

The Spectral Zoo of Networks: Embedding and Visualizing Networks with Spectral Moments

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

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

ABSTRACT

Network embedding methods have been widely and successfully used in network-based applications such as node classification and link prediction. However, an ideal network embedding should not only be useful for machine learning, but interpretable. We introduce a spectral embedding method for a network, its *Spectral Point*, which is basically the first few spectral moments of a network. Spectral moments are interpretable, where we prove their close relationships to network structure (e.g. number of triangles and squares) and various network properties (e.g. degree distribution, clustering coefficient, and network connectivity). Using spectral points, we introduce a visualizable and bounded 3D embedding space for all possible graphs, in which one can characterize various types of graphs (e.g., cycles), or real-world networks from different categories (e.g., social or biological networks). We demonstrate that spectral points can be used for network identification (i.e., what network is this subgraph sampled from?) and that by using just the first few moments one does not lose much predictive power.

KEYWORDS

Network Representation, Network Embedding, Spectral Graph Theory, Graph Spectrum, Network Visualization

ACM Reference Format:

Shengmin Jin and Reza Zafarani. 2020. The Spectral Zoo of Networks: Embedding and Visualizing Networks with Spectral Moments. In *The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'20)*, Aug 22–27, 2020, San Diego, CA, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Networks have been widely used for modeling complex data from science (e.g. interaction between proteins in biology [37]), engineering (e.g. power grids [24]), and our daily life (e.g. friendships in a social network [2]). An appropriate representation of networks plays a critical role in studying networks. An ideal network representation should not only be informative for machine learning use, but also be able to help users understand the network. Recent advancements in network representation such as node (or graph)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD '20, Aug 22–27, 2020, San Diego, CA, USA

© 2020 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

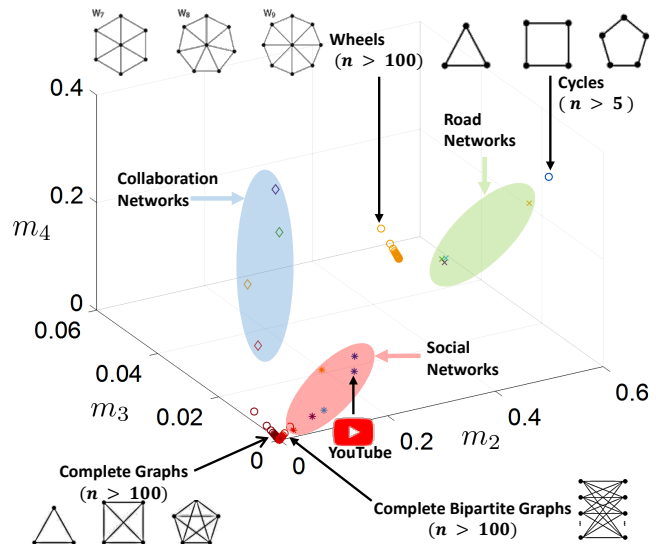


Figure 1: The Zoo of Networks

embedding techniques aim to learn a mapping from a graph, or its nodes, to points in a low-dimensional vector space [12, 13, 29, 36]. These methods have been successful in many network-based applications. However, these learned representations are often difficult to understand, mainly because of two reasons: (1) node (or graph) embedding techniques map nodes (or graphs) to points in a k -dimensional space, but no interpretation is often provided for each of these k dimensions. Particularly, one can hardly get exact structural properties from these k -dimensional vectors, for example, are there many triangles in the network? Or, is it a well-connected network?; (2) in practical machine learning tasks, the number of dimensions k is often set to a number such as 128 or 256. Though this k -dimensional space is a relatively low-dimensional compared to the high-dimensional network data (i.e., adjacency matrix), it is impractical for users to search among all these k dimensions to gain some insights from the network. Here, our goal is to address such issues by proposing an interpretable, visualizable, and compact embedding space for networks: “a zoo of networks”, where one can easily characterize networks based on their locations in this zoo, similar to how animals are grouped and located in different regions of a zoo. Figure 1 illustrates a big picture of this zoo of networks obtained using the method developed in this paper. In the figure, we plot the embeddings of real-world networks from three different categories: Social Networks, Collaboration Networks and Road Networks, and also plot various types of graphs including complete graphs, cycles, complete bipartite graphs, and wheels of different sizes (the number of nodes n are mostly above 100).

To build such an embedding space, we need the network embedding method to meet the following criteria: (1) **Easy to visualize**. We want an embedding space that can be easily visualized, so a 3D embedding of networks is needed; (2) **Capture network structure**. The embedding values should help users understand the network structure; (3) **Capture network properties**. The embedding values should shed light on different network properties such as the degree distribution or network connectivity; and (4) **Easy and fast to compute**. The method should be scalable for large networks. Spectral graph theory can help satisfy these constraints.

Spectral graph theory connects the structure of a network to the eigenvalues and eigenvectors of its associated matrices such as the adjacency matrix or the Laplacian. The extreme eigenvalues and associated eigenvectors are often used by various spectral methods. For example, the ratio between the largest and smallest eigenvalues can help estimate the chromatic number [7, 14]; the second-smallest eigenvalue of a graph Laplacian is related to graph connectivity and the associated eigenvector is used for spectral clustering [23]. Recently, more attention is paid to the overall distribution of eigenvalues, also known as the *spectral density* of the graph. Dong et al. [9] use methods from condensed matter physics to study spectral densities in networks, and they show that the spectral density is a practical tool to analyze large real-world networks. Inspired by the strong power of spectral density analysis, we aim to find a succinct way to represent the spectral density, so as to represent the network. Naturally, we propose to use spectral moments, as in statistics, moments are often used to capture the shape of a distribution. Specifically, we use the spectral moments of the random walk transition matrix as the embedding method, as (1) they have a very clear meaning, which is the expected return probability of a random walk; (2) these spectral moments, as we will see in the rest of the paper, are closely related to the network structure and various network properties such as the degree distribution and clustering coefficient; (3) using a few of these spectral moments (*truncated moments*), more specifically, the second m_2 , third m_3 , and fourth m_4 moments (the first moment is 0 as we only look into undirected graphs without self-loops), we can have a 3D embedding of a network that can be visualized. We will show that the error of using truncated spectral moments is bounded. We denote this 3D embedding as the *spectral point* of the network; (4) by definition, these spectral moments are all between 0 to 1, so we can have a compact embedding space for all possible graphs, which is a $1 \times 1 \times 1$ cube (see Figure 1). The points in Figure 1 are spectral points. In Section 5, we will go into further details on how graphs can be represented using spectral points.

