



# Comparing large-scale graphs based on quantum probability theory<sup>☆</sup>



Hayoung Choi<sup>a</sup>, Hosoo Lee<sup>b</sup>, Yifei Shen<sup>c</sup>, Yuanming Shi<sup>a,\*</sup>

<sup>a</sup> School of Information Science and Technology, ShanghaiTech University, Pudong District, Shanghai 200031, China

<sup>b</sup> Department of Mathematics Education, Teachers College Jeju National University, Jeju, Republic of Korea

<sup>c</sup> Department of Electronic & Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

## ARTICLE INFO

### Keywords:

Comparing graphs  
Large-scale datasets  
Quantum probability  
Moment matrix

## ABSTRACT

In this paper, a new measurement to compare two large-scale graphs based on the theory of quantum probability is proposed. An explicit form for the spectral distribution of the corresponding adjacency matrix of a graph is established. Our proposed distance between two graphs is defined as the distance between the corresponding moment matrices of their spectral distributions. It is shown that the spectral distributions of their adjacency matrices in a vector state includes information not only about their eigenvalues, but also about the corresponding eigenvectors. Moreover, we prove that the proposed distance is graph invariant and sub-structure invariant. Examples with various graphs are given, and distances between graphs with few vertices are checked. Computational results for real large-scale graphs show that its accuracy is better than any existing methods and time cost is extensively cheap.

© 2019 Elsevier Inc. All rights reserved.

## 1. Introduction

Graph is one of the most common representations of complex data and plays an crucial role in various research areas and many practical applications. Over the past several decades, enormous breakthroughs have been made while many fundamental problems about graphs are remaining to be solved. Comparing graphs is one of the most important problems with a very long history [41]. In practice, the similarity measure of graphs (or equivalently dissimilarity) is widely applied in social science, biology, chemistry, and many other fields. For instance, the similarity measure of graphs can be used to classify ego networks [33], distinguish between neurological disorders [6], identify physical designs of circuits [34], and discover molecules with similar properties [3]. In order to measure similarity between graphs effectively, several definitions of distance or similarity have been proposed [10–13,23]. For example, graph edit distances are the minimum cost for transforming one graph to another by the distortion of vertices and edges [19]. These definitions only pay attention to the similarities of the vertices and edges but lacks the information of topological structures of the graphs. For the purpose of addressing this limitation, frequency subgraph mining algorithms [36], graph kernels [31], and methods based on moments [27] have been

<sup>☆</sup> This article is partially based on preliminary results published in the proceeding of IEEE International Symposium on Information Theory from June 17 to 22, 2018 (ISIT2018Vail).

\* Corresponding author.

E-mail addresses: [hchoi@shanghaitech.edu.cn](mailto:hchoi@shanghaitech.edu.cn) (H. Choi), [hosoo@jejunu.ac.kr](mailto:hosoo@jejunu.ac.kr) (H. Lee), [yshenaw@connect.ust.hk](mailto:yshenaw@connect.ust.hk) (Y. Shen), [shiyim@shanghaitech.edu.cn](mailto:shiyim@shanghaitech.edu.cn) (Y. Shi).

proposed. Moreover, various distance between spectrums are used to measure similarity of graphs [1,9,21,24,26,39,40]. For a survey of similarity measures of graphs, please refer to [14] and to the literature cited therein. However, practically it is almost impossible to find the spectrum of a large-scale graph. Recently many different approaches are proposed for efficient algorithms [2,30,32]. However, these methods are not scalable to large-scale graphs containing millions of edges, which are common in today's applications. As a result, effective and scalable methods for large-scale graphs comparison are urgently needed.

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

## 2. Background and preliminary

Denote  $\mathbb{N}$  the set of nonnegative integer numbers. Let  $M_{m \times n} := M_{m \times n}(\mathbb{C})$  be a set of all  $m \times n$  matrices with entries in the field  $\mathbb{C}$  of complex numbers. We simply denote as  $M_n := M_{n \times n}$ . An  $\ell_2$  norm, known as the Frobenius norm, is defined by  $\|A\|_2 = [\text{tr}(A^*A)]^{1/2}$ .

### 2.1. Graph

Let  $V$  be the set of vertices, and let  $\{x, y\}$  denote the edge connecting two points  $x, y \in V$ . We say that two vertices  $x, y \in V$  are *adjacent* if  $\{x, y\} \in E$ , denoted by  $x \sim y$ . A graph  $G = (V, E)$  is called *finite* if  $V$  is a finite set. Otherwise, it is called *infinite*. In general  $E$  may contain *loops* which means that  $x = y$ . In this paper we consider a finite undirected graph with no loops. The *degree* of a vertex  $x \in V$  is defined by  $\deg(x) = |\{y \in V : y \sim x\}|$ . Two graphs  $G = (V, E)$  and  $G' = (V', E')$  are *isomorphic* if there is a bijection  $f : V \rightarrow V'$  such that  $u \sim v \iff f(u) \sim f(v)$ , denoted by  $G \cong G'$ . For  $m \in \mathbb{N}$ , a finite sequence of vertices  $x_0, x_1, \dots, x_m \in V$  is called a *walk of length  $m$*  if  $x_0 \sim x_1 \sim \dots \sim x_m$ , where some of  $x_0, x_1, \dots, x_m$  may coincide. A graph  $G = (V, E)$  is *connected* if every pair of distinct vertices  $x, y \in V$  ( $x \neq y$ ) is 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  $\partial(x, y)$ . 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  $G = G_1 \sqcup G_2$ . Without loss of generality we assume that  $V = \{1, 2, \dots, n\}$ . The *adjacency matrix* of a graph  $G = (V, E)$  is a  $n \times n$  matrix  $A \in \{0, 1\}^{n \times n}$  where  $A_{ij} = 1$  if and only if  $\{i, j\} \in E$  for all  $i, j \in V$ . Any graph  $G$  can be represented by an adjacency matrix. Every permutation  $\pi : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is associated with a corresponding *permutation matrix*  $P$ . The matrix operator  $P$  left multiplied to matrix  $A$  rearranges the rows according to  $\pi$  which right multiplication with  $P$  rearranges columns of the matrix  $A$ . Given an adjacency matrix  $A$ , graphs corresponding to adjacency matrix  $A$  and  $PAP^T$  are isomorphic for any permutation matrix  $P$ , 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 walks of length  $m$  connecting  $x$  and  $y$ . Remark that  $W_0(x, y) = 0$  if  $x \neq y$  and  $W_0(x, y) = 1$  if  $x = y$ .

**Theorem 2.1.** Let  $G = (V, E)$  be a graph and  $A$  the adjacency matrix. Then we have

$$(A^m)_{ij} = W_m(i, j)$$

for all  $i, j \in V$  and  $m \in \mathbb{N}$ .

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

It is clear that if  $G \cong G'$  then the corresponding eigenvalues of the adjacency matrices are identical. However, in general the converse is not true.

*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 Fig. 1. This is known that it is a unique pair of cospectral graphs among 34 non-isomorphic graphs on 5 vertices. Both have a common characteristic polynomial,  $x^3(x - 2)(x + 2)$ . For more examples for small graphs, see [37] and for more information about cospectral graphs see [5,18,22].

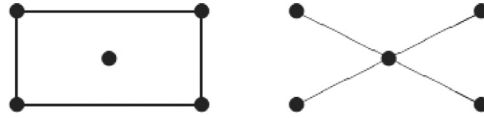


Fig. 1. Cospectral graphs:  $C_4 \cup K_1$  and  $S_5$ .

2.2. Quantum probability

For proof of each theorem and proposition, see [28] and references therein. To measure distance between two graphs we propose to compare the spectral distributions of their adjacency matrices.

**Definition 2.2.** Let  $\mathcal{A}$  be a unital  $*$ -algebra over the complex number field  $\mathbb{C}$  with the multiplication unit  $1_{\mathcal{A}}$ . A function  $\varphi : \mathcal{A} \rightarrow \mathbb{C}$  is called a state on  $\mathcal{A}$  if

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

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

**Proposition 2.3.** A state  $\varphi$  on a unital  $*$ -algebra  $\mathcal{A}$  is a  $*$ -map, i.e.,  $\varphi(a^*) = \overline{\varphi(a)}$ .

**Definition 2.4.** Let  $(\mathcal{A}, \varphi)$  be an algebraic probability space. An element  $a \in \mathcal{A}$  is called an algebraic random variable or a random variable for short. A random variable  $a \in \mathcal{A}$  is called real if  $a = a^*$ .

For a random variable  $a \in \mathcal{A}$  the quantity of the form:

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

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  $a$  in  $\mathcal{A}$  the mixed moments are reduced to the moment sequence:

$$\varphi(a^m), \quad m = 0, 1, 2, \dots,$$

where  $\varphi(a^m)$  is called the  $m$ th moment of  $a$ . By definition  $\varphi(a^0) = 1$ .

