A Spectral Representation of Networks: The Path of Subgraphs

Shengmin Jin
Data Lab, EECS Department
Syracuse University
shengmin@data.syr.edu

Hao Tian
Data Lab, EECS Department
Syracuse University
haotian@data.syr.edu

Jiayu Li
Data Lab, EECS Department
Syracuse University
jli221@data.syr.edu

Reza Zafarani
Data Lab, EECS Department
Syracuse University
reza@data.syr.edu

ABSTRACT

Network representation learning has played a critical role in studying networks. One way to study a graph is to focus on its spectrum, i.e., the eigenvalue distribution of its associated matrices. Recent advancements in spectral graph theory show that spectral moments of a network can be used to capture the network structure and various graph properties. However, sometimes networks with different structures or sizes can have the same or similar spectral moments, not to mention the existence of the cospectral graphs. To address such problems, we propose a 3D network representation that relies on the spectral information of subgraphs: the *Spectral Path*, a path connecting the spectral moments of the network and those of its subgraphs of different sizes. We show that the spectral path is interpretable and can capture relationship between a network and its subgraphs, for which we present a theoretical foundation. We demonstrate the effectiveness of the spectral path in applications such as network visualization and network identification.

CCS CONCEPTS

• **Mathematics of computing** → **Spectra of graphs**; • **Human-centered computing** → *Visualization techniques.*

KEYWORDS

Network Representations, Spectral Graphs

ACM Reference Format:
Shengmin Jin, Hao Tian, Jiayu Li, and Reza Zafarani. 2022. A Spectral Representation of Networks: The Path of Subgraphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), August 14–18, 2022, Washington, DC, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3534678.3539433

1 INTRODUCTION

Ideally, a network representation should be both informative for machine learning and interpretable for users. Therefore, spectral graph theory has been widely used to study a graph as it connects the structure of a network to the eigenvalues and eigenvectors of its associated matrices, e.g., the adjacency matrix or the Laplacian. Spectral methods have classically focused on the extreme eigenvalues and associated eigenvectors. Well-known examples include the Cheeger's inequality, which relates the second-smallest eigenvalue of a graph Laplacian to graph connectivity [5], and the inverse of the largest eigenvalue of the adjacency matrix (*epidemic threshold*), which determines whether a virus can spread or die out in a network [4]. Recently, instead of the extreme eigenvalues, the overall distribution of eigenvalues, also known as the *spectral density*, has received more attention [8, 17]. To clearly represent a network with its spectral density, one can rely on its *spectral point* [10]. The spectral point of a network represents it with the low-order spectral moments as moments are often used to capture the shape of a distribution in statistics. Specifically, spectral moments of the *random walk transition matrix* can be used, as these moments (1) have a very clear meaning, which is the expected return probability of a random walk; and (2) have theoretical connections to the network structure and various network properties such as the degree distribution and clustering coefficient [10]. However, there are a few drawbacks in using the spectral density (or spectral moments) as a representation of a graph: (1) in theory, different graphs can have the same spectral density. It is not difficult to construct such graphs. One way is to make a copy of a graph and take its union with the original graph. In this way, one doubles the size of the graph, but the spectral density does not change. Moreover, there exist non-isomorphic graphs sharing the same graph spectrum (i.e. *cospectral* or *isospectral* graphs) [19]; (2) the spectrum may dramatically change with a small change in graph structure [23].

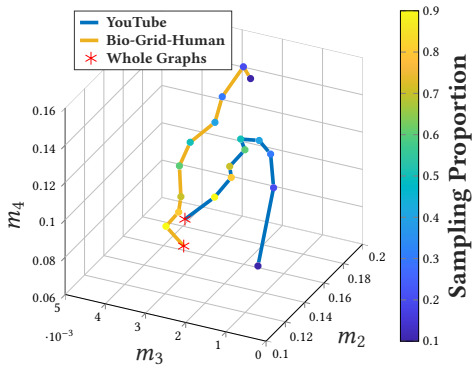


Figure 1: Spectral Paths of YouTube and Bio-Grid-Human

KDD '22, August 14–18, 2022, Washington, DC, USA

ACM ISBN 978-1-4503-9385-0/22/08…$15.00
https://doi.org/10.1145/3534678.3539433

To address these problems, we propose utilizing the spectral information of subgraphs. In this work, we propose to represent a graph using a 3D path in the spectral embedding space, which we denote as the *Spectral Path*. The spectral path connects the spectral moments of a network and its subgraphs. Figure 1 plots the spectral paths of a social network (YouTube) and a biological network. As we can see, though their spectral points of the whole networks (marked with *) are close, their spectral paths show different patterns.

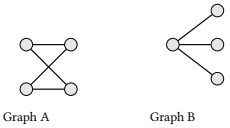


Figure 2: Cospectral Graphs

Our idea is inspired by the *Reconstruction Conjecture*, which is an open problem in theoretical computer science. The reconstruction conjecture — initially due to Kelly [11] and Ulam [11, 16, 21] — states that graphs are determined uniquely by their subgraphs. Bollobás has shown that the probability that a randomly chosen graph on n vertices is not reconstructible goes to 0 as $n \to \infty$ [1]. In other words, almost all graphs are reconstructible with their subgraphs. Naturally, by using the spectral density of subgraphs to capture subgraph information, one can better represent the whole graph. Figure 2 provides an example of two graphs that share the same spectral density but their subgraphs do not necessarily do so. In Figure 2, graphs G_A and G_B have the same graph spectrum of the random walk transition matrix: $[-1, 0, 0, 1]$. Hence, they also have the same spectral moments; for example, the second moment m_2 of their spectrum is $m_{2,A} = m_{2,B} = \frac{1}{2}$. If we randomly remove one node (and edges connected to it) from G_A, we get some subgraph $G_{A'}$. Subgraph $G_{A'}$ is 100% likely to be isomorphic to  , whose spectrum is $[-1, 0, 1]$ and its second spectral moment is $\frac{2}{3}$. However, if we randomly remove one node from G_B and get a subgraph $G_{B'}$, with 75% probability, we get  , and with 25% probability, we get an empty graph of 3 nodes whose spectrum is $[0, 0, 0]$ and its second spectral moment is 0. Therefore, though G_A and G_B are cospectral graphs, they have a different distribution of (second) spectral moments across all the subgraphs by randomly removing one node. If we take the expectation of the second spectral moment of $G_{A'}$ (or $G_{B'}$) over the distribution of the subgraphs, we get $\mathbb{E}(m_{2,A'}) = \frac{2}{3} \times 100\% = \frac{2}{3}$, but $\mathbb{E}(m_{2,B'}) = \frac{2}{3} \times 75\% + 0 \times 25\% = \frac{1}{2}$. In expectation, if we randomly remove one node, the second spectral moment of G_A will increase by $\frac{1}{6}$ but the second spectral moment of G_B will not change. In Section 6, we will provide a detailed theoretical analysis. These observations indicate that even if two networks have similar spectral density, one can capture their difference in terms of substructures by using the expected spectral points (moments) of their subgraphs.

Overall, our contributions are mainly the following:

1. Spectral Path. We propose representing a network using its *Spectral Path*: a path connecting the spectral moments of the network and its subgraphs. Spectral paths provide an *interpretable-by-design* network representation.

2. Interpretability of Spectral Path. We study the interpretability of spectral paths by investigating the *shapes* of spectral paths. We provide the theoretical relationship between the spectral moments of a network and those of its subgraphs. We show how this relationship is closely related to the network structure. To the best of our knowledge, this work is the first to explore spectral moments of subgraphs and to study the relationship between spectral moments of subgraphs and those of the whole network.

3. Spectral Path Applications. We show that spectral path can be used for applications such as network visualization and network identification, i.e., identifying the source of an anonymized graph.

4. Spectral Paths of Cospectral Graphs. We theoretically explore the possibility of using the expected spectral moments of subgraphs to help distinguish cospectral graphs.

Organization. In Section 2, we detail the preliminaries including the random walk transition matrix and its spectral moments. In Section 3, we present the algorithm to build a spectral path. In Section 4, we demonstrate the theoretical interpretation of spectral paths. We also present the relationship between the spectral moments of a network and its subgraphs, and how spectral paths are connected to the network structure. In Section 5, we show that spectral paths can be used for network visualization and network identification (that is, answering questions such as "Is this anonymized graph sampled from Twitter?"). Section 6 demonstrates how spectral paths provide a potential way to distinguish cospectral graphs. We review additional related work in Section 7 and conclude in Section 8.

2 PRELIMINARIES AND NOTATION

Random Walk Transition Matrix and Normalized Laplacian. For an undirected graph $G = (V, E)$ with vertices $V = \{v_1, v_2, ..., v_n\}$ and edges $E \subseteq V \times V$, its adjacency matrix $A \in \mathbb{R}^{n \times n}$ has $A_{ij} = 1$ if $(i, j) \in E$ and otherwise, $A_{ij} = 0$. The degree matrix $D \in \mathbb{R}^{n \times n}$ is a diagonal matrix with node degrees on its diagonal, i.e. $D_{ii} = \sum_{j=1}^{n} A_{ij}$. The transition matrix of the random walk on G is matrix $P = AD^{-1}$. The spectrum of a matrix is the set of its eigenvalues. As P is a stochastic matrix, it has a bounded spectrum: $1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n \geq -1$, where λ_i's are the eigenvalues of P. The normalized Laplacian of G is the matrix $L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. The spectrum of a matrix is the set of its eigenvalues. The spectrum of the normalized Laplacian is also bounded, i.e. $0 = \mu_1 \leq \mu_2 \leq \cdots \leq \mu_{n-1} \leq \mu_n \leq 2$, where μ_i's are the eigenvalues of L. As P is *similar* to $D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ (i.e., they have the same eigenvalues), it is easy to find the relationship between the eigenvalues of P and L: $\lambda_i = 1 - \mu_i$, for $1 \leq i \leq n$.

Spectral Moments and Spectral Points. We denote the ℓ-th spectral moment m_ℓ of a graph G using the spectrum of its random walk transition matrix P, $m_\ell = \mathbb{E}(\lambda^\ell)$, as $\frac{1}{n} \sum_{i=1}^{n} {\lambda_i}^\ell = \mathbb{E}(\lambda^\ell)$. Note that $m_1 = 0$. To allow interpretable visualization, one can use a 3D *spectral point* to represent a network, i.e., representing a network using its second, third, and forth spectral moments (m_2, m_3, m_4).

