Comparing Massive Networks via Moment Matrices

Hayoung Choi, Yifei Shen, and Yuanming Shi
School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
Email: {hchoi,shenyf,shiym}@shanghaitech.edu.cn

***Abstract*—In this paper, a novel similarity measure for comparing massive complex networks based on moment matrices is proposed. We consider the corresponding adjacency matrix of a graph as a real random variable of the algebraic probability space with a state. It is shown that the spectral distribution of the matrix can be expressed as a unique discrete probability measure. Then we use the geodesic distance between positive definite moment matrices for comparing massive networks. It is proved that this distance is graph invariant and sub-structure invariant. Numerical simulations demonstrate that the proposed method outperforms state-of-art method in collaboration network classification and its computational cost is extremely cheap.**

I. Introduction

Network is one of the most common representations of complex data and plays an indispensable role in diversified research areas. Over the past several decades, enormous breakthroughs have been made while many fundamental problems about networks are remaining to be solved. Comparing networks is one of the most important problems with a very long history [1]. In practice, the similarity measure of networks is widely applied in social science, biology, and chemistry. For instance, the similarity measure of networks can be used to classify ego networks [2], distinguish between neurological disorders [3], and discover molecules with similar properties [4]. In order to measure similarity between networks effectively, several definitions of distance and similarity have been proposed. Graph edit distances are the minimum cost for transforming one network to another by the distortion of nodes and edges [5]. These definitions only pay attention to the similarities of the nodes and edges but lacks the information of topological structures of the networks. For the purpose of addressing this limitation, frequency subgraph mining algorithms [6], graph kernels [7] and methods based on moments [8] have been proposed. However, these methods are not scalable to massive networks containing millions of edges, which are common in today's applications. As a result, effective and scalable methods for massive network comparison is urgently needed.

In this paper, we propose a novel similarity measure for comparing massive complex networks. We consider the adjacency matrix of the network as a probability random variable of the algebraic probability space with the proposed state. We show that the spectral distribution of the matrix can be expressed as a unique discrete probability measure. Then we propose an efficient and scalable method to measure the similarity between massive networks based on the spectral distribution of the corresponding adjacency matrix in the given state. Specifically, we first compute the corresponding positive definite moment matrix whose entries consist of the first few number of moments of the spectrum distribution. Our proposed distance between networks is obtained by the *geodesic distance of the moment matrices* (called GDMM). We show that this distance is graph invariant and sub-structure invariant. GDMM is scalable to extremely massive networks and highly parallelable. Numerical Simulations demonstrate that GDMM not only has better performance over the competing methods, but also outperforms the state-of-art method in collaboration network classification.

This work was partly supported by the National Nature Science Foundation of China under Grant No. 61601290, and the Shanghai Sailing Program under Grant No. 16YF1407700.

II. Background and Preliminary

Let $M(n,\mathbb{C})$ (resp. $M(n,\mathbb{R})$) be the set of $n\times n$ complex (resp. real) matrices. Denote $\mathbb{N}$ the set of nonnegative integer numbers. In general, we will follow the notations and definitions in [9].

A. Graph

Let V be the set of vertices, and let $\{x,y\}$ denote the edge connecting two points $x,y\in V$. A *undirected graph* is a pair $G=(V,E)$ where the set V of vertices is finite, and the set E of edges is a subset of the set $\{\{x,y\}:x,y\in V\}$. We say that two vertices $x,y\in V$ are *adjacent* if $\{x,y\}\in E$, denoted by $x\sim y$. A *degree* of a vertex $x\in V$ is defined by $\deg(x)=|\{y\in V:y\sim x\}|$. In this paper we consider a finite undirected graph. Two graphs $G=(V,E)$ and $G'=(V',E')$ are *isomorphic* if there is a bijection $f:V\longrightarrow V'$, such that any two vertices $u,v\in V$ are adjancent in G if and only if $f(u)$ and $f(v)$ are adjacent in V'. For $m\in\mathbb{N}$, a finite sequence of vertices $x_0,x_1,\ldots,x_m\in V$ is called a *walk of length* m if $x_0\sim x_1\sim\cdots\sim x_m$, where some of $x_0,x_1,\ldots,x_m$ may coincide. A graph $G=(V,E)$ is *connected* if every pair of distinct vertices $x,y\in V$ $(x\neq y)$ are connected by a walk. If there is a walk connecting two distinct vertices $x,y\in V$, the *graph distance* between x and y is the minimum length of a walk connecting x and y, denoted by $\partial(x,y)$. If there is no such a walk, we define $\partial(x,y)=\infty$. For $x=y$ we define $\partial(x,x)=0$. For graphs $G_i=(V_i,E_i)$, $i=1,2$ with $V_1\cap V_2=\emptyset$, the *direct sum* of G_1 and G_2 is defined as $G=(V_1\cup V_2,E_1\cup E_2)$, denoted by $G=G_1\sqcup G_2$. From now on, without loss of generality we assume that $V=\{1,2,\ldots,n\}$. Any graph $G=(V,E)$ is represented by the *adjacency matrix* $\boldsymbol{A}\in\{0,1\}^{n\times n}$ where $\boldsymbol{A}_{i,j}=1$ if and only if $\{i,j\}\in E$. Every permutation

