i,\boldsymbol{G}}^{t} = \{c_S^{\mathrm{WL,t}} | S \in \boldsymbol{G}_{\boldsymbol{G}_S^i}\}$. $c_{\boldsymbol{G}}^{t}$ represents the color of the graph $\boldsymbol{G}$ mapped by 1-WL after the t-th iteration.

Algorithm 1: The k-dimensional Weisfeiler-Lehman Algorithm

Input: Graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ and the number of iterations T
Output: The coloring of all k-tuples C
Initialization: The initial coloring C^0 is defined using the isomorphism type of each k-tuple
for $t \leftarrow 1$ **to** T **do**
  for *each k-tuple* $\boldsymbol{i}$ **do**
    $C_{\boldsymbol{i}}^{t} := \mathrm{HASH}(C_{\boldsymbol{i}}^{t-1}, (\{\{\mathrm{C}_{\boldsymbol{j}}^{t-1} | \boldsymbol{j} \in N_j(\boldsymbol{i})\}| j \in [k]\}))$
  end
end
Result: C

Theorem 3. ***(RayE-WL is more powerful than 1-WL.)*** *RayE-WL is strictly more powerful than 1-WL in distinguishing non-isomorphic graphs, which is upper bounded by 3-WL.*

Proof. Given two non-isomorphic graphs $\boldsymbol{G}$, $\boldsymbol{H}$, we first prove that RayE-WL is stronger than 1-WL by two terms. (i) If they can be distinguished by the 1-WL graph isomorphism test, RayE-WL can strictly distinguish them. (ii) There exist graphs that cannot be distinguished by 1-WL but can be distinguished by RayE-WL.

For the first term, $\boldsymbol{G}$ and $\boldsymbol{H}$ can be distinguished by the 1-WL graph isomorphism test, which means $c_{\boldsymbol{G}} \neq c_{\boldsymbol{H}}$. We can easily deduce $\{c_{\boldsymbol{G}_S^i} | \boldsymbol{G}_S^i \in \boldsymbol{G}_S\} \neq \{c_{\boldsymbol{H}_S^i} | \boldsymbol{H}_S^i \in \boldsymbol{H}_S\}$, where $\boldsymbol{G}_S$ and $\boldsymbol{H}_S$ are the subgraph sets extracted by $\boldsymbol{G}$ and $\boldsymbol{H}$ respectively. The final the color multiset of $\boldsymbol{G}$ and $\boldsymbol{H}$ is also distinguishable.

For the second term, $\boldsymbol{G}$ and $\boldsymbol{H}$ cannot be distinguished by the 1-WL graph isomorphism test, which means $c_{\boldsymbol{G}} = c_{\boldsymbol{H}}$. After $\boldsymbol{G}$ and $\boldsymbol{H}$ are partitioned into subgraphs $\boldsymbol{G}_S$ and $\boldsymbol{H}_S$, there are two existing cases that should be discussed. One is that each subgraph can be distinguished, i.e., $\{M_{\boldsymbol{G}_S^i,\boldsymbol{G}} | \boldsymbol{G}_S^i \in \boldsymbol{G}_S\} \neq \{M_{\boldsymbol{H}_S^i,\boldsymbol{H}} | \boldsymbol{H}_S^i \in \boldsymbol{H}_S\}$. In other words, the independent mapping results based on 1-WL in each subgraph can distinguish between $\boldsymbol{G}_S$ and $\boldsymbol{H}_S$. Thus, we can similarly deduce $\{c_{\boldsymbol{G}_S^i} | \boldsymbol{G}_S^i \in \boldsymbol{G}_S\} \neq \{c_{\boldsymbol{H}_S^i} | S_S^i \in \boldsymbol{H}_S\}$. The final color of each node $c_{v,S}$ is also distinguishable. More importantly, another case is that their generated subgraphs still cannot be distinguished, i.e., $\{M_{\boldsymbol{G}_S^i,\boldsymbol{G}} | \boldsymbol{G}_S^i \in \boldsymbol{G}_S\} = \{M_{\boldsymbol{H}_S^i,\boldsymbol{H}} | \boldsymbol{H}_S^i \in \boldsymbol{H}_S\}$. The bottom panel of Figure 1 intuitively describes this case. According to the comparison among their color refinement algorithms, we can observe our RayE-WL is a subclass of 3-WL. Thus, we can conclude that RayE-WL is upper bounded by 3-WL.

Algorithm 2: The k-dimensional Folklore Weisfeiler-Lehman Algorithm

Input: Graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ and the number of iterations T
Output: The coloring of all k-tuples C
Initialization: The initial coloring C^0 is defined using the isomorphism type of each k-tuple **for** $t \leftarrow 1$ **to** T **do**
  for *each* k*-tuple* $\boldsymbol{i}$ **do**
    $C_{\boldsymbol{i}}^t := \mathrm{HASH}(C_{\boldsymbol{i}}^{t-1}, \{\{(\mathrm{C}_{\boldsymbol{j}}^{t-1} | \boldsymbol{j} \in N_j^F(\boldsymbol{i})) | j \in [n]\}\})$
  end
end
Result: C

It's crucial to emphasize that despite the degradation of subgraphs, the expressive power of existing subgraph learning methods are strictly more powerful than 1-dimensional Weisfeiler-Lehman (1-WL). Actually, this degradation phenomenon indicates that node-based approaches have an unacceptable lower limit of expression ability. The reason lies in the following analysis. It is well accepted that Node-WL is stronger than 1-WL by proving following two terms. Given two non-isomorphic graphs $\boldsymbol{G}$, $\boldsymbol{H}$, we can demonstrate (i) if they can be distinguished by the 1-WL graph isomorphism test, Node-WL can strictly distinguish them; (ii) there exist graphs that cannot be distinguished by 1-WL but can be distinguished by Node-WL. Therefore, there will be some non-isomorphic pairs of graphs, those can neither be distinguished by Node-WL nor 1-WL. The cases of subgraph degradation corresponds to this scenario, which does not affect the proof of these two terms. Our work systematically studies this degradation phenomenon and further propose a more stable subgraph learning architecture to address this limited distinguiability. In terms of expressive power, our architecture is still bounded above by 3-WL, yet it can achieve a higher lower bound against existing methods. The specific method design and theoretical proof will be provided in the subsequent sections.