3 SPECTRAL PATH

Here, we introduce the proposed network representation: the *spectral path* of a graph, which represents a network with the path connecting the spectral moments of a network and its subgraphs. The following simple steps can help build a spectral path:

Step 1: **Sample** many subgraphs from the network
Step 2: **Estimate** the expected spectral points using the spectral points of samples
Step 3: **Form** a spectral path by connecting the expected spectral points

The pseudocode is in Algorithm 1. The algorithm uses *Random Node Sampling* [12] to sample subgraphs from the network by (1) varying the proportion of nodes from 0% to 100% with step size s and (2) taking t independent samples for each proportion. We use Random Node Sampling as the subgraph can be viewed as a result of randomly removing nodes from a graph. For each sample and the whole network, the algorithm computes its spectral point (m_2, m_3, m_4). For all samples of the same size (sampling proportion p), the algorithm takes the average of their spectral points to estimate the expected spectral points. Hence, we get one (expected) spectral point for each sampling proportion p. Finally, it draws a path connecting the expected spectral points from $100\% \to \cdots \to 2p\% \to p\%$. Figure 3 illustrates the spectral path of YouTube with the spectral points of the samples. The figure shows that the spectral path can capture structural variations in subgraphs of different sizes.

Time Complexity. The majority of the computation time is dedicated to sampling subgraphs and computing three spectral moments for each subgraph. For one subgraph, random node sampling takes $O(n+m)$ where $|V| = n$ and $|E| = m$. For large graphs, we compute accurate estimates of the low-order moments with the ApproxSpectralMoment algorithm [7]. The algorithm estimates the moments by simulating many random walks and computes the proportion of closed walks. To compute the ℓ-th spectral moment by simulating r random walks, it takes $O(r\ell)$ time. In our case, $\ell \leq 4$ and we set $r = 10,000$ following the empirical results of [7]. As the random walks can be taken in parallel, it only takes less than a few seconds to compute the three spectral moments even for large networks [7, 10]. Hence, for each subgraph, the time complexity is $O((n+m)r\ell)$. We have a total of $\frac{100}{s} \times t + 1$ graphs (a network and its subgraphs) for which we compute spectral points. Thus, the time complexity for computing a spectral path is $O(\frac{tr\ell}{s}(n+m))$, linear in the number of nodes and edges.

4 INTERPRETABILITY OF SPECTRAL PATHS

The interpretability of spectral path can be studied from two aspects (1) *the location of the spectral path in the 3D embedding space.* Here, we do not focus on this aspect. As mentioned, each component (dimension) of the spectral points is closely related to the network structure and network properties (see examples in [10]);

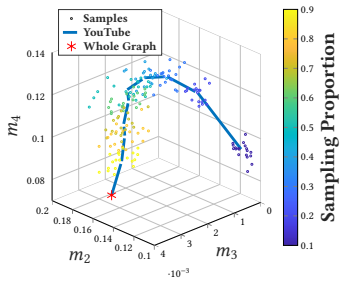


Figure 3: Spectral Path of YouTube and its sample points

Algorithm 1: Spectral Path algorithm

input : an undirected network graph: $G(V, E)$
output : the Spectral Path of G: SP_G
parameter : s : sampling proportion step size;
t : number of samples for one proportion;

```
Expected_spectral_points = { };
for ( p = s; p < 100%; p = p + s ) {
    Spectral_points = { }; %Spectral points for all the
     samples for proportion p
    for ( i = 1; i ≤ t; i = i + 1 ) {
        %Sample a p% subgraph G_p from G
        G_p = RandomNodeSampling(G, p);
        %Compute the spectral point of G_p (m2, m3, m4)
        Spectral_point = ComputeMoments(G_P);
        Spectral_points.add(Spectral_point);
    }
    %Compute the average m2, m3, m4 for all the samples
     for proportion p
    Expected_spectral_point = Average(Spectral_points);
    Expected_spectral_points.add(Expected_spectral_point);
}
%Compute the spectral point of G
Spectral_point = ComputeMoments(G);
Expected_spectral_points.add(Spectral_point);
%Form the spectral path of the (expected) spectral
 points, e.g., 100→90%. . . →10%
SP_G = Form_Path(Expected_spectral_points)
return SP_G;
```

(2) *the shape of the spectral paths*, which we will focus on here. The spectral points of different sampling sizes in the embedding space will determine the shape of a spectral path. Hence, we study the interpretability of spectral paths by investigating the relationship between the spectral points of subgraphs and that of the whole network. We start with the following questions:

$\mathcal{I}_1$) **Direction of the movement.** In which direction will the spectral point of a graph move if its nodes are randomly removed (equivalent to sampling a subgraph with random node sampling)? In other words, will spectral moments increase, decrease, or stay the same?;

$\mathcal{I}_2$) **Magnitude of the movement.** How far will the spectral point of a graph move under sampling?; and

$\mathcal{I}_3$) **Shape of spectral paths.** How do the shapes of spectral paths reveal the structural information of a network?

To answer these three questions ($\mathcal{I}_1$ to $\mathcal{I}_3$), we need to look into the spectral moments of subgraphs.

4.1 Second Spectral Moment of Subgraphs

We first look into the second spectral moment m_2 of subgraphs. In Theorem 4.1, we show what m_2 of a graph is expected to be when one node is removed, and the detailed proof is provided in Section 4.1.2. Then, we extend the result to removing k nodes ($k \geq 1$) from a graph in Theorem 4.4.

Theorem 4.1 (Expected second spectral moment m_2 of subgraphs after removing one node). *In undirected graph $G = (V, E)$, where $|V_G| = n, |E_G| = m$, subgraph G' of G is obtained by removing one node from G uniformly at random. The expected second*

moment of G' is

$$\mathbb{E}(m_{2,G'}) = m_{2,G} + \frac{2}{n(n-1)} \cdot \Big(\sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G \\ d_i=1, d_j=1}} \frac{1}{d_i d_j} + \delta_{triad} \Big),$$

where d_i, d_j denote the degree of node i and j, and

$$\delta_{triad} = \sum_{\substack{(i,j,k) \\ \text{is a triad in } G}} \Big(\frac{1}{d_i d_j (d_i-1)(d_j-1)} + \frac{1}{d_i d_k (d_i-1)(d_k-1)} + \frac{1}{d_j d_k (d_j-1)(d_k-1)} \Big).$$

Especially, if G is a triangle-free graph, then $\mathbb{E}(m_{2,G'})$ reduces to:

$$\mathbb{E}(m_{2,G'}) = m_{2,G} + \frac{2}{n(n-1)} \cdot \Big(\sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G \\ d_i=1, d_j=1}} \frac{1}{d_i d_j} \Big)$$

In Figure 4, we provide an example of a triangle-free graph to illustrate Theorem 4.1. We have a graph G with 7 nodes and 6 edges. The label on each edge is its value of $\frac{1}{d_i d_j}$. There are 7 subgraphs $(G_1, \ldots, G_7)$ by removing one node from G. Using Theorem 4.1, we get $\mathbb{E}(m_{2,G'}) = m_{2,G} + \frac{2}{n(n-1)} \cdot \Big(\sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G \\ d_i=1, d_j=1}} \frac{1}{d_i d_j} \Big) = \frac{13}{21} + \frac{2}{7\times 6}\big(\frac{1}{6} + \frac{1}{6} + \frac{1}{4} + \frac{1}{4} - 1\big) = \frac{11}{18}$, which is by definition equivalent to $\mathbb{E}(m_{2,G'}) = \frac{\sum_{i=1}^{7} m_{2,G_i}}{7} = \big[\frac{2}{3} + \frac{3}{4} + \frac{2}{3} + \frac{3}{4} + \frac{2}{3} + \frac{7}{18} + \frac{7}{18}\big]/7 = \frac{11}{18}$.

For an example of a graph with triangles, consider a complete graph $K_n (n > 3)$ as G. We get n subgraphs $(G_1, \ldots, G_n)$ by removing one node, each also being a complete graph K_{n-1}. We know that $m_{2,G} = \frac{1}{n-1}$, and using Theorem 4.1, we get $\mathbb{E}(m_{2,G'}) = m_{2,G} + \frac{2}{n(n-1)} \cdot \Big(\sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G \\ d_i=1, d_j=1}} \frac{1}{d_i d_j} + \delta_{triad} \Big) = \frac{1}{n-1} + \frac{2}{n(n-1)} \cdot \Big(\binom{n}{2} \frac{1}{(n-1)^2} + \binom{n}{3} \frac{3}{(n-1)^2(n-2)^2} \Big) = \frac{1}{n-2}$, which is the m_2 of K_{n-1}.

4.1.1 Interpretations of Theorem 4.1. We provide interpretations for Theorem 4.1 by answering the aforementioned questions $\mathcal{I}_1$ to $\mathcal{I}_3$. In the proof of Theorem 4.1 (see Section 4.1.2), we partition any edge (i, j) of G into three types based on its end-points i and j's node degrees d_i and d_j:

- **Type I**: $d_i = 1, d_j = 1$;
- **Type II**: $d_i > 1, d_j = 1$;
- **Type III**: $d_i > 1, d_j > 1$.

Theorem 4.1 shows the expected m_2 of subgraphs when removing one node is closely related to frequencies of these three types of edges. Next, we will take triangle-free graphs as an example to show the interpretability of spectral paths, in terms of m_2. By removing one node from a triangle-free graph G, we get a subgraph G'. Compared to the second spectral moment of G, $m_{2,G'}$ is expected to add the term $\frac{2}{n(n-1)} \cdot \Big(\sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G \\ d_i=1, d_j=1}} \frac{1}{d_i d_j} \Big)$.

$\mathcal{I}_1$) Direction of the movement of spectral point.

The direction of the movement of spectral point (moments) is basically the sign of the term, which is determined by the difference between the summation of $\frac{1}{d_i d_j}$ over Type III and Type I edges. If the difference is positive (or negative), then $m_{2,G'}$ is expected to increase (or decrease). Roughly speaking, if the graph G has more Type III edges and fewer Type I edges, $m_{2,G'}$ is expected to increase.

$\mathcal{I}_2$) Magnitude of the movement of spectral point.

Similarly, how far the spectral point will move is decided by the magnitude of the term. Besides the frequency of different types of edges, the size of a graph n also has an impact. When n is larger, the magnitude of the term is smaller, indicating that the more nodes G has, the smaller the impact on m_2 when one node is removed.

$\mathcal{I}_3$) Shape of spectral paths.

Assume that we start with a graph with many Type III edges. If we remove nodes one by one from it, it is very likely that (a) the second spectral moment will increase first, as there are still many Type III edges and the summation of $\frac{1}{d_i d_j}$ over them increases when the node degrees decrease; (b) if we keep removing more nodes, Type III edges get converted to Type II or Type I, which makes the difference negative, and the second spectral moment starts to decrease. In other words, we will see a *turning point*; (c) finally, the moment will converge to 0 as more nodes are removed and the graph becomes an empty graph. In this case, the spectral path will show an *increasing-decreasing* pattern. Technically, it is possible for the turning point to happen in samples smaller than those taken to compute the spectral path. In that case, the spectral path will show an *increasing-only* pattern. Similarly, if the starting graph has mostly Type I edges, then the trend will be *decreasing-only*.

For general graphs with triangles, as theorem 4.1 shows, we need to consider an extra term δ_{triad} but in general they follow a similar pattern. These patterns will be observed in our later experiments.