$\pi : \{1,2,\ldots,n\} \longrightarrow \{1,2,\ldots,n\}$ is associated with a corresponding *permutation matrix* $\boldsymbol{P}$. Given an adjacency matrix $\boldsymbol{A}$, graphs corresponding to adjacency matrix $\boldsymbol{A}$ and $\boldsymbol{PAP}^\top$ are isomorphic, i.e., they represent the same graph structure. A property of graph is called *graph invariant* if the property does not change under the transformation of reordering of vertices. Note that the adjacency matrix of a graph includes the full information about a graph. For $x, y \in V$ and $m \in \mathbb{N}$ let $W_m(x,y)$ denote the number of m-step walks connecting x and y. Remark that $W_0(x,y) = 0$ if $x \neq y$ and $W_0(x,y) = 1$ if $x = y$. It is noted that $(\boldsymbol{A}^m)_{ij} = W_m(i,j)$ for all $i, j \in V$ and $m \in \mathbb{N}$.

Let $\mathscr{A}(G)$ be the unital algebra generated by $\boldsymbol{A}$ (the algebra generated by $\boldsymbol{A}$ and the identity matrix $\boldsymbol{I} = \boldsymbol{A}^0$, i.e., $\mathscr{A}(G) = \{f(\boldsymbol{A}) : f \in \mathbb{C}[x]\}$, where $\mathbb{C}[x]$ is the set of all polynomials with complex coefficients. Moreover, the involution is defined by $(c\boldsymbol{A}^m)^* = \bar{c}\boldsymbol{A}^m$ for $c \in \mathbb{C}$. Then $\mathscr{A}(G)$ becomes a unital $*$-algebra. We call $\mathscr{A}(G)$ *adjacency algebra* of G.

Proposition 1. *Let $s(G)$ denote the number of distinct eigenvalue of G For a connected finite graph G we have*

$$s(G) = dim\mathscr{A}(G) \geq \mathrm{diam}(G) + 1.$$

B. Quantum Probability

Let $\mathscr{A}$ be a unital $*$-algebra over the complex number field $\mathbb{C}$ with the multiplication unit $1_{\mathscr{A}}$. A function $\varphi : \mathscr{A} \longrightarrow \mathbb{C}$ is called a *state* on $\mathscr{A}$ if

$$(i)\ \varphi \text{ is linear}; \quad (ii)\ \varphi(a^*a) \geq 0; \quad (iii)\ \varphi(1_{\mathscr{A}}) = 1.$$

The pair $(\mathscr{A}, \varphi)$ is called an algebraic probability space.

Note that a state φ on a unital $*$-algebra $\mathscr{A}$ is a $*$-map, i.e., $\varphi(a^*) = \overline{\varphi(a)}$. Let $(\mathscr{A}, \varphi)$ be an algebraic probability space. An element $a \in \mathscr{A}$ is called an *algebraic random variable* or a *random variable* for short. A random variable $a \in \mathscr{A}$ is called real if $a = a^*$. For a random variable $a \in \mathscr{A}$ the quantity of the form:

$$\varphi(a^{\varepsilon_1}a^{\varepsilon_2}\cdots a^{\varepsilon_m}), \quad \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m \in \{1, *\},$$

is called a *mixed moment* of order m. Statistical properties of an algebraic random variable are determined by its mixed moments. For a real random variable $a = a^*$ the mixed moments are reduced to the *moment sequence*: $\varphi(a^k)$, $k \in \mathbb{N}$, where $\varphi(a^k)$ is called the *kth moment* of a. By definition $\varphi(a^0) = 1$. Alternatively, $\varphi(a^k)$ is denoted by m_k.