Algorithm 3: The RayE-WL Algorithm

Input: Graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ and the number of iterations T
Output: The coloring of all nodes c^T
Initialization: Initialize the color of each node c^0, and the subgraph set $\boldsymbol{G}_S$ extracted from $\boldsymbol{G}$
for $t \leftarrow 1$ **to** T **do**
  for *each subgraph* S *in* $\boldsymbol{G}_S$ **do**
    for *each node* v *in* S **do**
      $c_S^t := \mathrm{HASH}(c_S^{t-1}, M_{S,G}^{t-1}, c_G^{WL,t-1})$
      $c_{v,S}^t := \mathrm{HASH}(c_{v,S}^{t-1}, N_{v,S}^{t-1}, c_S^t)$
    end
  end
end
Result: c^T

E. Efficiency Analysis

In this section, we provide an efficient approximation approach for calculating RD with the time complexity of $\mathcal{O}(|\boldsymbol{E}|)$. Specifically, the running time of the subgraphs partition stage mainly comes from calculating edge weight (w_{ab}) and resistance distance (RD_{ab}). It is obvious that the process of reweighting edge costs $\mathcal{O}(|\boldsymbol{E}|)$ time of computations. Given the major computational cost is introduced by RD, we are concerned with the more efficient calculation of resistance distance and do not need to follow the inefficient method with $\mathcal{O}(|V|^3)$. Inspired by Equation 45, we propose a resistance distance approximation method with $\mathcal{O}(|\boldsymbol{E}|)$ time complexity. Specifically, we first precalculate the resistance distance RD^* among all the nodes in the case that all edges are with the weight of 1. Then, with the weight w_{ab} of edge (a, b), we approximate the resistance distance $RD_{ab} = RD_{ab}^*/w_{ab}$. This proposal realizes the calculation of resistance distance in $\mathcal{O}(|\boldsymbol{E}|)$ time complexity. Note that this approximate method can only be used to calculate the resistance distance between nodes connected by edges. The cost time of the subgraph-level equivariant learning stage mostly stems from *Query* mechanism, where its time complexity is $\mathcal{O}(md^2)$, m represents the number of subgraphs, and d indicates the hidden dimension. Then, we can conclude the overall time complexity of ***RayE-Sub*** to be $\mathcal{O}(|\boldsymbol{E}|)$ due to d is a constant.

V. Understanding Rayleigh Quotient from Spectral Theory

Fundamentally, our approach is inspired by spectral graph theory. As mentioned earlier, the *Rayleigh quotient* intuitively indicates the stability of graphs, which provide a solid theoretical foundation to guide our subgraph partitioning. In this section, our goal is to offer a deep theoretical elucidation of

our design from the perspective of spectral theory. We first introduce Lemma 1 [24] to understand spectral theory from eigenvector space.

Lemma 1. *Let M be an n-dimensional real symmetric matrix. There exist numbers $\mu_1, ..., \mu_n$ and orthonormal vectors $\varphi_1, ..., \varphi_n$ such that $M\varphi_i = \mu_i\varphi_i$. Moreover,*

$$\varphi_1 \in \arg\max_{||x||=1} x^T M x, \tag{33}$$

and for $2 \leq i \leq n$,

$$\varphi_i \in \arg\max_{\substack{||x||=1 \\ x^T\varphi_j=0, j<i}} x^T M x. \tag{34}$$

Similarly,

$$\varphi_i \in \arg\min_{\substack{||x||=1 \\ x^T\varphi_j=0, j>i}} x^T M x. \tag{35}$$

We consider the case where M represents the Laplacian operator L_G of the graph $\boldsymbol{G}$. Let $L_{\boldsymbol{G}}$ be with eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. We can obtain,

$$\mu_k = \max_{\substack{S\subseteq\mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{x\in S \\ x\neq 0}} \frac{x^T L_G x}{x^T x} = \min_{\substack{T\subseteq\mathbb{R}^n \\ \dim(T)=n-k+1}} \max_{\substack{x\in T \\ x\neq 0}} \frac{x^T L_G x}{x^T x}, \tag{36}$$

where the maximization and minimization are over subspaces S and T of $\mathbb{R}^n$. Equation 36 reveals that the extreme values of *Rayleigh quotient* are equivalent with eigenvalues of L_G. Actually, the eigenvalues are also characterized by the resistance distance, which build a bridge connecting *Rayleigh quotient* and *resistance distance*.

The resistance distance between two vertices a and b in an electrical network is the resistance of the entire network when we treat it as one complex resistor. That is, we consider an electrical flow that delivers one unit of current into node a and outflows one unit of current from node b. We measure the potential difference between a and b as the resistance distance, denoted as RD_{ab},

$$RD_{ab} \overset{def}{=} (\delta_a - \delta_b)^T L^+ (\delta_a - \delta_b), \tag{37}$$

where δ_j is the elementary unit vector with $\mathbf{1}$ in coordinate j and L^+ is the pseudo-inverse of L,

$$\begin{aligned} RD_{ab} &\overset{def}{=} (\delta_a - \delta_b)^T L^+ (\delta_a - \delta_b) \\ &= (L^{+/2}(\delta_a - \delta_b))L^{+/2}(\delta_a - \delta_b) \\ &= ||L^{+/2}(\delta_a - \delta_b)||^2 \\ &= ||L^{+/2}\delta_a - L^{+/2}\delta_b||^2 \\ &= \mathrm{dist}(L^{+/2}\delta_a, L^{+/2}\delta_b)^2. \end{aligned} \tag{38}$$

Physical principle tells us that the vertices will settle into the position that is bound to minimize the potential energy. The potential energy ξ of an ideal linear spring with constant w when stretched to length l is given by,

$$\xi = \frac{1}{2} w l^2. \tag{39}$$

Thus, the potential energy in a configuration x is given by,

$$\xi(x) \overset{def}{=} \frac{1}{2} \sum_{(a,b)\in E} w_{ab}(x(a) - x(b))^2. \tag{40}$$

The lowest energy must be reached when the each variable of $\xi(x)$ is zero. The partial derivative with respect to $x(a)$ is,

$$\frac{1}{2} \sum_{(a,b)\in E} 2w_{ab}(x(a) - x(b)) = \sum_{(a,b)\in E} w_{ab}(x(a) - x(b)). \tag{41}$$

Setting this to zero gives the equations we can obtain,

$$x(a) = \frac{1}{d_a} \sum_{(a,b)\in E} w_{ab} x(b). \tag{42}$$

This result can be broadly generalized as the Lemma 2 [24].