For a real random variable  $a = a^*$ , a moment matrix with degree  $n$  is defined as

$$M_n := \begin{bmatrix} \varphi(a^0) & \varphi(a^1) & \dots & \varphi(a^n) \\ \varphi(a^1) & \varphi(a^2) & \dots & \varphi(a^{n+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi(a^n) & \varphi(a^{n+1}) & \dots & \varphi(a^{2n}) \end{bmatrix}. \tag{2.1}$$

**Definition 2.5.** Two real algebraic random variables  $a$  in  $(\mathcal{A}, \varphi)$  and  $b$  in  $(\mathcal{B}, \psi)$  are moment equivalent, denoted by  $a \circ_{eqb} b$ , if  $\varphi(a^m) = \psi(b^m)$  for all  $m \in \mathbb{N}$ .

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

**Theorem 2.6.** Let  $(\mathcal{A}, \varphi)$  be an algebraic probability space. For a real random variable  $a = a^* \in \mathcal{A}$  there exists a probability measure  $\mu \in \mathfrak{B}(\mathbb{R})$  such that

$$\varphi(a^k) = \int_{\mathbb{R}} x^k d\mu(x) \quad \text{for all } k \in \mathbb{N}_0. \tag{2.2}$$

Such  $\mu$  is called the spectral distribution of  $a$  in  $\varphi$ .

It is noted that  $M_n$  with the usual operators is a unital  $*$ -algebra. Recall that a matrix  $\rho \in M_n$  is called a density matrix if it is positive semidefinite and  $\text{tr } \rho = 1$ .

**Definition 2.7.** For  $A = [a_{ij}] \in M_n$ , the following are states on  $M_n$ , implying that  $(M_n, \varphi)$  is an algebraic probability space.

- (1) (Normalized trace) The normalized trace is defined by

$$\varphi_{\text{tr}}(A) = \frac{1}{n} \text{tr}(A) = \frac{1}{n} \sum_{i=1}^n a_{ii}.$$

- (2) (Vector state) For a unit vector  $\xi \in \mathbb{C}^n$ , we define

$$\varphi_{\xi}(A) = \langle \xi, A\xi \rangle, \quad A \in M_n,$$

where  $\langle \cdot, \cdot \rangle$  is the usual inner product in  $\mathbb{C}^n$ . Such a state is called a vector state with the state vector  $\xi$ .

- (3) (Density matrix state) For each density matrix  $\rho \in M_n$  we define

$$\varphi_{\rho}(A) = \text{tr}(\rho A), \quad A \in M_n.$$

Such a state is called a density matrix state with the density matrix  $\rho$ .

**Proposition 2.8.** For any state  $\varphi$  on  $M_n$  there exists a unique density matrix  $\rho$  such that  $\varphi = \varphi_{\rho}$ .

### 3. Main results

**Lemma 3.1.** Let  $A \in M_n$  have distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_s$  and let

$$q(t) = (t - \lambda_1)(t - \lambda_2) \dots (t - \lambda_s).$$

Then  $A$  is diagonalizable if and only if  $q(A) = 0$ .

Denote the Dirac measure at  $\lambda$  as  $\delta_\lambda$ . (i.e.,  $\delta_\lambda(S) = 1$  if  $\lambda \in S$  and  $\delta_\lambda(S) = 0$  if  $\lambda \notin S$ ). The support of measure,  $\mu$ , is denoted by  $\text{supp}(\mu)$ . A measure  $\mu$  is called a *measure with  $n_0$  mass point* if  $|\text{supp}(\mu)| = n_0$ .

**Theorem 3.2.** Let  $\{s_k\}$  be a real sequence and let

$$\mathcal{H}_n = \begin{bmatrix} s_0 & s_1 & \cdots & s_n \\ s_1 & s_2 & \cdots & s_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & \cdots & s_{2n} \end{bmatrix}. \tag{3.1}$$

be the Hankel matrix. If  $\det(\mathcal{H}_n) > 0$  for all  $n < s$  and  $\det(\mathcal{H}_n) = 0$  for all  $n \geq s$ . Then there exists unique discrete measure,  $\mu$ , with  $|\text{supp}(\mu)| = s$  such that

$$s_k = \int_{\mathbb{R}} x^k d\mu \quad \text{for all } k \in \mathbb{N}.$$

**Proof.** See [4, Theorem 1.1].  $\square$

A Hermitian matrix  $A \in M_n$  can be regarded as a real random variable in the algebraic probability space  $(M_n, \varphi_\xi)$  with a vector state  $\varphi_\xi$ , by Theorem 2.6 it follows that there exists the spectral distribution of  $A$  in  $\varphi_\xi$  satisfying (3.2). In the following theorem we provide an explicit form of such measure.

**Theorem 3.3.** Let  $(M_n, \varphi_\xi)$  be the algebraic probability space with a vector state  $\varphi_\xi$  and let  $A \in M_n$  be a Hermitian matrix whose all eigenvalues are distinct,  $\lambda_1, \dots, \lambda_n$ . There exists a unique probability discrete measure  $\mu$  such that

$$\varphi_\xi(A^k) = \int_{\mathbb{R}} x^k d\mu(x) \quad \text{for all } k \in \mathbb{N}. \tag{3.2}$$

Furthermore, the measure has an explicit form  $\mu = \sum_{i=1}^n \omega_i \delta_{\lambda_i}$ . Conversely, for a probability discrete measure  $\mu = \sum_{i=1}^n \omega_i \delta_{\lambda_i}$ , there exists a Hermitian matrix  $A \in M_n$  with eigenvalues  $\lambda_1, \dots, \lambda_n$  such that  $A$  holds the equality (3.2).

**Proof.** ( $\Rightarrow$ ) Let  $A \in \mathbb{C}^{n \times n}$  be a Hermitian matrix. By Spectral Theorem,  $A$  can be diagonalized by a unitary matrix  $U$ . That is,  $A = UDU^*$ . Put  $v = U^* \xi = [v_1, \dots, v_n]^T \in \mathbb{C}^n$ ,  $\omega_i = |v_i|^2$  for all  $i = 1, \dots, n$ , and  $\mu = \sum_{i=1}^n \omega_i \delta_{\lambda_i}$ . Then the  $k$ th moment of  $A$  is

$$\varphi_\xi(A^k) = \xi^* A^k \xi = v^* D^k v = \sum_{i=1}^n \omega_i \lambda_i^k = \int_{\mathbb{R}} x^k d\mu.$$

Since it holds that

$$\int_{\mathbb{R}} d\mu = \sum_{i=1}^n \omega_i = \sum_{i=1}^n v^* v = \sum_{i=1}^n \xi^* U U^* \xi = 1,$$

the measure  $\mu$  is a probability measure.

( $\Leftarrow$ ) Let  $\mu = \sum_{i=1}^n \omega_i \delta_{\lambda_i}$  with  $\omega_i \geq 0$  and  $\lambda_i \in \mathbb{R}$  for all  $i \in \mathbb{N}$ , and  $\sum_{i=1}^n \omega_i = 1$ . Let  $D$  be the  $n \times n$  diagonal matrix whose diagonal entries are  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Let  $v = [\sqrt{\omega_1} \dots \sqrt{\omega_n}]^T$ . Since  $v$  and  $\xi$  both are unit vectors, there exists a unitary matrix  $U$  such that  $Uv = \xi$ . Take  $A = UDU^*$ . Then  $A$  holds the equality (3.2).  $\square$

Note that  $\langle u_i, \xi \rangle = \cos \theta_i$  such that  $\theta_i$  is the angle between  $u_i$  and  $\xi$ , where  $u_i$  is  $i$ th column vector of  $U$ . Since  $\sum \omega_i = 1$ , it holds that  $\sum |\cos \theta_i|^2 = 1$ . So,  $\cos \theta_i$  are the direction cosines of  $\xi$  with respect to orthonormal eigenvectors,  $u_1, u_2, \dots, u_n$ . Especially, if  $\xi = u_i$ , then the spectral distribution is  $\mu = \delta_{\lambda_i}$ .

**Corollary 3.4.** Let  $(M_n, \varphi_{\text{tr}})$  be the algebraic probability space with the normalized trace state  $\varphi_{\text{tr}}$  and let  $A \in M_n$  be a Hermitian matrix whose all eigenvalues,  $\lambda_1, \dots, \lambda_n$ , are distinct. An explicit form of the unique probability discrete measure  $\mu$  such that  $\varphi_{\text{tr}}(A^k) = \int_{\mathbb{R}} x^k d\mu(x)$  for all  $k \in \mathbb{N}$  is  $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i}$ .

**Lemma 3.5.** If Hermitian matrices  $A, \tilde{A} \in M_n$  satisfy  $\varphi_{\text{tr}}(A^k) = \varphi_{\text{tr}}(\tilde{A}^k)$  for all  $k = 1, 2, \dots, n$ , then  $A$  and  $\tilde{A}$  have same spectrums.