Overall, our contributions are mainly the following:

1. Network Embedding with Spectral Moments. We introduce *Spectral Point*, a 3D network embedding method that uses the truncated spectral moments of the network. Spectral points have the following advantages: (i) each dimension is closely related to the network structure and various network properties, so it is easy to interpret; (ii) the embedding space can help characterize various types of networks; (iii) the embedding space provides easy network visualization; and (iv) the embeddings are easy to compute.

2. Spectral Moments and Network Structure. To the best of our knowledge, we are the first to study relationship between the

spectral moments of the random walk transition matrix (more importantly and equivalently, the spectral moments of the normalized Laplacian matrix) and network structure. We connect the spectral moments to basic subgraphs such as triangles and squares.

3. Spectral Moments and Network Properties. We find that the spectral moments provide various bounds on network properties such as the degree distribution and the global clustering coefficient. We define a measure that assesses global connectivity using the spectral moments, and we prove the relationship between the spectral moments of a network and those of its connected components.

4. Representing Various Graph Types. We mathematically derive spectral moments for various types of graphs such as complete graphs, cycles, star graphs, and complete bipartite graphs. For k -regular graphs, we derive the exact value for the second moment and the range of values that the third and fourth moments can take.

5. Representing Real-World Graphs. We compute the spectral moments of real-world graphs from various categories and show that their structure and properties identified in past research is captured by spectral moments. We show that spectral moments can help get a quick understanding of a real-world network at hand.

6. Spectral Network Identification. We demonstrate that spectral moments can be used for network identification, i.e., identifying the source of an anonymized graph. Our results indicate that truncated spectral moments do not lose much predictive power.

The rest of the paper is organized as follows. We first detail the preliminaries and notations used in the paper in Section 2. In Section 3, we provide proofs on the relationship between spectral moments and network structure, and Section 4 demonstrates the relationship between spectral moments and network properties. In Section 5, we analyze special graphs and real-world networks using their spectral moments. In Section 6, we use spectral moments for network identification. After reviewing additional related work in Section 7, we conclude the paper in Section 8.

2 PRELIMINARIES AND NOTATION

For an undirected graph $G = (V, E)$ with vertices $V = \{v_1, v_2, \dots, v_n\}$ 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. $D_{ij} = \begin{cases} d_i & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$. Various properties of graph G (e.g., connectivity or cuts) can be identified using matrices defined in terms of A and D . 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 normalized Laplacian has a bounded spectrum, i.e. $0 = \mu_1 \leq \mu_2 \leq \dots \leq \mu_n \leq 2$, where μ_i 's are the eigenvalues of L . The transition matrix of the random walk on G is matrix $P = A D^{-1}$. As P is a stochastic matrix, its spectrum is also bounded: $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq -1$, where λ_i 's are the eigenvalues of P . 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 $1 \leq i \leq n$.

In this work, we denote the i -th spectral moment m_i of a graph G using the spectrum of its random walk transition matrix P , $m_i = \sum_{j=1}^n \lambda_j^i$, as $\frac{1}{n} \sum_{j=1}^n \lambda_j^i = \text{Tr}(P^i)$. We look at the relationship between the spectral moments and network structure and use the first few spectral moments to represent networks.

3 RELATIONSHIP BETWEEN SPECTRAL MOMENTS AND NETWORK STRUCTURE

In this section, we aim to see how spectral moments are related to network structure. As we only look into undirected graphs without self-loops, it is easy to see that the first spectral moment m_1 is 0. Therefore, we start with the second spectral moment.

3.1 Second Spectral Moment

THEOREM 3.1. *The 2nd spectral moment m_2 of P is*

$$m_2 = E^1 \lambda^{2^0} = E^1 \mathbf{1} \cdot l^0;$$

where l follows $p^1 l j k^0$, the probability that a random neighbor of a node with degree k has degree l .

PROOF. The ℓ -th spectral moment of P can be viewed as the expected return probability of an ℓ -step random walk starting from node i , where i is chosen uniformly at random from all the nodes [8]. For any node i , its return probability of a 2-step random walk is equal to $\sum_{j:j \sim i} \frac{1}{d_i d_j}$, where d_i and d_j are the degrees of nodes i and j , respectively, and $j \sim i$ means that j is a neighbor of i . Hence, the 2nd spectral moment of P is equal to $E_{i \sim j} \sum_{j:j \sim i} \frac{1}{d_i d_j}$, and

$$\begin{aligned} E_{i \sim j} \sum_{j:j \sim i} \frac{1}{d_i d_j} &= E_{i \sim j} \sum_{j:j \sim i} \frac{1}{d_j} \\ &= E_{i \sim j} \sum_{j:j \sim i} \frac{1}{d_j} E_{j \sim i} \frac{1}{d_j} \\ &= E_{i \sim j} E_{j \sim i} \frac{1}{d_j} \end{aligned}$$

Note that d_i follows the degree distribution of the graph $p^1 k^0$, and d_j follows the conditional degree distribution $p^1 l j k^0$ as j is constrained to be a neighbor of i , so

$$E_{i \sim j} E_{j \sim i} \frac{1}{d_j} = E_{i \sim j} \sum_{d_j} \frac{1}{d_j} p^1 l = d_j j k = d_i^0 \quad (1)$$

$$= \sum_{d_j} p^1 k = d_i^0 \sum_{d_j} \frac{1}{d_j} p^1 l = d_j j k = d_i^0 \quad (2)$$

$$= \sum_{d_i} \sum_{d_j} p^1 k = d_i^0 \sum_{d_j} \frac{1}{d_j} p^1 l = d_j j k = d_i^0 \quad (3)$$

$$= \sum_{d_i} \sum_{d_j} \frac{1}{d_j} p^1 l = d_j j k = d_i^0 \quad (4)$$

$$= E^1 \frac{1}{d_j} \quad (5)$$

Note that in Theorem 3.1, $p^1 l j k^0$ is the joint degree distribution of nodes (k) and their neighbors (l) and the 2nd spectral moment is $E^1 \frac{1}{d_j}$ under this distribution. This joint distribution is not symmetric, i.e., $p^1 l = d_j j k = d_i^0 \neq p^1 l = d_i j k = d_j^0$, and can be cumbersome to compute; hence, the following theorem states the connection between the 2nd spectral moment P and the joint degree distribution $p^1 d_i j$, which is symmetric.

