An Exponential Lower Bound for Spectral Density Estimation on Unweighted Graphs

Pan Peng PPENG@USTC.EDU.CN

Yuyang Wang WANGYVYANG@MAIL.USTC.EDU.CN

School of Computer Science and Technology, University of Science and Technology of China, Hefei, China

Joy Qiping Yang YANGQP5@PROTONMAIL.COM

The University of Sydney

Yichun Yang YC.YANG@BIT.EDU.CN

Beijing Institute of Technology

Editors: Steve Hanneke and Tor Lattimore

Abstract

We study lower bounds for estimating the spectral density of the normalized adjacency matrix of a graph. Previously, Cohen-Steiner et al. [KDD 2018] proposed an algorithm for ε-approximate spectral density estimation in the Wasserstein-1 distance, using $2^{O(1/\varepsilon)}$ random walks initiated from uniformly random nodes in the graph. Later, Jin et al. [COLT 2023] established a nearly matching exponential lower bound for *weighted* graphs, assuming the algorithm has access to samples from random walks started at random nodes. It was left open whether this lower bound could be extended to *unweighted* graphs.

In this paper, we answer this question in the affirmative by proving an exponential lower bound for unweighted graphs. Specifically, we show that no algorithm can compute an ε-approximation to the spectrum of a normalized graph adjacency matrix with constant success probability, even when given the full transcripts of $2^{\Omega(1/\varepsilon^{1/6})}$ random walks, each of length $2^{\Omega(1/\varepsilon^{1/6})}$, started from uniformly random nodes.

Keywords: Spectral density estimation, Sublinear-time algorithms, Query complexity, Random walks

1. Introduction

We study the problem of estimating the spectral density of the normalized adjacency matrix of an undirected graph. Let $G = (V, E, w)$ be an undirected graph on n vertices with positive edge weights $w \in \mathbb{R}^E_{>0}$. Let $A_G \in \mathbb{R}^{V\times V}$ denote its adjacency matrix and let D_G be the diagonal degree matrix. The normalized adjacency matrix of G is defined as $N_G = D_G^{-1/2} A_G D_G^{-1/2} \in \mathbb{R}^{V\times V}$. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of N_G. The *spectral density* of N_G (or simply of G), denoted by ρ_{N_G}, is the probability measure

$$\rho_{N_G}(x) = \frac{1}{n}\sum_{i=1}^{n} \delta(x - \lambda_i),$$

where δ denotes the Dirac delta function. The *spectral density estimation (SDE)* problem asks for a probability distribution q that approximates ρ_{N_G} under a suitable metric. In this work, we focus on approximating ρ_{N_G} to accuracy ε in the *Wasserstein-1 distance* (also known as Earth Mover

distance; see Definition 3). Equivalently (see e.g. Cohen-Steiner et al. (2018)), the goal is to compute (succinct representations of) estimates $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ such that

$$\frac{1}{n}\sum_{i=1}^{n}|\tilde{\lambda}_i - \lambda_i| \leq \varepsilon.$$

We refer to (the succinct representation of) such a vector $\tilde{\lambda}_i$ as an ε-approximation, in Wasserstein-1 distance (abbreviated as W_1-distance below), to the spectral distribution on $[-1, 1]$ induced by N_G. More broadly, spectral density estimation is defined for any symmetric matrix and has found applications across a wide range of domains, including network science (Farkas et al., 2001; Eikmeier and Gleich, 2017; Cohen-Steiner et al., 2018; Dong et al., 2019; Chen et al., 2021; Braverman et al., 2022), numerical linear algebra (Lin et al., 2016; Li et al., 2019) and deep learning (Mahoney and Martin, 2019).

Sublinear-time algorithms for SDE Several works have developed *sublinear-time algorithms* for SDE, i.e., algorithms whose runtime is sublinear in the size of the graph's representation.

In the *random neighbor* model (also called the *adaptive random walk model* in (Jin et al., 2023)), the algorithm is given access to two operations: (i) sample a uniformly random vertex of G, and (ii) given any specified vertex v, query a uniformly random neighbor of v. Using these operations, Cohen-Steiner, Kong, Sohler and Valiant (Cohen-Steiner et al., 2018) gave a randomized algorithm with query and time complexity $2^{O(1/\varepsilon)}$. Their algorithm relies solely on the transcripts of $2^{O(1/\varepsilon)}$ random walks of length $O(1/\varepsilon)$ starting from uniformly random vertices. It outputs a succinct representation of approximate eigenvalues with only $O(1/\varepsilon)$ distinct values, independent of n.

Moreover, in a stronger model that additionally allows the algorithm to access all neighbors of a vertex v in time proportional to its degree[1], Braverman, Krishnan, and Musco (Braverman et al., 2022) gave a randomized algorithm for SDE with runtime $O(n\varepsilon^{-7})$ that outputs a length-n vector representing a discrete spectral density. This was improved by Jin, Karmarkar, Musco, Sidford, and Singh (Jin et al., 2024), who provided a randomized algorithm with $O(n\varepsilon^{-2})$ queries and $O(n\varepsilon^{-3})$ runtime, as well as a deterministic algorithm with running time $n \cdot 2^{\tilde{O}(1/\varepsilon)}$. Their approach can also be generalized to weighted graphs, though it requires a slightly stronger query model.

Given these two classes of randomized algorithms – one with query complexity $2^{O(1/\varepsilon)}$ and the other with $O(n \cdot \mathrm{poly}(1/\varepsilon))$ – it is natural to ask whether the exponential dependence on ε in the sublinear-time algorithm of Cohen-Steiner et al. (2018) can be improved, possibly to polynomial, while still avoiding any dependence on the graph size n.