**Proof.** Recall that a monic polynomial is a univariate polynomial in which the leading coefficient (the nonzero coefficient of highest degree) is equal to 1. Let  $\lambda_i, \tilde{\lambda}_i$  be eigenvalues of  $A, \tilde{A}$ , respectively. Let  $f$  and  $\tilde{f}$  be degree  $m$  monic polynomial functions whose roots consist of eigenvalues of  $A$  and  $\tilde{A}$ , respectively. Since  $\sum_i (\lambda_i)^k = \sum_i (\tilde{\lambda}_i)^k$ ,  $k = 1, 2, \dots, n$ , by Newton's

identities it follows that the coefficients of two polynomials  $f$  and  $\hat{f}$  are identical. Thus, the roots of  $f$  and  $\hat{f}$  are identical. Therefore,  $\lambda_i = \tilde{\lambda}_i$  for all  $i$ .  $\square$

Remark that while the spectral distribution of a Hermitian matrix  $A$  in  $\varphi_{tr}$  includes only information about eigenvalues for  $A$ , the spectral distribution of  $A$  in  $\varphi_\xi$  generally includes information about the corresponding eigenvectors as well as eigenvalues for  $A$ .

Now we generalize [Theorem 3.3](#) for any Hermitian matrices.

**Theorem 3.6.** *Let  $(M_n, \varphi_\xi)$  be the algebraic probability space with a vector state  $\varphi_\xi$  and let  $A \in M_n$  be a real random variable. Let  $\lambda_1, \lambda_2, \dots, \lambda_s$  with respective multiplicities  $n_1, \dots, n_s$ , and let  $\Lambda = \lambda_1 I_{n_1} \oplus \dots \oplus \lambda_s I_{n_s}$ . Supposed that  $U = [U_1 \ U_2 \ \dots \ U_s] \in M_n$  is unitary matrix such that  $A = U\Lambda U^*$  and for each  $i = 1, 2, \dots, s$ ,*

$$U_i = [u_1^{(i)} \ u_2^{(i)} \ \dots \ u_{n_i}^{(i)}] \in M_{n \times n_i},$$

where  $u_1^{(i)}, \dots, u_{n_i}^{(i)}$  are the corresponding unit eigenvectors of  $\lambda_i$ . Then there uniquely exists a probability discrete measure,  $\mu = \sum_{i=1}^s \omega_i \delta_{\lambda_i}$ , such that

$$\varphi_\xi(A^k) = \int_{\mathbb{R}} x^k d\mu(x) \quad \text{for all } k \in \mathbb{N}. \tag{3.3}$$

Furthermore,  $\omega_i = \sum_{j=1}^{n_i} |\cos \theta_j^{(i)}|^2$  where  $\theta_j^{(i)}$  is the angle between  $u_j^{(i)}$  and  $\xi$ .

**Proof.** Note that since  $A$  is a Hermitian matrix, by Spectral Theorem,  $A$  can be diagonalized by an unitary matrix. Let  $\omega_i = \sum_{j=1}^{n_i} |\langle u_j^{(i)}, \xi \rangle|^2$  and  $\mu = \sum_{i=1}^s \omega_i \delta_{\lambda_i}$ . Then it follows that for each  $k \in \mathbb{N}$

$$\varphi_\xi(A^k) = \xi^* U \Lambda^k U^* \xi = \xi^* U_1 \lambda_1^k I_{n_1} U_1^* \xi \oplus \dots \oplus \xi^* U_s \lambda_s^k I_{n_s} U_s^* \xi = \sum_{i=1}^s \omega_i \lambda_i^k = \int_{\mathbb{R}} x^k d\mu(x).$$

Since it holds that

$$\int_{\mathbb{R}} d\mu = \sum_{i=1}^s \omega_i = \sum_{i=1}^s \sum_{j=1}^{n_i} |\langle u_j^{(i)}, \xi \rangle|^2 = \sum_{i=1}^s |U_i^* \xi|^2 = \xi^* U U^* \xi = 1$$

the measure  $\mu$  is a probability measure. Since  $u_j^{(i)}$  for all  $i, j$  is a unit vector and  $\xi$  is also a unit vector, it follows that

$$\omega_i = \sum_{j=1}^{n_i} |\langle u_j^{(i)}, \xi \rangle|^2 = \sum_{j=1}^{n_i} |\cos \theta_j^{(i)}|^2,$$

where  $\theta_j^{(i)}$  is the angle between  $u_j^{(i)}$  and  $\xi$ .

**(Uniqueness)** (To apply [Theorem 3.2](#), we have to show that  $\det(\mathcal{M}_j)$  is nonnegative.) For the moment sequence  $\{\varphi_\xi(A^k)\}_k$ , the corresponding moment matrix  $\mathcal{M}_j$  with degree  $j$  is defined as

$$\mathcal{M}_j := \begin{bmatrix} \varphi_\xi(A^0) & \varphi_\xi(A^1) & \dots & \varphi_\xi(A^j) \\ \varphi_\xi(A^1) & \varphi_\xi(A^2) & \dots & \varphi_\xi(A^{j+1}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_\xi(A^j) & \varphi_\xi(A^{j+1}) & \dots & \varphi_\xi(A^{2j}) \end{bmatrix} = \begin{bmatrix} \xi^* I \xi & \xi^* A \xi & \dots & \xi^* A^j \xi \\ \xi^* A \xi & \xi^* A^2 \xi & \dots & \xi^* A^{j+1} \xi \\ \vdots & \vdots & \ddots & \vdots \\ \xi^* A^j \xi & \xi^* A^{j+1} \xi & \dots & \xi^* A^{2j} \xi \end{bmatrix}.$$

Since for any  $n$

$$\mathcal{M}_n = [I \xi \ A \xi \ \dots \ A^j \xi]^* [I \xi \ A \xi \ \dots \ A^j \xi],$$

$\mathcal{M}_n$  is positive semidefinite and  $\det(\mathcal{M}_j) \geq 0$ .

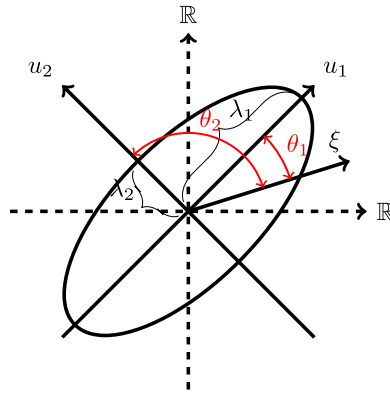
Let  $\lambda_1, \lambda_2, \dots, \lambda_s$  be the distinct eigenvalues of  $A$  and

$$q(t) = (t - \lambda_1)(t - \lambda_2) \dots (t - \lambda_s) = \sum_{j=1}^s c_j t^j$$

Since  $A$  is a Hermitian matrix, by [Lemma 3.1](#) it follows that  $q(A) = \sum_{j=1}^s c_j A^j = 0$ . Then it follows that

$$0 = \langle \xi, q(A) \xi \rangle = \xi^* \sum_{j=1}^s c_j A^j \xi = \sum_{j=1}^s c_j \varphi_\xi(A^j).$$

So, it is clear that  $\det(\mathcal{M}_j) = 0$  for  $j \geq s$  and  $\det(\mathcal{M}_j) > 0$  for  $j < s$ . By [Theorem 3.2](#), there exists a unique discrete measure with  $s$  number of point mass. Therefore,  $\mu$  is a unique probability discrete measure.  $\square$



**Fig. 2.** For a symmetric matrix  $A \in M_{2 \times 2}(\mathbb{R})$  and a unit vector  $\xi \in \mathbb{R}^2$ , there exists the probability discrete measure  $\mu = \cos^2 \theta_1 \delta_{\lambda_1} + \cos^2 \theta_2 \delta_{\lambda_2}$  such that  $\xi^T A^k \xi = \int_{\mathbb{R}} x^k d\mu$  for all  $k \in \mathbb{N}$ .

**Example 3.7.** Consider  $2 \times 2$  symmetric matrix  $A \in M_2$  as

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}.$$

Let  $\xi = [\xi_1 \ \xi_2]^T$  be a unit vector in  $\mathbb{R}^2$ . Then by [Theorem 3.2](#), there exists a unique probability measure  $\mu = \omega_1 \delta_{\lambda_1} + \omega_2 \delta_{\lambda_2}$  such that

$$\varphi_{\xi}(A^k) = \xi^T A^k \xi = \int_{\mathbb{R}} x^k d\mu(x) \quad \text{for all } k \in \mathbb{N}.$$

It is easy to check a closed form of the measure is  $\mu = (1/2 + \xi_1 \xi_2) \delta_3 + (1/2 - \xi_1 \xi_2) \delta_1$ . The weights  $\omega_1, \omega_2$  depends on  $\xi$ . Especially, when  $\xi = [1/\sqrt{2} \ 1/\sqrt{2}]^T$ , the spectral distribution of  $A$  is  $\mu = \delta_3$ . When  $\xi = [-1/\sqrt{2} \ 1/\sqrt{2}]^T$ , the spectral distribution of  $A$  is  $\mu = \delta_1$  (see [Fig. 2](#)).

Since  $\omega_i$  depends on  $\xi$  for all  $i$ , the spectral distribution of a given Hermitian matrix  $A \in M_n$  depends on a unit vector  $\xi \in \mathbb{C}^n$ . However, since the eigenvalues do not change, the Dirac measures  $\delta_{\lambda_i}$  do not change. The only weights  $\omega_i$  depends on a unit vector  $\xi \in \mathbb{C}^n$ .