A *moment matrix* with degree n is defined as

$$\boldsymbol{\mathcal{M}}_n := \begin{bmatrix} m_0 & m_1 & \cdots & m_n \\ m_1 & m_2 & \cdots & m_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ m_n & m_{n+1} & \cdots & m_{2n} \end{bmatrix}.$$

Let $\mathfrak{B}(\mathbb{R})$ denote the set of all probability measures having finite moments of all orders.

Theorem 2. *Let $(\mathscr{A}, \varphi)$ be an algebraic probability space. For a real random variable $a = a^* \in \mathscr{A}$ there exists a probability measure $\mu \in \mathfrak{B}(\mathbb{R})$ such that*

$$\varphi(a^k) = \int_{\mathbb{R}} x^k \mathrm{d}\mu(x) \quad \textit{for all } k \in \mathbb{N}_0.$$

Such μ is called the *spectral distribution* of a in φ [9].

It is noted that $M(n,\mathbb{C})$ with the usual operators is a unital $*$-algebra. The following is a typical example of a state. For $\boldsymbol{A} = [\boldsymbol{A}_{ij}] \in M(n,\mathbb{C})$, the *the normalized trace* is defined by

$$\varphi_{\mathrm{tr}}(\boldsymbol{A}) = \frac{1}{n}\mathrm{tr}(\boldsymbol{A}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}_{ii}. \tag{II.1}$$

Note that the normalized trace is a state on $M(n,\mathbb{C})$, implying $(M(n,\mathbb{C}), \varphi_{\mathrm{tr}})$ becomes an algebraic probability space.

III. Main results

Denote the vector of all ones by $\boldsymbol{e} \in \mathbb{C}^n$. Define a function $\varphi_{\boldsymbol{e}} : M(n,\mathbb{C}) \longrightarrow \mathbb{C}$ by

$$\varphi_{\boldsymbol{e}}(\boldsymbol{A}) = \frac{1}{n}\langle \boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}\rangle \tag{III.2}$$

for all $\boldsymbol{A} \in M(n,\mathbb{C})$. Then it is clear that $\varphi_{\boldsymbol{e}}$ is a vector state on $M(n,\mathbb{C})$, implying that $(M(n,\mathbb{C}), \varphi_{\boldsymbol{e}})$ is an algebraic probability space.

Let $G = (V, E)$ be a graph and let φ be a state given on the adjacency algebra $\mathscr{A}(G)$. Since the adjacency matrix $\boldsymbol{A} \in M(n,\mathbb{C})$ of G can be regarded as a real random variable of the algebraic probability space $(M(n,\mathbb{C}), \varphi_e)$, by Theorem 2 it follows that there exists the spectral distribution of $\boldsymbol{A}$ in the state $\varphi_{\boldsymbol{e}}$ such that

$$\varphi_{\boldsymbol{e}}(\boldsymbol{A}^k) = \int_{\mathbb{R}} x^k \mathrm{d}\mu(x) \quad \text{for all } k \in \mathbb{N}. \tag{III.3}$$

Note that

$$\varphi_{\boldsymbol{e}}(\boldsymbol{A}^k) = \frac{1}{n}\langle \boldsymbol{e}, \boldsymbol{A}^k\boldsymbol{e}\rangle = \mathbb{E}[\boldsymbol{A}^k\boldsymbol{e}],$$

where $\mathbb{E}(\boldsymbol{v}) = \frac{1}{n}\sum_{i=1}^{n} v_i$ is the average of entries of vector $\boldsymbol{v}$. Since $(\boldsymbol{A}^k)_{ij} = W_m(i,j)$ and $\boldsymbol{A}^k\boldsymbol{e}$ is the column vector whose i-th entry is equal to the sum of the number of all walks of length k from the vertex i, $\varphi_{\boldsymbol{e}}(\boldsymbol{A}^k)$ is the average of the the sum of the number of all walks of length k from each vertex.

Let δ_λ denote a dirac measure at λ (i.e., $\delta_\lambda(S) = 1$ if $\lambda \in S$ and $\delta_\lambda(S) = 0$ if $\lambda \notin S$).

Theorem 3. *Let $(M(n,\mathbb{C}), \varphi_{\boldsymbol{e}})$ be the algebraic probability space. For a real random variable $\boldsymbol{A} \in M(n,\mathbb{C})$, there exists a unique probability discrete measure $\mu = \sum_{i=1}^{s} \omega_i\delta_{\lambda_i}$ such that*

$$\varphi_{\boldsymbol{e}}(\boldsymbol{A}^k) = \int_{\mathbb{R}} x^k \mathrm{d}\mu(x) \quad \textit{for all } k \in \mathbb{N}. \tag{III.4}$$

Conversely, for a probability discrete measure $\mu = \sum_{i=1}^{s} \omega_i\delta_{\lambda_i}$, there exists a real random variable $\boldsymbol{A} \in (M(n,\mathbb{C}), \varphi_{\boldsymbol{e}})$ with s distinct eigenvalues such that $\varphi_{\boldsymbol{e}}(\boldsymbol{A}^k) = \int_{\mathbb{R}} x^k \mathrm{d}\mu(x)$ for all $k \in \mathbb{N}$.