Lemma 2. *Let $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E}, w)$ be a weighted graph, let $\boldsymbol{B} \subset \boldsymbol{V}$, and let $\boldsymbol{S}$ be $\boldsymbol{V} - \boldsymbol{B}$. Given $x(\boldsymbol{B})$, $\xi(x)$ is minimized by setting $x(\boldsymbol{S})$ so that x is harmonic on $\boldsymbol{S}$.*

Given this lemma, we focus on the resistance distance between a and b, thus x should be harmonic on $\boldsymbol{V} - \{a, b\}$. Fortunately, we already know how compute such a vector x. Thus, we can set,

$$y = L^+(\delta_a - \delta_b)/RD_{ab}. \tag{43}$$

We obtain,

$$y(a) - y(b) = (\delta_a - \delta_b)^T L^+ (\delta_a - \delta_b)/RD_{ab} = 1. \tag{44}$$

Thus, y is harmonic on $\boldsymbol{V} - \{a, b\}$. We further set $x = y - 1 * y(s)$. It is obvious that x satisfies $x(s) = 0$, $x(t) = 1$, and it is harmonic on $\boldsymbol{V} - \{a, b\}$. We compute the energy as,

$$\begin{aligned} x^T L x &= y^T L y \\ &= \tfrac{1}{(RD(a,b))^2}(L^+(\delta_a - \delta_b))^T L (L^+(\delta_a - \delta_b)) \\ &= \tfrac{1}{(RD(a,b))^2}(\delta_a - \delta_b)^T L^+ L L^+ (\delta_a - \delta_b) \\ &= \tfrac{1}{(RD(a,b))^2}(\delta_a - \delta_b)^T L^+ (\delta_a - \delta_b) \\ &= \tfrac{1}{RD(a,b)}. \end{aligned} \tag{45}$$

This derivation reveals a fact that the weights of edges are the reciprocals of their resistance distance. In practical implementation, we employ this understanding to achieve the approximate resistance distance with the time complexity of $\mathcal{O}(|\boldsymbol{E}|)$. More importantly, Equation 45 builds a theoretical bridge between *Rayleigh quotient* and *Resistance distance*. This verifies the rationale of utilizing resistance distance to partition graphs, and the effectiveness of our proposed *Rayleigh-resistance*.

VI. Experiments

A. Datasets

The selected datasets are two-fold, four real-world datasets and two synthetic datasets on graph classification tasks.

BA-2Motifs is a synthetic dataset created by [43] with two graph classes. House motifs and cycle motifs give class labels and thus are regarded as ground-truth explanations for the two classes respectively.

Spurious-Motif is a synthetic dataset proposed by [26] with three graph classes. Each graph is composed of one base S and one motif C. The motif C directly determines the label of the graph. We can create Spurious-Motif datasets with different spurious correlation, which represents the degree (b) between

the base S and the label. In our implementation, we choose $b = 0.5$, 0.7 and 0.9 to obtain datasets.

MUTAG [44] is a binary dataset of molecular property, where nodes represent atoms and edges denote chemical bonds. Each graph is associated with a binary label based on its mutagenic effect.

Open Graph Benchmark (OGB) [45] is a series of real, large-scale and diverse datasets which is utilized for machine learning on graphs. It covers almost all real-world tasks, including node-level, link-level and graph-level property prediction. We choose MOLHIV, BBBP and SIDER to verify our method.

B. Baselines

Our baselines are three-fold, including GNN backbones (i.e., GCN, Graph-SAGE, GIN), node-based subgraph learning methods (i.e., ESAN, GNN-AK, SUN, MAG-GNN, 2-DRFWL(2)) and partition-based subgraph learning models (i.e., IB-subgraph, GSAT, DIR, EdgeRWSE, MICRO-Graph). A detailed description of the datasets is provided below.

GCN [46] serves as a landmark model due to its ability to effectively handle node-level tasks in graph-structured data. It introduced the concept of graph convolutional layers, allowing for information propagation and aggregation across the nodes of a graph.

Graph-SAGE [47] is a graph neural network framework that extends the concept of graph convolutional networks (GCNs). It introduces the idea of "aggregators" to aggregate information from a node's neighborhood, allowing for a more flexible and scalable approach to graph representation learning.

GIN [10] is a graph neural network model that is designed to handle graph-structured data and capture important structural information. It is a provably maximally powerful GNN under the neighborhood aggregation framework.

ESAN [16] implements an subgraph equivariant learning architecture and achieve better expressiveness by per-layer aggregation across subgraphs.

GNN-AK [17] follows a similar manner to develop subgraph GNNs by considering the star-pattern as the pre-defined substructure.

SUN [15] falls into a practice of node-based subgraph learning. It aligns the permutation group of nodes and subgraphs, and models the symmetry with a smaller single permutation group.

IB-subgraph [25] is a pioneering implementation of the information bottleneck theory in the field of graph learning. It accomplishes subgraph partitioning through the utilization of edge drop technology.

GSAT [13] follows the practice of IB-subgraph and designs a subgraph extraction strategy with edge deletions based on stochastic attention mechanism.

DIR [26] disentangles the input graph into causal and non-causal subgraphs, and utilizes invariant features to construct interpretable GNNs from a causal perspective.

MAG-GNN [48] proposes a reinforcement learning-driven approach that efficiently identifies discriminative subgraphs, significantly reducing computational costs while preserving strong expressive capabilities.

d-DRFWL(2) [49] strikes a balance between expressive power and computational complexity by limiting message passing to a maximum distance of d between node pairs.

EdgeRWSE [50] focuses on enhancing the theoretical expressiveness and practical performance of GNNs through random walks on simplicial complexes.

MICRO-Graph [51] is designed to pretrain GNNs using self-supervised contrastive learning to achieve stronger predictive performance.

C. Backbone and Metrics

We exploit GIN as the backbone of ***RayE-Sub*** due to its extensive popularity. To ensure fair comparisons, we let GIN serve as the basic model in all baselines. We also explore the influence of different backbones on performance, which is provided in Section VI-I. For prediction performance, we employ ROC-AUC for OGB datasets (MOLHIV, BBBP and SIDER) and accuracy for the other datasets. All our experiments are conducted on a Tesla V100-PCIE-16GB GPU, and are repeated with 10 different random seeds of $[2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032]$. Therefore, the performance of our selected baselines may differ from that described in the original paper due to different random seed settings.

D. Main Results

We conduct a comparative analysis to evaluate the effectiveness of our ***RayE-Sub*** compared to these two lines of approaches respectively, as shown in Table I and II. On the whole, our approach exhibits strong competitiveness and clear advantages. Especially when compared to node-based methods, our approach demonstrates dominance, achieving 6 optimal performances. Next, we present the insights gained from comparing our approach with node-based and partition-based subgraph learning methods.

