Discrete Applied Mathematics 190–191 (2015) 56–74

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Spectral distances on graphs

Jiao Gu [a,b,c], Bobo Hua [c,d], Shiping Liu [c,e,*]

[a] School of Science, Jiangnan University, 214122, Wuxi, China
[b] Mathematics and Science college, Shanghai Normal University, 200234, Shanghai, China
[c] Max Planck Institute for Mathematics in the Sciences, 04103, Leipzig, Germany
[d] School of Mathematical Sciences, LMNS, Fudan University, Shanghai 200433, China
[e] Department of Mathematical Sciences, Durham University, DH1 3LE Durham, United Kingdom

ARTICLE INFO

Article history:
Received 28 October 2014
Received in revised form 20 March 2015
Accepted 16 April 2015
Available online 8 May 2015

Keywords:
Wasserstein distance
Spectral measure
Random rooted graph
Asymptotic behavior
Biological networks

ABSTRACT

By assigning a probability measure via the spectrum of the normalized Laplacian to each graph and using L^p Wasserstein distances between probability measures, we define the corresponding spectral distances d_p on the set of all graphs. This approach can even be extended to measuring the distances between infinite graphs. We prove that the diameter of the set of graphs, as a pseudo-metric space equipped with d_1, is one. We further study the behavior of d_1 when the size of graphs tends to infinity by interlacing inequalities aiming at exploring large real networks. A monotonic relation between d_1 and the evolutionary distance of biological networks is observed in simulations.

1. Introduction

One major interest in graph theory is to explore the differences of graphs in structure, that is, in the sense of graph isomorphism. In computational complexity theory, the subgraph isomorphism problem, like many combinational problems in graph theory, is NP hard. Therefore, a method that gives a quick and easy estimate of the difference between two graphs is desirable [34]. As we know, all the topological information of a graph can be found in its adjacency matrix. The spectral graph theory studies the relationship between the properties of graphs and the spectra of their representing matrices, such as adjacency matrices and Laplace matrices [14,18,17]. In particular, some important topological information of a graph can be extracted from its specific eigenvalue like the first or the largest one, see e.g. [18,17,39,11,25,12,10]. The approach of reading information from the entire spectrum of a graph was explored in [5–7,30,32] etc. In spite of the existence of co-spectral graphs (see [38, Chapter 3] for a general construction and the references therein), the spectra of graphs can support us one way on exploring problems that involve (sub-)graph isomorphism by the fast computation algorithms and the close relationship with the structure of graphs.

A spectral distance on the set of finite graphs of the same size, i.e. the same number of vertices, was suggested in a problem of Richard Brualdi in [37] to explore the so-called cospectrality of a graph. It was further studied in [26] using the spectra of adjacency matrices. Employing certain Gaussian measures associated to the spectra of normalized Laplacians and the corresponding L^1 distances, the first named author, Jost, the third named author and Stadler [21,20] explored a spectral distance well-defined on the set of all finite graphs without any constraint about sizes. In this paper, instead of

* Corresponding author at: Department of Mathematical Sciences, Durham University, DH1 3LE Durham, United Kingdom.
E-mail addresses: jiaogu@mis.mpg.de (J. Gu), bobohua@mis.mpg.de (B. Hua), shiping.liu@durham.ac.uk (S. Liu).

http://dx.doi.org/10.1016/j.dam.2015.04.011

the Gaussian measures, we assign Dirac measures to graphs through the spectra of normalized Laplacians and use the Wasserstein distances between probability measures to propose spectral distances between graphs. In fact, this notion of spectral distances provides a metrization of the notion of spectral classes of graphs introduced in [21] via the weak convergence of the corresponding Dirac measures. The spectral class can be considered as a weak notion of graph limits (see the concepts of graphon, graphing and related theories in the monograph of Lovász [33]). This notion of spectral distances is even adaptable for weighted infinite graphs. And we can prove diameter estimates with respect to these distances, which are sharp for certain cases.

A weighted graph G is a triple (V, E, θ) where V is the set of vertices, E is the set of edges and $\theta : E \to (0, \infty)$, $(x, y) \mapsto \theta_{xy}$, is the (symmetric) edge weight function. We write $x \sim y$ if $(x, y) \in E$. We assume that for any vertex x, the weighted degree defined by $\theta_x := \sum_{y\sim x} \theta_{xy}$ is finite and $\theta_{xx} = 0$ (i.e. there is no self-loops).

Let us first consider finite weighted graphs. The normalized Laplacian of $G = (V, E, \theta)$ is defined as, for any function $f : V \to \mathbb{R}$ and any $x \in V$,

$$\Delta_G f(x) = f(x) - \frac{1}{\theta_x} \sum_{y\sim x} f(y)\theta_{xy}. \tag{1}$$

This operator can be extended to an infinite weighted graph which has countable vertex set V but is not necessarily locally finite (see [27] or Section 2 below). As a matrix, Δ_G is unitarily equivalent to the Laplace matrix studied in [17].

If $x \in V$ is an isolated vertex, i.e. $\theta_x = 0$, (1) reads as $\Delta_G f(x) = f(x)$. This implies that an isolated vertex contributes an eigenvalue 1 to the spectrum of Δ_G, denoted by $\sigma(G)$. In this way, by the absence of the self-loops, the spectrum of any finite weighted graph $\sigma(G) = \{\lambda_i\}_{i=1}^N$, counting the multiplicity, satisfies the trace condition

$$\sum_{i=1}^{N} \lambda_i = N \tag{2}$$

where $N = |V|$. It is well-known that $\sigma(G)$ is contained in $[0, 2]$. We associate to $\sigma(G)$ a probability measure on $[0, 2]$ as follows:

$$\mu_{\sigma(G)} := \frac{1}{N} \sum_i \delta_{\lambda_i}, \tag{3}$$

where δ_{λ_i} is the Dirac measure concentrated on λ_i. We call $\mu_{\sigma(G)}$ the *spectral measure* for a finite weighted graph. (This is known as the empirical distribution of the eigenvalues in random matrix theory.) Denote by $P([0, 2])$ the set of probability measures on the interval $[0, 2]$. For any $\mu \in P([0, 2])$, the first moment of μ is defined as $m_1(\mu) := \int_{[0,2]} \lambda \, d\mu(\lambda)$. The trace condition (2) is then translated to

$$m_1(\mu_{\sigma(G)}) = 1. \tag{4}$$

This is a key property of the spectral measures for our further investigations.

Let d_p^W $(1 \le p < \infty)$ be the pth Wasserstein distance on $P([0, 2])$. That is, for any $\mu, \nu \in P([0, 2])$ (see e.g. [40]),

$$d_p^W(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{[0,2]\times[0,2]} d(x, y)^p d\pi(x, y) \right)^{1/p},$$

where $\Pi(\mu, \nu)$ denotes the collection of all measures on $[0, 2] \times [0, 2]$ with marginals μ and ν on the first and second factors respectively, i.e. $\pi \in \Pi(\mu, \nu)$ if and only if $\pi(A \times [0, 2]) = \mu(A)$ and $\pi([0, 2] \times B) = \nu(B)$ for all Borel subsets $A, B \subseteq [0, 2]$.

It is well-known that $(P([0, 2]), d_p^W)$ is a complete metric space for $p \in [1, \infty)$ which induces the weak topology of measures in $P([0, 2])$(see e.g. [40, Theorem 6.9]).

One can prove that $\operatorname{diam}(P([0, 2]), d_p^W) = 2$. Indeed, on one hand, for any $\mu, \nu \in P([0, 2])$ by the optimal transport interpretation of Wasserstein distance, $d_p^W(\mu, \nu) \le 2$. On the other hand, $d_p^W(\delta_0, \delta_2) = 2$. (Recall that δ_0, δ_2 are the Dirac measures concentrated on 0, 2, respectively.)

Definition 1.1. Given two finite weighted graphs $G = (V, E, \theta)$ and $G' = (V', E', \theta')$, the *spectral distance* between G and G' is defined as

$$d_p(G, G') := d_p^W(\mu_{\sigma(G)}, \mu_{\sigma(G')}). \tag{5}$$

We denote by $\mathcal{FG}$ the space of all finite weighted graphs. Then for any $1 \le p < \infty$, $(\mathcal{FG}, d_p)$ is a pseudo-metric space. This is not a metric space due to the existence of co-spectral graphs. However, in applications this spectral consideration leads to the simplification of measuring the discrepancy of graphs.

One of the main results of our paper is the following theorem.

Theorem 1.2. *For any $1 \leq p < \infty$, we have*

$$\operatorname{diam}(\mathcal{FG}, d_p) \leq 2^{1-\frac{1}{p}}.$$

Remark 1.3. (a) Embedded as a subspace of $P([0,2])$, $\mathcal{FG}$ is a proper subspace by considering the diameters.
(b) One can prove an upper bound directly by using Chebyshev inequality, see Theorem 2.4. Clearly, this theorem improves that estimate.
(c) This estimate is tight for $p = 1$, i.e. $\operatorname{diam}(\mathcal{FG}, d_1) = 1$, see Corollary 1.8.
(d) We do not claim the sharpness of upper bound estimates for $p \in (1, \infty)$.

In fact, Theorem 1.2 follows from the estimates on the Wasserstein distance of probability measures in condition of the first moments.

Theorem 1.4 (*Measure-theoretic Version*)**.** *For any $\mu, \nu \in P([0,2])$ with $m_1(\mu) = m_1(\nu) = 1$ and $p \in [1, \infty)$,*

$$d_p^W(\mu, \nu) \leq 2^{1-\frac{1}{p}}. \tag{6}$$

By Proposition 2.1 and Lemma 2.2 below, one easily shows that the above measure-theoretic estimate is equivalent to the following analytic estimate.