Proof. ($\Rightarrow$) Let $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. By Spectral Theorem, A can be diagonalized by a unitary matrix $\boldsymbol{U}$. That is, $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*$. Then the kth moment of $\boldsymbol{A}$ is

$$\varphi_e(\boldsymbol{A}^k) = \frac{1}{n}\boldsymbol{e}^*\boldsymbol{A}^k\boldsymbol{e} = \frac{1}{n}\boldsymbol{v}^*\boldsymbol{D}^k\boldsymbol{v} = \sum_{i=1}^{s}\omega_i\lambda_i^k = \int_{\mathbb{R}} x^k \mathrm{d}\mu,$$

where $\boldsymbol{v} = \boldsymbol{U}^*\boldsymbol{e}$. And it holds that

$$\sum \omega_i = \frac{1}{n}\sum \boldsymbol{v}^*\boldsymbol{v} = \frac{1}{n}\sum \boldsymbol{e}^*\boldsymbol{U}\boldsymbol{U}^*\boldsymbol{e} = 1.$$

($\Leftarrow$) Let $\mu = \sum_{i=1}^{s}\omega_i\delta_{\lambda_i}$ with $\omega_i \geq 0$ and $\lambda_i \in \mathbb{R}$ for all $i \in \mathbb{N}$, and $\sum_{i=1}\omega_i$=1. Let $\boldsymbol{D}$ be the $n \times n$ diagonal matrix whose diagonal entries are $\lambda_1, \lambda_2, \ldots, \lambda_s, \lambda_s, \ldots, \lambda_s$. Let $\boldsymbol{v} = [\sqrt{\omega_1}\ldots\sqrt{\omega_s}\ 0\ldots 0]^\top$. Since $\boldsymbol{v}$ and $\frac{1}{\sqrt{n}}\boldsymbol{e}$ both are unit vectors, there exists a unitary matrix $\boldsymbol{U}$ such that $\boldsymbol{U}\boldsymbol{v} = \frac{1}{\sqrt{n}}\boldsymbol{e}$. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^*$. Then $\boldsymbol{A}$ holds the equality (III.4). □

Theorem 4. *The k-th moment of $\boldsymbol{A}$, $\varphi_{\boldsymbol{e}}(\boldsymbol{A}^k)$, is a graph invariant.*

Proof. For any given permutation matrix $\boldsymbol{P}$ it holds that

$$\varphi_{\boldsymbol{e}}\left((\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^*)^k\right) = \frac{1}{n}\boldsymbol{e}^*(\boldsymbol{P}\boldsymbol{A}^k\boldsymbol{P}^*)\boldsymbol{e} = \varphi_{\boldsymbol{e}}(\boldsymbol{A}^k).$$

□

Hence, we will henceforth denote $\boldsymbol{\mathcal{M}}_n$ as $\boldsymbol{\mathcal{M}}_n(G)$ if a graph G is given. Remark that the probability measure includes information not only about the eigenvalues of $\boldsymbol{A}$, but also about the corresponding eigenvectors. To measure similarity between two large-scale networks, by Theorem 3 it is enough to compare the corresponding probability measures. There are various distances and divergences between two distributions such as Kullback-Leibler divergence, Bhattacharyya distance, etc (see [10], [11]). However, since a large-scale network in real world has a rich spectrum, to reconstruct the spectral distribution is impossible in practice. Instead, we can use moments of the distributions. Mathematically, all the moments up to infinity are requested in order to obtain a perfect reconstruction. However, the first few moments are only sufficient if the class of functions in which the reconstruction is sought is restricted appropriately. It has been mentioned in the literature that the most of the information about the measure is contained in the first few moments, and the higher-order ones providing only little additional information [12]–[14]. Since $\boldsymbol{\mathcal{M}}_n(G)$ has sufficient information about the distribution, a distance between moment matrices can be calculated to measure a similarity between two distributions.