THEOREM 3.2. *The 2nd spectral moment m_2 of P is*

$$m_2 = E^1 \lambda^{2^0} = E^1 d_i^0 E^1 \frac{1}{d_i d_j};$$

where $E^1 d_i^0$ denotes the average degree in the graph and $d_i d_j$ follows the joint degree distribution $p^1 d_i j$; d_j^0 : the probability that a node with degree d_j is connected to another node with degree d_j .

PROOF. Denote the joint degree distribution as

$$p^1 d_i j; d_j^0 = \frac{n_{d_i; d_j}}{E^1 d_i^0 n}; \quad (6)$$

where $n_{d_i; d_j}$ is the number of edges between nodes with degree d_i and nodes with degree d_j , n is the total number of nodes in the graph, and $E^1 d_i^0$ is the average degree. Let n_{d_i} denote the number of nodes with degree d_i . Then, $p^1 l = d_j j k = d_i^0 p^1 k = d_i^0$ can be stated as

$$p^1 l = d_j j k = d_i^0 p^1 k = d_i^0 \quad (7)$$

$$= \frac{n_{d_i; d_j}}{d_i n_{d_i}} p^1 k = d_i^0 \quad (8)$$

$$= \frac{p^1 d_i j; d_j^0 E^1 d_i^0 n}{d_i n_{d_i}} p^1 k = d_i^0 \quad (9)$$

$$= \frac{p^1 d_i j; d_j^0 E^1 d_i^0}{d_i p^1 k = d_i^0} p^1 k = d_i^0 \quad (10)$$

$$= \frac{p^1 d_i j; d_j^0 E^1 d_i^0}{d_i}; \quad (11)$$

using which Equation (4) can be restated as

$$\sum_{d_i} \sum_{d_j} \frac{1}{d_j} p^1 l = d_j j k = d_i^0 = \sum_{d_i} \sum_{d_j} \frac{1}{d_j} \frac{p^1 d_i j; d_j^0 E^1 d_i^0}{d_i} \quad (12)$$

$$= E^1 d_i^0 E^1 \frac{1}{d_i d_j}; \quad (13)$$

One can interpret Theorem 3.2 in this way: the expected return probability of a 2-step random walk is equal to (I) the average return probability through an edge, multiplied by (II) the average number of edges a node has, as $\frac{1}{d_i d_j}$ is the return probability of a 2-step random walk through a specific edge linking two nodes with degrees d_i and d_j , respectively, and $E^1 \frac{1}{d_i d_j}$ is the average return probability over all edges. This observation motivates us to extend Theorem 3.2 to higher moments.

3.2 Third Spectral Moment

THEOREM 3.3. *The 3rd spectral moment m_3 of P is*

$$m_3 = E^1 \lambda^{3^0} = 2 E^1 i^0 E^1 \frac{1}{d_h d_i d_j};$$

where $E^1 i^0$ is the average number of triads a node is in and $d_h d_i d_j$ follows the joint degree distribution of triads $p^1 d_h i j$; d_j^0 : the probability that a triad is formed by nodes with degrees d_h , d_i , and d_j .

PROOF. As mentioned, the 3rd spectral moment of P can be viewed as the expected return probability of a 3-step random walk, which is equal to the summation of the return probability of a 3-step random walk starting from any node i divided by the number of nodes. Assume that nodes h , i , and j are connected to each other and form a triad. The triad will increase the overall return probability by $6 \frac{1}{d_h d_i d_j}$ as it includes 6 closed walks: $h \rightarrow i \rightarrow j \rightarrow h$, $h \rightarrow j \rightarrow i \rightarrow h$, and so on. Denote n as the total number of triads

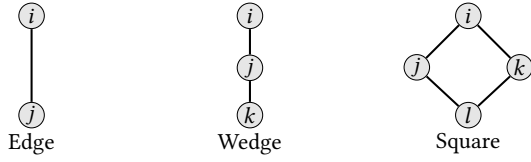


Figure 2: Structures related to the 4th spectral moment of P .

in the graph, and there are $p^1 d_h; d_i; d_j^0$ triads with nodes having degree $d_h; d_i$ and d_j . Therefore,

$$m_3 = E^1 \lambda^{3^0} = \frac{\int d_h; d_i; d_j \frac{6}{d_h d_i d_j} p^1 d_h; d_i; d_j^0}{n} \quad (14)$$

By definition, $E^1 i^0 = \frac{3}{n}$, so

$$m_3 = 2 E^1 i^0 E^1 \frac{1}{d_h d_i d_j}^0. \quad (15)$$

3.3 Higher-Order Spectral Moments

The proof for Theorem 3.3 can be extended to a general case:

THEOREM 3.4. *The ℓ -th spectral moment m_ℓ of P is*

$$m_\ell = E^1 \lambda^{\ell^0} = E^1 CW_{\ell; i^0} E^1 \frac{1}{d_1 d_2 \dots d_{\ell-1} d_\ell}^0;$$

where $E^1 CW_{\ell; i^0}$ denotes the average number of closed walks of length ℓ a node is in and $d_1 d_2 \dots d_{\ell-1} d_\ell$ follows the joint degree distribution of closed walk of length ℓ formed by nodes with degrees d_1, d_2, \dots, d_ℓ .

PROOF. The proof is straightforward:

$$E^1 \lambda^{\ell^0} = \frac{CW_{\ell; i^0} E^1 \frac{1}{d_1 d_2 \dots d_{\ell-1} d_\ell}^0}{n} \quad (16)$$

$$= E^1 CW_{\ell; i^0} E^1 \frac{1}{d_1 d_2 \dots d_{\ell-1} d_\ell}^0; \quad (17)$$

where $CW_{\ell; i^0}$ is the total number of closed walks of length ℓ .

Basically, when ℓ is small, we can connect the ℓ -th spectral moment of P with the local structure of the graph. Next, we look at the case of $\ell = 4$.

THEOREM 3.5. *The 4th spectral moment m_4 of P is*

$$m_4 = E^1 \lambda^{4^0} = E^1 d_i^0 + 4 E \frac{d_i}{2} + 2 E^1 i^0 E^1 \frac{1}{d_i d_j d_k d_l}^0;$$

where $E^1 d_i^0$ is average degree, $E^1 i^0$ 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 .