Answering this question turns out to be quite challenging. From a lower-bound perspective, it is difficult to construct pairs of graphs whose spectral densities are far apart, yet which are locally very similar and therefore hard to distinguish using a small number of queries. In this direction, Jin, Musco, Sidford, and Singh (Jin et al., 2023) proved a lower bound of $\Omega(1/\varepsilon^2)$ queries in the random neighbor model. They also studied the *non-adaptive random walk model*, in which an algorithm observes transcripts of a few random walks of short length starting from uniformly random vertices, and showed that no algorithm can compute an ε-accurate approximation of the spectral density with constant success probability even when given $2^{\Omega(1/\varepsilon)}$ walks of length $2^{\Omega(1/\varepsilon)}$. However, this lower bound applies only to weighted graphs.

1. This operation can be efficiently supported in the classical *adjacency list model* – where one can query vertex degrees and the i-th neighbor. It can also be simulated with high probability in the random neighbor model at an additional $\log n$ factor, provided the degree of the queried vertex is known (see, e.g., Footnote 8 in (Braverman et al., 2022)).

Our contribution In this paper, we study lower bounds on the query complexity of SDE, with a particular focus on *whether the exponential lower bound of Jin et al. (2023), previously established for weighted graphs, extends to unweighted graphs*. This question is well motivated: most real-world networks are naturally unweighted, and most sublinear-time SDE algorithms are designed for this setting (e.g., (Cohen-Steiner et al., 2018; Braverman et al., 2022)), so weighted-only lower bounds leave open the possibility of much faster algorithms for unweighted graphs. From a theoretical perspective, unweighted graphs impose strong structural constraints — in particular, the low-weight-edge mechanism used in prior weighted lower bounds is entirely unavailable — making it substantially harder to construct hard instances. Establishing lower bounds in this regime therefore clarifies the intrinsic limitations of sublinear access models and isolates the hardness of spectral estimation itself. Indeed, Jin et al. (2023) explicitly posed this as an open question.

In this work, we provide an affirmative answer to this question. Our result shows that even in this restricted and practically relevant setting, exponential query complexity barriers persist, thereby placing fundamental limits on what sublinear-time algorithms for SDE can achieve. Specifically, we show the following:

Theorem 1 *There exists universal constants $\varepsilon_0 > 0$ such that the following holds. For any $\varepsilon < \varepsilon_0$, no algorithm that is given access to the transcripts of m random walks of length T, each initiated at a uniformly random vertex of an* unweighted *graph G, can approximate the spectral density of G to ε accuracy in Wasserstein-1 distance with probability greater than $3/4$, unless*

$$m \cdot T > 2^{c/\varepsilon^{1/6}}$$

for some universal constant $c > 0$.

Consequently, even for unweighted graphs, no algorithm can compute an ε-accurate spectral density estimate with constant probability, even when given the transcripts of $2^{\Omega(1/\varepsilon^{1/6})}$ random walks of length $2^{\Omega(1/\varepsilon^{1/6})}$ started from uniformly random nodes.

We remark that our lower bound applies only to the non-adaptive random walk model. It remains an open question whether an exponential lower bound can be established for the random neighbor model (also called the adaptive random walk model in (Jin et al., 2023)), in which the algorithm is allowed to start a random walk either from a uniformly random node or from any node of its choice. At present, the best known lower bound in this adaptive model is the aforementioned $\Omega(1/\varepsilon^2)$.

1.1. Technical overview

We briefly review the lower bound for weighted graphs established by Jin et al. (2023). Their hard instances are constructed by combining cycles of different lengths. Specifically, they consider two graph families: R_1, consisting of $2n$ disjoint cycles of length $\ell := \Theta(1/\varepsilon)$, and R_2, consisting of n disjoint cycles of length 2ℓ. These two graphs satisfy $W_1(R_1, R_2) = \Omega(1/\ell)$, yet they can be easily distinguished by short random walks. To overcome this issue, they first pad both R_1 and R_2 with additional isolated vertices, and then augment the resulting graphs with a low-weight complete graph on the same vertex set. Since the added edges have very small weights, this modification does not significantly affect the W_1-distance, while simultaneously making the two graphs hard to distinguish via random walks. Intuitively, although each auxiliary edge has small weight, the *total* auxiliary weight incident to any vertex forms a constant fraction of its weighted degree. Consequently, a random walk leaves the original cycle with constant probability at each step, and with

high probability does not remain long enough to complete a full traversal of the cycle. At the same time, the spectral perturbation introduced by the auxiliary edges remains controlled, preserving the W_1-distance gap.

Extending this lower bound to unweighted graphs, as noted in (Jin et al., 2023), is "*surprisingly tricky*". For instance, a natural approach might be to replace the weighted complete graph with an unweighted expander; however, doing so would significantly alter the spectra of the graphs, making the analysis considerably more challenging.

Our approach. To extend the exponential lower bound to simple unweighted graphs and to bypass the difficulties above, we avoid directly augmenting disjoint cycles with unweighted expanders. Instead, we consider *Cartesian products* of unweighted expander graphs and cycle graphs. Specifically, we start with a random d-regular graph H (an expander) and construct two product graphs

$$G_1 = H \,\square\, C_1 \qquad \text{and} \qquad G_2 = H \,\square\, C_2,$$

where C_1 consists of two disjoint cycles of length ℓ, and C_2 is a single cycle of length 2ℓ. Here $\ell = \mathrm{poly}(1/\varepsilon)$. As before, the W_1-distance between the spectral densities of C_1 and C_2 is large, on the order of $\Omega(1/\ell)$. We next show the following:

1. The graphs G_1 and G_2 are ε-far in W_1-distance.
2. Given access to transcripts of m non-adaptive random walks of length T with $m \cdot T < 2^{c/\varepsilon^{1/6}}$, no algorithm can distinguish whether the transcripts were generated from G_1 or from G_2.