**Lemma 3.8.** Let  $A$  be a  $n \times n$  Hermitian matrix and  $\xi$  be a unit column vector in  $\mathbb{C}^n$ . Then  $\varphi_{\xi}(A^k) = \lambda^k$  for all  $k \in \mathbb{N}$  if and only if  $A\xi = \lambda\xi$ .

**Proof.** ( $\Leftarrow$ ) Since  $A\xi = \lambda\xi$  implies  $A^k \xi = \lambda^k \xi$  for all  $k \in \mathbb{N}$ , it follows that  $\xi^* A^k \xi = \xi^* \lambda^k \xi = \lambda^k$ .

( $\Rightarrow$ ) Since  $\varphi_{\xi}(A^k) = \lambda^k$ , by definition  $\xi^* A^k \xi = \lambda^k$ . Then by [Theorem 3.6](#) it follows that  $\sum_{i=1}^s \omega_i \lambda_i^k = \lambda^k$  for all  $k \in \mathbb{N}$ , where  $\lambda_1, \dots, \lambda_s$  are the distinct eigenvalues of  $A$ . By the uniqueness of the spectral distribution,  $\lambda = \lambda_{\ell}$  for some  $1 \leq \ell \leq s$ . The following linear system

$$\begin{bmatrix} 1 & 1 & \dots & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \dots & \lambda_s \\ \lambda_1^2 & \lambda_2^2 & \dots & \dots & \lambda_s^2 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \lambda_1^{s-1} & \lambda_2^{s-1} & \dots & \dots & \lambda_s^{s-1} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \vdots \\ \omega_s \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_{\ell}^1 \\ \lambda_{\ell}^2 \\ \vdots \\ \lambda_{\ell}^s \end{bmatrix}$$

has a unique solution  $\omega_i = 0$  for all  $i \neq \ell$  and  $\omega_{\ell} = 1$ . Thus,  $\xi$  is the corresponding eigenvector of  $\lambda_{\ell}$ .  $\square$

**Theorem 3.9.** The following statement are equivalent.

- (i) There exists a unique discrete measure  $\mu$  with  $n_0$  mass points such that  $m_k = \int x^k d\mu$  for all  $k \in \mathbb{N}$ ;
- (ii) There is a Hermitian matrix  $A \in \mathbb{C}^{m \times m}$  with  $n_0$  distinct eigenvalues and a unit vector  $\xi \in \mathbb{C}^m$  such that  $m_k = \xi^* A^k \xi$  for all  $k \in \mathbb{N}$ ;
- (iii)  $\mathcal{M}_n \geq 0$  for all  $n \in \mathbb{N}$ , and  $\mathcal{M}_n > 0$  if and only if  $n < n_0$ .

**Proof.** (i)  $\Rightarrow$  (ii) Let  $\mu = \sum_{i=1}^{n_0} \omega_i \delta_{\lambda_i}$  with  $\omega_i \geq 0$  and  $\lambda_i \in \mathbb{R}$  for all  $i \in \mathbb{N}$ , and  $\sum_{i=1}^{n_0} \omega_i = 1$  and  $\xi$  be a unit vector in  $\mathbb{C}^n$ . Let  $D$  be the  $n \times n$  diagonal matrix whose diagonal entries are  $\lambda_1, \lambda_2, \dots, \lambda_{n_0}, 0, \dots, 0$ . Let  $v = [\sqrt{\omega_1} \ \dots \ \sqrt{\omega_{n_0}} \ 0 \ \dots \ 0]^T$ . Since  $v$  and  $\xi$  both are unit vectors, there exists a unitary matrix  $U$  such that  $Uv = \xi$ . Let  $A = UDU^*$ . Then it follows that  $m_k = \xi^* A^k \xi$  for all  $k \in \mathbb{N}$ .

(ii)  $\Rightarrow$  (iii) Let  $c = [c_0, c_1, \dots, c_n]^*$  be a vector in  $\mathbb{C}^{n+1}$ . Then

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

Since  $c \in \mathbb{C}^{n+1}$  is arbitrary,  $\mathcal{M}_n$  is positive semidefinite for all  $n \in \mathbb{N}$ . Since  $A$  is Hermitian, the minimal polynomial is  $q(x) := (x - \lambda_1)(x - \lambda_2) \dots (x - \lambda_{n_0})$  where  $\lambda_i$  are all eigenvalues. So,  $\sum_{i=1}^{n_0} q_i A^i = 0$ , implying  $\| \sum_{i=0}^{n_0} q_i A^i e \| = 0$ . Thus  $\mathcal{M}_n$  is singular for  $n \geq n_0$ . Suppose that there exists a polynomial  $r(x) := \sum_{i=1}^m r_i x^i$  with  $m < n_0$  such that  $r(A)e = 0$ . Since  $n_0$  is the degree of minimal polynomial,  $r(x)$  is a zero function.

(iii)  $\Rightarrow$  (i) See [4, Theorem 1.1].  $\square$

Denote the set of  $n \times n$  permutation matrices as  $\mathcal{S}$ . Denote the identity matrix as  $I$  and the matrix whose all entries are 1 as  $J$ .

**Definition 3.10.** Let  $(M_n, \varphi)$  be an algebraic probability space. A state  $\varphi$  is called permutationally invariant on  $M_n$  if

$$\varphi(A) = \varphi(P^T A P) \quad \text{for all } A \in M_n, P \in \mathcal{S}. \tag{3.4}$$

**Lemma 3.11.** A necessary and sufficient condition that  $\varphi$  is permutationally invariant is that there exists a density matrix  $\rho$  such that  $\varphi(A) = \text{tr}(\rho A)$  satisfying

$$\rho = pI + qJ, \quad n(p + q) = 1, \quad p \geq 0, \quad p + qn \geq 0. \tag{3.5}$$

**Proof.** Recall that for any state  $\varphi$  on  $M_n$  there exists a unique density matrix  $\rho$  such that  $\varphi = \varphi_\rho$ .

( $\Rightarrow$ ) If  $\rho = pI + qJ$ , then  $\text{tr}(\rho P^T A P) = \text{tr}(P^T \rho P A) = \text{tr}(\rho A)$  for all  $A \in M_n, P \in \mathcal{S}$ .

( $\Leftarrow$ ) Let  $\rho = [\rho_{ij}]$  be a density matrix. Since  $\text{tr}(\rho P^T A^T P) = \text{tr}(\rho A^T)$  for all

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \in M_n,$$

it holds  $\rho_{11}a_{11} + \rho_{12}a_{12} + \rho_{21}a_{21} + \rho_{22}a_{22} = \rho_{ii}a_{11} + \rho_{ij}a_{12} + \rho_{ji}a_{21} + \rho_{jj}a_{22}$  for all  $i, j$ . Since  $a_{11}, a_{12}, a_{21}, a_{22}$  are arbitrary,  $\rho_{ii} = \rho_{jj}$  for all  $i, j$  and  $\rho_{ij} = \rho_{kl}$  for all  $i \neq j, k \neq l$ . So,  $\rho$  is of the form  $\rho = pI + qJ$ . Note that the eigenvalues of  $pI + qJ$  are  $p$  and  $p + qn$ . Since  $\text{tr}(pI + qJ) = 1$  and  $pI + qJ \geq 0$ , it follows that  $n(p + q) = 1, p \geq 0$ , and  $p + qn \geq 0$ .  $\square$

**Theorem 3.12.** Let  $A$  be the adjacency matrix of a given graph  $G$ . Then the  $k$ th moment of  $A$  in a permutationally invariant state is a graph invariant.

Let  $A$  be the adjacency matrix of a given graph  $G$ . By Theorem 2.1, it is easy to check that  $\varphi_{\text{tr}}(A^k)$  is the average of closed walks of length  $k$  in  $G$ . Denote the  $n$  dimensional all-ones column vector by  $1_n$ , and denote  $1_n / \|1_n\|$  by  $e$ . From now on, the vector state with the state vector  $\xi = e$  will be mainly used to compare two graphs. Specifically, the state  $\varphi_e : M_n \rightarrow \mathbb{C}$  is defined by

$$\varphi_e(A) = \langle e, Ae \rangle \tag{3.6}$$

for all  $A \in M_n$ . Then it is clear that  $\varphi_e$  is a state on  $M_n$ , implying that  $(M_n, \varphi_e)$  is an algebraic probability space. Note that it holds

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

where  $\mathbb{E}(v) = \frac{1}{n} \sum_{i=1}^n v_i$  is the average of entries of vector  $v$ . Since the value of  $(A^k)_{ij}$  is equal to the number of walks of length  $k$  from vertex  $i$  to vertex  $j$  and  $A^k 1_n$  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  $i$ ,  $\varphi_e(A^k)$  is the average of the the sum of the number of all walks of length  $k$  from each vertex.

**Proposition 3.13.** If  $\varphi_{\text{tr}}(A) = \varphi_{\text{tr}}(B)$  and  $\varphi_e(A) = \varphi_e(B)$  for all  $k \in \mathbb{N}$ , then  $\varphi(A) = \varphi(B)$  for all permutationally invariant state  $\varphi$ .