4.1.2 Proof of Theorem 4.1. To prove Theorems 4.1, we need Theorem 4.2 and Lemma 4.3.

Theorem 4.2 (Theorems 3.2, 3.3, 3.5 in [10]). *The 2^{nd}, 3^{rd}, and 4^{th} spectral moments (m_2, m_3, m_4) of random walk transition matrix P are $m_2 = \mathbb{E}(\lambda^2) = \mathbb{E}(d_i)\,\mathbb{E}(\frac{1}{d_i d_j})$, $m_3 = \mathbb{E}(\lambda^3) = 2\,\mathbb{E}(\Delta_i)\,\mathbb{E}(\frac{1}{d_h d_i d_j})$, and $m_4 = \mathbb{E}(\lambda^4) = \big[\,\mathbb{E}(d_i) + 4\,\mathbb{E}\binom{d_i}{2} + 2\,\mathbb{E}(\square_i)\,\big]\,\mathbb{E}(\frac{1}{d_i d_j d_k d_l})$, where $\mathbb{E}(d_i)$ is the average degree, $d_i d_j$ follows the joint degree distribution $p(d_i, d_j)$, $\mathbb{E}(\Delta_i)$ is the average number of triads a node is in, $d_h d_i d_j$ follows the joint degree distribution of triads $p(d_h, d_i, d_j)$, $\mathbb{E}(\square_i)$ is the average number of squares a node is in, and $d_i d_j d_k d_l$ follows the joint degree distribution of closed walks of length 4 formed by nodes with degrees d_i, d_j, d_k, d_l.*

Note that in this theorem, for example, the term $\mathbb{E}(\frac{1}{d_i d_j})$ is the expected value of $\frac{1}{d_i d_j}$ over all edges, where d_i and d_j are the degree of the nodes connected by some edge.

Lemma 4.3. *Graph $H = (V, E)$ is a disjoint union of k graphs $G_1, G_2, \ldots, G_{k-1}, G_k$, i.e., $H = \bigcup_{i=1}^{k} G_i$. Let m_{ℓ,G_i} denote the ℓ-th spectral moment for $G_i = (V_i, E_i)$. Then, the ℓ-th spectral moment of H is the weighted average of m_{ℓ,G_i}'s weighted by $|V_i|$'s, i.e., $m_{\ell,H} = \frac{\sum_i |V_i| m_{\ell,G_i}}{|V|}$.*

Proof. The lemma is a generalized version of Theorem 4.3 of [10], if we consider any graph as a disjoint union of its connected components. Therefore, the proof is similar. Note that one can view the transition matrix of the random walk on H as a block matrix where each block represents the transition matrix of some G_i. □

A special case of Lemma 4.3 is that if all graphs G_i have the same order, i.e., $|V_i| = c$, for some constant c, then $m_{\ell,H} = \frac{\sum_i m_{\ell,G_i}}{k}$.

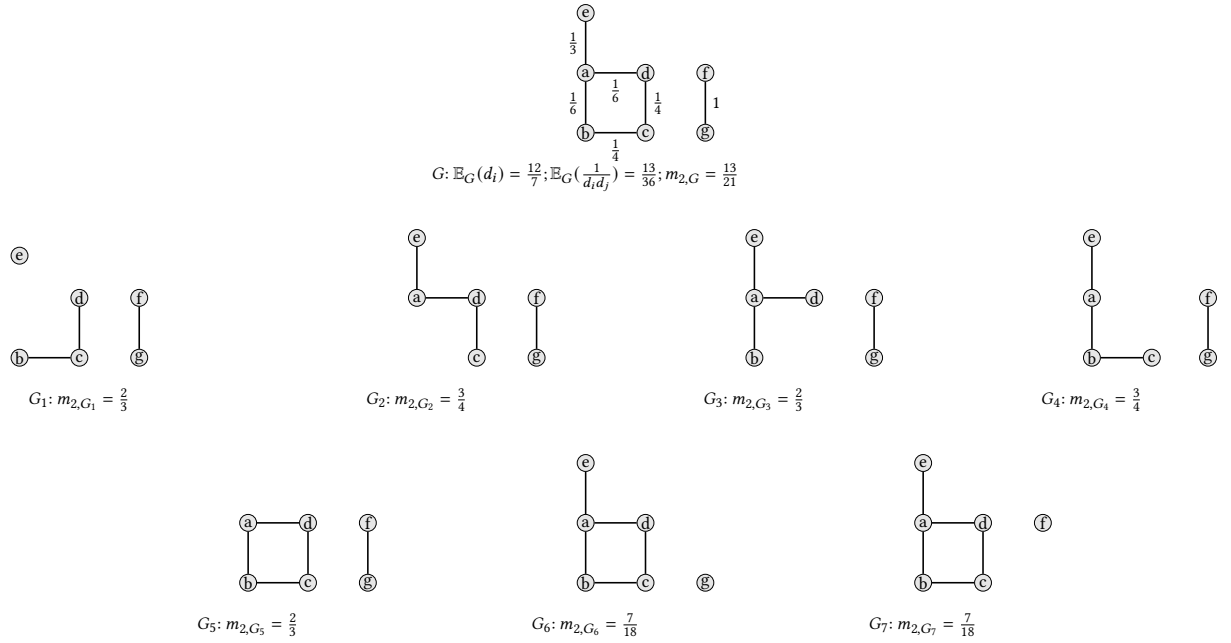


Figure 4: Example of Spectral Moment of Subgraphs

Proof of Theorem 4.1.

Proof. As G has n nodes, we have n choices to remove only one node, so we can get n possible subgraphs (G'). We denote these subgraphs as $G_1, G_2, \ldots, G_n$, e.g., $G_1, G_2, \ldots, G_7$ in Figure 4. Here, we aim to get the expected spectral moment of G', which is $\mathbb{E}(m_{2,G'}) = \frac{\sum_i m_{2,G_i}}{n}$.

We use proof by construction. Construct a new graph $H = (V_H, E_H)$ as a disjoint union of all these subgraphs G_i's, i.e., $H = \bigcup_{i=1}^{n} G_i$. From Lemma 4.3, we have $m_{2,H} = \frac{\sum_i m_{2,G_i}}{n}$ as each G_i has $n-1$ nodes. Hence, deriving $\mathbb{E}(m_{2,G'})$ is equivalent to finding $m_{2,H}$. From Theorem 4.2, we have $m_{2,H} = \mathbb{E}_H(d_i)\,\mathbb{E}_H(\frac{1}{d_i d_j})$. For $\mathbb{E}_H(d_i)$,

$$\mathbb{E}_H(d_i) = \frac{2 \cdot |E_H|}{|V_H|} = \frac{2 \cdot (n-2)m}{n(n-1)}$$

(as each edge of G appears $n-2$ times in H, and G_i has $n-1$ nodes.)

$$= \frac{n-2}{n-1} \cdot \mathbb{E}_G(d_i). \tag{1}$$

To derive $\mathbb{E}_H(\frac{1}{d_i d_j})$, we compare it to $\mathbb{E}_G(\frac{1}{d_i d_j})$: for any edge (i, j) of G, there are $n-2$ copies in H, but if the removed node is a neighbor of i (or j) in G, the degree d_i (or d_j) will decrease which leads to an increase of $\frac{1}{d_i d_j}$. Therefore, $\mathbb{E}_H(\frac{1}{d_i d_j}) > \mathbb{E}_G(\frac{1}{d_i d_j})$. To get the exact increase quantity, we partition any edge (i, j) of G into three types based on its node degree:

► **Type I**: $d_i = 1, d_j = 1$. For such an edge, all of its $n-2$ copies in H have $\frac{1}{d_i d_j} = 1$ as neither of i and j can lose other neighbors, so there will be no increment;

► **Type II**: $d_i > 1, d_j = 1$. For an edge of this type, if the removed node is a neighbor of i, then the edge contributes an increment $\frac{1}{(d_i-1)d_j} - \frac{1}{d_i d_j} = \frac{1}{d_i d_j (d_i-1)}$. Among the $n-2$ copies in H, there are $d_i - 1$ such cases as each neighbor of i gets removed once, so the overall contribution is $(d_i - 1) \cdot \frac{1}{d_i d_j (d_i-1)} = \frac{1}{d_i d_j}$;

► **Type III**: $d_i > 1, d_j > 1$. Assume nodes i and j have c_{ij} common neighbors. If the removed node is a neighbor of i but not j, then the edge contributes an increment $\frac{1}{(d_i-1)d_j} - \frac{1}{d_i d_j} = \frac{1}{d_i d_j (d_i-1)}$. Among its $n-2$ copies in H, there are $d_i - 1 - c_{ij}$ such cases (excluding j and the common neighbors), so the total contribution is $\frac{d_i-1-c_{ij}}{d_i d_j (d_i-1)}$. Similarly, if the removed node is a neighbor of j but not i, the total contribution for such cases is $\frac{d_j-1-c_{ij}}{d_i d_j (d_j-1)}$. If the removed node is a common neighbor of i and j, the increment is $\frac{1}{(d_i-1)(d_j-1)} - \frac{1}{d_i d_j} = \frac{1}{d_i d_j (d_i-1)} + \frac{1}{d_i d_j (d_j-1)} + \frac{1}{d_i d_j (d_i-1)(d_j-1)}$. As there are c_{ij} common neighbors, the contribution by such cases is $\frac{c_{ij}}{d_i d_j (d_i-1)} + \frac{c_{ij}}{d_i d_j (d_j-1)} + \frac{c_{ij}}{d_i d_j (d_i-1)(d_j-1)}$. Overall for one edge of Type III in G, it contributes the increment: $\frac{d_i-1-c_{ij}}{d_i d_j (d_i-1)} + \frac{d_j-1-c_{ij}}{d_i d_j (d_j-1)} + \frac{c_{ij}}{d_i d_j (d_i-1)} + \frac{c_{ij}}{d_i d_j (d_j-1)} + \frac{c_{ij}}{d_i d_j (d_i-1)(d_j-1)} = \frac{2}{d_i d_j} + \frac{c_{ij}}{d_i d_j (d_i-1)(d_j-1)}$.

Therefore, the total increment δ is $\delta = \sum_{\substack{(i,j)\in G \\ d_i>1, d_j=1}} \frac{1}{d_i d_j} + \sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} (\frac{2}{d_i d_j} + \frac{c_{ij}}{d_i d_j (d_i-1)(d_j-1)})$. We know that $\sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{c_{ij}}{d_i d_j (d_i-1)(d_j-1)} = \sum_{(i,j,k) \text{is a triad in } G} (\frac{1}{d_i d_j (d_i-1)(d_j-1)} + \frac{1}{d_i d_k (d_i-1)(d_k-1)} + \frac{1}{d_j d_k (d_j-1)(d_k-1)})$ and we denote it as δ_{triad}.