To prove Item (1), we exploit the fact that the spectral density of a Cartesian product satisfies a convolution property (Fact 9). Combined with recent eigenvalue concentration results for random d-regular graphs from (Huang et al., 2024) (Lemma 25), this enables us to analyze the Wasserstein distance between the spectral densities of the two product graphs after convolution:

$$\rho_{A_{G_1}} = \rho_{A_H} * \rho_{A_{C_1}}, \qquad \rho_{A_{G_2}} = \rho_{A_H} * \rho_{A_{C_2}},$$

where ρ_{A_K} denotes the spectral density of the adjacency matrix A_K of a graph K, and $*$ denotes the convolution product (Definition 8). By our construction, C_1 places two eigenvalues at -2 whereas C_2 places only one, with the next eigenvalue of C_2 at $-2\cos(\pi/\ell)$ (Fact 5), creating a CDF discrepancy of $\frac{1}{2\ell}$ over an interval of width $\Theta(1/\ell^2)$. Since $W_1(\rho_1, \rho_2) = \int |F_1 - F_2|\, dx$ is lower bounded by restricting the integral to any sub-interval, we focus on the interval $[-2, -2\cos(\pi/\ell)]$ where only this multiplicity difference contributes. Through a careful analysis (Lemmas 17 and 18), we show that a $\mathrm{poly}(1/\ell)$ gap persists after convolution with ρ_d, despite the fact that ρ_d vanishes at the endpoint $-2\sqrt{d-1}$ of its support. Consequently, the spectral densities of our constructed graphs G_1 and G_2 also exhibit a $\mathrm{poly}(1/\ell)$ gap under the W_1 distance. Finally, since G_1 and G_2 are $(d+2)$-regular graphs by our construction, we show that the gap between the spectral densities of their normalized adjacency matrix N_{G_1} and N_{G_2} remains: $W_1(\rho_{N_{G_1}}, \rho_{N_{G_2}}) = \frac{1}{d+2} W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}}) \geq \frac{1}{10\ell^6 d^2}$. See Theorem 13 and Corollary 15 for details. We ensure Item (1) by setting $2\varepsilon = \frac{1}{20\ell^6 d^2}$.

To prove Item (2), we construct a coupling $\mathcal{D}$ over the distributions of random walk transcripts (Definition 21) on G_1 and G_2 such that, with high probability, the two transcripts generated by $\mathcal{D}$ are identical. The claim then follows from the classical coupling lemma. While Jin et al. (2023) also employ a coupling argument to establish indistinguishability, extending their approach to our setting presents two key challenges due to the sparsity and lack of uniform local structure in our

unweighted graphs. First, the base graph H does not exhibit the uniform local connectivity of a complete graph. Second, after leaving a cycle, a random walk may return and continue exploring the same cycle, rather than escaping permanently as in (Jin et al., 2023). These issues make both the construction of the coupling and the analysis of coinciding transcripts substantially more delicate.

To address these challenges, we exploit the natural projection of vertices in the Cartesian product onto the base graph (Definition 19) and introduce the notion of *projected walks*. Using these notions, we derive sufficient conditions under which the coupling produces identical transcripts, which ultimately allows us to complete the proof.

Remark Finally, we note that there exists an algorithm that can distinguish G_1 from G_2 using $\mathrm{poly}(1/\varepsilon)$ queries in the *adaptive* random walk model. This follows from a structural property of the hard instances: for most vertices u in G_1 or G_2, there exist two edges e^+ and e^- incident to u that each belong to exactly d length-4 cycles. These edges are precisely those lying on the copy of C_1 or C_2 containing u. In the adaptive random walk model, such edges can be identified using only $\mathrm{poly}(d)$ queries (by running a breadth-first search of depth 4; since a $1 - o(1)$ fraction of vertices in a random d-regular expander are not contained in any length-4 cycles, any such cycle will be due to e^+ and e^-); completing the corresponding cycle then requires $\mathrm{poly}(d, \ell) = \mathrm{poly}(1/\varepsilon)$ queries, which suffices to distinguish whether u comes from G_1 or G_2.

1.2. Related work

The matrix SDE problem – computing the full eigendecomposition – requires at least $O(n^\omega)$ time, where $\omega < 2.373$ is the fast matrix multiplication exponent Parlett (1998); Banks et al. (2023). For a matrix A, methods for spectral density estimation that run in $o(n^\omega)$ time were first developed for applications in condensed matter physics and quantum chemistry Skilling (1989); Silver and Röder (1994); Moldovan et al. (2020). It is shown that for any $n \times n$ symmetric matrix A with spectral density s, the popular Stochastic Lanczos Quadrature (SLQ) method provably computes an approximate spectral density q satisfying $W_1(s, q) \leq \varepsilon \|A\|_2$ using only $O(1/\varepsilon)$ matrix-vector multiplications with A when $\varepsilon = \tilde{\Omega}(1/\sqrt{n})$ (Chen et al., 2021). The same error bound and query complexity also hold by using the Chebyshev moment matching method (Braverman et al., 2022). Later, Musco et al. (2025) removed the $\varepsilon = \tilde{\Omega}(1/\sqrt{n})$ assumption; see also (Bhattacharjee et al., 2025) for further improvements.

Our work is motivated by these advances, by the general importance of sublinear-time graph algorithms (Rubinfeld and Shapira, 2011), and by recent progress on sublinear-time algorithms for other spectral problems, including expander testing (Goldreich and Ron, 2011) and spectral clustering (Peng, 2020; Gluch et al., 2021), etc.

2. Preliminaries

Let $G = (V, E)$ be a graph with vertex set V and edge set E. Its adjacency matrix is denoted by A_G, its diagonal degree matrix by D_G, and its normalized adjacency matrix by $N_G = D_G^{-1/2} A_G D_G^{-1/2}$. We begin by defining the two types of spectral density functions that will be frequently used in our analysis.

Definition 2 (Empirical spectral density) *Given a graph G and its normalized adjacency matrix N_G with eigenvalues $\{\lambda_i\}_{1 \leq i \leq n}$, the spectral density function of N_G is defined by: $\rho_{N_G}(x) =$*

$\frac{1}{n}\sum_{i=1}^{n}\delta(x-\lambda_i)$, *where* $\delta(\cdot)$ *denotes the Dirac delta function, which can be heuristically represented as* $\int_{-\infty}^{\infty}\delta(x)dx=1$ *with* $\delta(0)=\infty$ *and* $\delta(x)=0$ *for all* $x\neq 0$.