**Proof.** Since  $\varphi_{\text{tr}}(A) = \frac{1}{n} \text{tr}(IA)$  and  $\varphi_e(A) = \frac{1}{n} \text{tr}(JA)$  for all  $A \in M_n$ , by Lemma 3.11 it holds.  $\square$

In other words, if the averages of closed walks and all walks of length  $k$  in two graphs are identical for all  $k$ , respectively, then their adjacency matrices are moment equivalent in  $(M_n, \varphi)$  for any permutationally invariant state  $\varphi$ . Especially,



if two Hermitian matrices  $A, B \in M_n$  have distinct eigenvalues, respectively, and  $\varphi_e(A^k) = \varphi_e(B^k)$  for all  $k \in \mathbb{N}$ , then for all permutation invariant state  $\varphi$ , we have  $\varphi(A^k) = \varphi(B^k)$  for all  $k \in \mathbb{N}$ .

Using several known results for the number of all walks of length  $k$ , it is easy to check the following results.

**Proposition 3.14.** *Let  $A \in M_n$  be an adjacency matrix of a given graph. Then the following are true.*

- (1)  $\varphi_e(A) \leq \frac{1}{n} \sum_{i=1}^n \deg(i)^k$  for all  $k \geq 1$ ,
- (2)  $\varphi_e(A^k) \leq \Delta^k$  for all  $k \geq 1$  and  $\varphi_e(A^k) \leq 2\varphi_e(A)\Delta^{k-1}$  for all  $k \geq 2$ ,
- (3)  $(\varphi_e(A))^k \leq \varphi_e(A^k)$  for all  $k \in \mathbb{N}$ ,
- (4)  $\varphi_e(A^{2a+b})\varphi_e(A^b) \leq \varphi_e(A^{2a+2b})$  for all  $a, b \in \mathbb{N}$ ,
- (5)  $\varphi_e(A^{a+b})\varphi_e(A^{a+b}) \leq \varphi_e(A^{2a})\varphi_e(A^{2b})$  for all  $a, b \in \mathbb{N}$ ,

where  $\Delta$  is the maximum degree.

**Proof.** For (1) see [16]. For (2) see [7, Theorem 2]. For (3)–(5) see [35, Theorem 1–3].  $\square$

Since  $\varphi_e$  is a permutationally invariant state, the  $k$ th moment of  $A$  in  $\varphi_e$  is a graph invariant.

**Theorem 3.15.** *Let  $A$  be the adjacency matrix of a given graph  $G$ . Then the moment matrix  $\mathcal{M}_n$  is a graph invariant.*

Hence, we will henceforth denote  $\mathcal{M}_n$  as  $\mathcal{M}_n(G)$  if a graph  $G$  is given.  $\mathcal{M}_n(G)$  is an informative representation for the given graph  $G$ . Indeed,  $\mathcal{M}_n(G)$  includes information about the spectral properties of the adjacent matrix of  $G$ .

To measure similarity between two large-scale graphs, we compare the spectral distributions of their adjacency matrices. There are various distances and divergences between two distributions such as Kullback-Leibler divergence, Bhattacharyya distance, etc [8,25]. However, since large-scale graphs in real world possibly have rich spectrums, to reconstruct the spectral distributions is almost impossible in practice. Instead, we can use moments of the distributions. In general, all the moments up to infinity are required 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 [17,20]. Since the moment matrix has sufficient information about the distribution, a distance between moment matrices can be calculated to measure a distance between two spectral distributions.

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

$$d(G, \tilde{G}) := \delta(\mathcal{M}_m(G), \mathcal{M}_m(\tilde{G})),$$

where  $m \in \mathbb{N}$  is fixed and  $\delta(\cdot, \cdot)$  is a distance between positive (semi)definite matrices. If  $\mathcal{M}_m$  is positive definite for some  $m$ , then it is a point on the Riemannian manifold of positive definite matrices (see [4, Theorem 1.1]). Denote the set of all  $n \times n$  positive definite matrices as  $\mathcal{P}^o$ . There are various distances between two positive definite matrices such as Frobenius, Cholesky-Frobenius, J-divergence, Affine-invariant, Log-Frobenius [38]. The affine-invariant distance on  $\mathcal{P}^o$  given by  $\delta(A, B) = \|\log(A^{-1/2}BA^{-1/2})\|_2$  for any  $A, B \in \mathcal{P}^o$ . Our computational results showed that the affine-invariant distance is slightly better than other distances. In this article, we mainly use the Affine-invariant metric. However, since the moment matrix for a graph with few vertices can be possibly singular positive semidefinite, we alternatively use the Frobenius distance in such case. It is questionable which distance is the best in some sense. We remain for the future work.

**Theorem 3.16.** *For graphs  $G, \tilde{G}, \hat{G}$ ,*

- (a) (Nonnegativity)  $d(G, \tilde{G}) \geq 0$ ,
- (b) (Identification)  $d(G, \tilde{G}) = 0$  if  $G = \tilde{G}$ ,
- (c) (Symmetry)  $d(G, \tilde{G}) = d(\tilde{G}, G)$ ,
- (d) (Triangle Inequality)  $d(G, \hat{G}) \leq d(G, \tilde{G}) + d(\tilde{G}, \hat{G})$ .

**Definition 3.17.** A property of graphs is called *sub-structure invariant* if the property of  $G$  holds for  $G \sqcup G \sqcup \dots \sqcup G$ .

Note that  $k$ th moments of adjacency matrices of  $G$  and  $G \sqcup G$  in  $\varphi_e$  are same for each  $k \in \mathbb{N}$ , so their distributions are identical. Then it is easy to show the following proposition. The proof is left to the reader.

**Proposition 3.18.** *Let  $A$  be the adjacency matrix of a given graph  $G$ . Then the moment matrix of  $A$  in  $\varphi_e$  is sub-structure invariant.*

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 (see Figs. 3 and 4).

#### 4. Cospectral graphs and various examples

The smallest pair of cospectral graphs is the graph union  $C_4 \cup K_1$  and star graph  $S_5$ , illustrated in Fig. 1. While the corresponding adjacency matrices are different, both have the same graph spectrum,  $-2, 0, 0, 0, 2$ . Let  $A$  and  $\tilde{A}$  be the adjacency matrix of  $C_4 \cup K_1$  and  $S_5$ , respectively. If  $A$  and  $\tilde{A}$  are considered as real algebraic random variables in  $(M_5, \varphi_{\text{tr}})$ , then it is easy to check that they are moment equivalent. Thus, their spectral distributions in  $\varphi_{\text{tr}}$  are identical (see Fig. 5).



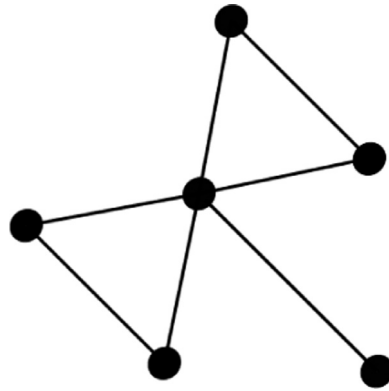


Fig. 3. Graph G.

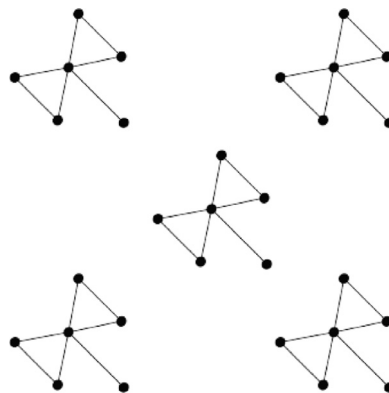


Fig. 4. Graph  $\tilde{G} = G \sqcup G \sqcup G \sqcup G \sqcup G$ .

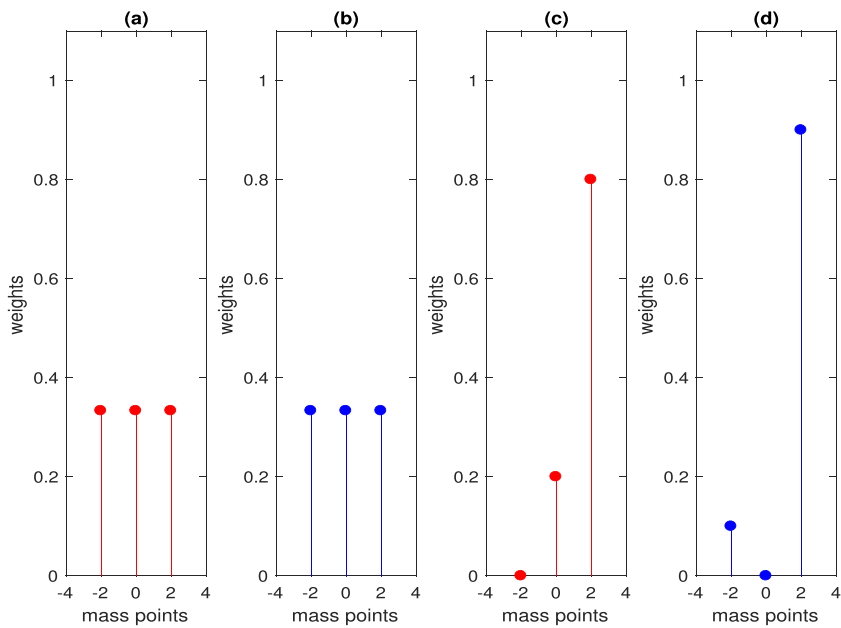


Fig. 5. (a) and (b) are spectral distributions of  $G_1$  and  $G_2$  in  $\varphi_{tr}$ ; (c) and (d) are spectral distributions of  $G_1$  and  $G_2$  in  $\varphi_e$ . It shows that (a) and (b) are identical while (c) and (d) are distinguishable.