Theorem 1.5 (*Analytic Version*)**.** *Let $f, g : [0,1] \to [0,2]$ be two nondecreasing functions such that $\int_0^1 f(x)dx = \int_0^1 g(x)dx = 1$. Then for any $p \in [1, \infty)$*

$$\left(\int_0^1 |f - g|^p(x)dx\right)^{\frac{1}{p}} \leq 2^{1-\frac{1}{p}}. \tag{7}$$

Section 3 is devoted to the proofs of Theorems 1.2, 1.4 and 1.5.

We extend our approach of the spectral distance to infinite graphs (with countable vertex set V) in Section 4. Note that in the above arguments we only use the normalization of the first moment of the spectral measures, i.e. $m_1(\mu_{\sigma(G)}) = 1$, our results generalize to all weighted graphs including infinite ones. For spectral measures with distinguished vertices on infinite graphs, we refer to Mohar–Woess [36]. We introduce two definitions of spectral measures for infinite graphs. One is defined via the exhaustion of the infinite graphs by the spectral measures of normalized Dirichlet Laplacians on subgraphs. The other is defined for random rooted graphs following Benjamini–Schramm [13], Aldous–Lyons [2] and Abért–Thom–Virág [1].

We denote by $\mathcal{G}$ the collection of all (possibly infinite) weighted graphs. For any $G \in \mathcal{G}$, we define $\mathrm{SM}(G)$ as the spectral measures of G by exhaustion, see Definition 4.1, which is a closed subset of $P([0,2])$. Then $\mathcal{G}$ endowed with the Hausdorff distance induced from the metric space $(P([0,2]), d_p^W)$, denoted by $d_{p,H}$, is a pseudo-metric space. A direct application of Theorem 1.4 yields the following corollary (recalled below as Theorem 4.2).

Corollary 1.6. *For $p \in [1, \infty)$,*

$$\operatorname{diam}(\mathcal{G}, d_{p,H}) \leq 2^{1-\frac{1}{p}}.$$

For any $D \geq 1$, we denote by $\mathcal{RRG}_D$ the collection of random rooted graphs of degree D, see Section 4.2 for definitions. Any finite weighted graph G gives rise to a random graph by assigning the root of G uniformly randomly. There are many interesting class of random rooted graphs such as unimodular and sofic ones, see [1]. For each random rooted graph $G \in \mathcal{RRG}_D$, we associate it with an expected spectral measure, denoted by μ_G. In this way, $\mathcal{RRG}_D$ endowed with d_p^W Wasserstein distance for expected spectral measures (d_p in short) is a pseudo-metric space. By Theorem 1.4, one can prove the following corollary (recalled below as Theorem 4.4).

Corollary 1.7. *For $p \in [1, \infty)$,*

$$\operatorname{diam}(\mathcal{RRG}_D, d_p) \leq 2^{1-\frac{1}{p}}.$$

In fact, there are examples of finite graphs which saturate the upper bounds for $p = 1$, see Examples 2.5 and 2.6.

Corollary 1.8. *All upper bounds on d_1 are tight, i.e.*

$$\operatorname{diam}(\mathcal{FG}, d_1) = \operatorname{diam}(\mathcal{G}, d_{1,H}) = \operatorname{diam}(\mathcal{RRG}_D, d_1) = 1.$$

We then concentrate on the spectral distance d_1. In Section 5, we calculate d_1 on several particular classes of graphs. For our purpose of applications to large real networks, we are more concerned with the behavior of d_1 when the size of graphs N tends to infinity. We observe convergence behaviors of order $\mathcal{O}(1/N^2)$, $\mathcal{O}(1/N)$ in those examples.

The asymptotic behavior of d_1 is studied in general in Section 6 by employing interlacing inequalities of the spectra of finite weighted graphs. For two graphs G and G', which differ from each other by some standard operations including e.g. edge deleting, vertex replication, vertices contraction and edge contraction, we prove

$$d_1(G, G') \leq \frac{C}{N}, \tag{8}$$

where C depends only on the operations and is independent of the size N of G (see Theorem 6.3). By this result, we further derive a convergence result of graphs under the d_1 distance.

In the last section, we apply the distance d_1 to study the evolutionary process of biological networks by simulations. We start from a Barabási–Albert scale-free network, which has proven to be a very common type of real large networks [8]. We then simulate the evolutionary process by the operations, edge-rewiring and duplication–divergence respectively. We observe a monotonic relation between d_1 and the evolutionary distance, which is a crucial point to anticipate further applications in exploring evolutionary history of biological networks.

2. Preliminaries, spectral measures and spectral distances

In this section, we recall basics about graph spectra and Wasserstein distances on the space of probability measures, and define the spectral distances of finite graphs. The spectral distances of infinite graphs and random graphs will be postponed to Section 4.

Let us consider a (possibly infinite) weighted graph $G = (V, E, \theta)$, where V is a countable set and $\theta : E \to (0, \infty)$ is a weight function on edges. For convenience, we extend θ to $V \times V, \theta : V \times V \to (0, \infty)$, such that $\theta_{xy} > 0$ iff $(x, y) \in E$. We require that the weight function θ satisfies

$$\sum_{y \in V} \theta_{xy} < \infty, \quad \forall x \in V.$$

The weighted degree of the vertex $x \in V$ is still defined as $\theta_x := \sum_{y \sim x} \theta_{xy}$. The graph is called connected if for every two vertices $x, y \in V$ there exists a finite path $x = x_0 \sim x_1 \sim \cdots \sim x_n = y$ connecting x and y.

We define the *(formal) normalized Laplacian* Δ on the *formal domain*

$$F(V) := \left\{ f : V \to \mathbb{R} \mid \sum_{y \in V} \theta_{xy} |f(y)| < \infty \text{ for all } x \in V \right\},$$

by

$$\Delta f(x) = \frac{1}{\theta_x} \sum_{y \in V} \theta_{xy} (f(x) - f(y)).$$

As a linear operator, its restriction to the Hilbert space $\ell^2(V, \theta) := \{f : V \to \mathbb{R} | \sum_{x \in V} |f(x)|^2 \theta_x < \infty\}$, denoted by Δ_G, coincides with the generator of the Dirichlet form

$$Q(f) = \frac{1}{2} \sum_{x, y \in V} \theta_{xy} |f(x) - f(y)|^2,$$

defined on $\ell^2(V, \theta)$, for details see [27].

If $G = (V, E, \theta)$ is a weighted graph without isolated vertices, i.e. $\theta_x > 0$ for all $x \in V$, then the normalized Laplacian of G can be rephrased as

$$\Delta_G := I - D^{-1} A,$$

where D is the degree operator and A is the adjacency operator (defined as $D\tau_x = \theta_x \tau_x$ and $A\tau_x = \sum_{y \sim x} \theta_{yx} \tau_y$, where $\tau_x(y) = 1$ if $y = x$ and 0 otherwise), i.e. for any finitely supported function $f : V \to \mathbb{R}$,

$$\Delta_G f(x) = f(x) - \frac{1}{\theta_x} \sum_{y \sim x} f(y) \theta_{xy}.$$

Since $D^{-1}A$ is a bounded selfadjoint operator with operator norm less than or equal to 1 on $\ell^2(V, \theta)$, the spectrum of Δ_G, denoted by $\sigma(G)$, is contained in the interval $[0, 2]$.

We order the spectrum of any finite weighted graph G in the nondecreasing way:

$$0 \leq \lambda_1 \leq \cdots \leq \lambda_N \leq 2,$$

where $N = |V|$. For convenience, we also denote the spectrum of G by a vector, called spectral vector of G, $\lambda_G := (\lambda_i)_{i=1}^N = (\lambda_1, \lambda_2, \ldots, \lambda_N) \in [0, 2]^N$.

2.1. Spectral measures

Let G be a finite weighted graph. We denote by F_G the cumulative distribution function associated to $\mu_{\sigma(G)}$ (recall (3)), and by

$$F_G^{-1}(x) := \inf\{t \in \mathbb{R} : F_G(t) > x\}$$

the inverse cumulative distribution function. Since $\sigma(G) \subseteq [0, 2]$, we have $F_G : [0, 2] \to [0, 1]$ and $F_G^{-1} : [0, 1] \to [0, 2]$. Recalling the trace condition (2), we have the following proposition.

Proposition 2.1. *Let $G = (V, E, \theta)$ be a finite weighted graph. Then the following are true:*

(a) *F_G and F_G^{-1} are nonnegative nondecreasing step functions;*
(b) $\int_0^2 F_G(x)dx = 1$;
(c) $\int_0^1 F_G^{-1}(x)dx = 1$.

Proof. (a) is trivial. (c) follows from the trace condition (2). (b) is equivalent to (c) since the total area of the rectangle $[0, 1] \times [0, 2]$ is 2. □

2.2. Spectral distances

Since the spectrum of the normalized Laplacian of a graph lies in the interval $[0, 2] \subset \mathbb{R}$, one may calculate the spectral distance (5) explicitly. This is an advantage of probability measures supported in the 1-dimensional space. In fact, the spectral distance between two finite weighted graphs G, G', i.e. the Wasserstein distance of two spectral measures $\mu_{\sigma(G)}$, $\mu_{\sigma(G')}$, can be calculated by the inverse cumulative distribution functions F_G^{-1} and $F_{G'}^{-1}$ thanks to the following lemma.