Similarly, the spectral density function of adjacency matrix A_G *is defined by:* $\rho_{A_G}(x)=\frac{1}{n}\sum_{i=1}^{n}\delta(x-\lambda_i')$, *where* $\{\lambda_i'\}_{1\leq i\leq n}$ *are the eigenvalues of* A_G.

In this paper we quantify the discrepancy between two density functions using the W_1 distance.

Definition 3 *Given two distributions* ρ_1,ρ_2, *define* Ψ *to be the set of all couplings* $\psi(x,y)$ *between* ρ_1,ρ_2. *Then the* W_1 *distance is defined as:* $W_1(\rho_1,\rho_2)=\inf_{\psi\in\Psi}\int_{\mathbb{R}}\int_{\mathbb{R}}|x-y|\cdot\psi(x,y)dxdy$.

Compared to the original definition, we use an alternative and more convenient formula for the W_1 distance.

Fact 4 (Panaretos and Zemel (2019)) *The* W_1 *distance between two density functions* $\rho_1,\rho_2:\mathbb{R}\to\mathbb{R}$ *can be expressed by:* $W_1(\rho_1,\rho_2)=\int_{\mathbb{R}}|F_1(x)-F_2(x)|dx$, *where* F_1,F_2 *are the cumulative distribution functions (CDFs) of* ρ_1,ρ_2 *respectively.*

We will use the following lemma, which characterizes the eigenvalues of the cycle graph.

Fact 5 (Theorem 5.5.1 in Spielman (2019)) *The eigenvalues of the adjacency matrix of length* ℓ *cycle is* $2\cos\left(\frac{2k}{\ell}\pi\right)$ *(with multiplicity) for* $0\leq k<\ell$.

2.1. Cartesian product of graphs

Definition 6 (Cartesian product) *The Cartesian product* $G\square H$ *of graphs* G *and* H *is the graph defined as follows:*

- *the vertex set of* $G\square H$ *is* $V(G)\times V(H)=\{(u,v)\mid u\in V(G),\ v\in V(H)\}$;
- *two vertices* (u,v) *and* (u',v') *are adjacent in* $G\square H$ *if and only if either*
  - *(a)* $u=u'$ *and* v *is adjacent to* v' *in* H, *or*
  - *(b)* $v=v'$ *and* u *is adjacent to* u' *in* G.

A well-known result states that the eigenvalues of the Cartesian product $G\square H$ are obtained by summing the eigenvalues of G and H respectively.

Fact 7 (Lemma 6.3.2 in Spielman (2019)) *Let* $\{\lambda_i\}_{1\leq i\leq n}$ *be the eigenvalues of adjacency matrix* A_G, $\{\mu_i\}_{1\leq i\leq m}$ *be the eigenvalues of adjacency matrix* A_H. *Then the adjacency matrix of the Cartesian product* $G\square H$ *has eigenvalues* $\{\lambda_i+\mu_j\}_{1\leq i\leq n;1\leq j\leq m}$.

Definition 8 *The convolution between two functions* f *and* g *is defined as:* $f*g(z)=\int_{\mathbb{R}}f(x)g(z-x)dx$.

We state the following fact about the spectral density $\rho_{A_{G\square H}}$ of the product graph $G\square H$: the spectral density of the product is the convolution of the spectral densities of the factors. We will defer its proof to Appendix A.1.

Fact 9 (Cartesian product and convolution) *It holds that:* $\rho_{A_{G\square H}}=\rho_{A_G}*\rho_{A_H}$.

2.2. Random d-regular graphs and their spectrum

Our hard instances are based on the random d-regular graphs defined as follows.

Definition 10 *A random d-regular graph on n vertices is chosen uniformly from the set of all n-vertex d-regular graphs.*

According to classical results, the expected spectral density of the adjacency matrix for a random d-regular graph with sufficiently large n vertices is the Kesten-McKay distribution (McKay, 1981):

$$\rho_d(x) = \begin{cases} \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)} & \text{if } x \in [-2\sqrt{d-1}, 2\sqrt{d-1}], \\ 0 & \text{otherwise} \end{cases}$$

Since ρ_{A_H} is a discrete measure, we approximate it by the continuous Kesten–McKay density ρ_d, to which the analytic arguments of Lemmas 17 and 18 apply. We now prove that the gap between density functions ρ_{A_H} and ρ_d is small, where H is the n-vertex random d-regular graph, using concentration results of Lemma 25. We defer its proof to Appendix A.2.

Lemma 11 *It holds that $W_1(\rho_{A_H}, \rho_d) \le 3\sqrt{d-1} \cdot n^{-1+o(1)} \le n^{-0.99}$ with probability $1 - n^{-\alpha}$ for constant $\alpha \ge 1$.*

3. The hard instances

We now define our construction of the hard instances. An illustration is provided in Figure 1.

Definition 12 (Hard instances) *Let d be some constant, and n and ℓ be integers such that $n \ge \ell$. Fix $H = (V_H, E_H)$ to be an n-vertex random d-regular graph. Let $C_1 = (V_{C_1}, E_{C_1})$ and $C_2 = (V_{C_2}, E_{C_2})$ be graphs on 2ℓ vertices, where C_1 consists of two vertex-disjoint cycles of length ℓ, while C_2 consists of a single cycle of length 2ℓ. We define the hard instances:*

- *$G_1 = H \square C_1$, the Cartesian product of H and C_1;*
- *$G_2 = H \square C_2$, the Cartesian product of H and C_2.*