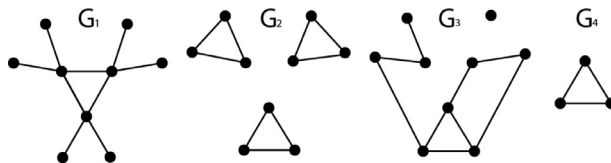


Fig. 6. Four different graphs with the same number of vertices and edges are shown in [30].

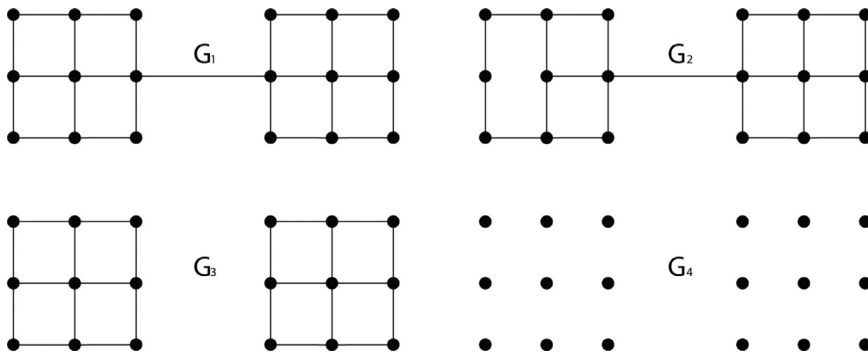


Fig. 7. Four different graphs with the same number of vertices in [39].

Table 1

To scale values of  $\|\mathcal{M}_2(G_i) - \mathcal{M}_2(G_j)\|_2$  we alternatively use the value of  $\log(\|\mathcal{M}_2(G_i) - \mathcal{M}_2(G_j)\|_2 + 1)$ . Note that if  $d(\cdot, \cdot)$  is a distance function, then so is  $\psi(d(\cdot, \cdot))$  for  $\psi(x) = \log(x + 1)$ .

	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$
$G_1$	0	2.5325	2.8009	4.4449	11.3354
$G_2$	2.5325	0	1.5889	4.2998	11.3355
$G_3$	2.8009	1.5889	0	4.2473	11.3356
$G_4$	4.4449	4.2998	4.2473	0	11.3363
$G_5$	11.3354	11.3355	11.3356	11.3363	0

However, If  $A$  and  $\tilde{A}$  are considered as real algebraic random variables in  $(M_4, \varphi_e)$ , then they are not moment equivalent. So, using the state  $\varphi_e$  allows us to distinguish two graphs. Indeed, the moment matrices

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

are different, implying that each spectral distributions are different (see Fig. 5).

Fig. 6 is introduced in [30]. Three graphs have the same number of vertices and edges. Hamming distance and graph edit distance do not capture relevant topological differences. However, our proposed measure perform a highly precise comparison as the measure in [30] does. It shows that  $d(G_1, G_2) = 14.0844$ ,  $d(G_1, G_3) = 30.3974$ , and  $d(G_2, G_3) = 16.3209$ . By the sub-structure invariant property,  $d(G_2, G_4) = 0$ .

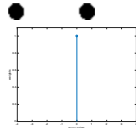
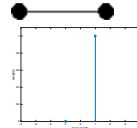
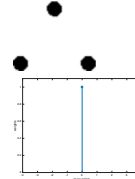
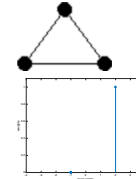
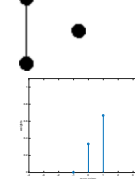
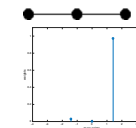
Five graphs  $G_1, G_2, G_3, G_4$  on 18 vertices in Fig. 7 and  $G_5 = K_{18}$  are considered in [39] to measure dissimilarity between them. However, using the measure in [39] the graphs  $G_1, G_2, G_3, G_5$  are not distinguishable from  $G_4$  with  $d(G_j, G_4) = 18$ ,  $j = 1, 2, 3, 5$ . Table 1 shows that our proposed measure overcomes such drawbacks. Note that the spectral distributions for the graphs  $G_1, G_2, G_3, G_4, G_5$  are all different. It shows that the distance between two graphs among them follows dissimilarities between the distributions.

There are two non-isomorphic simple graphs with 2 vertices and four non-isomorphic simple graphs with 3 vertices. The moment matrix distances  $\|\mathcal{M}_2(G) - \mathcal{M}_2(\tilde{G})\|_2$  between two graphs  $G$  and  $\tilde{G}$  among them are shown in Table 2. The spectral distributions of their adjacency matrices for each graphs are shown as well. There are 11 non-isomorphic simple graphs with 4 vertices. Their spectral distributions and their pairwise distances are shown as well in Tables 3 and 4, respectively.


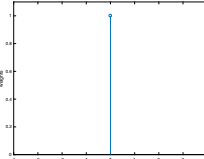

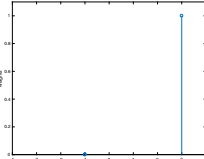

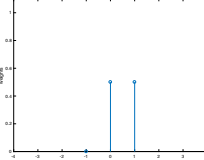

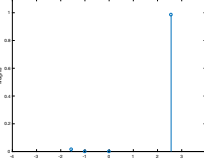

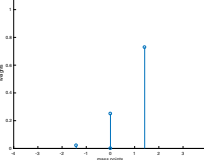
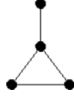
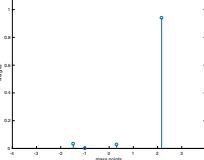

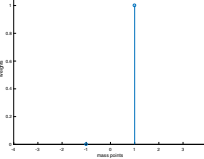

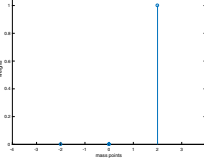

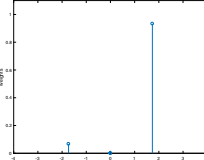

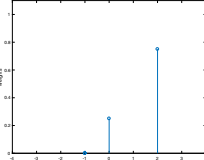

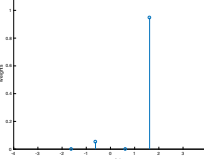
### 5. Complexity and parallelism

Our proposed method requires mainly two computational steps. Assume that we consider the moment matrix with fixed degree  $n$ , i.e.,  $(n + 1) \times (n + 1)$ . The first step is to obtain the moment matrix  $\mathcal{M}_n$  whose entries consist of the moment sequence  $\{m_k\}_{k=0}^{2n}$ . In the second step, we use the affine-invariant distance between positive (semi)definite matrices to

**Table 2**  
Dissimilarities between two graphs among non-isomorphic simple graphs with 2 or 3 vertices.

						
	$2K_1$	$K_2$	$3K_1$	$K_3$	$\bar{P}_3$	$P_3$
$2K_1$	0	2.8284	0	20.9762	1.8856	6.7659
$K_2$	2.8284	0	2.8284	18.7617	0.9428	4.2164
$3K_1$	0	2.8284	0	20.9762	1.8856	6.7659
$K_3$	20.9762	18.7617	20.9762	0	19.4822	14.6211
$\bar{P}_3$	1.8856	0.9428	1.8856	19.4822	0	5.0332
$P_3$	6.7659	4.2164	6.7659	14.6211	5.0332	0

**Table 3**  
Spectral distributions of 11 non-isomorphic simple graphs with 4 vertices.

Name	Graph	Spectral distribution	Name	Graph	Spectral distribution
$4K_1 = \bar{K}_4$			$K_4 = W_3$		
co-diamond			diamond		
co-paw			paw=3-pan		
$2K_2 = \bar{C}_4$			$C_4 = K_{2,2}$		
claw = $K_{1,3}$			co-claw		
$P_4$					

**Table 4**

The moment matrix distances  $\|\mathcal{M}_2(G) - \mathcal{M}_2(\tilde{G})\|_2$  between two graphs  $G$  and  $\tilde{G}$  among them are shown.