Thus, the total increment δ is $\delta = \sum_{\substack{(i,j)\in G \\ d_i>1, d_j=1}} \frac{1}{d_i d_j} + 2 \cdot \sum_{\substack{(i,j)\in G \\ d_i>1, d_j>1}} \frac{1}{d_i d_j} + \delta_{\text{triad}}$; normalized by $|E_H|$, we get

$$\mathbb{E}_H(\frac{1}{d_i d_j}) = \mathbb{E}_G(\frac{1}{d_i d_j}) + \frac{\delta}{(n-2)m}. \tag{2}$$

Next, we compute $m_{2,H}$ using Equations 1 and 2:

$$\begin{aligned}
m_{2,H} &= \mathbb{E}_H(d_i)\,\mathbb{E}_H(\frac{1}{d_i d_j}) \\
&= \frac{n-2}{n-1}\,\mathbb{E}_G(d_i)\cdot(\mathbb{E}_G(\frac{1}{d_i d_j}) + \frac{\delta}{(n-2)m}) \\
&= \frac{n-2}{n-1}\,\mathbb{E}_G(d_i)\,\mathbb{E}_G(\frac{1}{d_i d_j}) + \frac{\delta\cdot\mathbb{E}_G(d_i)}{(n-1)m} \\
&= \mathbb{E}_G(d_i)\,\mathbb{E}_G(\frac{1}{d_i d_j}) + \frac{\delta\cdot\mathbb{E}_G(d_i)}{(n-1)m} - \frac{\mathbb{E}_G(d_i)\,\mathbb{E}_G(\frac{1}{d_i d_j})}{n-1} \\
&= m_{2,G} + \frac{\mathbb{E}_G(d_i)}{(n-1)m}\cdot(\delta - m\cdot\mathbb{E}_G(\frac{1}{d_i d_j})) \\
&= m_{2,G} + \frac{2}{n(n-1)}\cdot(\delta - m\cdot\mathbb{E}_G(\frac{1}{d_i d_j})) \quad (\text{as } \mathbb{E}_G(d_i) = \frac{2m}{n})
\end{aligned}$$

Note that $m\cdot\mathbb{E}_G(\frac{1}{d_i d_j}) = \sum_{(i,j)\in G}\frac{1}{d_i d_j}$, so we obtain the following which finalizes the proof:

$$\begin{aligned}
\delta - m\cdot\mathbb{E}_G(\frac{1}{d_i d_j}) &= \sum_{\substack{(i,j)\in G\\ d_i>1, d_j=1}}\frac{1}{d_i d_j} + 2\cdot\sum_{\substack{(i,j)\in G\\ d_i>1, d_j>1}}\frac{1}{d_i d_j} + \delta_{\text{triad}} - \sum_{(i,j)\in G}\frac{1}{d_i d_j} \\
&= \sum_{\substack{(i,j)\in G\\ d_i>1, d_j>1}}\frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G\\ d_i=1, d_j=1}}\frac{1}{d_i d_j} + \delta_{\text{triad}},
\end{aligned}$$

If G is triangle-free, then $\delta_{\text{triad}} = 0$. The theorem is proved. □

4.1.3 A more general case. Next, we explore a more general case, so we provide a bound for the expected m_2 by removing k nodes from a triangle-free graph in Theorem 4.4. When $k = 1$, the theorem is reduced to Theorem 4.1, so the bound is tight.

Theorem 4.4. *(Expected m_2 by Removing k Nodes) In undirected triangle-free graph $G = (V, E)$, where $|V_G| = n, |E_G| = m$, subgraph G' of G is obtained by removing k nodes from G uniformly at random. For $\mathbb{E}(m_{2,G'})$, the expected second moment of G', we have*

$$\mathbb{E}(m_{2,G'}) \le m_{2,G} + \frac{2k}{n(n-1)}\cdot(\frac{n-1}{n-k}\cdot\sum_{\substack{(i,j)\in G\\ d_i>1, d_j>1}}\frac{1}{d_i d_j} - \sum_{\substack{(i,j)\in G\\ d_i=1, d_j=1}}\frac{1}{d_i d_j}),$$

where d_i and d_j denote the degrees of node i and j, respectively.

For the detailed proof of Theorem 4.4, please refer to the supplementary material. We use the same triangle-free graph G in Figure 4, and we list all the 21 subgraphs[1] by removing two nodes from G. By Theorem 4.4, $\mathbb{E}(m_{2,G'}) \le \frac{13}{21} + \frac{2\times2}{7\times6}(\frac{6}{5}\cdot(\frac{1}{6}+\frac{1}{6}+\frac{1}{4}+\frac{1}{4})-1) = \frac{13}{21} \approx 0.619$, while the actual $\mathbb{E}(m_{2,G'}) = \frac{\sum_{i=1}^{21} m_{2,G_i}}{21} \approx 0.594 < 0.619$.

4.2 Third and Fourth Spectral Moments

In this part, we provide two theorems for the expected third and fourth spectral moments of subgraphs ($\mathbb{E}(m_{3,G'})$ and $\mathbb{E}(m_{4,G'})$) when one node is removed. Theorem 4.5 provides an upper bound for $\mathbb{E}(m_{3,G'})$ indicating that whether $\mathbb{E}(m_{3,G'})$ is expected to increase or not over that of the original graph (m_3) depends on the weighted summation of $\frac{1}{d_i d_j d_k}$ over four different types of triads. In general, if there are more triads formed by higher degree nodes, $m_{3,G'}$ is expected to increase; if there are more triads with low degree nodes (i.e., with degree 2), $m_{3,G'}$ is expected to decrease. The forth moment m_4 is related to closed walks of length 4 (generated by edges, wedges, or squares). We provide a loose bound on $\mathbb{E}(m_{4,G'})$ in Theorem 4.6. Our experiments also show that m_4 has a high correlation with m_2. For detailed proofs, please refer to the supplementary material. In general, if we view triads or squares as higher-order edges of a network, a similar analysis on m_2 can be applied to m_3 and m_4, leading to increasing-decreasing, increasing-only, and decreasing-only patterns for these moments.

[1]Due to space limits, the figure is available at https://bit.ly/3zqkVVa

Table 1: Dataset Statistics

Type	Network	$\lvert V\rvert = n$	$\lvert E\rvert = m$	Average Degree	Clustering Coefficient
Social Networks	Brightkite [13]	58,228	214,078	7.353	0.1723
	Flixster [24]	2,523,386	7,918,801	6.276	0.0834
	Gowalla [13]	196,591	950,327	9.668	0.2367
	Hyves [24]	1,402,673	2,777,419	3.960	0.0448
	Livejournal [25]	3,017,286	85,654,976	56.78	0.1196
	MySpace [25]	854,498	5,635,296	13.19	0.0433
	Orkut [13]	3,072,441	117,185,083	76.28	0.1666
	YouTube [13]	1,134,890	2,987,624	5.265	0.0808
Collaboration Networks	Astro-Ph [13]	18,772	198,050	21.10	0.6306
	Cond-Mat [13]	23,133	93,439	8.078	0.6334
	Gr-Qc [13]	5,242	14,484	5.526	0.5296
	Hep-Th [13]	9,877	25,973	5.259	0.4714
Road Networks	Road-BEL [18]	1,441,295	1,549,970	2.143	0.0017
	Road-CA [13]	1,965,206	2,766,607	2.816	0.0464
	Road-PA [13]	1,088,092	1,541,898	2.834	0.0465
	Road-TX [13]	1,379,917	1,921,660	2.785	0.0470
Biological Networks	Bio-Dmela [18]	7,393	25,569	6.917	0.0119
	Bio-Grid-Human [18]	9,527	62,364	13.09	0.1094
	Bio-Grid-Yeast [18]	5,870	313,890	106.9	0.0516
	Human-Brain [18]	177,600	15,669,036	176.4	0.4580

Theorem 4.5 (Expected third spectral moment m_3 of subgraphs after removing one node). *In undirected graph $G = (V, E)$, where $|V_G| = n, |E_G| = m$, subgraph G' of G is obtained by removing one node from G uniformly at random. For $\mathbb{E}(m_{3,G'})$, the expected third moment of G', we have*

$$\begin{aligned}
\mathbb{E}(m_{3,G'}) < m_{3,G} + \frac{6}{n(n-1)}\cdot(2\sum_{\substack{(i,j,k)\in G\\ d_i>2, d_j>2, d_k>2}}\frac{1}{d_i d_j d_k} + \frac{1}{4}\sum_{\substack{(i,j,k)\in G\\ d_i>2, d_j>2, d_k=2}}\frac{1}{d_i d_j d_k} \\
- \sum_{\substack{(i,j,k)\in G\\ d_i>2, d_j=2, d_k=2}}\frac{1}{d_i d_j d_k} - 2\sum_{\substack{(i,j,k)\in G\\ d_i=2, d_j=2, d_k=2}}\frac{1}{d_i d_j d_k}).
\end{aligned}$$

Theorem 4.6 (Expected forth spectral moment m_4 of subgraphs after removing one node). *In undirected graph $G = (V, E)$, where $|V_G| = n, |E_G| = m$, subgraph G' of G is obtained by removing one node from G uniformly at random. For $\mathbb{E}(m_{4,G'})$, the expected fourth moment of G', we have*