PROOF. Figure 2 provides the three graph structures that result in closed walks of length 4: an edge, a wedge, and a square. Each edge contributes 2 closed walks; each wedge adds another 4 closed walks (without considering the walks that go through only one edge); similarly, each square contributes additional 8 closed walks. Let $|E|$ denote the number of edges, $|W|$ denote the number of wedges, and $|S|$ denote the number of squares in the graph. Hence,

$$E^1 CW_{4; i^0} = \frac{2|E| + 4|W| + 8|S|}{n} \quad (18)$$

As $E^1 d_i^0 = \frac{2|E|}{n}$, $|W| = \sum d_i \frac{d_i}{2} n d_i = n E \frac{d_i}{2}$ so $\frac{4|W|}{n} = 4 E \frac{d_i}{2}$, and $E^1 i^0 = \frac{4|S|}{n}$, using Theorem 3.4, the theorem is proved.

4 RELATIONSHIP BETWEEN SPECTRAL MOMENTS AND NETWORK PROPERTIES

4.1 Degree Distribution

Degree distribution plays a vital role in network analysis. Different degree distributions have helped define different network families. Perhaps the most well-studied network family with connections to real-world networks are scale-free networks that exhibit a power law degree distribution, at least asymptotically [3, 5]. Degree distribution is often used to assess how well a network model fits real-world networks. Here, we look into the relationship between the spectral moments and the degree distribution, and we prove that the degree distribution helps bound the spectral moments.

THEOREM 4.1. *The 2nd spectral moment m_2 of P is lower bounded by the degree distribution's expectation $E^1 d_i^0$ and variance $\text{Var}^1 d_i^0$:*

$$m_2 \geq \frac{E^3 d_i^0}{E^2 d_i^0 + \text{Var}^1 d_i^{002}};$$

PROOF. From Theorem 3.2, we have $m_2 = E^1 d_i^0 E^1 \frac{1}{d_i d_j}^0$. By Jensen's inequality, $E^1 \frac{1}{d_i d_j}^0 \geq \frac{E^2 d_i^0 E^2 d_j^0}{E^2 d_i d_j^0}$, and

$$E^1 \frac{1}{d_i d_j}^0 \geq \frac{E^1 d_i + d_j}{2|E|} = \frac{E^1 d_i^0 + E^1 d_j^0}{2|E|} \quad (19)$$

$$= \frac{E^1 d_i^0}{2|E|} \quad (\text{as each } d_i \text{ gets counted } d_i \text{ times}) \quad (20)$$

$$= \frac{n E^1 d_i^0}{2|E|} = \frac{E^1 d_i^0}{E^1 d_i^0}; \quad (21)$$

Therefore, $E^1 \frac{1}{d_i d_j}^0 \geq \frac{E^2 d_i^0}{E^2 d_i^0}$ and $m_2 \geq \frac{E^3 d_i^0}{E^2 d_i^0 + \text{Var}^1 d_i^0}$. As $E^1 d_i^0 = \frac{E^2 d_i^0}{E^1 d_i^0 + \text{Var}^1 d_i^0}$.

In Theorem 4.1, if $\text{Var}^1 d_i^0 = 0$, the bound is tight: $m_2 = \frac{1}{E^1 d_i^0}$. The bound can be stated as $m_2 \geq \frac{E^3 d_i^0}{E^2 d_i^0 + \text{Var}^1 d_i^{002}} = \frac{1}{E^1 d_i^0 + \frac{2 \text{Var}^1 d_i^0}{E^1 d_i^0} + \frac{\text{Var}^2 d_i^0}{E^3 d_i^0}}$.

Hence, given a fixed average degree $E^1 d_i^0$, a smaller degree variance $\text{Var}^1 d_i^0$ leads to a greater lower bound for m_2 . In statistics, the value $\frac{\text{Var}^1 d_i^0}{E^1 d_i^0}$ is defined as the *index of dispersion* D for the degree distribution, so we can rewrite the bound as $m_2 \geq \frac{1}{E^1 d_i^0 + 2D + \frac{D^2}{E^1 d_i^0}}$.

4.2 Clustering Coefficient

The clustering coefficient helps measure the degree to which nodes in a graph tend to cluster together. Clustering coefficient is specifically used to analyze transitivity (triangles) in an undirected graph. Theorem 3.3 states that $m_3 = 2 E^1 i^0 E^1 \frac{1}{d_h d_i d_j}^0$, indicating that m_3 is closely related to the number of triangles. Next, we prove that m_3 and the degree distribution's expectation and variance ($E^1 d_i^0$ and $\text{Var}^1 d_i^0$) provide a lower bound for the global clustering coefficient C . Also, m_3 is upper bounded by C , $E^1 d_i^0$ and $\text{Var}^1 d_i^0$.

COROLLARY 4.1.1. *The 3rd moment m_3 is upper bounded by $\frac{1}{4} E^1 i^0$.*

PROOF. Because any node in a triangle has a degree of at least 2, $E^1 \frac{1}{d_h d_i d_j}^0 \geq \frac{1}{8}$. Therefore, $m_3 = 2 E^1 i^0 E^1 \frac{1}{d_h d_i d_j}^0 \geq \frac{1}{4} E^1 i^0$.

Theorem 4.2. Given the 3rd spectral moment m_3 of P , $E^1 d_i^0$ and $\text{Var}^1 d_i^0$, the global clustering coefficient C can be lower bounded as

$$C \geq \frac{8m_3}{E^2 d_i^0 E^1 d_i^0 + \text{Var}^1 d_i^0}$$

Proof. Starting from the definition of clustering coefficient C :

$$C = \frac{\sum_i \frac{j_i(j_i-1)}{2} \frac{2}{j_i} \frac{1}{2} \frac{1}{2}}{\sum_i \frac{j_i(j_i-1)}{2} \frac{1}{2} \frac{1}{2}} = \frac{2 \sum_i \frac{j_i(j_i-1)}{2} \frac{1}{2}}{\sum_i \frac{j_i(j_i-1)}{2} \frac{1}{2}} \quad (22)$$

$$= \frac{E^1 i^0}{E^1 d_i^0} \quad (23)$$

Using Corollary 4.1.1, we get $\frac{C}{4} E^1 d_i^0 \geq \frac{4m_3}{E^1 d_i^0}$. As

$$E^1 d_i^0 = \frac{E^1 d_i^0 E^1 d_i^0}{2} = \frac{E^2 d_i^0 E^1 d_i^0 + \text{Var}^1 d_i^0}{2}$$
, the proof is complete.