Let $\boldsymbol{c} = [c_0, c_1, \ldots, c_n]^*$ be a vector in $\mathbb{C}^n$. Then

$$\boldsymbol{c}^*\boldsymbol{\mathcal{M}}_n\boldsymbol{c} = \sum_{i,j=0}^{n}(\boldsymbol{e}^*\boldsymbol{A}^{i+j}\boldsymbol{e})c_i^*c_j = \left\|\sum_{i=0}^{n}c_i\boldsymbol{A}^i\boldsymbol{e}\right\|^2 \geq 0.$$

Thus, the moment matrix in (II-B) is positive semi-definite matrix for all $n \in \mathbb{N}$. However, the corresponding moment matrix $\boldsymbol{\mathcal{M}}_n$ can possibly be a singular positive semi-definite matrix, which is not a positive definite matrix. We assume that the diameter of a large-scale network is always greater then 5. By Proposition 1, it follows that $s(G) \geq 6$, equivalently $\boldsymbol{I}, \boldsymbol{A}, \ldots, \boldsymbol{A}^6$ are linearly independent. Thus $\boldsymbol{\mathcal{M}}_n$ is positive definite for $n = 1, \ldots, 6$, which is a point on the Riemannian manifold of positive definite matrices (see [15, Theorem 1.1]).

Denote the set of positive definite matrices as $\mathcal{P}^o$. There are various distances between two positive definite matrices ([16], [17]). The Frobenius norm $\|\cdot\|_2$ gives rise to the affine-invariant metric on $\mathcal{P}^o$ given by $\delta(\boldsymbol{A}, \boldsymbol{B}) = \|\log(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})\|_2$ for any $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{P}^o$. Then $\mathcal{P}^o$ is a Cartan-Hadamard manifold, a simply connected complete Riemannian manifold with non-positive sectional curvature. The geodesic curve has a parametrization $\gamma(t) = \boldsymbol{A}^{1/2}(\boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2})^t\boldsymbol{A}^{1/2}$, $0 \leq t \leq 1$, which is the unique geodesic from $\boldsymbol{A}$ to $\boldsymbol{B}$ (see [18]).

For two graphs G and $\tilde{G}$, we propose new distance between G and $\tilde{G}$ as the geodesic distance between the corresponding moment matrices, i.e.,

$$d_n(G, \tilde{G}) := \delta(\boldsymbol{\mathcal{M}}_n(G), \boldsymbol{\mathcal{M}}_n(\tilde{G})), \quad n \in \mathbb{N}.$$

Assume $n \in \mathbb{N}$ is fixed such that the corresponding moment matrices are positive definite. Denote $d(G, \tilde{G})$ instead of $d_n(G, \tilde{G})$.

Theorem 5. *For graphs $G, \tilde{G}, \hat{G}$,*

(i) *(Nonnegativity) $d(G, \tilde{G}) \geq 0$,*
(ii) *(Identification) $d(G, \tilde{G}) = 0$ if $G = \tilde{G}$,*
(iii) *(Symmetry) $d(G, \tilde{G}) = d(\tilde{G}, G)$,*
(iv) *(Triangle Inequality) $d(G, \hat{G}) \leq d(G, \tilde{G}) + d(\tilde{G}, \hat{G})$.*

Cospectral graphs, also called *isospectral graphs*, are graphs that share the same graph spectrum. The smallest pair of cospectral graphs is the graph union $C_4 \cup K_1$ and star graph S_5, illustrated in Figure 1. Both have the same graph spectrum, $-2, 0, 0, 0, 2$. If the adjacency matrices A and $\tilde{A}$ are considered as real algebraic random variables in $(M(5, \mathbb{C}), \varphi_{\mathrm{tr}})$, then two algebraic random variables $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are moment equivalent, since $\varphi_{\mathrm{tr}}(\boldsymbol{A}^k) = \frac{1}{n}\operatorname{tr}(\boldsymbol{A}^k) = \frac{1}{n}\operatorname{tr}(\tilde{\boldsymbol{A}}^k) = \varphi_{\mathrm{tr}}(\tilde{\boldsymbol{A}}^k)$. However, if $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are considered as real algebraic random variables in $(M(5, \mathbb{C}), \varphi_{\boldsymbol{e}})$, then two algebraic random variables $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are not moment equivalent. So, using the state $\varphi_{\boldsymbol{e}}$ allows us to distinguish two graphs. Indeed, the moment matrices

$$\boldsymbol{\mathcal{M}}_1(C_4 \cup K_1) = \begin{pmatrix} 1 & 1.6 \\ 1.6 & 3.2 \end{pmatrix}, \quad \boldsymbol{\mathcal{M}}_1(S_5) = \begin{pmatrix} 1 & 1.6 \\ 1.6 & 4 \end{pmatrix}$$

are different.