Comparing with node-based subgraph learning methods. As shown in Table I, compared to node-based baselines, our approach demonstrates dominance, achieving 6 optimal performances. Node-based baselines are relatively simple in design, emphasizing direct intervention in the receptive field during the message-passing process. However, they often lack the incorporation of inductive biases specific to downstream tasks (e.g., chemical knowledge). Therefore, in tasks with strong inductive biases, such as MUTAG and BA-2Motifs, their performance tends to be suboptimal. Although our ***RayE-Sub*** also focuses on filtering and learning receptive fields within the structure, it enhances representation by incorporating rich information through a reweighting strategy. This becomes the main reasons why our model achieves better performance. Note that the performance of GSAT differs from the published version because we retrained it by aligning it with the random seeds of RayE-Sub.

Comparing with partition-based subgraph learning methods. As shown in Table II, compared to partition-based baselines, our approach is competitive, achieving nearly all optimal and sub-optimal results. We find that our method is robust across all datasets,with only minor disadvantages even

TABLE I
PERFORMANCE COMPARISONS WITH BACKBONES AND NODE-BASED BASELINES. THE BEST RESULTS ARE IN **BOLD** AND THE SECOND BEST IS UNDERLINED.

	MOLHIV	BBBP	SIDER	MUTAG	BA-2Motifs	Spurious-Motif 0.5	Spurious-Motif 0.7	Spurious-Motif 0.9
GCN	75.5 ± 1.6	65.3 ± 1.9	52.1 ± 2.0	83.7 ± 4.7	86.8 ± 1.7	33.2 ± 1.8	31.6 ± 1.7	29.6 ± 6.2
Graph-SAGE	74.8 ± 3.4	64.1 ± 2.8	52.5 ± 1.6	84.6 ± 5.3	85.7 ± 2.3	34.8 ± 2.0	31.5 ± 2.5	30.4 ± 3.4
GIN	75.8 ± 1.3	66.4 ± 2.0	56.2 ± 1.6	89.4 ± 5.6	89.5 ± 2.1	39.9 ± 1.3	39.0 ± 1.6	38.6 ± 2.3
ESAN	77.2 ± 1.3	68.8 ± 1.3	58.1 ± 1.8	92.0 ± 5.0	92.9 ± 2.9	**56.1 ± 1.7**	47.9 ± 1.5	44.8 ± 2.9
GNN-AK	76.8 ± 1.2	67.7 ± 4.2	57.5 ± 1.4	92.3 ± 6.8	91.6 ± 3.3	54.2 ± 1.2	44.8 ± 1.7	42.6 ± 1.8
SUN	76.6 ± 0.9	66.4 ± 1.5	56.7 ± 2.0	94.7 ± 5.2	93.6 ± 4.1	55.6 ± 3.2	45.2 ± 2.4	43.2 ± 1.6
MAG-GNN	77.1 ± 1.1	70.9 ± 1.0	57.9 ± 2.3	95.1 ± 2.0	97.2 ± 1.6	53.6 ± 2.5	47.2 ± 1.2	44.5 ± 1.3
2-DRFWL(2)	**78.2 ± 2.2**	71.6 ± 1.3	56.7 ± 2.0	94.7 ± 5.2	93.6 ± 4.1	52.9 ± 2.3	48.6 ± 1.6	45.1 ± 1.3
RayE-Sub	77.6 ± 1.0	**72.2 ± 1.1**	**58.4 ± 1.9**	**95.6 ± 2.4**	**98.5 ± 1.0**	53.8 ± 2.0	**49.6 ± 2.9**	**45.8 ± 2.2**

TABLE II
PERFORMANCE COMPARISONS WITH PARTITION-BASED BASELINES. THE BEST RESULTS ARE IN **BOLD** AND THE SECOND BEST IS UNDERLINED.

	MOLHIV	BBBP	SIDER	MUTAG	BA-2Motifs	Spurious-Motif 0.5	Spurious-Motif 0.7	Spurious-Motif 0.9
IB-subgraph	76.4 ± 2.6	68.1 ± 1.1	57.7 ± 2.1	91.1 ± 6.4	90.1 ± 6.5	**54.4 ± 7.0**	48.5 ± 5.8	**46.2 ± 5.7**
GSAT	76.5 ± 1.5	69.0 ± 1.2	57.2 ± 1.3	96.7 ± 2.1	97.4 ± 1.9	46.6 ± 2.9	49.1 ± 3.0	39.8 ± 2.4
DIR	76.3 ± 1.1	68.2 ± 1.4	57.8 ± 1.8	92.1 ± 2.3	93.8 ± 9.6	45.5 ± 3.8	41.1 ± 2.6	37.6 ± 2.0
EdgeRWSE	**78.9 ± 1.2**	71.9 ± 1.0	58.0 ± 1.3	**97.4 ± 2.0**	96.9 ± 2.9	51.2 ± 3.3	46.9 ± 2.2	42.1 ± 1.4
MICRO-Graph	75.1 ± 1.1	**84.4 ± 1.1**	56.7 ± 0.9	97.0 ± 2.8	97.8 ± 3.6	50.4 ± 2.8	47.6 ± 1.9	43.7 ± 2.1
RayE-Sub	77.6 ± 1.0	72.2 ± 1.1	**58.4 ± 1.9**	95.6 ± 2.4	**98.5 ± 1.0**	53.8 ± 2.0	**49.6 ± 2.9**	45.8 ± 2.2

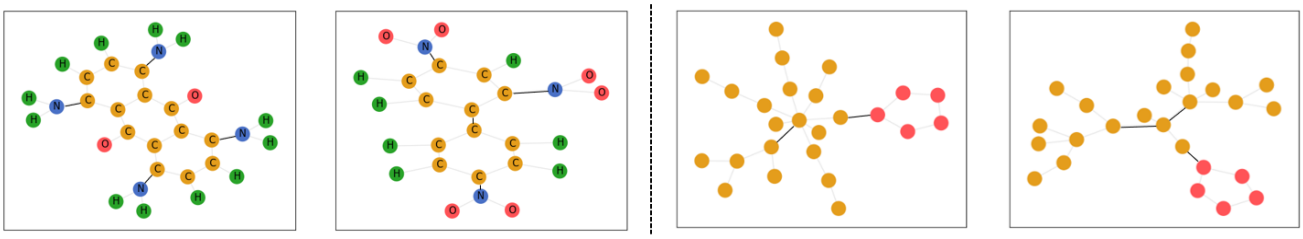


Fig. 4. Visualizing the boundary discovered by ***RayE-Sub***, where the **bold connections** are the edges of boundary. **Left panel:** the boundary obtained by ***RayE-Sub*** for MUTAG. **Right panel:** the boundary discovered by ***RayE-Sub*** for BA-2Motifs.

for suboptimal results. We argue that this is primarily due to the relatively minor subgraph degradation caused by partition-based methods in the existing datasets. For instance, the low proportion of isomers in these datasets enables current methods to perform well on the available samples. In our future work, we aim to explore more challenging domains that better align with the application scenarios of our method and construct such "subgraph degradation datasets" for further evaluation. Moreover, we also observe a significant performance improvement of MICRO-Graph on the BBBP dataset comparing all existing methods. Actully, this improvement can be attributed to the pre-training learning mechanism of MICRO-Graph. Therefore, this motivates us to further apply the same pre-training mechanism to further enhance performance and generalization in the future work.