Therefore, given $\text{Var}^1 d_i^0$ and $\text{Var}^1 d_i^0$, when m_3 is large, C has a greater lower bound. Also, m_3 can be upper bounded by $E^1 d_i^0$ and $\text{Var}^1 d_i^0$ using Theorem 4.2. $m_3 \leq \frac{C}{8} E^1 d_i^0 E^2 d_i^0 + \text{Var}^1 d_i^0$. So, given $\text{Var}^1 d_i^0$ and $\text{Var}^1 d_i^0$, when the global clustering coefficient is small, m_3 has a smaller upper bound, which makes sense as the number of closed walk of length 3 relies on the number of triangles.

4.3 Connectivity

Because C is the expected return probability of an n -step random walk, it is equal to the fraction of walks of length that are closed. Hence, the graph is globally well-connected when C is small, as it is more likely to have a walk which travels far away from the starting node. On the other hand, a large C indicates that more walks are closed and it is difficult to travel from a node to another one which is far away. Here, we link the spectral moments to graph connectivity by extending the Estrada index [10]. The Estrada index of a graph G is defined as $EEG^0 = \sum_{j=1}^n e^{j^2}$, where j 's are the eigenvalues of the adjacency matrix A . As $EEG^0 = \text{trace} e^{A^0} = \sum_{k=0}^{\infty} \frac{\text{trace} A^k}{k!}$, the Estrada index actually counts the number of closed walks, discounting longer walks; hence, it is sometimes used to measure the global connectivity of a graph. Therefore, we propose a variation of the Estrada index using the random walk transition matrix P , which we denote as $EE^1 P^0 = \sum_{j=1}^n e^{j^2} = \sum_{k=0}^{\infty} \frac{\text{trace} P^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{i=1}^n m_k$. Different from the Estrada index $EE^1 P^0$ computes the expected return probability of a random walk of any length, discounting longer walks. The smaller the $EE^1 P^0$ value, the more well-connected the graph.

4.4 Connected Components

In Section 4.3, we discussed the spectral moments and network connectivity. Here, we look into the relationship between the spectral moments of a network and those of its connected components.

Theorem 4.3. Consider graph $G = \cup V; E^0$ with k connected components $G_1; G_2; \dots; G_{k-1}; G_k$. For each $G_i = \cup V_i; E_i^0$, denote its s -th spectral moment as m_s^i . Then, the s -th spectral moment m_s is the weighted average of m_s^i 's weighted by $|V_i|$'s, i.e. $m_s = \frac{\sum_i |V_i| m_s^i}{|V|}$.

Proof. The theorem can be proved in two ways, both of which are straightforward. The first way is that one can view the transition matrix of the random walk on G as a block matrix where each

block represents the transition matrix of a connected component. The second way is that a walk starting from node i cannot reach the nodes in other connected components, so the overall return probability of an s -step random walk is the weighted sum of the expected return probability for each connected component.

5 REPRESENTING NETWORKS WITH SPECTRAL MOMENTS

The results in Sections 3 and 4 show that spectral moments of a network are closely related to its structure and various properties. As discussed, we propose to represent a graph as a point in the 3D space using its truncated spectral moments $(m_2; m_3; m_4)$, where we denote this point as the spectral point of the graph.

5.1 Spectral Points of Various Types of Graphs

In this section, we first explore the spectral points of various types of graphs including complete graphs, cycles, star graphs, complete bipartite graphs $K_{a;b}$, and wheels, which as we will show their spectral moments are functions of the number of nodes n . We derive the spectral moments for all such graphs. Table 1 lists the spectral moments of these graphs and the related information. We observe that (1) For complete graphs, the three spectral moments decreases when the number of nodes increases and when they all converge to 0; (2) For cycles, the spectral point is fixed at $(0.5, 0, 0.375)$, independent of n ; (3) Star graphs and complete bipartite graphs share the same spectral moments, as star graphs are a special case of complete bipartite graphs. Their third spectral moment is 0 and the other two moments are $\frac{2}{n}$. For wheels, the three spectral moments decreases when the number of nodes increases and when $n \rightarrow \infty$ they converge to $(\frac{2}{3}, 0, \frac{2}{27})$. We vary n from 100 to 10,000 and we plot the spectral points of these types of graphs in Figure 3. As expected, we find that (1) complete graphs, stars, and complete bipartite graphs are lines converging to $(0, 0, 0)$; (2) wheels form a line which converges to $(\frac{2}{3}, 0, \frac{2}{27})$.

Next, we look into spectral points of the k -regular graphs ($k > 0$), where each node has the same degree k . We find that the three spectral moments of k -regular graphs are upper bounded by

Theorem 5.1. For k -regular graph, $(m_2; m_3; m_4)$ satisfies

$$m_2 = \frac{1}{k}; \quad (24)$$

$$m_3 = \frac{k-1}{k^2} < \frac{1}{k}; \quad (25)$$

$$m_4 = \frac{1}{k}; \quad (26)$$

Proof. Form m_2 : as $E^1 d_i^0 = k$ and $E^1 \frac{1}{d_i d_j} = \frac{1}{k^2}$, $m_2 = \frac{1}{k}$.

For m_3 : clearly, $E^1 \frac{1}{d_i d_j d_k} = \frac{1}{k^3}$. As each node has k neighbors, it is in at most $\frac{k}{2}$ triangles, i.e. $E^1 i^0 = \frac{k}{2}$. Hence $m_3 = 2 E^1 i^0 E^1 \frac{1}{d_i d_j d_k} = \frac{k-1}{k^2} < \frac{1}{k}$.

Form m_4 : $E^1 \frac{1}{d_i d_j d_k d_l} = \frac{1}{k^4}$, and based on Theorem 4.1 [6] the total number of closed walks of length 4 is $4nk^3$, so $E^1 CW_{4;i}^0 = 4nk^3$. Using Theorem 3.4 [4], $\frac{1}{k}$.