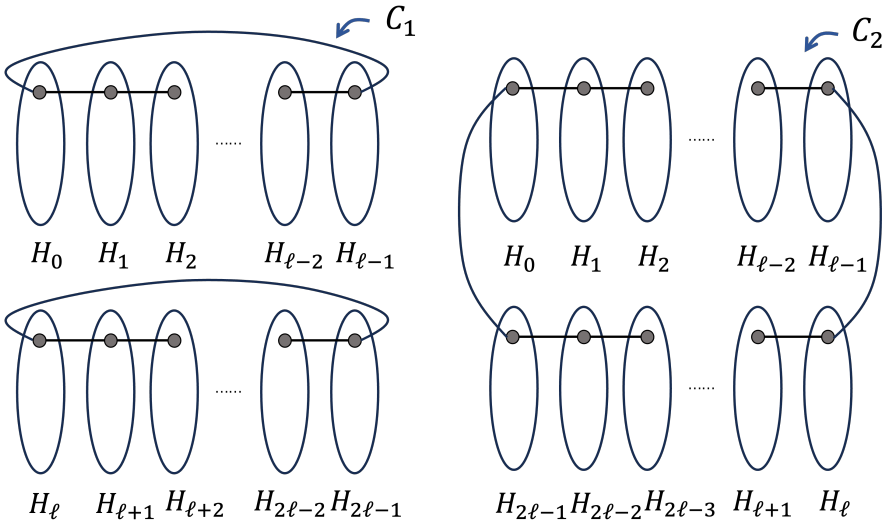
We remark that in our final construction, we fix $d \ge 7$ as a constant, and for a sufficiently small constant ε, we set $\ell = \mathrm{poly}(1/\varepsilon)$ and $n = 2^{\Omega(\ell)}$ (see Section 6 for details).

4. Large W_1-distance between spectral densities of G_1 and G_2

In this section we prove the following theorem, showing that there is a polynomial large gap between $\rho_{A_{G_1}}$ and $\rho_{A_{G_2}}$ in terms of Wasserstein-1 distance with high probability.

Theorem 13 *Let G_1 and G_2 be as defined in Definition 12 with constant $d \ge 7$ and $n \ge \ell^7 d^2$. Then it holds that $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}}) \ge \frac{\pi}{20\ell^6 d^{3/4}} - O(n^{-0.99})$, with probability $1 - n^{-\Omega(1)}$.*

Remark 14 *It is likely that the above W_1-distance lower bound of $\Omega(1/\ell^6)$ is not optimal. In fact, the lower bound in Theorem 13 may be improvable to $\Omega(1/\ell^2)$; see the simulation results in Appendix D for supporting evidence. Determining the optimal lower bound of $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}})$ remains an open problem.*



(*a*) G_1 (*b*) G_2

Figure 1: Illustration of our constructions G_1 and G_2.

We have the following corollary of the W_1-distance of $\rho_{N_{G_1}}$ and $\rho_{N_{G_2}}$.

Corollary 15 *Let G_1 and G_2 be as defined in Definition 12. Suppose that $d \geq 7$ and $n \geq \ell^7 d^2$. Then it holds that $W_1(\rho_{N_{G_1}}, \rho_{N_{G_2}}) \geq \frac{1}{10\ell^6 d^2}$, with probability $1 - n^{-\Omega(1)}$.*

Proof Note that G_1 and G_2 are $(d+2)$-regular graphs. By the relation between the eigenvalues of normalized adjacency matrix and eigenvalues of adjacency matrix, the fact that d is constant and $n \geq \ell^7 d^2$, and Theorem 13, we have

$$W_1(\rho_{N_{G_1}}, \rho_{N_{G_2}}) \geqslant \tfrac{1}{d+2} \cdot \tfrac{\pi}{20\ell^6 d^{3/4}} - O(n^{-0.99}) \geq \tfrac{1}{10\ell^6 d^2}.$$ ■

4.1. Proof of Theorem 13

Now we prove Theorem 13. In the following, we assume that $d \geq 7$ and $n \geq \ell^7 d^2$. We provide the following Lemma, where we defer its proof to Appendix A.3.

Lemma 16 *With probability at least $1 - n^{-1}$, it holds that*

$$W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}}) \geq W_1\big(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}\big) - O\big(n^{-0.99}\big).$$

In the following, we compute the lower bound for the W_1 distance $W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}})$. First, we show two lemmas (Lemma 17 and Lemma 18) that when combined give a lower bound on $W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}})$. We defer their proofs to Appendix A.4 and A.5, respectively.

Lemma 17 *It holds that*

$$W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}) \geq \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} \frac{(t+2)^2 + 4\sqrt{d-1}(t+2) + 4(d-1)}{2\left(2\sqrt{d-1} + t + 2\right)} \cdot \rho_d(t+2)dt.$$

Lemma 18 *For any $d \geq 3$, it holds that*

$$\rho_d(x) \;\geq\; \frac{\sqrt{x+2\sqrt{d-1}}}{2\pi d^{3/4}} \quad \textit{for all } x \in \left(-2\sqrt{d-1},\, -1.99\sqrt{d-1}\right).$$

By Lemma 17 and Lemma 18, we have

$$\begin{aligned}
W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}) &\geq \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} \frac{(t+2)^2 + 4\sqrt{d-1}(t+2) + 4(d-1)}{2\left(2\sqrt{d-1}+t+2\right)} \cdot \rho_d(t+2)dt. \\
&= \int_{-2\sqrt{d-1}}^{-2\sqrt{d-1}+2-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} \frac{t^2 + 4\sqrt{d-1}\cdot t + 4(d-1)}{2\left(2\sqrt{d-1}+t\right)} \cdot \rho_d(t)dt \\
&\geqslant \int_{-2\sqrt{d-1}}^{-2\sqrt{d-1}+2-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} \frac{t^2 + 4\sqrt{d-1}\cdot t + 4(d-1)}{2\left(2\sqrt{d-1}+t\right)} \cdot \frac{\sqrt{t+2\sqrt{d-1}}}{2\pi d^{3/4}} dt \\
&= \frac{1}{2\ell} \int_{-2\sqrt{d-1}}^{-2\sqrt{d-1}+2-2\cos\left(\frac{\pi}{\ell}\right)} \frac{\left(t+2\sqrt{d-1}\right)^2}{2\left(\sqrt{t+2\sqrt{d-1}}\right)} \cdot \frac{1}{2\pi d^{3/4}} dt \\
&= \frac{1}{2\ell} \int_0^{2-2\cos\left(\frac{\pi}{\ell}\right)} \frac{t^2}{2\sqrt{t}} \cdot \frac{1}{2\pi d^{3/4}} dt = \frac{1}{8\ell\cdot\pi d^{3/4}} \int_0^{2-2\cos\left(\frac{\pi}{\ell}\right)} t^{3/2} dt \\
&= \frac{1}{8\ell\cdot\pi d^{3/4}} \cdot \frac{2}{5}\left(2-2\cos\left(\frac{\pi}{\ell}\right)\right)^{5/2} \geq \frac{\pi}{20\ell^6 d^{3/4}}.
\end{aligned}$$