	$4K_1$	$K_4$	co-diamond	diamond	co-paw	paw	$2K_2$	$C_4$	claw	co-claw	$P_4$
$4K_1$	0	90.9945	1.4142	49.8999	5.0744	26.2726	2.8284	20.9762	12.3693	15.7321	9.8742
$K_4$	90.9945	0	90.0777	41.4970	86.6646	65.3854	89.1740	70.8802	79.4292	75.8650	82.0945
co-diamond	1.4142	90.0777	0	48.9081	3.7749	25.1942	1.4142	19.8494	11.1803	14.6116	8.6313
diamond	49.8999	41.4970	48.9081	0	45.3900	23.9322	47.9375	29.4279	38.0657	34.4891	40.7247
co-paw	5.0744	86.6646	3.7749	45.3900	0	21.5754	2.5981	16.1787	7.4666	10.9659	4.8734
paw	26.2726	65.3854	25.1942	23.9322	21.5754	0	24.1506	5.5000	14.1863	10.6184	16.8300
$2K_2$	2.8284	89.1740	1.4142	47.9375	2.5981	24.1506	0	18.7617	10.0499	13.5462	7.4498
$C_4$	20.9762	70.8802	19.8494	29.4279	16.1787	5.5000	18.7617	0	8.7750	5.2440	11.3798
claw	12.3693	79.4292	11.1803	38.0657	7.4666	14.1863	10.0499	8.7750	0	3.6742	2.7386
co-claw	15.7321	75.8650	14.6116	34.4891	10.9659	10.6184	13.5462	5.2440	3.6742	0	6.2450
$P_4$	9.8742	82.0945	8.6313	40.7247	4.8734	16.8300	7.4498	11.3798	2.7386	6.2450	0

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.

### 5.1. 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 vertices 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 [17,20]. Thus the time complexity and space complexity of proposed method are both  $\mathcal{O}(|E|)$ .

### 5.2. 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.

## 6. Experiments

### 6.1. Clustering random graphs

We first demonstrate the efficacy of our method and other methods utilizing moment via clustering random graphs. Specifically, we generate four sets of Erdős-Rényi random graphs [15]. The parameters are  $\Theta_1 = \{|V| = 1000, |E| = 10000, \rho = 0.1\}$ ,  $\Theta_2 = \{|V| = 1000, |E| = 20000, \rho = 0.1\}$ ,  $\Theta_3 = \{|V| = 1000, |E| = 10,000, \rho = 0.9\}$ ,  $\Theta_4 = \{|V| = 1000, |E| = 20,000, \rho = 0.9\}$ , in which  $|V|$  denotes the number of vertices,  $|E|$  denotes the number of edges,  $\rho$  denotes the rewiring probability, i.e. randomness. For example, when  $\rho = 0$  (resp.  $\rho = 1$ ), the graph is a regular graph (resp. a completely random graph). For each parameter setting, we generate 25 random graphs and label the graphs according to their parameter settings. The benchmark algorithms are as following:

- **Cov [33]**: Covariance method computes the covariance matrix of the vector  $\left[\frac{A^i e}{\|A^i e\|_{\ell_1}}\right]_{i=1}^n$ , in which  $A$  is the adjacency matrix and  $e$  is the vector of all ones. Then Bhattacharya similarity between the corresponding covariance matrices was considered as the distance between two graphs. According to [33], we take the size of moment matrix  $n = 4, 5, 6$  and choose the best one as the benchmark.
- **NCLM [27]**: NCLM first computes the log moment vector  $[\log(\text{tr}(A/n^i))]_{i=2}^7$ , where  $A$  is the adjacency matrix, and then uses the Euclidean distance between two log moment vectors as a distance between corresponding graphs.
- **GK [31]**: Graphlet kernel computes a distance between graphs by counting subgraphs with  $k$  vertices. For  $k = 3$  (resp.  $k = 4$ ) we denote GK3 (resp. GK4).
- **EIGS-10**: The eigenvalues of the adjacency matrix contains much information about the graph, and the spectrum is graph invariant. As a result, we take the biggest 10 eigenvalues for each graph. Then we employ the Euclidean distance between the eigenvalues of corresponding adjacency matrices as a distance between two graphs.

To compare Cov, NCLM, EIGS, GK3, GK4 with our proposed method, we first compute the distance matrix  $D$ , where  $D_{ij}$  is the distance between  $i$ th graph and  $j$ th graph. Then we construct the kernel matrix  $\mathcal{K} = \exp(-D)$ . Finally we apply kernel K-means algorithm to get the clustering result. The clustering performance is shown in Table 5.

**Table 5**

Accuracy for our proposed method and covariance method in random graph clustering.

	Proposed Method	Cov	NCLM	EIGS-10	GK3	GK4
ACCURACY	1	1	1	0.76	0.5	0.5

**Table 6**

Accuracy for Proposed method and other benchmark methods in collaboration network classification. Best results marked in bold.

	HEP Vs CM	HEP Vs ASTRO	ASTRO Vs CM	Full
Proposed	<b>0.991</b>	<b>0.913</b>	<b>0.904</b>	<b>0.905</b>
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

We see that our proposed method, Cov and NCLM perform best. The interesting point is that these methods all involves moment values. GK3 and GK4 have trouble in separating the graphs the same number of vertices and same number of edges which are generated by different randomness. It is shown that EIGS can not distinguish between the parameter setting  $\Theta_3$  and  $\Theta_4$ . This demonstrates that the methods based on moment values are able to capture the feature of edge distribution and randomness of the graph.

## 6.2. Classifying networks

We apply our method to classify networks. We adopt the setting in [33]. 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. Thus, social network classification can be used for recommendations such as job recommendations and citation recommendations.

Three datasets from Rossi and Ahmed [29] are considered: 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 Shrivastava and Li [33] 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 10 times to compute the average accuracy.

In the classification tasks, we use  $k$ -nearest-neighbor(KNN) classifier. We set the degree 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 Cov, NCLM, and EIGS-10. In addition, we add the state-of-art method in collaboration network classification, Cov with SVM [33], which employs SVM as the classifier, as the last benchmark algorithm. The performance of our method and the benchmark algorithm is shown in Table 6.

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 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 proposed method. This also shows that a few moments can provide enough information for collaboration classification. Besides, proposed method has a significant improvement over the state-of-art method in three collaboration network classification task. This shows proposed method is suitable to classification tasks for sophisticated networks.

## 6.3. Time comparison

In this section, we show the efficiency of our algorithm by comparing the running time of proposed method and other methods via a set of experiments. Specifically, in each experiment, we generate 100 Erdős-Rényi random graphs with the same number of vertices and edges. Then we employ the proposed method 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 vertices, number of edges, and the time consumed by different methods are shown in Table 7. Here, we use  $4 \times 4$  moment matrix in the proposed method,  $4 \times 4$  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 7**

Running time for computing pairwise distance among 100 random graphs (in seconds). Fastest method is marked in bold.

$ V $	$ E $	Proposed	Cov	EIGS-10	NCLM	GK3
2,000	2,000,000	<b>7.31</b>	7.32	18.92	39.92	99.63
5,000	1,000,000	<b>1.38</b>	1.48	85.88	533	797
10,000	2,000,000	<b>3.9</b>	4.7	353.5	27340	1757
50,000	15,000,000	<b>50</b>	68	11687	N/A	N/A

As shown in the table, the time cost of proposed method is cheaper than all the comparing methods. For example, it can compute pairwise distances of 100 random graphs with 50,000 vertices and 15,000,000 edges in 50 sec, which has 1.36 times speed up to Covariance method and 233 times speed up to EIGS-10. Besides, the computational cost of proposed method is almost linear in terms of the number of edges as shown in Table 7. This demonstrates our proposed method is scalable to large-scale graphs.

## 7. Discussion and conclusions

We considered the adjacency matrix of a graph as a real random variable and proposed a new similarity measure for graphs with a distance between corresponding moment matrices of their spectral distributions. Our proposed method demonstrated state-of-art results in collaboration network classification and turned out to be scalable to large-scale graphs. In the future, our main work is to study the proposed distance between different random graphs. Further work would also be to extend the proposed distance from undirected to directed graphs.

## Acknowledgement

The authors wish to express their gratitude to the anonymous referees for their careful reading of the manuscript and their helpful suggestions. This work of H. Choi and Y. Shi was partially supported by Shanghai Sailing Program (Grant No. 16YF1407700), and National Nature Science Foundation of China (NSFC) (Grant No. 61601290). The work of H. Lee was supported by the National Research Foundation of Korea grant funded by the Korea government (MIST) (Grant No. NRF-2018R1D1A1B07049948).