Corollary 5.1.1. The spectral points for k -regular graphs with $k \geq 2$ are within the $(\frac{2}{3}, 0; 0.5, 0; 0.5, 0)$ cube.

Table 1: Spectral Moments of Various Types of Graphs with n nodes ($n > 5$)

Graphs	$E^1 d_1^0$	$E^1 \frac{1}{d_i d_j^0}$	2 nd moment m_2	$E^1 i^0$	$E^1 \frac{1}{d_i d_j d_k^0}$	3 rd moment m_3	$E^1 d_2^1$	$E^1 i^0$	$E^1 \frac{1}{d_i d_j d_k d_l^0}$	4 th moment m_4
Complete Graphs	$n-1$	$\frac{1}{n-1}$	$\frac{1}{n-1}$	$n-1$	$\frac{1}{n-1}$	$\frac{n-2}{n-1}$	$\frac{n-1}{2}$	$3 \frac{n-1}{3}$	$\frac{1}{n-1}$	$\frac{n^2-3n+3}{n-1}$
Cycles	2	$\frac{1}{4}$	$\frac{1}{2}$	0	NA	0	1	0	$\frac{1}{16}$	$\frac{3}{8}$
Star Graphs	$\frac{2n-1}{n}$	$\frac{1}{n-1}$	$\frac{2}{n}$	0	NA	0	$\frac{1}{2n}$	0	$\frac{1}{n-1}$	$\frac{2}{n}$
Complete Bipartite Graph $K_{a,b}$	$\frac{2ab}{a+b}$	$\frac{1}{ab}$	$\frac{2}{a+b} = \frac{2}{n}$	0	NA	0	$\frac{ab^2+a^2b}{2n}$	$\frac{4ab^2}{2n}$	$\frac{1}{a^2b^2}$	$\frac{2}{a+b} = \frac{2}{n}$
Wheels	$\frac{4n-1}{n}$	$\frac{n+2}{18n-1}$	$\frac{2(n+2)}{9n}$	$\frac{3n-1}{n}$	$\frac{1}{9n-1}$	$\frac{2}{3n}$	$\frac{1}{2n}$	$\frac{1}{n}$	$\frac{1}{27n-1}$	$\frac{2n+20}{27n}$

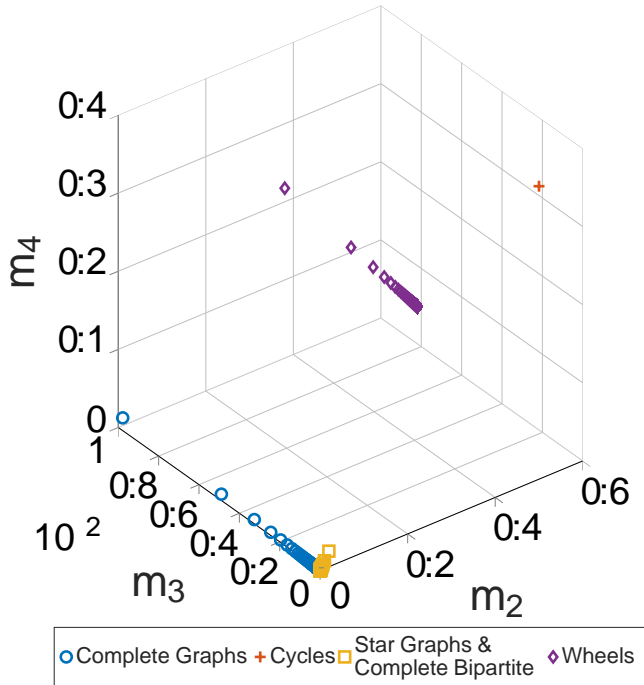


Figure 3: Spectral Moments of Various Types of Graphs

5.2 Representing Real-World Networks

Next, we move to real-world networks. We first introduce the datasets used in our experiments.

5.2.1 Datasets: For our experiments, we use twenty real-world networks from four general network categories: social networks, collaboration networks, road networks, and biological networks. The data statistics are in Table 2.

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

- (1) Brightkite [19]: was a location-based social networking site where users shared their locations by checking-in.
- (2) Flixster [38]: a social movie site allowing users to buy, rent, or watch movies, share ratings, and find new movies.
- (3) Gowalla [19]: similar to Brightkite, was a location-based social networking site where users shared their locations.
- (4) Hyves [38]: the most popular social networking site in the Netherlands with mainly Dutch visitors. It competes with sites such as Facebook and MySpace in that country.
- (5) Livejournal [40]: a social network where users can keep a blog or journal. Users can form friendship or follow others. Here, edges represent friendships (undirected).

Table 2: Dataset Statistics

Type	Network	$\sum_j j = n$	$\sum_j E_j = m$	Average Degree	Density ($\times 10^{-4}$)	Average Clustering Coefficient
Social Networks	Brightkite	58,228	214,078	7.353	1.263	0.1723
	Flixster	2,523,386	7,918,801	6.276	0.025	0.0834
	Gowalla	196,591	950,327	9.668	0.246	0.2367
	Hyves	1,402,673	2,777,419	3.960	0.028	0.0448
	Livejournal	3,017,286	85,654,976	56.78	0.188	0.1196
	MySpace	854,498	5,635,296	13.19	0.154	0.0433
	Orkut	3,072,441	117,185,083	76.28	0.248	0.1666
	YouTube	1,134,890	2,987,624	5.265	0.046	0.0808
Collaboration Networks	Astro-Ph	18,772	198,050	21.10	11.24	0.6306
	Cond-Mat	23,133	93,439	8.078	3.492	0.6334
	Gr-Qc	5,242	14,484	5.526	10.54	0.5296
	Hep-Th	9,877	25,973	5.259	5.324	0.4714
Road Networks	Road-BEL	1,441,295	1,549,970	2.143	0.014	0.0017
	Road-CA	1,965,206	2,766,607	2.816	0.014	0.0464
	Road-PA	1,088,092	1,541,898	2.834	0.026	0.0465
	Road-TX	1,379,917	1,921,660	2.785	0.020	0.0470
Biological Networks	Bio-Dmela	7,393	25,569	6.917	9.356	0.0119
	Bio-Grid-Human	9,527	62,364	13.09	13.74	0.1094
	Bio-Grid-Yeast	5,870	313,890	106.9	177.2	0.0516
	Human-Brain	177,600	15,669,036	176.4	9.910	0.4580

- (6) MySpace [40]: a social network having a significant influence on pop culture and music.
- (7) Orkut [19]: was a social networking website owned and operated by Google, shutdown in 2014.
- (8) YouTube [19]: a video-sharing site with a social network.