$$\mathbb{E}(m_{4,G'}) \le \frac{16(n-2)}{n-1}m_{4,G}.$$

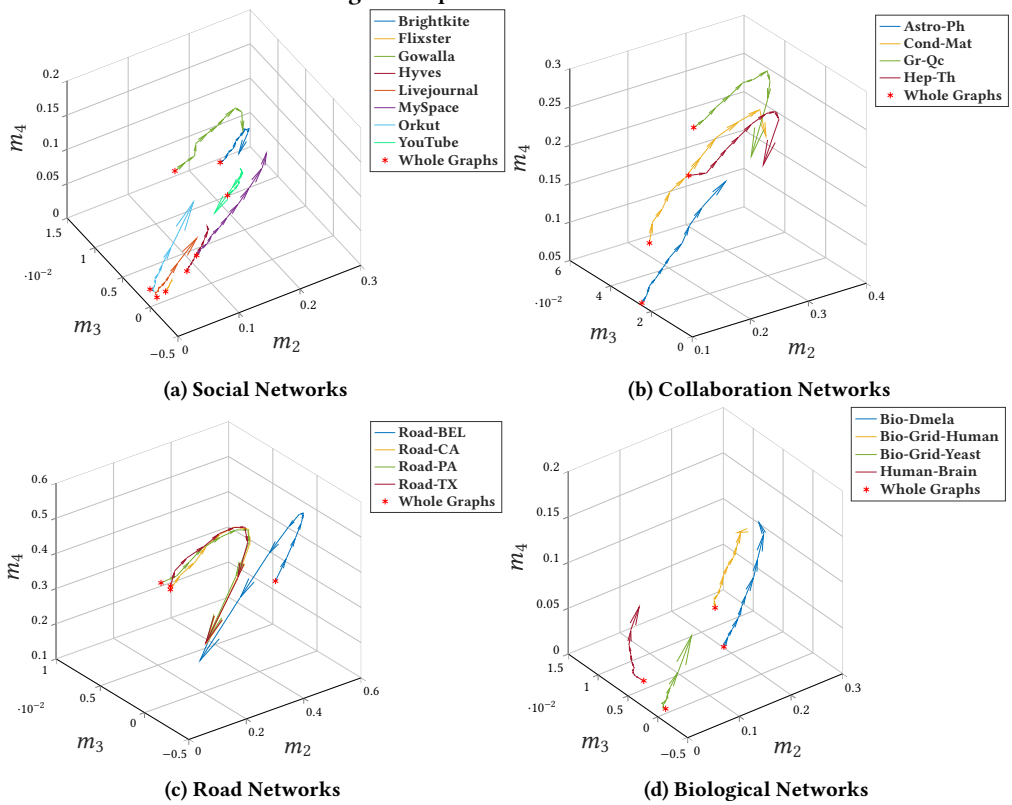
5 APPLICATIONS

5.1 Experimental Setup

In our experiments, we generate spectral path for each network by varying the proportion of nodes from 0% to 100% with step size 10%, i.e., $s = 10\%$ in Algorithm 1; for each proportion (except for 100% which represents the whole graph), we generate 20 independently sampled subgraphs, i.e., $t = 20$ (for which we have a discussion in Section 8). In total, we generate $20 \times 9 + 1 = 181$ spectral points for each network, and we compute the expected spectral points for each sampling proportion. Hence, for each network, a spectral path connects 10 expected spectral points from (100% to 10%). Code and datasets are publicly available.[2] Next, we summarize the datasets.

[2]https://github.com/shengminjin/SpectralPath

Figure 5: Spectral Paths of Networks



(a) Social Networks **(b) Collaboration Networks**

(c) Road Networks **(d) Biological Networks**

Datasets. We use 20 real-world networks from four network categories: social networks, collaboration networks, road networks, and biological networks. The data statistics are in Table 1.

Social Networks: In total, we have eight social networks.

(1) *Brightkite* [13]: was a location-based social networking site where users shared their locations by checking-in.
(2) *Flixster* [24]: a social movie site to buy or rent movies.
(3) *Gowalla* [13]: similar to Brightkite, was a location-based social networking site where users shared their locations.
(4) *Hyves* [24]: the most popular social networking site in the Netherlands with mainly Dutch visitors.
(5) *Livejournal* [25]: a social networking site where users can keep a blog or journal. Here, edges represent friendships (undirected).
(6) *MySpace* [25]: a social network having a significant influence on pop culture and music.
(7) *Orkut* [13]: was a social networking site shutdown in 2014.
(8) *YouTube* [13]: a video-sharing site with a social network.

Collaboration Networks: We have four collaboration networks from arXiv.org, including scientific collaborations between authors with different scientific interests.

(9) *Astro-Ph* [13]: a collaboration network in Astro physics.
(10) *Cond-Mat* [13]: a collaboration network in Condense matter.
(11) *Gr-Qc* [13]: a collaboration network in General relativity and quantum cosmology.
(12) *Hep-Th* [13]: a collaboration network in High energy physics.

Road Networks: We include four road networks.

(13) *Road-BEL* [13]: the OpenStreetMap road network of Belgium.
(14) *Road-CA* [13]: the road network of California.
(15) *Road-PA* [13]: the road network of Pennsylvania.
(16) *Road-TX* [13]: the road network of Texas.

Biological Networks: We include four biological networks.

(17) *Bio-Dmela* [18]: a protein-protein interaction network.
(18) *Bio-Grid-Human*[18]: a protein-protein interaction network.
(19) *Bio-Grid-Yeast*[18]: a protein-protein interaction network.
(20) *Human-Brain* [18]: the network of human brain.

5.2 Network Visualization

As a spectral path is a path connecting several 3D spectral points of a network and its subgraphs, we are able to plot them for visualization and to capture the network properties. In Figure 5, we plot the spectral paths of 20 real-world networks from four different categories. We have the following observations: (1) For m_2 and m_4, we see two common patterns: *increasing-decreasing*, and *increasing-only* which indicates the turning point happens before 10% samples of the graph. The observation shows that for most real-world graphs more edges are Type III edges; (2) among the eight social networks, Brightkite, Gowalla, and YouTube show the increasing-decreasing pattern for both m_2 and m_4. It can be explained by their relatively low average degree and smaller graph size, so in small samples like 10% or 20%, most of the edges become Type I edges. Moreover, Brightkite and Gowalla show an increasing-decreasing trend on m_3 while Orkut shows an increasing-only trend on m_3, as Orkut has a much higher average degree so most of its triads are composed of

high degree nodes. For the remaining networks, there is not much change on m_3 as they have a low clustering coefficient so they do not have as many triads as in those three; (3) for Collaboration networks, only Astro-Ph shows the increasing-only trend as it has a much higher average degree than other networks, so even 10% of the graph still has more edges among high-degree nodes. We observe the large change on m_3 for Cond-Mat, Gr-Qc and Hep-Th as in general collaboration networks have a high clustering coefficient and small samples of these sparse networks lose most triads or only have triads with low-degree nodes; (4) for road networks, we can see an early turning point (40%) on all m_2, m_3, and m_4. This can be explained by their low average degree so most edges and triads are among low degree nodes; (5) all biological networks show increasing-only trend as they all have a high average degree or high edge density. Overall, using the spectral path, one can get various insights on the graph structure.

5.3 Network Identification

Network identification [9] aims to identify the source network from which an anonymized graph is sampled, to find the *identity* of a subgraph. Network identification can be formulated as follows: given a set of networks $N = \{N_1, N_2, \ldots, N_n\}$, and a subgraph G sampled from $N_i \in N$ using a sampling strategy S, we want to identify G, i.e., the network N_i from which G is sampled. In the problem setting, there are a few assumptions: (1) The networks are not isomorphic, i.e., N_i and N_j are isomorphic $\implies i = j$, as isomorphic graphs are basically the same graph after anonymization; and (2) Subgraph G is not too small to lose its identity. It does not make much sense to verify the identity of a small subgraph such as a triad, since it can be found in most networks.

5.3.1 Experimental Setup. From each of the 20 real-world networks, we sample many subgraphs representing graphs G which are to be identified. We vary the sampling proportion from 10% to 99% by using random node sampling with step size 1%. For each proportion, we sample two subgraphs. Hence, for each network, we have 90×2 = 180 subgraphs, and in total, 180×20 = 3,600 samples to be identified.

5.3.2 Experiments. As the spectral path has both spectral points of subgraphs and their relationship, we aim to explore whether one can improve the network identification performance by using the distances between an unidentified subgraph and the spectral path. For each subgraph, we compute the euclidean distances from its spectral point to the 10 expected spectral points of the spectral path, respectively. As there are 20 networks in total, for each subgraph we use the 10×20 = 200 distances as features and the name of the source networks as the class label, to train a multiclass classifier. We use 10-fold cross validation, and decision trees, SVM, k-NN, and bagged trees as our classifiers. For evaluation, we compare to the following four baselines: (1) **Three Spectral Moments**, where the spectral point (m_2, m_3, m_4) of each subgraph is used as features; (2) **First Twenty Spectral Moments**, where the first 20 spectral moments of each subgraph are used as features; (3) **Kronecker Hull** [9], which uses Stochastic Kronecker Graph model to embed a network and its subgraphs and then a convex hull to represent the distribution of the embeddings. We use the distances between a subgraph to the convex hull of the whole network as features; (4) **Graph2Vec** is a graph embedding method which views a graph as a document and the rooted subgraphs around each node as words. It uses document embedding neural networks to embed a graph as a vector, which we use as a feature [15].

Table 2: Network Identification Accuracy with Spectral Path

Type	Spectral Path	Baselines			
		Three Spectral Moments	First 20 Spectral Moments	Kronecker Hull	Graph2Vec
All Networks	**96.3%**	82.0%	86.5%	84.4%	81.7%
Social Networks	**100.0%**	95.5%	96.1%	96.4%	83.7%
Collaboration Networks	**99.9%**	94.8%	97.1%	84.2%	97.4%
Road Networks	**86.8%**	51.2%	53.3%	76.8%	86.3%
Biological Networks	**100.0%**	99.7%	99.6%	80.4%	89.9%

We evaluate the methods for all networks and within each network category. We report the performance for the best classifier in Table 2, where spectral path significantly outperforms the baselines.

6 SPECTRAL PATH OF COSPECTRAL GRAPHS

Two non-isomorphic graphs are said to be cospectral with respect to a given matrix if they share the same graph spectrum. Well-known examples of cospectral graphs for the normalized Laplacian (as well as the random walk transition matrix) are complete bipartite graphs [6]. Butler et al. [2] propose constructing cospectral graphs by swapping in a bipartite subgraph with a *cospectral mate*. In general, cospectral graphs for the random walk transition matrix are related to the bipartite (sub)graphs.

Here, we aim to show that using the expected spectral moments (or spectral paths) of cospectral graphs may provide a potential way to distinguish two cospectral graphs. In this paper, we use complete bipartite graphs as an example. It is known that the spectrum of a complete bipartite graph $K_{a,b}$ is $-1^{[1]}, 0^{[n-2]}, 1^{[1]}$, where $n = a + b$ and the exponent indicates multiplicity. Hence, its spectral moments are $m_i = 0$ for an odd i, and $m_i = \frac{2}{n}$ for an even i. It is easy to see that complete bipartite graphs of the same order are all cospectral, i.e., one can find another complete bipartite graph $K_{a',b'}$ where $a' + b' = a + b$. In Theorem 6.1, we prove that if one samples a subgraph from a complete bipartite graph using random node sampling, the expectation of its spectral moments are not only related to n but also related to the values of a and b.