## References

- [1] A. Abdollahi, S. Janbaz, M.R. Oboudi, Distance between spectra of graphs, *Linear Algebra Appl.* 466 (2015) 401–408, doi:10.1016/j.laa.2014.10.020.
- [2] W. Ali, A.E. Wegner, R.E. Gaunt, C.M. Deane, G. Reinert, Comparison of large networks with sub-sampling strategies, *Scie. Rep.* 6 (2016) 38955.
- [3] S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, Determining structural similarity of chemicals using graph-theoretic indices, *Discrete Appl. Math.* 19 (1–3) (1988) 17–44. Applications of graphs in chemistry and physics. doi: 10.1016/0166-218X(88)90004-2.
- [4] C. Berg, R. Szwarc, A determinant characterization of moment sequences with finitely many mass points, *Linear Multilinear Algebra* 63 (8) (2015) 1568–1576, doi:10.1080/03081087.2014.954516.
- [5] S. Butler, A note about cospectral graphs for the adjacency and normalized Laplacian matrices, *Linear Multilinear Algebra* 58 (3–4) (2010) 387–390, doi:10.1080/03081080902722741.
- [6] A. Calderone, M. Formenti, F. Aprea, M. Papa, L. Alberghina, A.M. Colangelo, P. Bertolazzi, Comparing alzheimer's and parkinson's diseases networks using graph communities structure, *BMC Syst. Biol.* 10 (1) (2016) 25–34.
- [7] X. Chen, J. Qian, Bounds on the number of closed walks in a graph and its applications, *J. Inequal. Appl.* (2014) 2014:199,9, doi:10.1186/1029-242X-2014-199.
- [8] J.K. Chung, P. Kannappan, C.T. Ng, P.K. Sahoo, Measures of distance between probability distributions, *J. Math. Anal. Appl.* 138 (1) (1989) 280–292, doi:10.1016/0022-247X(89)90335-1.
- [9] K.C. Das, S. Sun, Distance between the normalized Laplacian spectra of two graphs, *Linear Algebra Appl.* 530 (2017) 305–321, doi:10.1016/j.laa.2017.05.025.
- [10] M. Dehmer, F. Emmert-Streib, Comparing large graphs efficiently by margins of feature vectors, *Appl. Math. Comput.* 188 (2) (2007) 1699–1710, doi:10.1016/j.amc.2006.11.185.
- [11] M. Dehmer, F. Emmert-Streib, J. Kilian, A similarity measure for graphs with low computational complexity, *Appl. Math. Comput.* 182 (1) (2006) 447–459, doi:10.1016/j.amc.2006.04.006.
- [12] M. Dehmer, F. Emmert-Streib, Y. Shi, Graph distance measures based on topological indices revisited, *Appl. Math. Comput.* 266 (2015) 623–633, doi:10.1016/j.amc.2015.05.072.
- [13] M. Dehmer, K. Varmuza, A comparative analysis of the Tanimoto index and graph edit distance for measuring the topological similarity of trees, *Appl. Math. Comput.* 259 (2015) 242–250, doi:10.1016/j.amc.2015.02.042.
- [14] F. Emmert-Streib, M. Dehmer, Y. Shi, Fifty years of graph matching, network alignment and network comparison, *Inf. Sci.* 346/347 (2016) 180–197, doi:10.1016/j.ins.2016.01.074.
- [15] P. ERDdS, A. R&Wl, On random graphs I, *Publ. Math. Debrecen* 6 (1959) 290–297.
- [16] M.A. Fiol, E. Garriga, Number of walks and degree powers in a graph, *Discrete Math.* 309 (8) (2009) 2613–2614, doi:10.1016/j.disc.2008.03.025.
- [17] J.B. French, Elementary principles of spectral distributions, in: B.J. Dalton, S.M. Grimes, J.P. Vary, S.A. Williams (Eds.), *Theory and Applications of Moment Methods in Many-Fermion Systems*, Springer US, Boston, MA, 1980, pp. 1–16, doi:10.1007/978-1-4613-3120-9\_1.
- [18] H. Fujii, A. Katsuda, Isospectral graphs and isoperimetric constants, *Discrete Math.* 207 (1–3) (1999) 33–52, doi:10.1016/S0012-365X(99)00133-8.
- [19] X. Gao, B. Xiao, D. Tao, X. Li, A survey of graph edit distance, *IEEE Pattern Anal. Applicat.* 13 (1) (2010) 113–129.
- [20] P. Gavrilidis, G. Athanassoulis, Moment data can be analytically completed, *Probabilistic Eng. Mech.* 18 (4) (2003) 329–338.
- [21] J. Gu, B. Hua, S. Liu, Spectral distances on graphs, *Discrete Appl. Math.* 190/191 (2015) 56–74, doi:10.1016/j.dam.2015.04.011.
- [22] W.H. Haemers, E. Spence, Enumeration of cospectral graphs, *European J. Combin.* 25 (2) (2004) 199–211, doi:10.1016/S0195-6698(03)00100-8.

- [23] S.-Y. Hsieh, C.-W. Huang, H.-H. Chou, A DNA-based graph encoding scheme with its applications to graph isomorphism problems, *Appl. Math. Comput.* 203 (2) (2008) 502–512, doi:[10.1016/j.amc.2008.04.041](https://doi.org/10.1016/j.amc.2008.04.041).
- [24] I. Jovanović, Z. Stanić, Spectral distances of graphs, *Linear Algebra Appl.* 436 (5) (2012) 1425–1435, doi:[10.1016/j.laa.2011.08.019](https://doi.org/10.1016/j.laa.2011.08.019).
- [25] S. Kullback, R.A. Leibler, On information and sufficiency, *Ann. Math. Stat.* 22 (1951) 79–86, doi:[10.1214/aoms/1177729694](https://doi.org/10.1214/aoms/1177729694).
- [26] H. Lin, D. Li, K.C. Das, Distance between distance spectra of graphs, *Linear Multilinear Algebra* 65 (12) (2017) 2538–2550, doi:[10.1080/03081087.2017.1278737](https://doi.org/10.1080/03081087.2017.1278737).
- [27] S.S. Mukherjee, P. Sarkar, L. Lin, On clustering network-valued data, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30*, Curran Associates, Inc., 2017, pp. 7071–7081.
- [28] N. Obata, Spectral analysis of growing graphs, *SpringerBriefs in Mathematical Physics*, 20, Springer, Singapore, 2017. A quantum probability point of view. doi: [10.1007/978-981-10-3506-7](https://doi.org/10.1007/978-981-10-3506-7).
- [29] R.A. Rossi, N.K. Ahmed, The network data repository with interactive graph analytics and visualization, in: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, in: AAAI'15, AAAI Press, 2015, pp. 4292–4293.
- [30] T.A. Schieber, L. Carpi, A. Díaz-Guilera, P.M. Pardalos, C. Masoller, M.G. Ravetti, Quantification of network structural dissimilarities, *Nature Commun.* 8 (2017) 13928.
- [31] N. Shervashidze, S. Vishwanathan, T. Petri, K. Mehlhorn, K. Borgwardt, Efficient graphlet kernels for large graph comparison, in: D. van Dyk, M. Welling (Eds.), *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, *Proceedings of Machine Learning Research*, 5, PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 2009, pp. 488–495.
- [32] Y. Shimada, Y. Hirata, T. Ikeguchi, K. Aihara, Graph distance for complex networks, *Scient. Rep.* 6 (2016) 34944.
- [33] A. Shrivastava, P. Li, A new space for comparing graphs, in: *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, in: ASONAM '14, IEEE Press, Piscataway, NJ, USA, 2014, pp. 62–71.
- [34] H. Tan, M. Peng, C.K. Tse, F. Wu, Global similarity tests of physical designs of circuits: a complex network approach, *Appl. Math. Comput.* 230 (2014) 96–103, doi:[10.1016/j.amc.2013.12.083](https://doi.org/10.1016/j.amc.2013.12.083).
- [35] H. Täubig, J. Weihmann, S. Kosub, R. Hemmecke, E.W. Mayr, Inequalities for the number of walks in graphs, *Algorithmica* 66 (4) (2013) 804–828, doi:[10.1007/s00453-013-9766-3](https://doi.org/10.1007/s00453-013-9766-3).
- [36] J. Ugander, L. Backstrom, J. Kleinberg, Subgraph frequencies: Mapping the empirical and extremal geography of large graph collections, in: *Proceedings of the International Conference World Wide Web*, May 2013, pp. 1307–1318.
- [37] E.R. van Dam, W.H. Haemers, J.H. Koolen, Cospectral graphs and the generalized adjacency matrix, *Linear Algebra Appl.* 423 (1) (2007) 33–41, doi:[10.1016/j.laa.2006.07.017](https://doi.org/10.1016/j.laa.2006.07.017).
- [38] R. Vemulapalli, D.W. Jacobs, Riemannian metric learning for symmetric positive definite matrices. CoRR, [abs/1501.02393](https://arxiv.org/abs/1501.02393), 2015. URL <http://arxiv.org/abs/1501.02393>.
- [39] N. Wicker, C.H. Nguyen, H. Mamitsuka, A new dissimilarity measure for comparing labeled graphs, *Linear Algebra Appl.* 438 (5) (2013) 2331–2338, doi:[10.1016/j.laa.2012.10.021](https://doi.org/10.1016/j.laa.2012.10.021).
- [40] R.C. Wilson, P. Zhu, A study of graph spectra for comparing graphs and trees, *Pattern Recogn.* 41 (9) (2008) 2833–2841, doi:[10.1016/j.patcog.2008.03.011](https://doi.org/10.1016/j.patcog.2008.03.011).
- [41] L. Zager, *Graph Similarity and Matching*, Massachusetts Institute of Technology, 2005 Ph.D. thesis.