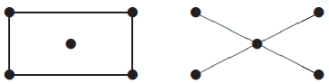

Fig. 1: Cospectral graphs: $C_4 \cup K_1$ and S_5

Fig. 2 is introduced in [19]. The first three graphs have the same number of vertices and edges. Table I shows distances between the graphs based on Hamming distance, graph edit distance, and our proposed distance. As mentioned in [19], a good measure should return a higher distance value between

G_1 and G_2, then between G_1 and G_3. Hamming distance and graph edit distance do not capture relevant topological differences. However, our proposed measure perform a highly precise comparison. Remark that $\boldsymbol{\mathcal{M}}_1(G_3)$ is a singular positive semi-definite matrix. So, the geodesic distance between $\boldsymbol{\mathcal{M}}_1(G_3)$ and any positive definite matrices is not finite. Alternatively, Bures-wasserstein distance between positive semi-definite matrices allows to overcome this situation [20].

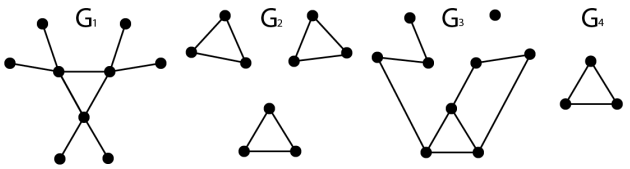

Fig. 2: The first three different networks with the same number of vertices and edges are shown in [19].

TABLE I: (i) H: Hamming distance; (ii) GED: graph edit distance; (iii) our propose method.

Dissimilarity	H	GED	Proposed Measure
$d(G_1, G_2)$	12	6	2.7333
$d(G_1, G_3)$	12	6	1.2103
$d(G_2, G_3)$	12	6	1.6815

Theorem 6. *Let $G_1, G_2, \ldots, G_K$ be given mutually disjoint graphs. Then*

$$\boldsymbol{\mathcal{M}}_n(G_1 \sqcup \ldots \sqcup G_K) = \sum_{j=1}^{K} \frac{c_j}{c} \boldsymbol{\mathcal{M}}_n(G_j)$$

where c_j is the number of vertices in G_j and $c = \sum_{j=1}^{K} c_j$.

Especially, $\boldsymbol{\mathcal{M}}_n(G \sqcup \ldots \sqcup G) = \boldsymbol{\mathcal{M}}_n(G)$ for all $n \in \mathbb{N}$. If a graph consists of identical subgraphs, then the moment matrix of a given graph is equal to one of its subgraph. In other words, a moment matrix of graph can preserve information regardless of repetition of structure, called *sub-structure invariant*.

IV. Complexity and Parallelism

Our proposed method, GDMM, has two steps. Consider the moment matrix with degree n whose size is $(n+1) \times (n+1)$. The first step is to obtain the moment matrix $\boldsymbol{\mathcal{M}}_n$ whose entries consist of the moment sequence $\{m_k\}_{k=0}^{2n}$. In the second step, we use the geodesic distance between positive definite matrices to compute the distance between two moment matrices. In the following, we will show the time complexity, space complexity, and parallelism of each step and those of the overall algorithm.

A. Complexity

We consider comparing two graphs G_1 and G_2. Let $|V_1|, |E_1|$ and $|V_2|, |E_2|$ denote the number of nodes and edges of graph G_1 and G_2 respectively. Let $|E| = max(|E_1|, |E_2|)$ and $|V| = max(|V_1|, |V_2|)$. The first step of the algorithm can be computed in $\mathcal{O}(n|E|)$ time and $\mathcal{O}(|E|)$ space using sparse matrix-vector multiplication. The second step mainly involves eigenvalue decomposition, which can be computed in $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ space. The time complexity of the total algorithm is $\mathcal{O}(n|E| + n^3)$ and space complexity is $\mathcal{O}(|E| + n^2)$. However, n is relatively small, say 4 or 5, in practical problems because most of the information about a distribution is contained in the first few moments [12]–[14]. Thus the time complexity and space complexity of GDMM are $\mathcal{O}(|E|)$.

B. Parallelism

As we discussed before, the first step is sparse matrix-vector multiplication. This operation can be completely paralleled on CPU or GPU. As n is small, the second step takes much less time than the first step. As a result, our algorithm can be paralleled efficiently.