Theorem 6.1. *Given an undirected complete bipartite graph $G = (U, V, E)$ where $|U| = a$, $|V| = b$, $a + b = n$, and $a \le b$, G' is a subgraph of G by removing k nodes from G uniformly at random. Then, when i is odd, $\mathbb{E}(m_{i,G'}) = 0$, and when i is even,*

$$\mathbb{E}(m_{i,G'}) = \begin{cases} \frac{2}{n-k} & k < a \\ \frac{2}{n-k}\left(1 - \frac{\binom{b}{k-a}}{\binom{n}{k}}\right) & a \le k < b \\ \frac{2}{n-k}\left(1 - \frac{\binom{b}{k-a}+\binom{a}{k-b}}{\binom{n}{k}}\right) & k \ge b \end{cases}$$

Proof. As G' is a subgraph by removing k nodes from a complete bipartite graph G, G' is either a complete bipartite graph or an empty graph. Hence, $m_{i,G'}$ is always 0 when i is odd. When i is even, if $k < a$, G' is always a complete bipartite graph of $n-k$ nodes, so $\mathbb{E}(m_{i,G'}) = \frac{2}{n-k}$; if $a \le k < b$, among all $\binom{n}{k}$ possible subgraphs, there are $\binom{b}{k-a}$ cases of G' being an empty graph where all the nodes in U are removed, and in the remaining cases, G' is a complete bipartite graph. Hence, $\mathbb{E}(m_{i,G'}) = \frac{\frac{2}{n-k}\cdot\left(\binom{n}{k}-\binom{b}{k-a}\right)+0\cdot\binom{b}{k-a}}{\binom{n}{k}} =$

$\frac{2}{n-k}(1-\frac{\binom{b}{k-a}}{\binom{n}{k}})$; finally, if $k \geq b$, there are $\binom{b}{k-a}$ cases where all the nodes in U are removed, and $\binom{a}{k-b}$ cases when all the nodes in V are removed, where in both cases G' becomes an empty graph so we get $\mathbb{E}(m_{i,G'}) = \frac{2}{n-k}(1-\frac{\binom{b}{k-a}+\binom{a}{k-b}}{\binom{n}{k}})$. □

Using Theorem 6.1, we get the following corollary:

Corollary 6.1.1. *Given two complete bipartite graphs $G_1 = (U_1, V_1, E_1)$ where $|U_1| = a_1, |V_1| = b_1$, and $G_2 = (U_2, V_2, E_2)$ where $|U_2| = a_2, |V_2| = b_2$, and $a_1 + b_1 = a_2 + b_2 = n$, $a_1 < a_2 \leq b_2 < b_1$. Let G_1' (or G_2') be a subgraph of G_1 (or G_2) by removing k nodes from G_1 (or G_2) uniformly at random. Then, for an even i, we have*

$$\mathbb{E}(m_{i,G_1'} - m_{i,G_2'}) = \begin{cases} 0 & k < a_1 \\ -\frac{2}{n-k} \cdot \frac{\binom{b_1}{k-a_1}}{\binom{n}{k}} & a_1 \leq k < a_2 \\ \frac{2}{n-k} \cdot \frac{\binom{b_2}{k-a_2} - \binom{b_1}{k-a_1}}{\binom{n}{k}} & a_2 \leq k < b_2 \\ \frac{2}{n-k} \cdot \frac{\binom{b_2}{k-a_2} + \binom{a_2}{k-b_2} - \binom{b_1}{k-a_1}}{\binom{n}{k}} & b_2 \leq k < b_1 \\ \frac{2}{n-k} \cdot \frac{\binom{b_2}{k-a_2} + \binom{a_2}{k-b_2} - \binom{b_1}{k-a_1} - \binom{a_1}{k-b_1}}{\binom{n}{k}} & k \geq b_1 \end{cases}$$

From Corollary 6.1.1, we notice that in general $\mathbb{E}(m_{i,G_1'} - m_{i,G_2'})$ is nonzero as long as $k \geq a_1$ (the special case is when $k = n$ or $n-1$, in which $\mathbb{E}(m_{i,G_1'} - m_{i,G_2'}) = 0$). Hence, if we remove more than a_1 nodes from two cospectral graphs (G_1, G_2) respectively, the expected (even) spectral moments of the corresponding subgraphs are different. Moreover, $a_1 < \frac{n}{2}$ as $a_1 + b_1 = a_2 + b_2 = n$, $a_1 < a_2 \leq b_2 < b_1$. Hence, when $k \geq \frac{n}{2}$, $\mathbb{E}(m_{i,G_1'} - m_{i,G_2'})$ can be used to distinguish between two complete bipartite graphs. In other words, one can use random node sampling to sample subgraphs of less than $\frac{n}{2}$ nodes, and estimate $\mathbb{E}(m_{i,G_1'})$ and $\mathbb{E}(m_{i,G_2'})$, to distinguish two complete bipartite graphs. The idea can be extended to general cospectral graphs, as long as they have an explicit form of spectrum.

7 ADDITIONAL RELATED WORK

I. Subgraph Spectrum. Past studies on the spectrum of subgraphs often focus on the Cauchy's interlacing theorem which bounds subgraph eigenvalues with the eigenvalues of whole graphs [3, 14]. However, interlacing theorem can only provide loose bounds on the spectral moments, whereas our bounds are either exact or tight.

II. Spectral Embedding. Recently, spectral information is used for different network embedding methods, such as FGSD [22] and NetLSD [20]. Compared to them, spectral path is easy to be interpreted and utilizes the spectral information of subgraphs.

III. Reconstruction Conjecture. As mentioned, Bollobás has shown that almost all graphs are reconstructible [1]. Moreover, not all subgraphs are necessary to reconstruct them: for almost all graphs, there exist three subgraphs that uniquely identify the graph. Our setting matches that of the theoretical findings of Bollobás. Specifically, by taking samples from the graph, we aim to capture the subgraphs that can uniquely represent a graph.

8 CONCLUSION

We propose representing a network with a 3D spectral path: a path connecting the spectral moments of a network to the expected spectral moments of its subgraphs. We demonstrate the interpretability of spectral paths. We show the utility of spectral paths in network visualization, network identification and distinguishing cospectral graphs. To the best of our knowledge, this is the first study to explore spectral moments of subgraphs and study the relationship between spectral moments of subgraphs and those of the whole network.

Limitations and Future Work. For a graph of n nodes, there are $\binom{n}{k}$ subgraphs of size k. When k is around $\frac{n}{2}$, the number of subgraphs is the largest. Hence, one may consider taking different numbers of samples to estimate the expected moments. In our experiments on real-world networks, we take 20 samples for each sampling proportion, and the standard deviation of spectral moments is often less than 5%. In the future, we aim to study the distribution of spectral moments of subgraphs.

Acknowledgements. This research was supported in part by the National Science Foundation under award CAREER IIS-1942929.

REFERENCES

[1] Béla Bollobás. 1990. Almost every graph has reconstruction number three. *Journal of Graph Theory* 14, 1 (1990), 1–4.
[2] Steve Butler and Jason Grout. 2010. A construction of cospectral graphs for the normalized Laplacian. *arXiv preprint arXiv:1008.3646* (2010).
[3] Steven Kay Butler. 2008. *Eigenvalues and structures of graphs.* Ph. D. Dissertation. UC San Diego.
[4] Deepayan Chakrabarti, Yang Wang, Chenxi Wang, Jurij Leskovec, and Christos Faloutsos. 2008. Epidemic thresholds in real networks. *ACM Transactions on Information and System Security (TISSEC)* 10, 4 (2008), 1–26.
[5] Jeff Cheeger. 2015. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis.* Princeton University Press, 195–200.
[6] Fan RK Chung and Fan Chung Graham. 1997. *Spectral graph theory.* Number 92. American Mathematical Soc.
[7] David Cohen-Steiner, Weihao Kong, Christian Sohler, and Gregory Valiant. 2018. Approximating the Spectrum of a Graph. In *SIGKDD*. ACM, 1263–1271.
[8] Kun Dong, Austin R Benson, and David Bindel. 2019. Network density of states. In *Proceedings of the 25th ACM SIGKDD.* 1152–1161.
[9] Shengmin Jin, Vir V Phoha, and Reza Zafarani. 2019. Network Identification and Authentication. In *2019 International Conference on Data Mining (ICDM). IEEE.*
[10] Shengmin Jin and Reza Zafarani. 2020. The Spectral Zoo of Networks: Embedding and Visualizing Networks with Spectral Moments. In *Proceedings of the SIGKDD.*
[11] Paul J Kelly et al. 1957. A congruence theorem for trees. *Pacific J. Math.* 7, 1 (1957), 961–968.
[12] Jure Leskovec and Christos Faloutsos. 2006. Sampling from large graphs. In *Proceedings of the SIGKDD.* ACM, 631–636.
[13] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.
[14] Jan R Magnus and Heinz Neudecker. 2019. *Matrix differential calculus with applications in statistics and econometrics.* John Wiley & Sons.
[15] Annamalai Narayanan, Mahinthan Chandramohan, Rajasekar Venkatesan, Lihui Chen, Yang Liu, and Shantanu Jaiswal. 2017. graph2vec: Learning distributed representations of graphs. *arXiv preprint arXiv:1707.05005* (2017).
[16] Peter V O'Neil. 1970. Ulam's conjecture and graph reconstructions. *The American Mathematical Monthly* 77, 1 (1970), 35–43.
[17] Edouard Pineau. 2019. Using Laplacian Spectrum as Graph Feature Representation. *arXiv preprint arXiv:1912.00735* (2019).
[18] Ryan Rossi and Nesreen Ahmed. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization.. In *AAAI*, Vol. 15. 4292–4293.
[19] Allen J Schwenk. 1973. Almost all trees are cospectral. *New directions in the theory of graphs* (1973), 275–307.
[20] Anton Tsitsulin, Davide Mottin, Panagiotis Karras, Alexander Bronstein, and Emmanuel Müller. 2018. Netlsd: hearing the shape of a graph. In *KDD*. 2347–2356.
[21] Stanislaw M Ulam. 1960. *A collection of mathematical problems.* Vol. 8. Interscience Publishers.
[22] Saurabh Verma and Zhi-Li Zhang. 2017. Hunt for the unique, stable, sparse and fast feature learning on graphs. In *NeurIPS*. 88–98.
[23] Richard C Wilson and Ping Zhu. 2008. A study of graph spectra for comparing graphs and trees. *Pattern Recognition* 41, 9 (2008), 2833–2841.
[24] R. Zafarani and H. Liu. 2009. Social Computing Data Repository at ASU. http://socialcomputing.asu.edu
[25] Yutao Zhang, Jie Tang, Zhilin Yang, Jian Pei, and Philip S Yu. 2015. Cosnet: Connecting heterogeneous social networks with local and global consistency. In *Proceedings of the 21th ACM SIGKDD.* ACM, 1485–1494.