This finishes the proof of Theorem 13.

5. Indistinguishability of G_1 and G_2

5.1. Basic tools

Recall from Definition 6 that in $G \,\square\, H$, for any fixed $u \in V(G)$, the vertex set $\{(u,v) \mid v \in V(H)\}$ induces a copy (layer) of H, and symmetrically for G. For the hard instances G_1 and G_2 (Definition 12), each is the Cartesian product of a d-regular graph H (expander) and a 2-regular graph (cycle), and hence is $(d+2)$-regular. Every vertex lies in a unique H-layer and a unique cycle layer, which we refer to as the expander and the cycle of the vertex, respectively. We define the projection accordingly.

Definition 19 (Projection onto the base graph) *Let $G_1 = H \,\square\, C_1$. The projection onto H is the mapping $p_H^1 : V(G_1) \to V(H)$ defined by $p_H^1(u,v) = u$ for all $(u,v) \in V(H) \times V(C_1)$.*

The projection p_H^2 is defined analogously for $G_2 = H \,\square\, C_2$.

Remark 20 *Slightly abusing notation, although we* relabel the vertices *using a uniformly random permutation of the integers $\{1, \ldots, 2n\ell\}$, the projections p_H^1 and p_H^2 are always understood with respect to the underlying Cartesian product structure, rather than the specific vertex labels.*

Now we define the distribution of random walk transcripts on G_1 and G_2.

Definition 21 (Transcript distribution) *For m non-adaptive random walks, each of length T, a random walk transcript S is a collection of m individual walks, $S = \{S_1, \ldots, S_m\}$, where each S_i is a sequence of T node labels $v_{i,0}, \ldots, v_{i,T}$ corresponding to the nodes visited by the walk. Let $\mathcal{D}_{G_1}$ and $\mathcal{D}_{G_2}$ denote the probability distributions over transcripts generated by performing the walks on G_1 and G_2, respectively, where the nodes are labeled using a uniformly random permutation of the integers $1, \ldots, 2n\ell$.*

5.2. Indistinguishability via coupling

This section bounds the TV distance between the transcript distributions using the coupling method, indicating indistinguishability of the hard instances under the non-adaptive random walk model.

Lemma 22 *Let $\ell \geq 10 \log_d n$. For m non-adaptive walks of length T, the TV distance between $\mathcal{D}_{G_1}$ and $\mathcal{D}_{G_2}$ is bounded by $d_{TV}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2}) \leq \frac{m^2 T^3}{n^c}$, where c is a small universal constant.*

To prove this, we define a coupling $\mathcal{D}$ between $\mathcal{D}_{G_1}$ and $\mathcal{D}_{G_2}$. The coupling $\mathcal{D}$ is a joint distribution over a pair of random walk transcripts S^1 and S^2, such that the marginal distribution of S^1 equals $\mathcal{D}_{G_1}$, and the marginal distribution of S^2 equals $\mathcal{D}_{G_2}$. We show that S^1 and S^2 are identical with high probability, which establishes the closeness by the classical coupling lemma:

$$d_{TV}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2}) \leq \Pr_{\mathcal{D}}[S^1 \neq S^2].$$

Projected walk. To introduce the coupling, we describe the random walk on the constructed graph G_1. The i-th walk starts from $v_{i,1}$ chosen uniformly at random from $V(G_1)$. At each step, the current vertex u has $d+2$ neighbors: d in its expander and 2 on its cycle. Hence with probability $\frac{d}{d+2}$ the walk performs an *expander step*, and with probability $\frac{2}{d+2}$ a *cycle step*.

We project the walk onto H via p_H^1 (Definition 19), obtaining the *projected walk*. Starting from $p_H^1(v_{i,1})$, each expander step $v_{i,k} \to v_{i,k+1}$ induces a step $p_H^1(v_{i,k}) \to p_H^1(v_{i,k+1})$. In contrast, a cycle step moves to a vertex u' aligned with u on H, implying $p_H^1(u) = p_H^1(u')$, so the projected walk stays. The walk on G_2 can be interpreted analogously. Thus, a random walk on G_1 (resp. G_2) corresponds, under projection p_H^1 (resp. p_H^2), to a lazy random walk on H with laziness $\frac{2}{d+2}$. This insight contributes to both the design of coupling and the proof of Lemma 22.

Coupling. We next define the coupling $\mathcal{D}$ by describing a random process that explicitly generates random walk transcripts S^1 and S^2. The process ensures that S^1 and S^2 are distributed according to $\mathcal{D}_{G_1}$ and $\mathcal{D}_{G_2}$. To this end, we employ a *lazy labeling* procedure, which assigns labels to vertices upon their first visit during the random walks. We maintain dictionaries $L_1 : V_1 \to \{1, \ldots, 2n\ell\}$ and $L_2 : V_2 \to \{1, \ldots, 2n\ell\}$; initially $L_i(v) = \mathbf{NULL}$ for every $v \in V_i$ and $i \in \{1, 2\}$. Once a label j is assigned to v, i.e. $L_i(v) \leftarrow j$, all subsequent queries return j. In describing the coupling, we also refer to the cycle containing a vertex v and to its *left* and *right* neighbors on the cycle.