Lemma 2.2 (*See [35, Theorem 8.1]*)**.** *Let $\mu, \nu \in P([0, 2])$ and F_μ^{-1}, F_ν^{-1} be their inverse cumulative distribution functions. Then for any $p \in [1, \infty)$,*

$$d_p^W(\mu, \nu) = \left(\int_0^1 |F_\mu^{-1}(x) - F_\nu^{-1}(x)|^p dx\right)^{1/p}.$$

One can show that if two graphs having the same number of vertices, say N, then the spectral distance between them is reduced to the ℓ^p distance between the spectral vectors, i.e. for any $1 \le p < \infty$,

$$d_p(G, G') = \frac{1}{N}\|\lambda_G - \lambda_{G'}\|_{\ell^p}.$$

In this paper, we are interested in the diameter of the pseudo-metric space $(\mathcal{FG}, d_p)$ for $p \in [1, \infty)$. Recall that we naturally have

$$\operatorname{diam}(\mathcal{FG}, d_p) \le \operatorname{diam}(P([0, 2]), d_p^W) = 2.$$

We denote by $\{\cdot\}$ a graph consisting of a single vertex with no edge. Then by our convention, $\sigma(\{\cdot\}) = \{1\}$. Clearly, for any weighted graph G,

$$d_p(G, \{\cdot\}) \le 1, \quad \text{for } 1 \le p < \infty.$$

In the following, we use (integral) Chebyshev inequality to derive a refined upper bound for the diameter.

Lemma 2.3 (*Chebyshev Inequality, See [22, Section 2.17] or [19]*)**.** *For any nonnegative, monotonically increasing integrable functions $f, g : [0, 1] \to [0, \infty)$, we have*

$$\int_0^1 f(x)g(x)dx \ge \int_0^1 f(x)dx \int_0^1 g(x)dx. \tag{9}$$

Theorem 2.4. *For any $1 \le p \le 2$, we have*

$$\operatorname{diam}(\mathcal{FG}, d_p) \le \sqrt{2},$$

i.e. for any finite weighted graphs G and G',

$$d_p(G, G') \le \sqrt{2}.$$

Proof. Let us denote $f = F_G^{-1}$ and $g = F_{G'}^{-1}$. Then by Chebyshev inequality (9) and Proposition 2.1(c),

$$\int_0^1 fg \geq \int_0^1 f \int_0^1 g = 1.$$

Hence, for any $1 \leq p \leq 2$ we have

$$\begin{aligned}\left(\int_0^1 |f-g|^p\right)^{2/p} &\leq \int_0^1 |f-g|^2 = \int_0^1 f^2 + \int_0^1 g^2 - 2\int_0^1 fg \\ &\leq 2\int_0^1 f + 2\int_0^1 g - 2 = 2,\end{aligned}$$

where we have used that $f \leq 2$ and $g \leq 2$. This proves the theorem. □

In the next section, we will give a tighter upper bound for the diameter estimates. In particular, in the case of $p = 1$, we derive an optimal upper bound, that is, we will prove that $\operatorname{diam}(\mathcal{F}\mathcal{G}, d_1) = 1$. The tightness of this estimate can be seen from the following two examples.

Example 2.5. Let $G = \{\cdot\}$ and $G' = P_2$ be the path on two vertices. Then $\sigma(G') = \{0, 2\}$. Hence we have

$$d_p(G, G') = 1, \quad p \in [1, \infty).$$

The following example is more convincing.

Example 2.6. Let $G' = P_2$ be the path on two vertices and G_N an unweighted (i.e. $\theta_{xy} = 1$ for every edge xy) complete graph on N vertices. Then it is known that

$$\sigma(G_N) = \left\{0, \underbrace{\frac{N}{N-1}, \ldots, \frac{N}{N-1}}_{N-1}\right\}. \tag{10}$$

Therefore we have

$$d_p(G_N, G') = \left[\left(\frac{1}{2} - \frac{1}{N}\right)\frac{N^p}{(N-1)^p} + \frac{1}{2}\left(2 - \frac{N}{N-1}\right)^p\right]^{\frac{1}{p}}.$$

In particular, $d_1(G_N, G') = 1 - \frac{1}{N-1}$. Observe that

$$\lim_{N\to+\infty} d_p(G_N, G') = 1.$$

3. The proof of the diameter estimate

This section is devoted to the proofs of Theorems 1.2, 1.4 and 1.5. We first prove some lemmata.
We call a function $f : [0, 1] \to [0, 2]$ an *admissible 2-step function* if there exist $a \in [0, \frac{1}{2}]$ and $b \in [\frac{1}{2}, 1]$ such that

$$f(x) = \begin{cases} 0, & 0 \leq x < a, \\ \dfrac{2b-1}{b-a}, & a \leq x < b, \\ 2, & b \leq x \leq 1. \end{cases} \tag{11}$$

In particular, we say f jumps at a and b. Clearly, $\int_0^1 f(x)dx = 1$ and $\int_0^2 f^{-1}(x)dx = 1$. The name for a 2-step function is evident from the graph of the function. In particular, any inverse function F_G^{-1} of a cumulative distribution function of a graph G with 3 vertices is an admissible 2-step function.

Lemma 3.1. *Let f, g be admissible 2 -step functions on* $[0, 1]$. *Then we have*

$$\int_0^1 |f-g|(x)dx \leq 1, \tag{12}$$

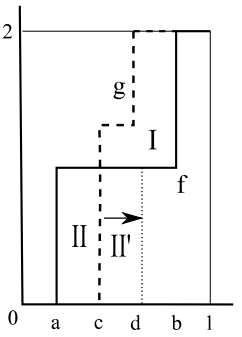

Fig. 1. Subcase 1.1.

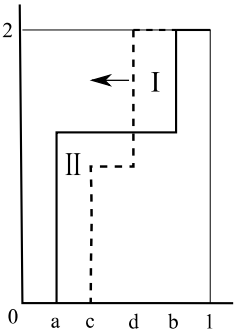

Fig. 2. Subcase 1.2.

where "=" holds if and only if (ignoring the order of f, g)

$$f(x) = \begin{cases} 0, & 0 \le x < \dfrac{1}{2}; \\ 2, & \dfrac{1}{2} \le x \le 1, \end{cases} \quad g(x) = 1,\ 0 \le x \le 1. \tag{13}$$

Observe that the inverse cumulative distribution functions in Example 2.5 are exactly the two functions in (13).

Proof. Let $f : [0, 1] \to [0, 2]$ ($g : [0, 1] \to [0, 2]$ resp.) be an admissible 2-step function jumping at a and b (c and d resp.). Denote the height of the first jump of f and g by $h_1 := \frac{2b-1}{b-a}$ and $h_2 := \frac{2d-1}{d-c}$ respectively.

The proof is divided into four cases and several subcases as follows:

***Case* 1.** $0 \le a \le c \le \frac{1}{2} \le d \le b \le 2$.

Subcase 1.1. $h_2 \ge h_1$. See Fig. 1.

For each domain I (II resp.) in Fig. 1, we denote by |I| (|II| resp.) the area of that domain. We reflect the domain II along the line $\{x = c\}$ to obtain a new domain II′. By the fact that $c \le \frac{1}{2}$, we have

$$\int_0^1 |f - g| = |\mathrm{I}| + |\mathrm{II}| = |\mathrm{I}| + |\mathrm{II}'| \le \int_0^1 g = 1.$$

Subcase 1.2. $h_2 < h_1$. See Fig. 2.

Reflect the domain I along the line $\{x = d\}$ to obtain I′. Then

$$\int_0^1 |f - g| = |\mathrm{I}| + |\mathrm{II}| = |\mathrm{I}'| + |\mathrm{II}| \le \int_0^2 g^{-1}(y)dy = 1.$$

***Case* 2.** $0 \le a \le c \le \frac{1}{2} \le b < d \le 2$.

We claim that $h_1 \le h_2$. Suppose not, by Fig. 3, we have

$$1 = \int_0^1 f > \int_0^1 g = 1,$$

which is a contradiction. This proves the claim.

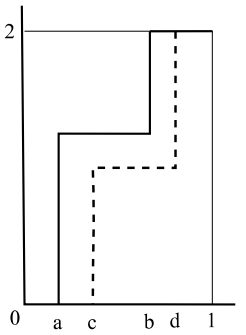

Fig. 3. Case 2: the proof of $h_1 \le h_2$.

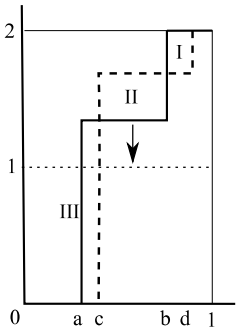

Fig. 4. Subcase 2.1.

Subcase 2.1. $h_1 \ge 1$, see Fig. 4.
Reflect the domain II along the line $\{y = h_1\}$ to get II′. Since $h_1 \ge 1$,

$$\int_0^1 |f - g| = |\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}| = |\mathrm{I}| + |\mathrm{II}'| + |\mathrm{III}| \le \int_0^1 f = 1.$$

Subcase 2.2. $h_1 < 1$. Further, we divide it into more subcases.

Subcase 2.2.1. $h_2 \le 1$, see Fig. 5.
Reflect the domain II along the line $\{y = h_2\}$ to have II′. By $h_2 \le 1$,

$$\int_0^1 |f - g| = |\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}| = |\mathrm{I}| + |\mathrm{II}'| + |\mathrm{III}| \le \int_0^2 g^{-1}(y)dy = 1.$$

Subcase 2.2.2. $h_2 > 1$. Moreover,

Subcase 2.2.2.1. $h_2 - h_1 \le 1$.
Then by the basic estimate,

$$\begin{aligned}\int_0^1 |f - g| &= |\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}| = (2 - h_2)(d - b) + (h_2 - h_1)(b - c) + h_1(c - a)\\ &\le d - b + b - c + c - a = d - a \quad (\text{by } \max\{2 - h_2, h_2 - h_1, h_1\} \le 1)\\ &\le 1.\end{aligned}$$

Subcase 2.2.2.2. $h_2 - h_1 > 1$, see Fig. 6.
Reflect I along the line $\{y = h_2\}$ to obtain I′, and III along the line $\{x = c\}$ to obtain III′. Then by the fact $h_2 - h_1 \ge 1 \ge 2 - h_2$, $\mathrm{I}' \cap \mathrm{III}' = \emptyset$. Thus,

$$\int_0^1 |f - g| = |\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}| = |\mathrm{I}'| + |\mathrm{II}| + |\mathrm{III}'| \le \int_0^1 g = 1.$$

***Case* 3.** $0 \le c < a \le \frac{1}{2} \le b < d \le 2$. By interchanging the role of a, b and c, d, this reduces to the Case 1.