Collaboration Networks: We include four collaboration networks from arXiv.org, which include scientific collaborations between authors with different scientific interests. In a collaboration network, an undirected edge between nodes i and j exists, if authors i and j have co-authored at least one paper.

- (9) Astro-Ph [19]: Astrophysics.
- (10) Cond-Mat [19]: Condensed matter physics.
- (11) Gr-Qc [19]: General relativity and quantum cosmology.
- (12) Hep-Th [19]: High energy physics theory.

Road Networks: We include four road networks. In road networks, nodes are intersections/endpoints and undirected edges are the roads connecting these intersections/road endpoints.

- (13) Road-BEL [19]: the OpenStreetMap road network of Belgium.
- (14) Road-CA [19]: the road network of California.
- (15) Road-PA [19]: the road network of Pennsylvania.
- (16) Road-TX [19]: the road network of Texas.

Biological Networks: We include four biological networks.

- (17) Bio-Dmela [28]: a protein-protein interaction (PPI) network.
- (18) Bio-Grid-Human [28]: a PPI network.
- (19) Bio-Grid-Yeast [28]: a PPI network.
- (20) Human-Brain [28]: the network of human brain.

5.2.2 Spectral Points of Real-World Networks: Figure 4 plots the spectral points of real-world networks. We observe the following for spectral moments of networks from different categories.

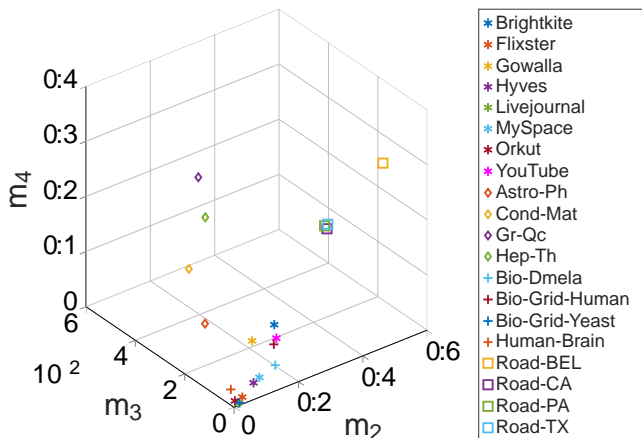


Figure 4: Spectral Points of Real-World Networks. Here, 'o's are social networks, 's are collaboration networks, 't are road networks, and '+' are biological networks.

Social Networks. Social networks are often weakly scale free [1] with a log-normal degree distribution [1, 11] and exhibit a core-periphery structure [4, 39]. Hence, social networks have a relatively large degree variance, which makes the value generally smaller than that of road networks and collaboration networks. In general, social networks have relatively small spectral moments, which shows they have better global connectivity. This observation accords to the fact that social networks exhibit the small-world phenomenon [21, 35] (e.g. in May 2011 the average path length between individuals in the Facebook graph was 4.7 [33]).

Collaboration Networks. Compared to networks from other categories, collaboration networks have greater values as their clustering coefficient is much higher.

Road Networks. Compared to other categories, road networks have (1) a small degree variance, as often not many roads intersect at the same point; (2) more squares, as many parts of road networks resemble rectangular grids; (3) relatively low clustering coefficient, as triangles are uncommon. Due to these properties, road networks have large m_2 and m_4 values, but small m_3 values.

Biological Networks. Biological networks are often strongly scale free with a power-law degree distribution [15, 37] and a core-periphery structure [20]. In general, they share similar patterns with social networks.

5.3 Error Bound on Truncated Moments

Here, we provide an error bound on using the truncated spectral moments to represent a graph. We discuss the problem from two views: (I) the spectral distribution and (II) the network structure.

I. Spectral Distribution. Truncated spectral moments have been used to approximate the spectral density [9]. Here, we provide a theoretical error bound on the maximum possible difference between the spectral distribution of two graphs that have the same truncated moments, i.e., worst-case scenario.

Lemma 5.2 Given two graphs G_1 and G_2 with respective spectral density functions p_1, p_2 of their random-walk transition matrices P_1 and P_2 , if G_1 and G_2 have the same spectral points (m_2, m_3, m_4) , then the Wasserstein distance $W_1(p_1, p_2)$ is upper bounded by

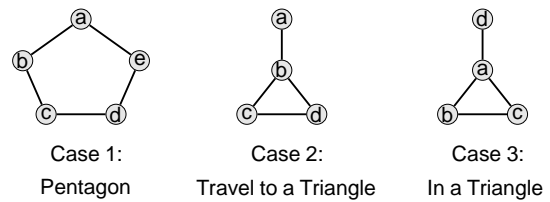


Figure 5: Structures related to the m_5 .

Proof. As we have discussed, both m_1 and p_2 are supported on $[0, 1]$. As $m_1 = 0$, p_1 and p_2 share the same first four moments. Based on Theorem 3 of [10], $W_1(p_1, p_2) \leq \frac{1}{4} + 3 \sum_{j=5}^{\infty} |m_j - m_j|$.

Lemma 5.2 holds in the worst-case sense, as the first few moments can be good enough to determine the distribution with high accuracy. When one approximates the spectral density with the kernel polynomial method, only a few moments are needed by using smoothing techniques as the spectrum approximation error will decay exponentially [9, 31]. The empirical studies also show that recovery of spectrum of real-world networks using truncated moments performs much better than the theory would suggest [31].

II. Network Structure. As mentioned, m_k represents the return probability of an k -step random walk. We notice that a long closed walk (e.g. $k=5$) can be a cycle of length 5, but in more cases the closed walk is composed of multiple short closed walks. Let us see the cases of $k=5$ in Figure 5. There are three cases which lead to a closed walk of length 5. Case 1: the starting node is in a pentagon, so the pentagon will provide two closed walks of length 5; Case 2: The starting node a travels to b which is in a triangle, then b takes a closed walk of length 3, and finally, goes back to a ; Case 3: a is in a triangle so it can have a closed walk of length 3, and during this walk any node b or c takes a closed walk of 2. In both Cases 2 and 3, the closed walk of length 5 is decomposed into a closed walk of 2 and a closed walk of 3. In real-world networks, we find that most long closed walks belong to the latter situation, as higher-order structures like a long cycle is less frequent than lower-order structures such as edges, wedges, or triangles [25]. As m_2, m_3, m_4 can capture these short closed walks, we basically only lose the information on uncommon structures like long cycles.