Combining all two lines of subgraph learning methods, we also have the following **Obs**ervations.

Obs 1: Partition-based subgraph discovery approaches have better predictive ability than node-based methods. In all eight datasets, partition-based subgraph learning methods including ***RayE-Sub*** obtain seven best results. Specifically, the top-2 performances on all datasets are almost achieved by partition-based approaches. From the perspective of invariant learning, we argue that the partition-based methods can often obtain the casual subgraphs to improve the representation ability of the model. Actually, this powerful ability to capture causal information is especially crucial for predicting the properties of chemical molecules. These results verify that partition-based methods not only have better interpretability, but also are with better expressiveness. Therefore, our framework, ***RayE-Sub***, follows the partition-based subgraph learning pathway, which is also backed by causal theory.

Obs 2: Our *RayE-Sub* consistently outperform conventional backbone models on all datasets. Compared with conventional backbone methods, our approach achieves significant improvements across all datasets with a maximum performance margin of 9%. This improvement empirically verifies that subgraph learning can effectively boost the expressive power

TABLE III
THE p-VALUE OF SIGNIFICANCE TEST. THE METHOD IN PARENTHESES INDICATES THE BASELINE THAT IS SUB-OPTIMAL IN THE GIVEN DATASET.

	BBBP(2-DRFWL(2))	BA-2Motifs (MAG-GNN)	BA-2Motifs (MICRO-Graph)
p-value	0.046	0.044	0.035

TABLE IV
THE EXECUTION TIME OF FOUR BASELINES ON THREE DATASETS (S/EPOCH).

	BBBP	MUTAG	BA-2Motifs
GIN	0.79	1.07	0.64
PNA	1.30	2.62	0.91
GNN-AK	2.76	4.32	1.23
GSAT	3.88	6.49	2.98
RayE-Sub	4.06	9.86	2.65

of graph learning. Actually, this viewpoint is also supported by the remarkable performance of existing subgraph learning methods, such as GNN-AK, IB-subgraph, GSAT. Moreover, we note that GIN stands out among backbone models as it achieves the best performance among these methods. This observation aligns with both theoretical proofs and empirical analysis. Therefore, for the implementation of ***RayE-Sub***, we choose GIN as the backbone.

Obs 3: Compared with existing subgraph learning methods, *RayE-Sub* achieves competitive results in both real and synthetic datasets. Encouragingly, ***RayE-Sub*** obtains the SOTA on five datasets. Specifically, our ***RayE-Sub*** outperforms best baselines by 3.2% and 1.1% respectively on BBBP and BA-2Motifs. Such performance superiority can be explicitly attributed to the coupling effects of both two objectives, i.e., *Rayleigh-resistance* based subgraph partition and *Siamese-Query* based equivariant subgraph learning. We also note that GSAT obtains several suboptimal results. Actually, both ***RayE-Sub*** and GSAT are partition-based subgraph learning methods. Supported by the visual analysis presented below, we argue that this phenomenon stems from the inherent inductive bias in the real-world prediction tasks. Specifically, it suggests that a consistent causal substructure can always be extracted through partition of the original graph. This further validates the effectiveness of our ***RayE-Sub***.

E. Quantitative Analysis

We further evaluate whether RayE-Sub significantly outperforms the second-best node-based baselines on the BBBP and BA-2Motifs, and the second-best partition-based baselines on BA-2Motifs. The Null Hypothesis is that Seman does not show a significant performance improvement. As shown in Tab. III, our method significantly outperforms the second-best method ($p < 0.05$) on above datasets.

F. Visual Analysis

We provide visualizations of the boundary discovered by ***RayE-Sub*** on two datasets (MUTAG and BA-2Motifs) as shown in Figure 4. According to the results of visualization, we explore (i) whether our method can partition original graph into label-relevant subgraphs, (ii) whether the setup of $\beta = 0.05$ matches the real scenario of the datasets.

Following [13], $-\mathrm{NO_2}$ and $-\mathrm{NH_2}$ in MUTAG dataset are labeled as ground-truth explanations. In our practice, we observe that our ***RayE-Sub*** always partition $-\mathrm{NO_2}$ or $-\mathrm{NH_2}$ into a subgraph as shown in Figure 4. Excitingly, our partition tends to divide the molecule into a bag of functional groups, which indicates that our method is suitable for real-world tasks. Besides, we find the number of label-relevant subgraphs in each graph is small, resulting in only a few edge partition (drop) can extract them. Therefore, the setting of $\beta = 0.05$ is appropriate for real-world tasks.

Following [43], house motifs and cycle motifs give class labels and thus are regarded as ground-truth explanations for the two classes respectively. Specifically, each graph is generated using the BA model and will be attached with two house motifs or three house motifs randomly. The number of house motifs represents the graph class. The primary goal of this task is to identify house motifs. We observe that our ***RayE-Sub*** always partition house motif (pink nodes) into a subgraph. Similar to MUTAG, BA-2Motifs also needs only a few number of boundary edges to achieve accurate prediction.

In conclusion, our ***RayE-Sub*** can partition original graph into label-relevant subgraphs, and the setup of $\beta = 0.05$ is appropriate for the tasks of these datasets.

G. Efficiency Analysis

The above analysis has confirmed that ***RayE-Sub*** maintains a reasonable time complexity, and we further investigate its efficiency in the actual training process. We compare the execution time (s/epoch) of four baselines (GIN, PNA, GNN-AK and GSAT) on three datasets (BBBP, MUTAG and BA-2Motifs) as shown in Table IV. Note that all experiments are conducted on a Tesla V100-PCIE-16GB GPU, and The backbones of all models are 2-layer GIN, same as ***RayE-Sub***. The result shows that the running efficiency of ***RayE-Sub*** is competitive, and it achieves interpretability and performance improvements within an acceptable time consumption.