***Case* 4.** $0 \le c < a \le \frac{1}{2} \le d \le b \le 2$. This reduces to Case 2 by the same change as in Case 3.

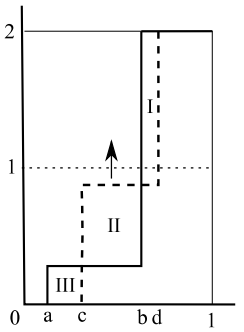

Fig. 5. Subcase 2.2.1.

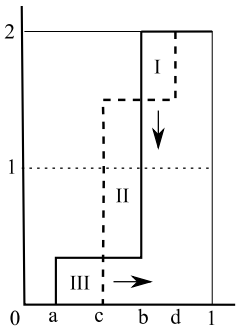

Fig. 6. Subcase 2.2.2.

Combining all the cases and subcases, we prove (12). Finally, we can check case by case that the equality in (12) can be achieved only when f and g are the functions given by the relation (13). This completes the proof. □

Before proving the next lemma, we recall some basic facts from the convex analysis. Let Ω be a convex subset of $\mathbb{R}^N$, possibly having lower Hausdorff dimension. A function $f : \Omega \to \mathbb{R}$ is called *convex* if for any $x, y \in \Omega$ and $0 \le t \le 1$,

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y).$$

In particular, for any norm $\|\cdot\|$ on $\mathbb{R}^N$, the function $f : \mathbb{R}^N \to \mathbb{R}$ defined by $f(x) = \|x - x_0\|$ for some fixed x_0 is a convex function. We say a point $x \in \Omega$ is *extremal* if it cannot be written as the nontrivial convex combination of two other points in Ω, i.e. if $x = tx_1 + (1-t)x_2$ for some $0 < t < 1$ and $x_1, x_2 \in \Omega$, then $x = x_1 = x_2$. The set of extremal points of a convex set Ω is denoted by $\mathrm{Ext}(\Omega)$. A subset $P \subset \mathbb{R}^N$ is called a (closed) convex polytope if it is the intersection of finite many half spaces, i.e. there exist $K \in \mathbb{N}$ linear functions $\{L_j\}_{j=1}^K$ on $\mathbb{R}^N$ such that

$$P = \bigcap_{j=1}^{K} \{x \in \mathbb{R}^N : L_j(x) \le 0\}.$$

We state a well-known fact which will be used to prove the next lemma.

Fact 3.2. *Let P be a compact convex polytope in $\mathbb{R}^N$ and $f : P \to \mathbb{R}$ a convex function. Then*

$$\max_P f = \max_{\mathrm{Ext}(P)} f. \tag{14}$$

The following lemma is the special case of Theorem 1.2 when two graphs have the same number of vertices.

Lemma 3.3. *Let $N \ge 1$. Assume that $\alpha = (\alpha_i)_{i=1}^N$ and $\beta = (\beta_i)_{i=1}^N$ satisfy $0 \le \alpha_1 \le \cdots \le \alpha_N \le 2$ and $0 \le \beta_1 \le \cdots \le \beta_N \le 2$ and*

$$\|\alpha\|_{\ell^1} = \|\beta\|_{\ell^1} = N.$$

Then we have

$$\|\alpha - \beta\|_{\ell^1} \le N.$$

Proof. Let P denote the compact convex polytope $\{\alpha \in \mathbb{R}^N : 0 \le \alpha_1 \le \cdots \le \alpha_N \le 2, \|\alpha\|_{\ell^1} = N\}$. Then by the induction on N, one can show that the set of extremal points of P is

$$\mathrm{Ext}(P) = \left\{ (\underbrace{0, \ldots, 0}_{k}, \underbrace{a, \ldots, a}_{N-k-l}, \underbrace{2, \ldots, 2}_{l}) : 0 \le k, l \le \frac{N}{2}, a = \frac{N-2l}{N-k-l} \right\}.$$

We divide the interval $[0, 1]$ equally into N subintervals $\{[\frac{i-1}{N}, \frac{i}{N}]\}_{i=1}^N$. Then for any $\alpha \in P$, we define a step function $f_\alpha : [0, 1] \to [0, 2]$ by

$$f_\alpha|_{[\frac{i-1}{N}, \frac{i}{N}]} = \alpha_i.$$

Clearly, $\int_0^1 f_\alpha = \frac{1}{N}\|\alpha\|_{\ell^1} = 1$. In addition, for any $\gamma \in \mathrm{Ext}(P)$, f_γ is an admissible 2-step function defined in (11).

Note that for any fixed $\beta_0 \in \mathbb{R}^N$, the function $F : \mathbb{R}^N \ni \alpha \mapsto \|\alpha - \beta_0\|_{\ell^1} \in \mathbb{R}$ is a convex function on $\mathbb{R}^N$. We claim that

$$\max_{\substack{\alpha \in P \\ \beta \in P}} \|\alpha - \beta\|_{\ell^1} = \max_{\substack{\gamma \in \mathrm{Ext}(P) \\ \theta \in \mathrm{Ext}(P)}} \|\gamma - \theta\|_{\ell^1}. \tag{15}$$

By Fact 3.2,

$$\begin{aligned} \max_{\substack{\alpha \in P \\ \beta \in P}} \|\alpha - \beta\|_{\ell^1} &= \max_{\beta \in P} \max_{\alpha \in P} \|\alpha - \beta\|_{\ell^1} = \max_{\beta \in P} \max_{\gamma \in \mathrm{Ext}(P)} \|\gamma - \beta\|_{\ell^1} \\ &= \max_{\gamma \in \mathrm{Ext}(P)} \max_{\beta \in P} \|\gamma - \beta\|_{\ell^1} = \max_{\gamma \in \mathrm{Ext}(P)} \max_{\theta \in \mathrm{Ext}(P)} \|\gamma - \theta\|_{\ell^1}. \end{aligned}$$

This proves the claim.

For any $\gamma, \theta \in \mathrm{Ext}(P)$, noting that f_γ and f_θ are admissible 2-step functions, by Lemma 3.1, we have

$$\|\gamma - \theta\|_{\ell^1} = N \int_0^1 |f_\gamma - f_\theta| \le N.$$

Combining this with (15), we prove the lemma. □

Now we can prove Theorem 1.5. A function $f : [0, 1] \to [0, 2]$ is called *a rationally distributed step function* if there is a (rational) partition $0 = r_0 < r_1 < r_2 < \cdots < r_N = 1$ with $r_i \in \mathbb{Q}$ for all $0 \le i \le N$ and an increasing sequence $0 \le a_1 < \cdots < a_N \le 2$ such that

$$f(x) = \begin{cases} a_1, & 0 \le x < r_1, \\ a_2, & r_1 \le x < r_2, \\ \vdots & \\ a_N, & r_{N-1} \le x \le 1. \end{cases}$$

Proof of Theorem 1.5. First, we consider $p = 1$. By the standard approximation argument, any such functions, f and g, can be approximated in L^1 norm by a sequence of rationally distributed step functions, say $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$, satisfying $\int_0^1 f_n = \int_0^1 g_n = 1$. Hence it suffices to prove the theorem for rationally distributed step functions.

W.l.o.g., we may assume f and g are rationally distributed step functions, say $f|_{[r_{i-1}, r_i]} = a_i$ for $1 \le i \le L$ and $g|_{[t_{j-1}, t_j]} = b_j$ for $1 \le j \le K$ where $L, K \in \mathbb{N}$. Let N denote the least common multiple of $\{m_i\}_{i=1}^L \cup \{n_j\}_{j=1}^K$ where m_i, n_j are the denominators of $r_i = \frac{c_i}{m_i}$ and $t_j = \frac{d_j}{n_j}$ $(c_i, m_i, d_j, n_j \in \mathbb{N})$, $1 \le i \le L$, $1 \le j \le K$. Then we have for any $1 \le p \le N$

$$\begin{aligned} f|_{[\frac{p-1}{N}, \frac{p}{N}]} &= \alpha_p, \\ g|_{[\frac{p-1}{N}, \frac{p}{N}]} &= \beta_p, \end{aligned}$$

where $\alpha_p = a_l$ and $\beta_p = b_k$ for some $1 \le l \le L$, $1 \le k \le K$. Obviously, $0 \le \alpha_1 \le \cdots \le \alpha_N \le 2$, $0 \le \beta_1 \le \cdots \le \beta_N \le 2$ and

$$\|\alpha\|_{\ell^1} = \|\beta\|_{\ell^1} = N.$$

Hence Lemma 3.3 implies that

$$\|\alpha - \beta\|_{\ell^1} \le N.$$

That is,

$$\int_0^1 |f - g| \le 1.$$

For $p \in (1, \infty)$, it can be easily derived from the result for $p = 1$.

$$\int_0^1 |f - g|^p \le 2^{p-1} \int_0^1 |f - g| \le 2^{p-1}.$$

This proves the theorem. □

Theorem 1.4 then follows directly.

Proof of Theorem 1.4. Let F_μ and F_ν denote the cumulative distribution functions of the measures μ and ν respectively. Since the total area of the square $[0, 1] \times [0, 2]$ is equal to 2, by the assumption $m_1(\mu) = m_1(\nu) = 1$ we have

$$\int_0^1 F_\mu^{-1}(x)dx = \int_0^1 F_\nu^{-1}(x)dx = 1.$$

Then our theorem follows from Theorem 1.5 and Lemma 2.2. □

Now we can prove Theorem 1.2.

Proof of Theorem 1.2. This follows from Theorem 1.4 directly. □

4. Spectral distances of infinite graphs

In this section, we introduce two definitions of spectral measures for infinite weighted graphs with countable vertex set and extend our approach of spectral distance to this setting.