A SUPPLEMENTARY MATERIAL

A.1 Proof of Theorem 4.4

Proof. As we are removing k nodes, there are $\binom{n}{k}$ possible subgraphs, denoted as $G_1, G_2, \ldots, G_{\binom{n}{k}}$. Each G_i has $n-k$ nodes. We construct $H = \bigcup_{i=1}^{\binom{n}{k}} G_i$. For each edge (i,j) in G, there will be $\binom{n-2}{k}$ copies in H when neither of the ending nodes is removed. Therefore,

$$E_H(d_i) = \frac{2\cdot|E_H|}{|V_H|} = \frac{2m\cdot\binom{n-2}{k}}{(n-k)\binom{n}{k}} = \frac{2m(n-k-1)}{n(n-1)} = \frac{n-k-1}{n-1}\cdot\mathbb{E}_G(d_i). \tag{3}$$

Similarly, to get $E_H(\frac{1}{d_id_j})$, we need to analyze the three types of edges. (1) **Type I**: $d_i = d_j = 1$. There will be no increment for edges of Type I; (2) **Type II**: $d_i > 1, d_j = 1$. If we consider x neighbors of node i are removed, the increment on the edge is $\frac{1}{(d_i-x)d_j} - \frac{1}{d_id_j} = \frac{x}{(d_i-x)d_id_j}$. Among the $\binom{n-2}{k}$ copies, there are $\binom{d_i-1}{x}\cdot\binom{n-2-(d_i-1)}{k-x}$ cases where x neighbors of i is removed. Moreover, x varies from 0 to $\min(d_i-1,k)$, so the overall increment of an edge of Type II is $\sum_{x=0}^{\min(d_i-1,k)} \frac{x}{(d_i-x)d_id_j}\binom{d_i-1}{x}\cdot\binom{n-2-(d_i-1)}{k-x} = \frac{1}{d_id_j}\sum_{x=1}^{\min(d_i-1,k)}\binom{d_i-1}{x-1}\cdot\binom{n-2-(d_i-1)}{k-x} \le \frac{1}{d_id_j}\cdot\binom{n-2}{k-1}$ (as $\frac{x}{d_i-x}\binom{d_i-1}{x} = \binom{d_i-1}{x-1}$ and the bound is tight when $k \le (d_i-1)$); (3) **Type III**: $d_i > 1, d_j > 1$. Assume x neighbors of i and y neighbors of j are removed, then the increment is $\frac{1}{(d_i-x)(d_j-y)} - \frac{1}{d_id_j} = \frac{x}{(d_i-x)d_id_j} + \frac{y}{(d_j-y)d_id_j} + \frac{xy}{(d_i-x)(d_j-y)d_id_j}$. As G is a triangle-free graph, H is also triangle-free, and i and j have no common neighbors. Therefore, among the $\binom{n-2}{k}$ copies, there are $\binom{d_i-1}{x}\binom{d_j-1}{y}\binom{n-d_i-d_j}{k-x-y}$ cases when x neighbors of i and y neighbors of j are removed. Moreover, x varies from 0 to $\min(d_i-1,k)$, y varies from 0 to $\min(d_j-1,k)$ and $x+y \le k$, so the overall increment of an edge of Type III is $\sum_{x=0}^{\min(d_i-1,k)}\sum_{y=0}^{\min(d_j-1,k-x)}\binom{d_i-1}{x}\binom{d_j-1}{y}\binom{n-d_i-d_j}{k-x-y}(\frac{x}{(d_i-x)d_id_j} + \frac{y}{(d_j-y)d_id_j} + \frac{xy}{(d_i-x)(d_j-y)d_id_j})$. Note that

$$\begin{aligned}
&\sum_{x=0}^{\min(d_i-1,k)}\sum_{y=0}^{\min(d_j-1,k-x)}\binom{d_i-1}{x}\binom{d_j-1}{y}\binom{n-d_i-d_j}{k-x-y}\frac{x}{(d_i-x)d_id_j}\\
&= \sum_{x=0}^{\min(d_i-1,k)}\binom{d_i-1}{x}\frac{x}{(d_i-x)d_id_j}\cdot\sum_{y=0}^{\min(d_j-1,k-x)}\binom{d_j-1}{y}\binom{n-d_i-d_j}{k-x-y}\\
&= \sum_{x=1}^{\min(d_i-1,k)}\binom{d_i-1}{x-1}\frac{1}{d_id_j}\cdot\binom{n-d_i-1}{k-x}\\
&\le \frac{1}{d_id_j}\cdot\binom{n-2}{k-1}; \text{(the bound is tight when } k \le (d_i-1)).
\end{aligned}$$

Due to the symmetry, the summation over $\frac{y}{(d_j-y)d_id_j}$ is also less or equal to $\frac{1}{d_id_j}\cdot\binom{n-2}{k-1}$. For the summation over $\frac{xy}{(d_i-x)(d_j-y)d_id_j}$:

$$\begin{aligned}
&\sum_{x=0}^{\min(d_i-1,k)}\sum_{y=0}^{\min(d_j-1,k-x)}\binom{d_i-1}{x}\binom{d_j-1}{y}\binom{n-d_i-d_j}{k-x-y}\frac{xy}{(d_i-x)(d_j-y)d_id_j}\\
&= \sum_{x=1}^{\min(d_i-1,k)}\sum_{y=1}^{\min(d_j-1,k-x)}\binom{d_i-1}{x-1}\binom{d_j-1}{y-1}\binom{n-d_i-d_j}{k-x-y}\frac{1}{d_id_j} \le \frac{1}{d_id_j}\cdot\binom{n-2}{k-2}
\end{aligned}$$

The increment of an edge of Type III is $\le \frac{1}{d_id_j}\cdot(2\binom{n-2}{k-1} + \binom{n-2}{k-2})$, so the total increment over all the edges is less or equal to $\delta = \binom{n-2}{k-1}\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j=1}}\frac{1}{d_id_j} + (2\binom{n-2}{k-1} + \binom{n-2}{k-2})\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j}$, and we normalize it by $|E_H|$:

$$\begin{aligned}
\frac{\delta}{|E_H|} &= \frac{\binom{n-2}{k-1}\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j=1}}\frac{1}{d_id_j} + (2\binom{n-2}{k-1} + \binom{n-2}{k-2})\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j}}{\binom{n-2}{k}m}\\
&= \frac{k}{(n-k-1)m}\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j=1}}\frac{1}{d_id_j}\\
&\quad + \left(\frac{2k}{(n-k-1)m} + \frac{k(k-1)}{(n-k)(n-k-1)m}\right)\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j}\\
&= \frac{k}{(n-k-1)m}\cdot\left(\sum_{\substack{(i,j)\in G\\ d_i>1,d_j=1}}\frac{1}{d_id_j} + \left(2+\frac{k-1}{n-k}\right)\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j}\right)
\end{aligned}$$

For simplicity, we denote $\delta' = \sum_{\substack{(i,j)\in G\\ d_i>1,d_j=1}}\frac{1}{d_id_j} + (2+\frac{k-1}{n-k})\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j}$, and we get:

$$\mathbb{E}_H(\frac{1}{d_id_j}) \le \mathbb{E}_G(\frac{1}{d_id_j}) + \frac{k}{(n-k-1)m}\delta' \tag{4}$$

Next, we compute $m_{2,H}$ using Equations 3 and 4:

$$\begin{aligned}
m_{2,H} &= \mathbb{E}_H(d_i)\,\mathbb{E}_H(\frac{1}{d_id_j})\\
&\le \frac{n-k-1}{n-1}\,\mathbb{E}_G(d_i)\cdot(\mathbb{E}_G(\frac{1}{d_id_j}) + \frac{k}{(n-k-1)m}\delta')\\
&= \frac{n-k-1}{n-1}\,\mathbb{E}_G(d_i)\,\mathbb{E}_G(\frac{1}{d_id_j}) + \frac{k\delta'\cdot\mathbb{E}_G(d_i)}{(n-1)m}\\
&= \mathbb{E}_G(d_i)\,\mathbb{E}_G(\frac{1}{d_id_j}) + \frac{k\delta'\cdot\mathbb{E}_G(d_i)}{(n-1)m} - \frac{k\,\mathbb{E}_G(d_i)\,\mathbb{E}_G(\frac{1}{d_id_j})}{n-1}\\
&= m_{2,G} + \frac{k\,\mathbb{E}_G(d_i)}{(n-1)m}\cdot(\delta' - m\cdot\mathbb{E}_G(\frac{1}{d_id_j}))\\
&= m_{2,G} + \frac{2k}{n(n-1)}\cdot(\delta' - m\cdot\mathbb{E}_G(\frac{1}{d_id_j})) \quad (\text{as } \mathbb{E}_G(d_i) = \frac{2m}{n})
\end{aligned}$$

Note that $m\cdot\mathbb{E}_G(\frac{1}{d_id_j}) = \sum_{(i,j)\in G}\frac{1}{d_id_j}$, so

$$\begin{aligned}
\delta' - m\cdot\mathbb{E}_G(\frac{1}{d_id_j}) &= \sum_{\substack{(i,j)\in G\\ d_i>1,d_j=1}}\frac{1}{d_id_j} + (2+\frac{k-1}{n-k})\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j} - \sum_{(i,j)\in G}\frac{1}{d_id_j}\\
&= \frac{n-1}{n-k}\cdot\sum_{\substack{(i,j)\in G\\ d_i>1,d_j>1}}\frac{1}{d_id_j} - \sum_{\substack{(i,j)\in G\\ d_i=1,d_j=1}}\frac{1}{d_id_j},
\end{aligned}$$

which finalizes the proof. □

A.2 Proof of Theorem 4.5

Proof. The idea of the proof is similar to that of Theorem 4.1. There are n subgraphs $G_1, G_2, \ldots, G_n$. We construct $H = \bigcup_{i=1}^{n} G_i$, so $m_{3,H} = \frac{\sum_i m_{3,G_i}}{n} = \mathbb{E}(m_{3,G'})$. From Theorem 4.2, $m_3 = \mathbb{E}(\lambda^3) = 2\,\mathbb{E}(\Delta_i)\,\mathbb{E}(\frac{1}{d_id_jd_k})$. In the formula, $\mathbb{E}(\Delta_i)$ is the average number of triads a node is in, and by definition $\mathbb{E}(\Delta_i) = \frac{3\Delta}{n}$, where Δ is the

number of triads of a graph. $\mathbb{E}(\frac{1}{d_i d_j d_k})$ is the expected value of $\frac{1}{d_i d_j d_k}$ over all triads. Hence, $m_{3,H} = 2\,\mathbb{E}_H(\Delta_i)\,\mathbb{E}_H(\frac{1}{d_i d_j d_k})$. For any triad (i, j, k) in G, it has $n-3$ copies in which no member of i, j, k is removed, so $\mathbb{E}_H(\Delta_i) = \frac{3\Delta_H}{|V_H|} = \frac{3(n-3)\Delta_G}{n(n-1)} = \frac{n-3}{n-1}\,\mathbb{E}_G(\Delta_i)$.

To get $\mathbb{E}_H(\frac{1}{d_i d_j d_k})$, we consider four types of triads based on their node degrees.

► **Type I**: $d_i = 2, d_j = 2, d_k = 2$. For such a triad, all of its $n-3$ copies have $\frac{1}{d_i d_j d_k} = \frac{1}{8}$ as neither of i and j can lose other neighbors, so there will be no increment;

► **Type II**: $d_i > 2, d_j = 2, d_k = 2$. For a triad of this type, if the removed node is a neighbor of i, then the triad contributes an increment $\frac{1}{(d_i-1)d_j d_k} - \frac{1}{d_i d_j d_k} = \frac{1}{d_i d_j d_k (d_i-1)}$. Among the $n-3$ copies, there are $d_i - 2$ such cases, so the overall contribution is $(d_i-2)\cdot\frac{1}{d_i d_j d_k (d_i-1)} = \frac{1}{d_i d_j d_k}\cdot\frac{d_i-2}{d_i-1} < \frac{1}{d_i d_j d_k}$.