V. Experiments

A. Classifying Networks

We apply our method to classify networks. We follow the system setting of [2]. Specifically, we classify one's research area using the information of the graph structure of one's collaboration network. Because researchers in one area usually tightly connected with researchers in that area compared to other areas, it is possible to determine to which area a researcher belongs considering one's collaboration networks. This information can be used for recommendations such as job recommendations and citation recommendations.

Three datasets from [21] are used: high energy physics collaboration network(HEP), condensed matter collaboration network(CM), and astro physics collaboration network(ASTRO). In the network, an undirected edge from u to v means that the author u and the author v are co-authored. We use the method from [2] to generate subgraphs and obtain 415 subgraphs for CM and 1000 subgraphs for HEP and ASTRO respectively. Then we label each sub-graph according to the dataset which it belongs to. The tasks are classifications between each two datasets and among three datasets. For each task, we first split the dataset into 10 folds of the same size. We then combine 9 of the folds as the training set, the left 1 fold as the test set. We repeat this ten times to compute the average accuracy.

In the classification tasks, we use k-nearest-neighbor(KNN) classifier. We set the size of moment matrix n from 2 to 7 and k in KNN from 1 to 10 and choose the best one. The first three benchmark algorithms are Covariance [2], NCLM [8], and Top-10 eigenvalues(EIGS-10). Specifically, Covariance computes the covariance matrix of the vector $[\frac{\boldsymbol{A}^i e}{|\boldsymbol{A}^i e|}]_{i=1}^{5}$ and uses Bhattacharyya similarity between two covariance matrices as the distance between the corresponding networks. NCLM first computes the log moment sequence vector $[\log(\frac{\mathrm{tr}(\boldsymbol{A}^i)}{n^i})]_{i=2}^{7}$ and uses the Euclidean distance between two moment vectors as the distance between corresponding networks. EIGS-10 takes the largest 10 eigenvalues as a vector and uses the Euclidean distance between two vectors as the distance between corresponding networks. In addition, we add the state-of-art method in collaboration network classification, Covariance with SVM [2], which employs SVM as the classifier, as the

last benchmark algorithm. The performance of our method and the benchmark algorithm is shown in Table. II.

TABLE II: Accuracy for GDMM and other benchmark methods in collaboration network classification. Best results marked in bold.

	HEP Vs CM	HEP Vs ASTRO	ASTRO Vs CM	Full
GDMM	**0.991**	**0.913**	**0.904**	**0.905**
EIGS-10	0.981	0.879	0.861	0.820
NCLM	0.982	0.850	0.865	0.804
Covariance	0.976	0.857	0.861	0.819
Covariance with SVM	0.987	0.889	0.887	0.849

From the table, we see that with KNN classifier, Covariance, EIGS, and NCLM have similar performance in each task. We also notice that Covariance with SVM performs better than Covariance with KNN. This shows that SVM classifier is more suitable to Covariance method. On top of that, our proposed method, GDMM, not only outperforms various of benchmarks with KNN classifier, but also performs better than Covariance with SVM, the state-of-art method in collaboration classification task in every classification task. This demonstrates the effectiveness of GDMM. This also shows that a few moments can provide enough information for collaboration classification. Besides, GDMM has a significant improvement over the state-of-art method in three collaboration network classification tasks. This shows GDMM is suitable to classification tasks for sophisticated networks.

B. Time Comparison

In this section, we show the efficiency of our algorithm by comparing the running time of GDMM and other methods via a set of experiments. Specifically, in each experiment, we generate 100 Erdős–Rényi random graphs [22] with the same number of nodes and edges. Then we employ GDMM and other methods to get pairwise distances among all possible pairs. For each method, we run 10 times and take the average running time. The number of nodes, number of edges, and the time consumed by different methods are shown in Table III. Here, we use 4×4 moment matrix in GDMM, 4×4 covariance matrix in Covariance and 6 moments in NCLM. All of these experiments are done in MATLAB on the server with an Intel Xeon 2.80 GHz CPU and 64 GB RAM.

TABLE III: Running time for computing pairwise distance among 100 random networks(in seconds). Fastest method marked in bold.

$\|V\|$	$\|E\|$	GDMM	Covariance	EIGS-10	NCLM
2000	2000000	**7.31**	7.32	18.92	39.92
5000	1000000	**1.38**	1.48	85.88	533
10000	2000000	**3.9**	4.7	353.5	27340
50000	15000000	**50**	68	11687	N/A

As shown in the table, the time cost of GDMM is cheaper than all the comparing methods. For example, it can compute pairwise distances of 100 random networks with 500000 nodes and 15000000 edges in 50 seconds, which has $1.36\times$ speed up to Covariance method and $233\times$ speed up to EIGS-10. Besides, from the table, GDMM is almost linear in terms of the number of edges. This demonstrates GDMM is scalable to massive networks.