4.1. Spectral measures by exhaustion

Let $G = (V, E, \theta)$ be an infinite weighted graph and $G_\Omega := (\Omega, E_{|\Omega}, \theta_{|\Omega\times\Omega})$ a finite connected subgraph of G induced by a subset $\Omega \subset V$. We introduce the Dirichlet boundary problem of the normalized Laplacian on Ω, see e.g. [10]. Let $\ell^2(\Omega, \theta)$ denote the space of real-valued functions on Ω. Note that every function $f \in \ell^2(\Omega, \theta)$ can be extended to a function $\tilde{f} \in \ell^2(V, \theta)$ by setting $\tilde{f}(x) = 0$ for all $x \in V \setminus \Omega$. The normalized Laplacian with the Dirichlet boundary condition on Ω, denoted by Δ_{G_Ω}, is defined as $\Delta_{G_\Omega} : \ell^2(\Omega, \theta) \to \ell^2(\Omega, \theta)$,

$$\Delta_{G_\Omega} f = (\Delta_G \tilde{f})_{|\Omega}.$$

Thus for $x \in \Omega$ the Dirichlet normalized Laplacian is pointwise defined by

$$\Delta_{G_\Omega} f(x) = f(x) - \frac{1}{\theta_x} \sum_{y\in\Omega} \theta_{xy} f(y) = \tilde{f}(x) - \frac{1}{\theta_x} \sum_{y\in V} \theta_{xy} \tilde{f}(y),$$

where $\theta(x)$ is the weighted degree of the entire graph. A simple calculation shows that Δ_{G_Ω} is a positive self-adjoint operator. We arrange the eigenvalues of the Dirichlet Laplace operator Δ_{G_Ω} in nondecreasing order, i.e. $\lambda_1(\Omega) \le \lambda_2(\Omega) \le \cdots \le \lambda_N(\Omega)$, where N is the cardinality of the set Ω, i.e. $N = |\Omega|$. By the trace condition, we also have the key property

$$\sum_{i=1}^{N} \lambda_i(\Omega) = N.$$

As same as finite graphs, we associate it with the spectral measure,

$$\mu_\Omega = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(\Omega)}.$$

Hence $m_1(\mu_\Omega) = 1$.

A sequence of finite connected subgraphs $\{\Omega_n\}_{n=1}^\infty$ is called an *exhaustion* of the infinite graph G if $\Omega_n \subset \Omega_{n+1}$ for all $n \in \mathbb{N}$ and $\cup_{n=1}^\infty \Omega_n = V$. Hence we have a sequence of probability measures $\{\mu_{\Omega_n}\}_{n=1}^\infty$ on $[0, 2]$. Since $P([0, 2])$ is compact under the weak topology, up to a subsequence, w.l.o.g. we have $\mu_{\Omega_n} \rightharpoonup \mu$ for some $\mu \in P([0, 2])$. Note that any subsequence of an exhaustion is still an exhaustion. Therefore we define the spectral measures of an infinite graph by all possible exhaustions. Note that the convergence of the spectral structure was studied in more general setting by Kuwae–Shioya [29].

Definition 4.1. Let G be an infinite weighted graph. We define the *spectral measures* of G by exhaustion as

$$\mathrm{SM}(G) := \{\mu \in P([0, 2]) : \text{there is an exhaustion } \{\Omega_n\}_{n=1}^\infty \, s.t. \, \mu_{\Omega_n} \rightharpoonup \mu\}.$$

One can show that $\mathrm{SM}(G)$ is a closed subset of $P([0,2])$. Since $m_1(\mu_{\Omega_n}) = 1$ for any $n \in \mathbb{N}$, by the weak convergence, we have $m_1(\mu) = 1$ for any $\mu \in \mathrm{SM}(G)$.

For any metric space (X, d), one can define the *Hausdorff distance* between the subsets of X. For any subset $A \subset X$, we define the distance function to the subset A as $X \ni x \mapsto d(x, A) = \inf\{d(x, y)|y \in A\}$, and the r-neighborhood of A as $U_r(A) := \{y \in X|d(y, A) < r\}$, $r > 0$. Given two subsets $A, B \subset X$, the Hausdorff distance between them is defined as

$$d_H(A, B) := \inf\{r > 0|A \subset U_r(B), B \subset U_r(A)\}.$$

One can show that the set of closed subsets of X endowed with the Hausdorff distance is a metric space.

Note that for $p \in [1, \infty)$, $P([0, 2])$ endowed with the pth Wasserstein distance is a metric space and $\mathrm{SM}(G)$ is a closed subset of $P([0, 2])$ for any weighted graph G. We denote by $\mathcal{G}$ the collection of all (possibly infinite) weighted graphs. Hence $\mathcal{G}$ endowed with the Hausdorff distance induced from $(P([0, 2]), d_p^W)$, denoted by $d_{p,H}$, is a pseudo-metric space.

A direct application of Theorem 1.4 yields

Theorem 4.2. *For $p \in [1, \infty)$,*

$$\mathrm{diam}(\mathcal{G}, d_{p,H}) \le 2^{1-\frac{1}{p}}.$$

4.2. Spectral measures for random rooted graphs

We follow Benjamini–Schramm [13], Aldous–Lyons [2] and Abért–Thom–Virág [1] to define random rooted graphs.

For any $D \ge 1$, we define a subcollection of $\mathcal{G}$, $\mathcal{G}_D := \{(V, E, \theta) \in \mathcal{G}|\deg_x \le D, \theta_{xy} \le D \text{ for all } x, y \in V\}$ where $\deg_x = |\{y \in V|y \sim x\}|$, i.e. the set of weighted graphs with bounded (unweighted) degree ($\le D$) and bounded edge weights ($\le D$). Let $\mathcal{RG}_D$ denote the set of graphs G in $\mathcal{G}_D$ with a distinguished vertex, called the root of G.

For any $x, y \in V$ of $G = (V, E, \theta)$, we denote by $d_G(x, y)$ the distance between x and y, i.e. $d_G(x, y) := \inf\{n|$ there exist $\{x_i\}_{i=0}^n$ $s.t.$ $x = x_0 \sim x_1 \sim \cdots \sim x_n = y\}$, and by $B_k(x) := \{z \in V|d_G(x, z) \le k\}$, $k \in \mathbb{N} \cup \{0\}$, the ball of radius k centered at x. Let (G_1, o_1) and (G_2, o_2) be two rooted graphs with distinguished vertices o_1 and o_2, respectively. We call that $B_k(o_1)$ is isomorphic to $B_k(o_2)$ if there exists a bijective map $f : B_k(o_1) \to B_k(o_2)$ such that $f(o_1) = f(o_2)$ and $x \sim y$ for $x, y \in B_k(o_1)$ if and only if $f(x) \sim f(y)$. For $(G_1, o_1), (G_2, o_2) \in \mathcal{RG}_D$ with $G_1 = (V_1, E_1, \theta_1)$ and $G_2 = (V_2, E_2, \theta_2)$, we define the *rooted distance* between G_1 and G_2 as $1/K$ where

$$K = \max\left\{k \in \mathbb{N}|\ \exists \text{ an isomorphism } f : B_k(o_1) \to B_k(o_2) \right.$$
$$\left. \text{such that } \sup_{x,y \in B_k(o_1)} |\theta_{1,xy} - \theta_{2,f(x)f(y)}| \le \frac{1}{k}\right\},$$

$\theta_{1,xy}, \theta_{2,f(x)f(y)}$ are edge weights of $xy \in E_1, f(x)f(y) \in E_2$, respectively. One can prove that $\mathcal{RG}_D$ endowed with the rooted distance is a compact metric space.

By a *random rooted graph* of degree D we mean a Borel probability distribution on $\mathcal{RG}_D$. We denote by $\mathcal{RRG}_D$ the collection of random rooted graphs of degree D. Any finite weighted graph G gives rise to a random rooted graph by assigning the root of G uniformly randomly.

For a rooted weighted graph $(G, o) \in \mathcal{RG}_D$ with $G = (V, E, \theta)$, the normalized Laplacian is a bounded self-adjoint operator on $\ell^2(V, \theta)$ which is independent of o. By spectral theorem, there is a projection-valued measure, denoted by $P_\bullet$, on $[0, 2]$, i.e. P_A is a projection on $\ell^2(V, \theta)$ for any Borel $A \subset [0, 2]$, such that for any continuous function $f \in C([0, 2])$ we have the functional calculus

$$f(\Delta_G) = \int_{[0,2]} f(x) dP_x \tag{16}$$

where $P_x = P_{[0,x]}$. We define the *spectral measure* of the rooted graph (G, o) as

$$\mu_{G,o}(A) = \frac{1}{\theta_o}\langle P_A\delta_o, \delta_o\rangle, \quad \forall A \subset [0, 2],$$

where $\langle\cdot, \cdot\rangle$ is the inner product for $\ell^2(V, \theta)$. One can easily show that $\mu_{G,o}$ is a probability measure on $[0, 2]$. Further calculation by using (16) yields $m_1(\mu_{G,o}) = 1$. Now we can define the expected spectral measure for rooted random graphs.

Definition 4.3. Let G be a random rooted graph. We define the *expected spectral measure* of G as

$$\mu_G = E(\mu_{G,o})$$

where the expectation is taken over the distribution on $\mathcal{RG}_D$.

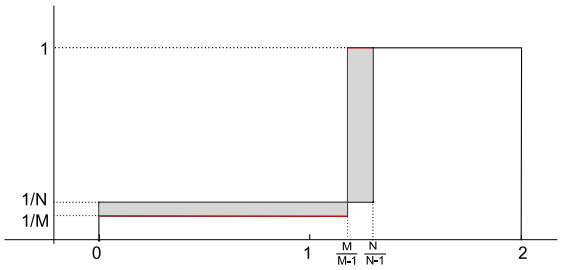

Fig. 7. Two complete graphs of size N and M.