► **Type III**: $d_i > 2, d_j > 2, d_k = 2$. If the removed node is a neighbor of i but not connected to j (or a neighbor of j but not connected to i) and not including k, the triad contributes an increment $\frac{1}{(d_i-1)d_j d_k} - \frac{1}{d_i d_j d_k} = \frac{1}{d_i d_j d_k (d_i-1)}$ (or $\frac{1}{d_i d_j d_k (d_j-1)}$). If the removed node is a common neighbor of i and j, the increment is $\frac{1}{(d_i-1)(d_j-1)d_k} - \frac{1}{d_i d_j d_k} = \frac{1}{d_i d_j d_k (d_i-1)} + \frac{1}{d_i d_j d_k (d_j-1)} + \frac{1}{d_i d_j d_k (d_i-1)(d_j-1)}$. Assume i and j have c_{ij} common neighbors not including k, then the overall increment of a triad of Type III is $(d_i - 2 - c_{ij})\cdot\frac{1}{d_i d_j d_k (d_i-1)} + (d_j - 2 - c_{ij})\cdot\frac{1}{d_i d_j d_k (d_j-1)} + c_{ij}\cdot(\frac{1}{d_i d_j d_k (d_i-1)} + \frac{1}{d_i d_j d_k (d_j-1)} + \frac{1}{d_i d_j d_k (d_i-1)(d_j-1)}) = \frac{1}{d_i d_j d_k}(\frac{d_i-2}{d_i-1} + \frac{d_j-2}{d_j-1} + \frac{c_{ij}}{(d_i-1)(d_j-1)})$. Note that c varies from 0 to $\min(d_i, d_j) - 2$, so the overall increment is less than $\frac{9}{4}\cdot\frac{1}{d_i d_j d_k}$.

► **Type IV**: $d_i > 2, d_j > 2, d_k > 2$. If the removed node is a neighbor of i but not connected to j or k, the increment is $\frac{1}{d_i d_j d_k}\cdot\frac{1}{d_i-1}$; if the removed node is a common neighbor of i and j but not connected to k, then the increment is $\frac{1}{d_i d_j d_k}\cdot(\frac{1}{d_i-1} + \frac{1}{d_j-1} + \frac{1}{(d_i-1)(d_j-1)})$; if the removed node is a common neighbor of i, j and k, the increment is $\frac{1}{(d_i-1)(d_j-1)(d_k-1)} - \frac{1}{d_i d_j d_k} = \frac{1}{d_i d_j d_k}\cdot(\frac{1}{d_i-1} + \frac{1}{d_j-1} + \frac{1}{d_k-1} + \frac{1}{(d_i-1)(d_j-1)} + \frac{1}{(d_i-1)(d_k-1)} + \frac{1}{(d_j-1)(d_k-1)} + \frac{1}{(d_i-1)(d_j-1)(d_k-1)})$. Similarly, we can get the overall increment for a triad of Type IV: $\frac{1}{d_i d_j d_k}\cdot(\frac{d_i-2}{d_i-1} + \frac{d_j-2}{d_j-1} + \frac{d_k-2}{d_k-1} + \frac{c_{ij}}{(d_i-1)(d_j-1)} + \frac{c_{ik}}{(d_i-1)(d_k-1)} + \frac{c_{jk}}{(d_j-1)(d_k-1)} + \frac{c_{ijk}}{(d_i-1)(d_j-1)(d_k-1)})$, where c_{ij} (or c_{ik}, c_{jk}) is the number of common neighbors of i and j (i and k, j and k), not including i, j, k; and c_{ijk} is the number of common neighbors of i, j and k. The increment is less than $4\cdot\frac{1}{d_i d_j d_k}$.

Therefore, the total increment over all the triad is less than $\delta = \sum_{\text{Type II}}\frac{1}{d_i d_j d_k} + \frac{9}{4}\sum_{\text{Type III}}\frac{1}{d_i d_j d_k} + 4\sum_{\text{Type IV}}\frac{1}{d_i d_j d_k}$ and after normalizing it by $|\Delta_H|$ we get: $\mathbb{E}_H(\frac{1}{d_i d_j d_k}) < \mathbb{E}_G(\frac{1}{d_i d_j d_k}) + \frac{\delta}{(n-3)\Delta_G}$.

Next, we compute $m_{3,H}$:

$$
\begin{aligned}
m_{3,H} &= 2\,\mathbb{E}_H(\Delta_i)\,\mathbb{E}_H(\tfrac{1}{d_i d_j d_k}) \\
&< 2\tfrac{n-3}{n-1}\,\mathbb{E}_G(\Delta_i)(\mathbb{E}_G(\tfrac{1}{d_i d_j d_k}) + \tfrac{\delta}{(n-3)\Delta_G}) \\
&= 2(\tfrac{n-3}{n-1}\,\mathbb{E}_G(\Delta_i)\,\mathbb{E}_G(\tfrac{1}{d_i d_j d_k}) + \tfrac{\mathbb{E}_G(\Delta_i)\delta}{(n-1)\Delta_G}) \\
&= 2(\mathbb{E}_G(\Delta_i)\,\mathbb{E}_G(\tfrac{1}{d_i d_j d_k}) + \tfrac{\mathbb{E}_G(\Delta_i)\delta}{(n-1)\Delta_G} - \tfrac{2}{n-1}\,\mathbb{E}_G(\Delta_i)\,\mathbb{E}_G(\tfrac{1}{d_i d_j d_k})) \\
&= m_{3,G} + \tfrac{2\,\mathbb{E}_G(\Delta_i)}{(n-1)\Delta_G}(\delta - 2\Delta_G\,\mathbb{E}_G(\tfrac{1}{d_i d_j d_k})) \\
&= m_{3,G} + \tfrac{6}{n(n-1)}(\delta - 2\Delta_G\,\mathbb{E}_G(\tfrac{1}{d_i d_j d_k})) \quad (\text{as } E_G(\Delta_i) = \tfrac{3\Delta_G}{n}).
\end{aligned}
$$

Notice that $\Delta_G\,\mathbb{E}_G(\frac{1}{d_i d_j d_k}) = \sum_{(i,j,k)\in G}\frac{1}{d_i d_j d_k}$, so $\delta - 2\Delta_G\,\mathbb{E}_G(\frac{1}{d_i d_j d_k})$ $= \frac{1}{4}\sum_{\text{Type III}}\frac{1}{d_i d_j d_k} + 2\sum_{\text{Type IV}}\frac{1}{d_i d_j d_k} - \sum_{\text{Type II}}\frac{1}{d_i d_j d_k} - 2\sum_{\text{Type I}}\frac{1}{d_i d_j d_k}$. The theorem is proved. □

A.3 Proof of Theorem 4.6

Proof. Similarly, there are n subgraphs $G_1, G_2, \ldots, G_n$. We construct $H = \bigcup_{i=1}^n G_i$, so $m_{4,H} = \frac{\sum_i m_{4,G_i}}{n} = \mathbb{E}(m_{4,G'})$. From Theorem 4.2, $m_4 = [\,\mathbb{E}(d_i) + 4\,\mathbb{E}\binom{d_i}{2} + 2\,\mathbb{E}(\square_i)\,]\,\mathbb{E}(\frac{1}{d_i d_j d_k d_l})$. In the formula, $\mathbb{E}(d_i)$ is the average degree; $\mathbb{E}\binom{d_i}{2}$ is the average number of wedges a node is in, so it equals to $\frac{w}{n}$ where w is the number of wedges; $\mathbb{E}(\square_i)$ is the average number of squares a node is in. $\mathbb{E}(\frac{1}{d_i d_j d_k d_l})$ is the expected value of $\frac{1}{d_i d_j d_k d_l}$ over all the closed walks of length 4. Hence, for graph H, we have $m_{4,H} = [\,\mathbb{E}_H(d_i) + 4\,\mathbb{E}_H\binom{d_i}{2} + 2\,\mathbb{E}_H(\square_i)\,]\,\mathbb{E}_H(\frac{1}{d_i d_j d_k d_l})$.

From Equation 1, we have $\mathbb{E}_H(d_i) = \frac{n-2}{n-1}\cdot\mathbb{E}_G(d_i)$; As each wedge in G has $n-3$ copies in H (when none of the three nodes is removed), $\mathbb{E}_H\binom{d_i}{2} = \frac{w_H}{n(n-1)} = \frac{(n-3)w_G}{n(n-1)} = \frac{n-3}{n-1}\cdot\mathbb{E}_G\binom{d_i}{2}$; Similarly, $\mathbb{E}_H(\square_i) = \frac{n-4}{n-1}\cdot\mathbb{E}_G(\square_i)$ as each square in G has $n-4$ copies in H.

To get $\mathbb{E}_H(\frac{1}{d_i d_j d_k d_l})$, of course we can discuss all the possible closed walks of length 4 (generated by edges, wedges or squares), as we have done in the previous theorems. However, for a simpler exposition, we provide the following upper bound. Notice that $\mathbb{E}_H(\frac{1}{d_i d_j d_k d_l}) \le \mathbb{E}_G(\frac{1}{(d_i-1)(d_j-1)(d_k-1)(d_l-1)})$ and the bound is tight when G is a complete graph ($n \ge 3$) or all of its components are k-cliques ($k \ge 3$), and $\frac{1}{(d_i-1)(d_j-1)(d_k-1)(d_l-1)} = \frac{1}{d_i d_j d_k d_l}\cdot\prod_{x\in\{i,j,k,l\}}\frac{d_x}{d_x-1} \le \frac{16}{d_i d_j d_k d_l}$, so $\mathbb{E}_H(\frac{1}{d_i d_j d_k d_l}) \le 16\cdot\mathbb{E}_G(\frac{1}{d_i d_j d_k d_l})$. Therefore,

$$
\begin{aligned}
m_{4,H} &= [\,\mathbb{E}_H(d_i) + 4\,\mathbb{E}_H\binom{d_i}{2} + 2\,\mathbb{E}_H(\square_i)\,]\,\mathbb{E}_H(\tfrac{1}{d_i d_j d_k d_l}) \\
&\le [\,\tfrac{n-2}{n-1}\,\mathbb{E}_G(d_i) + 4\tfrac{n-3}{n-1}\,\mathbb{E}_G\binom{d_i}{2} + 2\tfrac{n-4}{n-1}\,\mathbb{E}_G(\square_i)\,] \\
&\quad\times 16\cdot\mathbb{E}_G(\tfrac{1}{d_i d_j d_k d_l}) \\
&\le \tfrac{n-2}{n-1}[\,\mathbb{E}_G(d_i) + 4\,\mathbb{E}_G\binom{d_i}{2} + 2\,\mathbb{E}_G(\square_i)\,]\times 16\cdot\mathbb{E}_G(\tfrac{1}{d_i d_j d_k d_l}) \\
&= \tfrac{16(n-2)}{n-1}m_{4,G}
\end{aligned}
$$

□