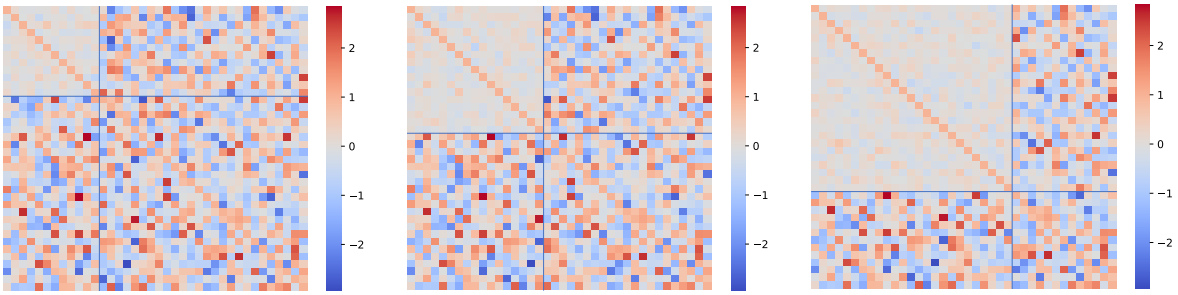
H. Sensitivity Analysis

We conduct sensitivity analysis of model performance about β, the results are in Figure 6(d). Our model achieves best results when β is within $[0.05, 0.1]$. In other words, about 5%-10% of the edges in two molecular datasets are with fragile connections, and their role is simply to bond different functional groups. Therefore, we set $\beta = 0.05$ on all datasets.

I. Ablation Studies

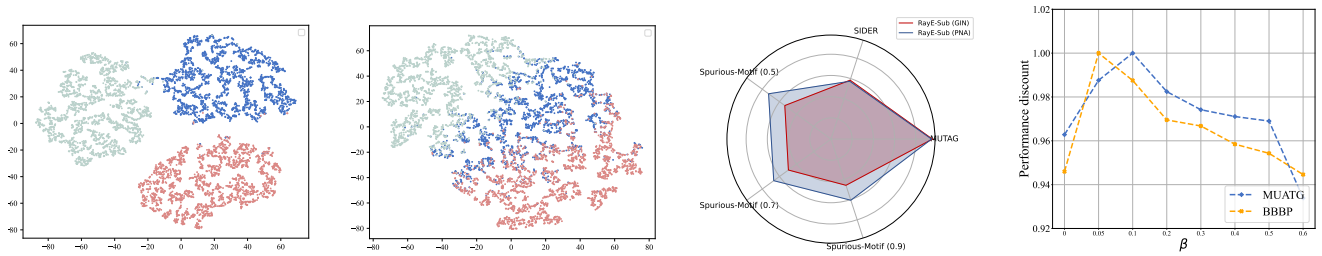
Our ablation studies are conducted from the following three perspectives: (i) The superiority of *Rayleish-resistance* metric in subgraph discovery. (ii) The effectiveness of *Siamese-Query* architecture for graph prediction. (iii) The impacts of different backbones on performances of ***RayE-Sub***.

Our primary focus lies in investigating whether *Rayleigh-resistance* can yield more interpretable subgraphs. We designed



(a) The ground-truth causal invariant subgraph. (b) Extracted subgraph by RayE-Sub with *Rayleigh-resistance*. (c) Extracted subgraph by the variant of RayE-Sub.

Fig. 5. Comparison of subgraph extracted by different methods with ground-truth causal invariant subgraph, where the light-colored region in the upper left corner represents the extracted causal invariant subgraph.



(a) The ***RayE-Sub*** with *Siamese-Query* module. (b) The variant of ***RayE-Sub*** without *Siamese-Query* module. (c) The performance of ***RayE-Sub*** with different backbones. (d) Sensitivity analysis of β on MUTAG and BBBP.

Fig. 6. Ablation studies and sensitivity analysis. Left panel: the rationale distributions of the inspection points from Spurious-Motif(0.33), which are visualized by t-SNE [52], and the backbone sensitivity analysis. Right panel: sensitivity analysis of β on MUTAG and BBBP, where $y\text{-}axis$ represents the discount of each setting to the best performance.

a variant of ***RayE-Sub***, using the hop-based distance instead of *Rayleigh-resistance*. Figure 5 shows the comparison of subgraph extracted by different methods with ground-truth causal invariant subgraph on Spurious-Motif dataset. We can observe that ***RayE-Sub*** can extract accurate invariant subgraph. Although the subgraph extracted by the variant of ***RayE-Sub*** can cover invariant subgraph, its precision is far less than that of ***RayE-Sub***. Therefore, we can intuitively verify the validity of *Rayleigh-resistance*. In addition, we also provide visualizations of the subgraphs discovered by *Rayleigh-resistance* to verify its superiority in Section VI-F.

Next, we explore the effectiveness of *Siamese-Query*. As shown in Figure 6, we employ t-SNE to visualize 6000 representations of samples from the Spurious-Motif dataset [52]. The variant of ***RayE-Sub*** is implemented by using a backbone GNN instead of *Siamese-Query*. We can observe that the representation space obtained by ***RayE-Sub*** is better discriminative than the variant. Actually, subgraph-level equivariant design is crucial for the prediction performance. The effectiveness of *Siamese-Query* is demonstrated.

Our final ablation study is designed to explore the effects of ***RayE-Sub*** with different backbones. Specifically, we compare the performances of ***RayE-Sub*** when GIN and PNA are utilized as base encoder respectively. As shown in Figure 6(c), we observe that PNA-based architecture has strong expression ability on synthetic datasets. We conclude that the design of PNA using multiple aggregators can better exploit common sub-units of graphs, which matches the generation principles of synthetic datasets [26]. GIN has an advantage on more complex tasks in real-world datasets.

VII. Conclusion and Future Work

In this paper, we conduct a systematic investigation of two mainstream subgraph learning approaches from the unified perspective of *perfect reconstruction* and propose a novel architecture ***RayE-Sub***. Our motivation stems from a significant observation that existing subgraph learning methods commonly encounter the issue of subgraph degradation. Through our theoretical and empirical investigation, we discover that the performance degradation is primarily caused by the redundancy and inadequacy of the extracted subgraph information. As a result, we develop a unified standard to evaluate the performance of subgraph learning methods, i.e., *perfect reconstruction*. Our aim lies in proposing a novel subgraph learning framework that can effectively extract subgraphs with *perfect reconstruction* ability, addressing the existing problems associated with performance degradation. Specifically, we propose a subgraph extraction method *Rayleigh-resistance* based on spectral theory and a subgraph-level equivariant learning architecture *Siamese-Query*, both of them jointly tackle existing technical limitation to guarantee perfect reconstruction of extracted subgraphs. Experiments on both synthetic and real-world datasets demonstrate the effectiveness of ***RayE-Sub***. Moreover, theoretical analysis and practical observations profoundly guarantee the superiority of our architecture.

In future, we can improve our work from following two aspects. In the subgraphs partition stage, how to set an personalized and optimal partition rate via domain knowledge across different datasets is still unexplored. And in the subgraph aggregation stage, it is also interesting to investigate more powerful equivariant architectures for effective aggregations.