Let G be a random rooted graph rising from a finite weighted graph with uniform distribution on its vertices. A similar calculation as in Abért–Thom–Virág [1] shows that

$$\mu_G = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$$

where $\{\lambda_i\}_{i=1}^{N}$ is the spectrum of the finite graph. Hence the expected spectral measure of random rooted graphs generalizes the spectral measure of finite graphs. There are other interesting classes of random rooted graphs such as unimodular and sofic ones, see e.g. [1].

The set of random rooted graphs of degree D, $\mathcal{RRG}_D$, endowed with d_p^W Wasserstein distance for expected spectral measures (d_p in short) is a pseudo-metric space. By Theorem 1.4, one can prove the following theorem.

Theorem 4.4. *For $p \in [1, \infty)$,*

$$\operatorname{diam}(\mathcal{RRG}_D, d_p) \le 2^{1-\frac{1}{p}}.$$

5. Calculation of examples

From now on, we will concentrate on the study of the spectral distance d_1. We calculate this distance for several classes of graphs in this section. Rather than the exact value of the d_1 distance between two graphs, we are more concerned with the asymptotical behavior of the distance between two sequences of graphs which become larger and larger, as the sizes of real networks in practice nowadays are typically huge. All example graphs we consider in this section are unweighted.

Proposition 5.1. *For two complete graphs G and G' with N and M ($M > N$) vertices respectively, we have*

$$d_1(G, G') = 2\frac{M-N}{N(M-1)}.$$

Proof. Recall the spectrum (10) of a complete graph. We then calculate the distance (i.e. the area of the gray region shown in Fig. 7),

$$d_1(G, G') = \frac{M}{M-1}\left(\frac{1}{N}-\frac{1}{M}\right)+\left(\frac{N}{N-1}-\frac{M}{M-1}\right)\left(1-\frac{1}{N}\right) = 2\frac{M-N}{N(M-1)}. \quad \square$$

Remark 5.2. When the size difference $M - N$ of two complete graphs is a fixed constant C, we have

$$d_1(G, G') = \mathcal{O}(1/N^2) \quad \text{as } N \to \infty.$$

Proposition 5.3. *For two connected complete bipartite graphs G and G' of size N and M ($M > N$) respectively, we have*

$$d_1(G, G') = 2\frac{M-N}{NM}.$$

Proof. The spectrum of a complete bipartite graph G with N vertices is

$$\sigma(G) = \{0, \underbrace{1, \ldots, 1}_{N-2}, 2\}.$$

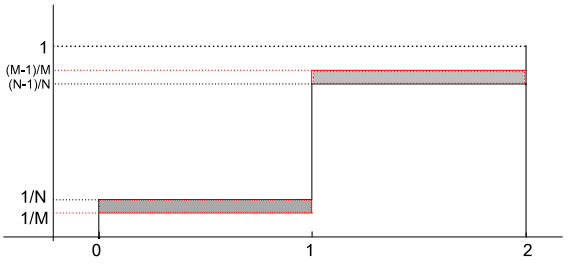

Fig. 8. Two complete bipartite graphs of size N and M.

Then the distance is (the area of the gray region shown in Fig. 8)

$$d_1(G, G') = \left(\frac{1}{N} - \frac{1}{M}\right) + \left(\frac{M-1}{M} - \frac{N-1}{N-1}\right) = 2\frac{M-N}{NM}. \quad \square$$

Remark 5.4. If the size difference $M - N$ of two complete bipartite graphs is a fixed constant C, we again observe the behavior

$$d_1(G, G') = \mathcal{O}(1/N^2) \quad \text{as } N \to \infty.$$

Proposition 5.5. *For two cubes G and G' of size 2^N and 2^{N+1} respectively, we have*

$$d_1(G, G') = \frac{1}{N+1}.$$

Proof. The spectrum of the cube G with 2^N vertices is

$$\left\{\frac{2i}{N} \text{ with multiplicity } \binom{N}{i}, i = 0, \ldots, N\right\}.$$

Firstly, observe $\frac{2i}{N} = \frac{2j}{N+1}$ when $i = j = 0$ or $i = N, j = N+1$. And for $j = i$, we have

$$\frac{2(i-1)}{N} < \frac{2j}{N+1} < \frac{2i}{N}, \quad \text{for } 1 \le i \le N.$$

Secondly, by the recursive formula $\binom{N+1}{k} = \binom{N}{k-1} + \binom{N}{k}$, for $1 \le k \le N$, we derive

$$\frac{1}{2^{N+1}}\sum_{i=0}^{k}\binom{N+1}{i} < \frac{1}{2^N}\sum_{i=0}^{k}\binom{N}{i} < \frac{1}{2^{N+1}}\sum_{i=0}^{k+1}\binom{N+1}{i}, \quad \text{for } 0 \le k \le N-1.$$

Therefore the distance between G and G' equals the area of the gray region depicted in Fig. 9. Again by the recursive formula of binomial numbers, we calculate,

$$\begin{aligned}
d_1(G, G') &= \sum_{k=1}^{N}\left\{\left(\frac{2k}{N+1} - \frac{2(k-1)}{N}\right)\left[\frac{1}{2^N}\sum_{i=0}^{k-1}\binom{N}{i} - \frac{1}{2^{N+1}}\sum_{i=0}^{k-1}\binom{N+1}{i}\right]\right. \\
&\quad \left. + \left(\frac{2k}{N} - \frac{2k}{N+1}\right)\left[\frac{1}{2^{N+1}}\sum_{i=0}^{k}\binom{N+1}{i} - \frac{1}{2^N}\sum_{i=0}^{k-1}\binom{N}{i}\right]\right\} \\
&= \frac{1}{2^N N(N+1)}\left\{\sum_{k=1}^{N}(N-k+1)\left[2\sum_{i=0}^{k-1}\binom{N}{i} - \sum_{i=0}^{k-1}\binom{N+1}{i}\right]\right. \\
&\quad \left. + k\left[\sum_{i=0}^{k}\binom{N+1}{i} - 2\sum_{i=0}^{k-1}\binom{N}{i}\right]\right\} \\
&= \frac{1}{2^N N(N+1)}\sum_{k=1}^{N}\left[(N-k+1)\binom{N}{k-1} + k\binom{N}{k}\right] \\
&= \frac{2}{2^N N(N+1)}\sum_{k=1}^{N}k\binom{N}{k} = \frac{2}{2^N N(N+1)} \cdot N \cdot 2^{N-1} = \frac{1}{N+1}. \quad \square
\end{aligned}$$

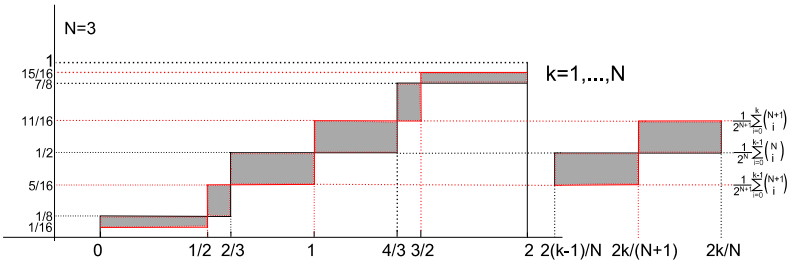

Fig. 9. An example of two neighboring cubes with $N = 3$ and $N + 1 = 4$.

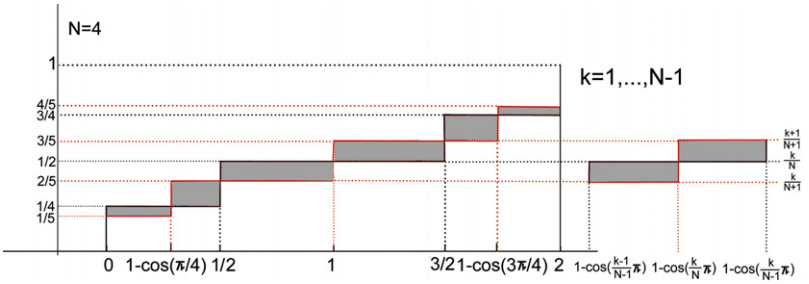

Fig. 10. An example of two neighboring paths with $N = 4$ and $N + 1 = 5$.

Remark 5.6. The distance between two neighboring cubes (N-cube and $(N + 1)$-cube) is $\mathcal{O}(1/N)$ as N tends to infinity. Recall a crucial difference of this example from previous ones is that the size difference, 2^N, is not uniformly bounded as $N \to \infty$.

Proposition 5.7. *For two paths G and G' of size N and $N + 1$ respectively, we have*

$$d_1(G, G') = \frac{1}{N(N+1)}\left(\cot^2\frac{\pi}{2N} - \cot^2\frac{\pi}{2(N-1)} + 1\right).$$

Proof. The spectrum of the path G with N vertices is

$$\left\{1 - \cos\frac{\pi i}{N-1}, i = 0, 1, \ldots, N-1\right\}.$$

Since $1 - \cos\frac{i\pi}{N-1} < 1 - \cos\frac{(i+1)\pi}{N} < 1 - \cos\frac{(i+1)\pi}{N-1}$ for $i = 0, \ldots, N-2$, and every eigenvalue of a path has multiplicity one, the situation is similar to Proposition 5.5, as shown in Fig. 10.

$$\begin{aligned} d_1(G, G') &= \sum_{k=1}^{N-1}\left\{\left(\cos\frac{k-1}{N-1}\pi - \cos\frac{k}{N}\pi\right)\left(\frac{k}{N} - \frac{k}{N+1}\right)\right. \\ &\quad \left. + \left(\cos\frac{k}{N} - \cos\frac{k}{N-1}\pi\pi\right)\left(\frac{k+1}{N+1} - \frac{k}{N}\right)\right\} \\ &= \frac{2}{N(N+1)}\sum_{k=1}^{N-1} k\left(\cos\frac{k-1}{N-1}\pi - \cos\frac{k}{N}\pi\right) \\ &= \frac{1}{N(N+1)}\left(\cot^2\frac{\pi}{2N} - \cot^2\frac{\pi}{2(N-1)} + 1\right). \end{aligned}$$

For the last equality above we use Lagrange's trigonometric identities

$$\sum_{k=1}^{N}\sin kx = \frac{\cos\frac{1}{2}x - \cos(n+\frac{1}{2})x}{2\sin\frac{1}{2}x}, \qquad \sum_{k=1}^{N}\cos kx = \frac{\sin(n+\frac{1}{2})x - \sin\frac{1}{2}x}{2\sin\frac{1}{2}x},$$

and their derivatives. □

Remark 5.8. By a Taylor expansion argument, we observe that

$$\cot^2 \frac{\pi}{2N} - \cot^2 \frac{\pi}{2(N-1)} = \mathcal{O}(N), \quad \text{as } N \to +\infty.$$

Therefore in this example, we have $d_1(G, G') = \mathcal{O}(1/N)$ as N tends to infinity.