We now describe the process.

1. Choose a uniformly random permutation Π of $\{1, \cdots, 2n\ell\}$. Let $\Pi(j)$ denote the j-th label in the permutation. Initialize $j \leftarrow 1$.

2. For $k = 1, \cdots, m$:

(a) Choose independent, uniformly random vertices $v^1_{k,0} \in G_1$ and $v^2_{k,0} \in G_2$. For $i \in \{1,2\}$, if $L_i(v^i_{k,0}) = \mathbf{NULL}$, set $L_i(v^i_{k,0}) \leftarrow \Pi(j)$. Increment $j \leftarrow j+1$.

(b) During this iteration, maintain a variable x tracking the *depth* of the current vertex. Initialize $x = 0$. Every cycle step increments (right) or decrements (left) x by 1.

(c) Define $p^1_{k,i} := p^1_H(v^1_{k,i})$ and $p^2_{k,i} := p^2_H(v^2_{k,i})$. Initialize H^1_k (resp. H^2_k) as a subgraph of H containing only the isolated vertex $p^1_{k,0}$ (resp. $p^2_{k,0}$).

We use H^1_k and H^2_k to maintain the subgraphs induced by the projected walks (ignoring self-loops), updating them after each step. For every $i \in \mathbb{Z}$, let $H^1_{k,i}$ (resp. $H^2_{k,i}$) denote the subgraph induced by projected walk steps taken at depth $x = i$.

We aim to maintain that H^1_k and H^2_k remain isomorphic trees. Accordingly, we maintain a *consistent isomorphic mapping* $f_k : V^1_k \to V^2_k$, initially mapping the two starting vertices. During this iteration, we maintain the invariant

$$(\bigstar) \quad f_k(p^1_{k,i}) = p^2_{k,i} \quad \text{for all } i.$$

If this invariant cannot be preserved, we say that the mapping f_k *fails*; otherwise H^1_k and H^2_k remain isomorphic.

(d) For $i = 1, \cdots, T$:

i. If f_k fails, choose $v^1_{k,i+1}$ and $v^2_{k,i+1}$ uniformly from the neighbors of $v^1_{k,i}$ and $v^2_{k,i}$, respectively. Otherwise, **with probability** $\frac{2}{d+2}$ perform a cycle step and **with probability** $\frac{d}{d+2}$ perform an expander step, defined as follows.

- **(Cycle step)**
  - **With probability $\frac{1}{2}$:** set $v^1_{k,i+1}$ (resp. $v^2_{k,i+1}$) to be the left neighbor of $v^1_{k,i}$ (resp. $v^2_{k,i}$), and update $x \leftarrow x-1$.
  - **With probability $\frac{1}{2}$:** set $v^1_{k,i+1}$ (resp. $v^2_{k,i+1}$) to be the right neighbor of $v^1_{k,i}$ (resp. $v^2_{k,i}$), and update $x \leftarrow x+1$.
- **(Expander step)** Define

$$R^1 := \{u \in V_1 : (v^1_{k,i}, u) \in E_1 \text{ and } (p^1_{k,i}, p^1_H(u)) \in H^1_k\},$$

  that is, neighbors of $v^1_{k,i}$ whose projections have been visited. Define R^2 analogously for $v^2_{k,i}$.
  Since f_k has not failed, $f_k(p^1_{k,i}) = p^2_{k,i}$ and H^1_k, H^2_k are isomorphic. Let $r := |R^1| = |R^2|$. For each $a \in R^1$, there is a unique $b \in R^2$ with $f_k(p^1_H(a)) = p^2_H(b)$.
  The next step is performed as follows:
  - **With probability $\frac{r}{d}$:** Choose $v^1_{k,i+1}$ uniformly at random from R^1. Then, deterministically set $v^2_{k,i+1}$ to be the unique vertex in R^2 satisfying $p^2_H(v^2_{k,i+1}) = f_k(p^1_H(v^1_{k,i+1}))$.
  - **With probability $\frac{d-r}{d}$:** Choose $v^1_{k,i+1}$ and $v^2_{k,i+1}$ independently and uniformly at random from the remaining $(d-r)$ expander neighbors of $v^1_{k,i}$ and $v^2_{k,i}$, respectively.
    If $p^1_H(v^1_{k,i+1}) \notin V^1_k$ and $p^2_H(v^2_{k,i+1}) \notin V^2_k$, then update $f_k(p^1_H(v^1_{k,i+1})) := p^2_H(v^2_{k,i+1})$; otherwise, say f_k fails.

ii. If $L_1(v^1_{k,i+1}) = \mathbf{NULL}$, set $L_1(v^1_{k,i+1}) \leftarrow \Pi(j)$. Likewise, if $L_2(v^2_{k,i+1}) = \mathbf{NULL}$, set $L_2(v^2_{k,i+1}) \leftarrow \Pi(j)$. Update f_k, and increment $j \leftarrow j+1$.

3. Return

$$S^1 = \{\{L_1(v^1_{1,0}), \ldots, L_1(v^1_{1,T})\}, \ldots, \{L_1(v^1_{m,0}), \ldots, L_1(v^1_{m,T})\}\},$$
$$S^2 = \{\{L_2(v^2_{1,0}), \ldots, L_2(v^2_{1,T})\}, \ldots, \{L_2(v^2_{m,0}), \ldots, L_2(v^2_{m,T})\}\}.$$

Since the next vertex is chosen uniformly at random from the neighborhood of current vertex at every step, the above procedure is a coupling over random walk transcripts. Next we argue that $S^1 = S^2$ with high probability. To achieve this, we define the following good events and show that $S^1 = S^2$ if these events occur simultaneously.