REFERENCES

[1] N. Yang, K. Zeng, Q. Wu, X. Jia, and J. Yan, "Learning substructure invariance for out-of-distribution molecular representations," *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 964–12 978, 2022.

[2] Y. Zhang, H. Gao, J. Pei, and H. Huang, "Robust self-supervised structural graph neural network for social network prediction," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 1352–1361.

[3] Z. Zhou, Y. Wang, X. Xie, L. Chen, and C. Zhu, "Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[4] S. Wu, F. Sun, W. Zhang, X. Xie, and B. Cui, "Graph neural networks in recommender systems: a survey," *ACM Computing Surveys*, vol. 55, no. 5, pp. 1–37, 2022.

[5] M. Chen, C. Huang, L. Xia, W. Wei, Y. Xu, and R. Luo, "Heterogeneous graph contrastive learning for recommendation," in *Proceedings of the sixteenth ACM international conference on web search and data mining*, 2023, pp. 544–552.

[6] A. Sharma, A. Sharma, P. Nikashina, V. Gavrilenko, A. Tselykh, A. Bozhenyuk, M. Masud, and H. Meshref, "A graph neural network (gnn)-based approach for real-time estimation of traffic speed in sustainable smart cities," *Sustainability*, vol. 15, no. 15, p. 11893, 2023.

[7] K.-H. N. Bui, J. Cho, and H. Yi, "Spatial-temporal graph neural network for traffic forecasting: An overview and open research issues," *Applied Intelligence*, vol. 52, no. 3, pp. 2763–2774, 2022.

[8] Q. Wang, K. Huang, P. Chandak, M. Zitnik, and N. Gehlenborg, "Extending the nested model for user-centric xai: A design study on gnn-based drug repurposing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 1, pp. 1266–1276, 2022.

[9] R. Sun, H. Dai, and A. W. Yu, "Does gnn pretraining help molecular representation?" *Advances in Neural Information Processing Systems*, vol. 35, pp. 12 096–12 109, 2022.

[10] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[11] C. Morris, M. Ritzert, M. Fey, W. L. Hamilton, J. E. Lenssen, G. Rattan, and M. Grohe, "Weisfeiler and leman go neural: Higher-order graph neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 4602–4609.

[12] B. Zhang, S. Luo, L. Wang, and D. He, "Rethinking the expressive power of gnns via graph biconnectivity," *arXiv preprint arXiv:2301.09505*, 2023.

[13] S. Miao, M. Liu, and P. Li, "Interpretable and generalizable graph learning via stochastic attention mechanism," in *International Conference on Machine Learning*. PMLR, 2022, pp. 15 524–15 543.

[14] L. Cotta, C. Morris, and B. Ribeiro, "Reconstruction for powerful graph representations," *Advances in Neural Information Processing Systems*, vol. 34, pp. 1713–1726, 2021.

[15] F. Frasca, B. Bevilacqua, M. Bronstein, and H. Maron, "Understanding and extending subgraph gnns by rethinking their symmetries," *Advances in Neural Information Processing Systems*, vol. 35, pp. 31 376–31 390, 2022.

[16] B. Bevilacqua, F. Frasca, D. Lim, B. Srinivasan, C. Cai, G. Balamurugan, M. M. Bronstein, and H. Maron, "Equivariant subgraph aggregation networks," in *International Conference on Learning Representations*, 2022.

[17] L. Zhao, W. Jin, L. Akoglu, and N. Shah, "From stars to subgraphs: Uplifting any gnn with local structure awareness," *arXiv preprint arXiv:2110.03753*, 2021.

[18] D. Kreuzer, D. Beaini, W. Hamilton, V. Létourneau, and P. Tossou, "Rethinking graph transformers with spectral attention," *Advances in Neural Information Processing Systems*, vol. 34, pp. 21 618–21 629, 2021.

[19] J. Feng, Y. Chen, F. Li, A. Sarkar, and M. Zhang, "How powerful are k-hop message passing graph neural networks," *Advances in Neural Information Processing Systems*, vol. 35, pp. 4776–4790, 2022.

[20] G. Bouritsas, F. Frasca, S. Zafeiriou, and M. M. Bronstein, "Improving graph neural network expressivity via subgraph isomorphism counting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 657–668, 2022.

[21] G. Preti, G. De Francisci Morales, and M. Riondato, "Maniacs: Approximate mining of frequent subgraph patterns through sampling," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 3, pp. 1–29, 2023.

[22] Y. Chen, Y. Zhang, Y. Bian, H. Yang, M. Kaili, B. Xie, T. Liu, B. Han, and J. Cheng, "Learning causally invariant representations for out-of-distribution generalization on graphs," *Advances in Neural Information Processing Systems*, vol. 35, pp. 22 131–22 148, 2022.

[23] W. Jin, R. Barzilay, and T. Jaakkola, "Hierarchical generation of molecular graphs using structural motifs," in *International conference on machine learning*. PMLR, 2020, pp. 4839–4848.

[24] D. Spielman, "Spectral and algebraic graph theory," *Yale lecture notes, draft of December*, vol. 4, p. 47, 2019.

[25] J. Yu, T. Xu, Y. Rong, Y. Bian, J. Huang, and R. He, "Graph information bottleneck for subgraph recognition," *arXiv preprint arXiv:2010.05563*, 2020.

[26] Y.-X. Wu, X. Wang, A. Zhang, X. He, and T.-S. Chua, "Discovering invariant rationales for graph neural networks," *arXiv preprint arXiv:2201.12872*, 2022.

[27] H. Li, Z. Zhang, X. Wang, and W. Zhu, "Learning invariant graph representations for out-of-distribution generalization," *Advances in Neural Information Processing Systems*, vol. 35, pp. 11 828–11 841, 2022.

[28] S. Gui, M. Liu, X. Li, Y. Luo, and S. Ji, "Joint learning of label and environment causal independence for graph out-of-distribution generalization," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[29] K. Yang, Z. Zhou, W. Sun, P. Wang, X. Wang, and Y. Wang, "Extract and refine: Finding a support subgraph set for graph representation," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 2953–2964.

[30] C. Bodnar, F. Frasca, Y. Wang, N. Otter, G. F. Montufar, P. Lio, and M. Bronstein, "Weisfeiler and lehman go topological: Message passing simplicial networks," in *International Conference on Machine Learning*. PMLR, 2021, pp. 1026–1037.

[31] C. Morris, G. Rattan, and P. Mutzel, "Weisfeiler and leman go sparse: Towards scalable higher-order graph embeddings," *Advances in Neural Information Processing Systems*, vol. 33, pp. 21 824–21 840, 2020.