We can calculate the example of cycles similarly.

Proposition 5.9. *For two cycles G and G' of size N and $N+1$ respectively, we have*

$$d_1(G, G') = \begin{cases} \dfrac{1}{N} + \dfrac{1}{N(N+1)} \left(\dfrac{1}{1-\cos(\frac{\pi}{N+1})} - \dfrac{4}{1-\cos(\frac{2\pi}{N})} \right), & \text{if } N \text{ is even;} \\ \dfrac{1}{N+1} - \dfrac{1}{N(N+1)} \left(\dfrac{1}{1-\cos(\frac{\pi}{N})} - \dfrac{4}{1-\cos(\frac{2\pi}{N+1})} \right), & \text{if } N \text{ is odd.} \end{cases}$$

Remark 5.10. For N- and $(N+1)$-cycles, we also have $d_1(G, G') = \mathcal{O}(1/N)$ as N tends to infinity.

6. Distance between large graphs

In this section we explore the behaviors of the spectral distance d_1 between large graphs in general. We require two large graphs are different from each other only by finite steps of operations which will be made clear in Remark 6.1. The main tool we employ is the so-called interlacing inequalities, which describe the changes of the spectrum when we perform some operations on the underlying graph. Such kind of results for normalized Laplacian of a graph have been studied in [16,31, 15,23,3]. In fact, we can observe the interlacing phenomena of eigenvalues for paths and cycles in Propositions 5.7 and 5.9.

Let the cardinality of vertices of G and G' be N and $N-j$ respectively, where $j \in \mathbb{Z}$ can be either negative or positive. Assume

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_N \quad \text{and} \quad 0 \le \lambda'_1 \le \lambda'_2 \le \cdots \le \lambda'_{N-j}$$

are the spectra of the corresponding normalized Laplacian Δ_G and $\Delta_{G'}$. Then interlacing inequalities have the following general form.

$$\lambda_{i-k_1} \le \lambda'_i \le \lambda_{i+k_2}, \quad \text{for each } i = 1, 2, \ldots, N-j, \tag{17}$$

with the notation that $\lambda_i = 0$ for $i \le 0$ and $\lambda_i = 2$ for $i > N$, and k_1, k_2 are constants independent of the index i.

Remark 6.1. G' can be obtained from G by performing the following operations.

- G' is the proper difference of G and one of its subgraph L. We say L is a subgraph of G if the weights $\theta_{L,uv} \le \theta_{G,uv}$ for all u, v. And the proper difference of G and L is a weighted graph with weights $\theta_G - \theta_L$. In this case,

  $$k_1 = \text{number of vertices in } L - \text{number of connected components of } L$$

  and

  $$k_2 = \text{number of vertices in } L$$

  (Horak–Jost [23, Corollary 2.11], see also Butler [15]). This includes the operation of deleting an edge (see Chen et al. [16] for the result for this particular operation). Symmetrically, this also covers the operation of adding a graph, see Butler [15] for particular results and Atay–Tunçel [3] for vertex replication.
- G' is the image of an edge-preserving map $\varphi : G \to G'$. By an edge-preserving map here we mean an onto map from the vertices of G to vertices of G', such that

  $$\theta_{G',xy} = \sum_{\substack{u \in \varphi^{-1}(x) \\ v \in \varphi^{-1}(y)}} \theta_{G,uv}$$

  for all vertices x, y of G', and the degree of vertices are defined according to the edge weights as usual in both graphs. Notice that for our purpose, we do not allow φ maps two neighboring vertices in G to the same vertex in G' in order to avoid self-loops. In this case,

  $$k_1 = 0 \quad \text{and} \quad k_2 = j.$$

  (Horak–Jost [23, Theorem 3.8].) This includes the operation of contracting vertices u, v such that $N(u) \cap (N(v) \cup \{v\}) = \emptyset$ (see Chen et al. [16]), where $N(u)$ stands for the neighborhood of u.

- G' is obtained from G by contracting an edge. We only consider edges uv in G such that $d_u, d_v > 1$. By edge contracting we mean deleting the edge (u, v) and identifying u and v (Horak–Jost [23, Definition 4.2]). Denote the number of common neighbors of u, v by m. Then

$$\text{when } m \neq 0, \quad k_1 = 2m, \qquad k_2 = 1 + 2m; \qquad \text{when } m = 0, k_1 = 0, k_2 = 2.$$

(Horak–Jost [23, Theorem 4.1], where the unweighted normalized Laplacian case was discussed. We do not know whether it is also true for weighted normalized Laplacian.)

Remark 6.2. To the knowledge of the authors, the above three classes of operations includes all the operations discussed in the literature for interlacing results of normalized Laplacian.

We prove the following result.

Theorem 6.3. *Let G, G' be two graphs, for which the spectra of corresponding normalized Laplacians satisfy (17). Then we have*

$$d_1(G, G') \leq C(k_1, k_2, j)\frac{1}{N}. \tag{18}$$

Proof. By definition, we have

$$d_1(G, G') = d_1^W\left(\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}, \frac{1}{N+j}\sum_{i+1}^{N+j}\delta_{\lambda_i'}\right).$$

By symmetry, w.l.o.g., we can suppose $j \geq 0$. We use a particular transport plan to derive the upper bound estimate. We move the mass $\frac{1}{N}$ from λ_i to λ_i' for $i = 1, 2, \ldots, N - j$. We then move the mass at the remaining positions $\lambda_{N-j+1}, \ldots, \lambda_N$ to fill the gaps at $\lambda_1', \lambda_2', \ldots, \lambda_{N-j}'$ with a cost for every transportation at most 2. That is, we have

$$\begin{aligned} d_1(G, G') &\leq \frac{1}{N}\sum_{i=1}^{N-j}|\lambda_i - \lambda_i'| + \frac{j}{N} \times 2 \\ &\leq \frac{1}{N}\sum_{i=1}^{N+j}|\lambda_{i+k_2} - \lambda_{i-k_1}| + \frac{2j}{N} \\ &\leq \frac{k_1 + k_2 + 1}{N}\sum_{i=1}^{N}|\lambda_{i+1} - \lambda_i| + \frac{2j}{N} \\ &\leq 2(k_1 + k_2 + j + 1)\frac{1}{N}. \end{aligned}$$

In the second inequality above, we used interlacing inequalities (17). This complete the proof. □

Remark 6.4. The disjoint union of a path of size N and an isolated vertex can be obtained from a path of size $N + 1$ by deleting an edge. A cycle of size N can be obtained from a cycle of size $N + 1$ by contracting an edge. Recall our calculation in Propositions 5.7 and 5.9, we see the estimate (18) is sharp in the order of $1/N$.

Remark 6.5. This theorem tells that if two large graphs share similar structure, then the spectral distance between them is small.

If G' is the graph obtained from G by performing operations such that k_1, k_2 are bounded (then j is also bounded), we say G' differs from G by a bounded operation.

Corollary 6.6. *Let $\{G_i\}_{i=1}^{\infty}$ be a sequence of graphs with size N_i tending to infinity. Assume that for any i, G_i' differs from G_i by a uniformly bounded operation, then*

$$\lim_{i\to\infty} d_1(G_i, G_i') = 0.$$

7. Applications to biological networks

In real biological networks, such as protein–protein interaction networks, edge-rewiring and duplication–divergence are two edit operations which have been proven to be closely related to some evolutionary mechanism, see [24,28]. For a spectral analysis of the effect of such operations on protein–protein interaction networks, we refer to [4]. In this section, we apply the spectral distance d_1 to capture evolutionary signals in protein–protein interaction networks through detecting

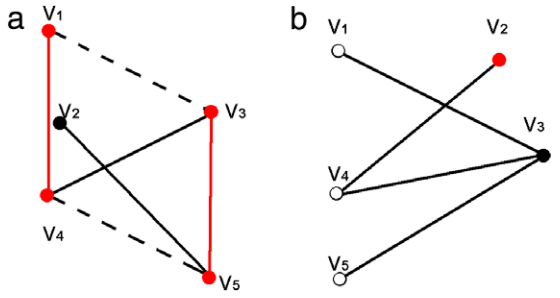

Fig. 11. (a) Edge-rewiring; (b) duplication–divergence.

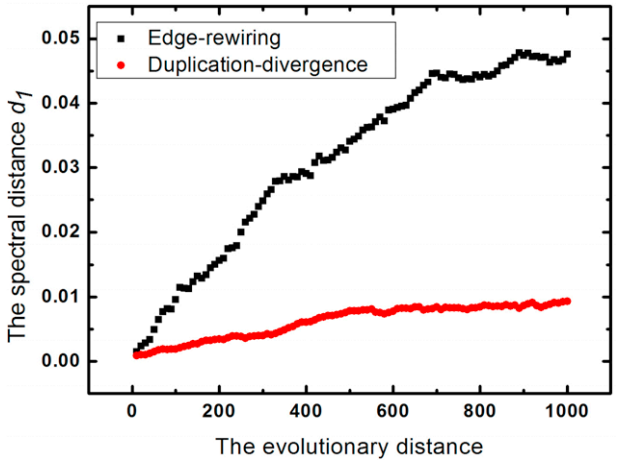

Fig. 12. The relation between the spectral distance d_1 and the evolutionary distance. Edit operations includes (a) edge-rewiring; (b) duplication–divergence.

their structural differences. We evolve graphs by operations of edge-rewiring and duplication–divergence, and then check the connection between the spectral distance d_1 and the evolutionary distance (i.e. the number of evolutionary operation steps). We restrict our simulations in the following to unweighted graphs.