5.4 Time Complexity

In this paper, we compute the accurate estimates of the low-order moments with the ApproxSpectralMoment algorithm proposed by [8]. The algorithm estimates the moments by simulating many random walks and computes the proportion of closed walks. To compute the k -th spectral moment by simulating random walks, it takes $O(s^k)$ time. In our case, $k=4$ and following the empirical results by [8], we set $s=10,000$. For all networks used in our datasets, it takes only a few seconds to compute the three spectral moments. The Python code for computing moments has been released.

6 SPECTRAL NETWORK IDENTIFICATION

In Section 5, we showed how networks can be represented and visualized using their truncated moments. Here, we demonstrate that spectral moments can be used for network identification [16], a problem which we will briefly review next.

¹<https://github.com/shengminjin/EstimateSpectralMoments>

6.1 Network Identification

Network identification [16] aims to identify the source 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 $N = \{N_1; N_2; \dots; N_n\}$, and a subgraph G sampled from $N_j \in N$ using a sampling strategy S , we want to identify G , i.e., the network N_j from which G is sampled.

Based on the perturbation analysis of spectral density [9], the Wasserstein distance between the spectral density of a graph and the perturbed graph is bounded by the Frobenius norm of the perturbation. More specifically, suppose $\mathbf{A} = A + \Delta A$ is the perturbed graph matrix with spectral density \mathbf{f} , then $W_1(\mu; \mathbf{f}) \leq \|\Delta A\|_F$. As one can view a subgraph as a result of a perturbation (removing nodes/edges) on the whole graph, the spectral moments of the subgraphs and the whole graphs should also be close. Therefore, we use spectral moments for network identification as they can capture the similarity between subgraphs and the whole graph. We use spectral moments as features directly to find the graph identities. Before detailing the experiments, we present the experimental setup.

6.2 Experimental Setup

From each real-world network, we sample many subgraphs representing graphs G which are to be identified. We vary the sampling proportion from 10% to 99% and sample using random node sampling. For each proportion, we sample two subgraphs. Hence, for each network we have $90 \times 2 = 180$ subgraphs, and for twenty networks, we have $180 \times 20 = 3,600$ samples to be identified.

6.3 Experiments

We use the spectral moments of each subgraph as its features and the name of the source networks as the class label, to train a multiclass classification model. We use 10-fold cross validation, and decision tree, SVM, k -NN and bagged trees as our classifiers. For spectral moments, we consider using the three spectral moments ($m_2; m_3; m_4$) and using the first 20 spectral moments. For evaluation, we provide the following two baselines for comparison.

- (1) **Top Eigenvalues.** Top eigenvalues have been used to study graph similarity [18]. We compute the top 5 eigenvalues of each sample as features for classification.
- (2) **Random Prediction.** A simple *random prediction*, so the accuracy will be $1/n$ where n is the number of networks.

We evaluate the methods for all networks and within each network category and report the performance for the best classifier in Table 3. The results show that spectral moments significantly outperform the baselines. The first 20 spectral moments perform best, but using three spectral moments one does not lose much predictive power, which confirms our discussion that the truncated spectral moments can keep most information on the spectral distribution and network structure of real-world networks.

7 ADDITIONAL RELATED WORK

Additionally, our work has links to the following areas:

I. 3D Network Embedding. Jin et al. [17] propose a 3D embedding method using Stochastic Kronecker Graph model. Their embedding method can capture the core-periphery structure of a network, and

Table 3: Network Identification with Spectral Moments

Type	Three Spectral Moments	First 20 Spectral Moments	Baselines	
			Top Eigenvalues	Random Prediction ($1/n$)
All Networks	82.0%	86.5%	62.4%	5%
Social Networks	95.5%	96.1%	74.8%	12.5%
Collaboration Networks	94.8%	97.1%	70.9%	25%
Road Networks	51.2%	53.3%	44.9%	25%
Biological Networks	99.7%	99.6%	90.4%	25%

the embedding values can quantify the core strength of the network to some extent. Compared to their approach, spectral points carry further information on the network structure and properties.

II. Spectral Moments of other Associated Matrices. Preciado and colleagues [26, 27] have provided detailed analysis on the spectral moments of the combinatorial Laplacian matrix ($D - A$) of graphs to connect the spectral moments with network structure. Here, we choose not to use the spectral moments of combinatorial Laplacian as the bounds on its eigenvalues are related to the size of the graph, and we prefer using the random walk transition matrix to have a compact embedding space.

III. Spectral Embedding. Recently, spectral information is used for different network embedding methods. One example is FGSD [34], which proposed a family of graph spectral distances that embeds the information as histograms and computes histograms on the biharmonic kernel of the graph. Another example is NetLSD [32], which computes and samples the heat or wave trace over the eigenvalues of a graph's normalized Laplacian to build embeddings. By using the spectral information, these embeddings are more focused on the predictive power, but are still relatively difficult to be interpreted.

8 CONCLUSION AND DISCUSSION

In this paper, we introduce an approach to map all possible networks to a bounded 3D embedding space using the spectral moments of networks. We propose a 3D network embedding method, the *Spectral Point*, by using the truncated spectral moments of the network. We prove that spectral moments are closely related to network structure and various network properties. To the best of our knowledge, we are the first to study relationship between the spectral moments of the random walk transition matrix and network structure. We prove that the spectral moments are bounded by network properties such as the degree distribution and the global clustering coefficient, and spectral moments can be used as a measure for global connectivity. We prove the relationship between the spectral moments of the network and those of its connected components. We derive the spectral points of various types of graphs such as complete graphs, cycles, star graphs, complete bipartite graphs, and wheels. For k -regular graphs, we prove each dimension of the spectral points are bounded by $\frac{1}{k}$. We analyze the spectral points of real-world graphs and show that their structure and properties identified in past literature are often captured by spectral points. Finally, we demonstrate that spectral moments can be used for network identification, i.e., to identify the source of an anonymized graph. The result shows that the spectral moments outperform the baselines and the truncated spectral moments do not lose much predictive power. We believe the spectral embedding space (i.e., the