VI. Conclusion

We considered the adjacency matrix of the network as a random variable and proposed a new network similarity measure based on the geodesic distance of corresponding positive definite moment matrix. Our proposed method demonstrated state-of-art results in collaboration network classification and turned out to be scalable to massive networks.

References

[1] L. Zager, *Graph similarity and matching*. PhD thesis, Massachusetts Institute of Technology, 2005.

[2] A. Shrivastava and P. Li, "A new space for comparing graphs," in *Proc. IEEE Int. Conf. Social Networks Anal. Mining.(ASONAM)*, pp. 62–71, Aug. 2014.

[3] A. Calderone, M. Formenti, F. Aprea, M. Papa, L. Alberghina, A. M. Colangelo, and P. Bertolazzi, "Comparing alzheimer's and parkinson's diseases networks using graph communities structure," *BMC Syst. Biol.*, vol. 10, pp. 25–34, Mar. 2016.

[4] J. K. Morrow, L. Tian, and S. Zhang, "Molecular networks in drug discovery," *Crit. Rev. Biomed. Eng.*, vol. 38, Nov. 2010.

[5] X. Gao, B. Xiao, D. Tao, and X. Li, "A survey of graph edit distance," *IEEE Pattern Anal. Applicat.*, vol. 13, pp. 113–129, Feb. 2010.

[6] J. Ugander, L. Backstrom, and J. Kleinberg, "Subgraph frequencies: Mapping the empirical and extremal geography of large graph collections," in *Proc. Int. Conf. World Wide Web*, pp. 1307–1318, May 2013.

[7] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, and K. Borgwardt, "Efficient graphlet kernels for large graph comparison," in *Proc. IEEE Int. Conf. Artificial Intell. and Stat.(AISTATS)*, pp. 488–495, Apr. 2009.

[8] S. S. Mukherjee, P. Sarkar, and L. Lin, "On clustering network-valued data," in *Adv. Neural Inf. Process. Syst. (NIPS)*, pp. 7074–7084, Dec. 2017.

[9] N. Obata, *Spectral Analysis of Growing Graphs: A Quantum Probability Point of View*. Springer, 2017.

[10] J. Chung, P. Kannappan, C. Ng, and P. Sahoo, "Measures of distance between probability distributions," *J. Math. Anal. Appl.*, vol. 138, pp. 280 – 292, Feb. 1989.

[11] S. Kullback and R. A. Leibler, "On information and sufficiency," *IMS Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.

[12] D. Fasino and G. Inglese, "Recovering a probabilty density from a finite number of moments and local a priori information," *Rendiconti dell'Istituto di matematica dell'Universitá di Trieste: an International Journal of Mathematics*, 1996.

[13] F. JB, "Elementary principles of spectral distributions," in *Theory and applications of moment methods in many fermion systems*, pp. 1–16, Plenum Press, 1980.

[14] P. Gavriliadis and G. Athanassoulis, "Moment data can be analytically completed," *Probabilistic Eng. Mech.*, vol. 18, pp. 329–338, Oct. 2003.

[15] C. Berg and R. Szwarc, "A determinant characterization of moment sequences with finitely many mass points," *Linear Multilinear Algebra*, vol. 63, pp. 1568–1576, Sep. 2015.

[16] R. Bhatia and J. Holbrook, "Riemannian geometry and matrix geometric means," *Linear Algebra Appl.*, vol. 413, pp. 594–618, Mar. 2006.

[17] S. Sra, "Positive definite matrices and the s-divergence," *Proc. Amer. Math. Soc.*, vol. 144, pp. 2787–2797, Oct. 2016.

[18] R. Bhatia, *Positive Definite Matrices*. Princeton, 2007.

[19] T. A. Schieber, L. Carpi, A. Díaz-Guilera, P. M. Pardalos, C. Masoller, and M. G. Ravetti, "Quantification of network structural dissimilarities," *Nature Commun.*, vol. 8, pp. 13928–13937, Jan. 2017.

[20] R. Bhatia, T. Jain, and Y. Lim, "On the Bures-Wasserstein distance between positive definite matrices," *arXiv preprint arXiv:1712.01504*, 2017.

[21] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proc. AAAI Conf. Artificial Intell.*, 2015.

[22] P. ERDdS and A. R&WI, "On random graphs I," *Publ. Math. Debrecen*, vol. 6, pp. 290–297, 1959.