Let us first explain the two edit operations on an unweighted graph $G = (V, E)$ explicitly.

- Edge-rewiring: Select randomly two edges (v_1, v_3), $(v_4, v_5) \in E$ on four distinct vertices $v_1, v_3, v_4, v_5 \in V$ (see Fig. 11(a)). Delete these two edges (v_1, v_3), (v_4, v_5) and add new edges (v_1, v_4), (v_3, v_5). The size of the graph is preserved by this operation, and so is the degree sequence.
- Duplication–divergence: Select randomly a target vertex $v_3 \in V$. Add a replica v_2 of v_3 and new potential edges connecting v_2 with every neighbor of v_3. Each of these potential edges is activated with certain probability (0.5 in our simulations). Then if at least one of these potential edges is established, keep the replica v_2; otherwise, delete the replica v_2 (see Fig. 11(b)).

Our simulations are designed as follows. We start form a Barabási–Albert scale-free graph with 1000 vertices. This is obtained through a mechanism incorporating growth and preferential attachment from a small complete graph of size 10, see [8,9]. For each step of preferential attachment, we add one vertex with two edges. We remark that the Barabási–Albert scale-free graph is not necessarily the best starting model for any biological network. However, it is closer to biological networks in many cases than the other two popular models, the Erdős–Rényi random graph and the Watts–Strogatz small-world graph. Therefore, we use it as our starting point here. We carry out the operation of edge-rewiring (duplication-divergence, resp.) on this graph iteratively, and plot the relationship of the spectral distance and the evolutionary distance between new obtained graphs and the original one.

In the plot of Fig. 12, we observe that the spectral distance between graphs obtained by edge-rewiring operations and the original one increases more quickly than that obtained by duplication–divergence operations. This indicates that, after the same number of operation steps, edge-rewiring brings in more randomness to the graph than duplication–divergence. Recall also the fact that the sizes of graphs are invariant in the former case and vary in the later case.

Although there is no strictly linear relation between the two distances, the spectral distance increases monotonically with respect to the evolutionary distance. Based on this crucial point, the spectral distance is very useful for exploring the hiding evolutionary history of large real networks.

Acknowledgments

The authors thank Jürgen Jost for inspiring discussions on the topic of spectral distance of large networks. BH thanks Andreas Thom and Balint Virág for their patient explanations of random graphs. SL thanks Willem Haemers for helpful discussions and pointing out Ref. [37]. The authors thank the anonymous referees for their valuable comments which improved the presentation of the manuscript.

JG and SL were supported by the International Max Planck Research School "Mathematics in the Sciences". BH was partially supported from the funding of the European Research Council under the European Union's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement n° 267087. SL was partially supported by the EPSRC Grant EP/K016687/1 "Topology, Geometry and Laplacians of Simplicial Complexes".

References

[1] M. Abért, A.B. Thom, B. Virág, Benjamini–Schramm convergence and pointwise convergence of the spectral measure, preprint.
[2] D. Aldous, R. Lyons, Processes on unimodular random networks, Electron. J. Probab. 12 (54) (2007) 1454–1508.
[3] F.M. Atay, H. Tunçel, On the spectrum of the normalized Laplacian for signed graphs: Interlacing, contraction, and replication, Linear Algebra Appl. 442 (2014) 165–177.
[4] A. Banerjee, J. Jost, Laplacian spectrum and protein–protein interaction networks, arXiv: 0705:3373 2007.
[5] A. Banerjee, J. Jost, Spectral plot properties: towards a qualitative classification of networks, Netw. Heterog. Media 3 (2) (2008) 395–411.
[6] A. Banerjee, J. Jost, On the spectrum of the normalized graph Laplacian, Linear Algebra Appl. 428 (2008) 3015–3022.
[7] A. Banerjee, J. Jost, Graph spectra as a systematic tool in computational biology, Discrete Appl. Math. 157 (10) (2009) 2425–2431.
[8] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[9] A.-L. Barabási, E. Bonabeau, Scale-Free Networks, Scientific American, 2003.
[10] F. Bauer, B. Hua, J. Jost, The dual Cheeger constant and spectra of infinite graphs, Adv. Math. 251 (2014) 147–194.
[11] F. Bauer, J. Jost, Bipartite and neighborhood graphs and the spectrum of the normalized graph Laplacian, Comm. Anal. Geom. 21 (4) (2013) 787–845.
[12] F. Bauer, J. Jost, S. Liu, Ollivier–Ricci curvature and the spectrum of the normalized graph Laplace operator, Math. Res. Lett. 19 (6) (2012) 1185–1205.
[13] I. Benjamini, O. Schramm, Recurrence of distributional limits of finite planar graphs, Electron. J. Probab. 6 (23) (2001) 13pp.
[14] N. Biggs, Algebraic Graph Theory, Second ed., Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1993.
[15] S. Butler, Interlacing for weighted graphs using the normalized Laplacian, Electron. J. Linear Algebra 16 (2007) 90–98.
[16] G. Chen, G. Davis, F. Hall, Z. Li, K. Patel, M. Stewart, An interlacing result on normalized Laplacians, SIAM J. Discrete Math. 18 (2) (2004) 353–361.
[17] F.R.K. Chung, Spectral graph theory, in: CBMS Regional Conference Series in Mathematics, vol. 92, American Mathematical Society, Providence, RI, 1997.
[18] D. Cvetković, M. Doob, H. Sachs, Spectra of graphs, in: Theory and Applications, Third ed., Johann Ambrosius Barth, Heidelberg, 1995.
[19] A.M. Fink, M. Jodeit Jr., On Chebyshev's other inequality, in: Inequalities in Statistics and Probability (Lincoln, Neb., 1982), in: IMS Lecture Notes Monogr. Ser., vol. 5, Inst. Math. Statist, Hayward, CA, 1984, pp. 115–120.
[20] J. Gu, The Spectral Distance Based on the Normalized Laplacian and Applications to Large Networks (Ph.D. Thesis), University of Leipzig, 2014.
[21] J. Gu, J. Jost, S. Liu, P.F. Stadler, Spectral classes of regular, random, and empirical graphs, 2014. arXiv: 1406.6454.
[22] G.H. Hardy, J.E. Littlewood, G. Pólya, Inequalities, second ed., University Press, 1952.
[23] D. Horak, J. Jost, Interlacing inequalities for eigenvalues of discrete Laplace operators, Ann. Global Anal. Geom. 43 (2) (2013) 177–207.
[24] I. Ispolatov, P.L. Krapivsky, A. Yuryev, Duplication-divergence model of protein interaction network, Phys. Rev. E 71 (6) (2005) 061911.
[25] J. Jost, S. Liu, Ollivier's Ricci curvature, local clustering and curvature dimension inequalities on graphs, Discrete Comput. Geom. 51 (2) (2014) 300–322.
[26] I. Jovanović, Z. Stanić, Spectral distances of graphs, Linear Algebra Appl. 436 (5) (2012) 1425–1435.
[27] M. Keller, D. Lenz, Dirichlet forms and stochastic completeness of graphs and subgraphs, J. Reine Angew. Math. 666 (2012) 189–223.
[28] J. Kim, I. Kim, S.K. Han, J.U. Bowie, S. Kim, Network rewiring is an important mechanism of gene essentiality change, Sci. Rep. 2 (2012).
[29] K. Kuwae, T. Shioya, Convergence of spectral structures: a functional analytic theory and its applications to spectral geometry, Comm. Anal. Geom. 11 (4) (2003) 599–673.
[30] J.R. Lee, S. Oveis Gharan, L. Trevisan, Multi-way spectral partitioning and higher-order Cheeger inequalities, in: STOC'12-Proceedings of the 2012 ACM Symposium on Theory of Computing, 1117-1130, ACM, New York, 2012; J. ACM 61 (2014), no. 6, 37:1–30.
[31] C.-K. Li, A short proof of interlacing inequalities on normalized Laplacians, Linear Algebra Appl. 414 (2–3) (2006) 425–427.
[32] S. Liu, Multi-way dual Cheeger constants and spectral bounds of graphs, Adv. Math. 268 (2015) 306–338.
[33] L. Lovász, Large networks and graph limits, in: American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012.
[34] O. Macindoe, W. Richards, Graph comparison using fine structure analysis, in: SOCIALCOM'10 Proceedings of the 2010 IEEE Second International Conference on Social Computing, IEEE Computer Society Washington, DC, USA, 2010, pp. 193–200.
[35] P. Major, On the invariance principle for sums of independent identically distributed random variables, J. Multivariate Anal. 8 (4) (1978) 487–517.
[36] B. Mohar, W. Woess, A survey on spectra of infinite graphs, Bull. Lond. Math. Soc. 21 (3) (1989) 209–234.
[37] D. Stevanović, Research problems from the Aveiro Workshop on graph spectra, Linear Algebra Appl. 423 (1) (2007) 172–181.
[38] M. Thüne, Eigenvalues of Matrices and Graphs (Ph.D. thesis), University of Leipzig, 2012.
[39] L. Trevisan, Max cut and the smallest eigenvalue, in: STOC'09-Proceedings of the 2009 ACM International Symposium on Theory of Computing, ACM, New York, 2009, pp. 263–271;
SIAM J. Comput. 41 (6) (2012) 1769–1786.
[40] C. Villani, Topics in optimal transportation, in: Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.