[32] Z. Chen, L. Chen, S. Villar, and J. Bruna, "Can graph neural networks count substructures?" *Advances in neural information processing systems*, vol. 33, pp. 10 383–10 395, 2020.

[33] O. Puny, H. Ben-Hamu, and Y. Lipman, "Global attention improves graph networks generalization," *arXiv preprint arXiv:2006.07846*, 2020.

[34] V. P. Dwivedi, A. T. Luu, T. Laurent, Y. Bengio, and X. Bresson, "Graph neural networks with learnable structural and positional representations," *arXiv preprint arXiv:2110.07875*, 2021.

[35] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu, "Do transformers really perform badly for graph representation?" *Advances in Neural Information Processing Systems*, vol. 34, pp. 28 877–28 888, 2021.

[36] J. Sun, L. Zhang, G. Chen, P. Xu, K. Zhang, and Y. Yang, "Feature expansion for graph neural networks," in *International Conference on Machine Learning*. PMLR, 2023, pp. 33 156–33 176.

[37] E. Alsentzer, S. Finlayson, M. Li, and M. Zitnik, "Subgraph neural networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 8017–8029, 2020.

[38] H. Yuan, H. Yu, J. Wang, K. Li, and S. Ji, "On explainability of graph neural networks via subgraph explorations," in *International Conference on Machine Learning*. PMLR, 2021, pp. 12 241–12 252.

[39] P. Li, Y. Wang, H. Wang, and J. Leskovec, "Distance encoding: Design provably more powerful neural networks for graph representation learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4465–4478, 2020.

[40] T. Wu, H. Ren, P. Li, and J. Leskovec, "Graph information bottleneck," *Advances in Neural Information Processing Systems*, vol. 33, pp. 20 437–20 448, 2020.

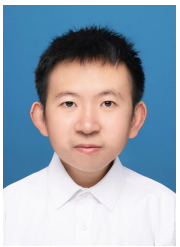
[41] W. Xiao and I. Gutman, "Resistance distance and laplacian spectrum," *Theoretical chemistry accounts*, vol. 110, pp. 284–289, 2003.

[42] B. Zhang, G. Feng, Y. Du, D. He, and L. Wang, "A complete expressiveness hierarchy for subgraph gnns via subgraph weisfeiler-lehman tests," in *International Conference on Machine Learning*. PMLR, 2023, pp. 41 019–41 077.

[43] D. Luo, W. Cheng, D. Xu, W. Yu, B. Zong, H. Chen, and X. Zhang, "Parameterized explainer for graph neural network," *Advances in neural information processing systems*, vol. 33, pp. 19 620–19 631, 2020.
[44] A. K. Debnath, R. L. Lopez de Compadre, G. Debnath, A. J. Shusterman, and C. Hansch, "Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity," *Journal of medicinal chemistry*, vol. 34, no. 2, pp. 786–797, 1991.
[45] W. Hu, M. Fey, M. Zitnik, Y. Dong, H. Ren, B. Liu, M. Catasta, and J. Leskovec, "Open graph benchmark: Datasets for machine learning on graphs," *Advances in neural information processing systems*, vol. 33, pp. 22 118–22 133, 2020.
[46] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
[47] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.
[48] L. Kong, J. Feng, H. Liu, D. Tao, Y. Chen, and M. Zhang, "Mag-gnn: Reinforcement learning boosted graph neural network," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[49] J. Zhou, J. Feng, X. Wang, and M. Zhang, "Distance-restricted folklore weisfeiler-leman gnns with provable cycle counting power," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[50] C. Zhou, X. Wang, and M. Zhang, "Facilitating graph neural networks with random walk on simplicial complexes," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
[51] S. Zhang, Z. Hu, A. Subramonian, and Y. Sun, "Motif-driven contrastive learning of graph representations," *IEEE Transactions on Knowledge and Data Engineering*, 2024.
[52] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

VIII. Biography Section

Kuo Yang received his BE degree from Northeastern University at Qinhuangdao, China, in 2021. He is now working toward a PhD degree at School of Data Science, USTC. He has published many high-quality papers on top conferences and journals such as KDD, ICLR, and TMC. His mainly research interests are graph representation learning, spatiotemporal data mining and especially subgraph-driven spatiotemporal graph learning.



Zhengyang Zhou is now an associate researcher at Suzhou Institute for Advanced Research, University of Science and Technology of China (USTC). He got his Ph.D. degree at University of Science and Technology of China in 2023. He has published over 30 papers on top conferences and journals such as NeurIPS, ICLR, KDD, TKDE, WWW, AAAI and ICDE. His mainly research interests include spatiotemporal data minining, human-centered urban computing, and deep learning generalization with model behavior analysis. He is now especially interested in improving generalization capacity of neural networks for streaming and spatiotemporal data.

Xu Wang is now an associate researcher at Suzhou Institute for Advanced Research, University of Science and Technology of china (USTC). He got his Ph.D. degree at University of Science and Technology of China in 2023. He has published over 15 papers on top conferences and journals such as NeurIPS, ICLR, KDD, TKDE, WWW and AAAI. His mainly research interests include spatiotemporal data minining, transfer learning and AI for chemistry.

Pengkun Wang (Member, IEEE) is now an associate researcher at Suzhou Institute for Advanced Research, University of Science and Technology of China (USTC). He got his Ph.D. degree at University of Science and Technology of China in 2023. He has published over 30 papers on top conferences and journals such as IEEE TKDE, AAAI, ICLR, KDD, and WWW. His mainly research interests include imbalanced/long-tailed learning, interpretable data augmentation, multi-objective optimization, and open environment data mining. He also works on applying generalized deep learning to real-world tasks such as urban computing and AI4Science.

Limin Li is now a PhD candidate at School of Software Engineering, University of Science and Technology of China (USTC). He received his BE degree from Northeastern University at Qinhuangdao, China, in 2007 and got his master's degree University of Science and Technology of China in 2011. His research interests are mainly focused on molecular data mining and AI for Science. He also works on applying generalized deep learning to real-world tasks such as urban computing and AI4Science.

Yang Wang is now an associate professor at School of Computer Science and Technology, School of Software Engineering, and School of Data Science in USTC. He got his Ph.D. degree at University of Science and Technology of China in 2007. Since then, he keeps working at USTC till now as a postdoc and an associate professor successively. Meanwhile, he also serves as the vice dean of school of software engineering of USTC. His research interest mainly includes wireless (sensor) networks, distribute systems, data mining, and machine learning, and he is also interested in all kinds of applications of AI and data mining technologies especially in urban computing and AI4Science.