Event 1: For all $k \in [m]$, H^1_k and H^2_k remain trees.
Event 2: For all $k \in [m]$ and depth $x \neq y$, $H^1_{k,x} \cap H^1_{k,y} = \emptyset$ and $H^2_{k,x} \cap H^2_{k,y} = \emptyset$ if $\ell \mid (x - y)$.
Event 3: For all $j \neq k$, $H^1_j \cap H^1_k = \emptyset$ and $H^2_j \cap H^2_k = \emptyset$.

In the following, we assume that **Events 1, 2, 3** hold. To show that $S^1 = S^2$, it suffices to establish the following *identical labeling* property: at each step, the labels of the destination vertices $L_1(v^1_{k,i+1})$ and $L_2(v^2_{k,i+1})$ are either both $\mathbf{NULL}$ or are both defined and equal. We begin by stating a few auxiliary results.

The mapping f_k does not fail. Observe that f_k can fail only after an expander step (say, on G_1), and only if this step reaches a vertex $v^1_{k,i+1} \notin R^1$ whose projection $p^1_{k,i+1}$ has already been visited by the projected walk. Since $v^1_{k,i+1} \notin R^1$, the projected edge $(p^1_{k,i}, p^1_{k,i+1})$ has not been visited before. After taking this step, the edge $(p^1_{k,i}, p^1_{k,i+1})$ is added to H^1_k. As H^1_k is already a connected tree prior to adding this edge, this creates a cycle in H^1_k, contradicting **Event 1**.

Identical labeling. To show the identical labeling property, we use the following claim.

Claim 23 *For the k-th random walk, we have $v^1_{k,i} = v^1_{k,j}$ if and only if $v^2_{k,i} = v^2_{k,j}$.*

The claim, together with **Event 3** (the transcripts of different random walks are disjoint), implies the identical labeling property. It therefore suffices to prove the claim.
Proof [of Claim 23] We prove one direction. Suppose $v^1_{k,i} = v^1_{k,j}$; the other direction is symmetric. Let the depths of $v^1_{k,i}, v^2_{k,i}$ be x, and those of $v^1_{k,j}, v^2_{k,j}$ be y. Since $v^1_{k,i} = v^1_{k,j}$, we have $p^1_{k,i} = p^1_{k,j}$. By the invariant ($\bigstar$) $f_k(p^1_{k,i}) = p^2_{k,i}$, we have

$$p^2_{k,i} = f_k(p^1_{k,i}) = f_k(p^1_{k,j}) = p^2_{k,j}.$$

Now $p^2_{k,i}$ is visited at depth x, while $p^2_{k,j}$ is visited at depth y, implying $H^2_{k,x}$ and $H^2_{k,y}$ intersect. By **Event 2**, this can occur only when $x = y$. Hence $v^2_{k,i}$ and $v^2_{k,j}$ lie in the same H-copy and have the same projection, implying $v^2_{k,i} = v^2_{k,j}$. ■

Finally, we show that all the aforementioned good events occur simultaneously w.p. $1 - \frac{m^2T^3}{n^c}$ and we defer the corresponding analysis and the remaining proof of Lemma 22 to Appendix B.

6. Putting things together: Proof of Theorem 1

In this section, we complete the proof of our main theorem. We first record a standard two-point reduction showing that transcript indistinguishability, together with a separation in spectral density, implies a lower bound for SDE.

Lemma 24 (Transcript indistinguishability implies SDE hardness) *Let G_1, G_2 be two labeled graphs on the same vertex set $[N]$, and suppose that*

$$W_1\big(\rho_{N_{G_1}}, \rho_{N_{G_2}}\big) > 2\varepsilon.$$

Let $\mathcal{D}_{G_i}$ denote the probability distribution over transcripts generated by performing the walks on G_i, for $i = 1, 2$. If

$$d_{\mathrm{TV}}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2}) < \frac{1}{2},$$

then no algorithm which receives a transcript from $\mathcal{D}_{G_i}$ can output an ε-approximation to $\rho_{N_{G_i}}$ with success probability at least $3/4$ for both $i = 1$ and $i = 2$.

More generally, if $d_{\mathrm{TV}}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2}) \le \tau$, then every such algorithm has success probability at most $1/2 + \tau/2$ on at least one of the two inputs.

We defer the proof of the above lemma to Section C.1. We will apply Lemma 24 to the two hard transcript distributions induced by randomly labeled versions of G_1 and G_2. Since relabeling does not change the spectral density, the same two-point argument applies to these mixture transcript distributions. By Yao's minimax principle, a distributional lower bound for these hard input distributions rules out any randomized algorithm that succeeds on every labeled graph in the corresponding worst-case input class.

Proof [Proof of Theorem 1] For the hard instances in Definition 12, fix any constant $d \ge 7$. For any sufficiently small constant $\varepsilon > 0$, choose ℓ and n such that $2\varepsilon = \frac{1}{20\ell^6 d^2}$ and $n = 2^{\ell \log d/10} \ge \ell^7 d^2$. Let G_1 and G_2 denote the resulting pair of graphs. Substituting $\ell = (40\varepsilon d^2)^{-1/6}$ and rearranging constants, we obtain

$$n = 2^{\ell \log d/10} \ge 2^{c_1/\varepsilon^{1/6}},$$

for some constant $c_1 > 0$.

On the one hand, by Corollary 15, we have $W_1(\rho_{N_{G_1}}, \rho_{N_{G_2}}) \ge \frac{1}{10\ell^6 d^2} > 2\varepsilon$. Therefore, by Lemma 24, any algorithm that outputs an ε-approximation to the spectral density with probability at least $3/4$ on both inputs must be able to distinguish the corresponding transcript distributions.

On the other hand, Lemma 22 implies that the total variation distance between the corresponding transcript distributions satisfies $d_{\mathrm{TV}}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2}) \le \frac{m^2 T^3}{n^c}$.

Consequently, by Lemma 24, given access only to the transcripts of m non-adaptive random walks of length T, no algorithm can distinguish G_1 from G_2 with probability at least $3/4$, whenever $m \cdot T \le 0.01\, n^{c/3}$. Since

$$0.01\, n^{c/3} \ge 2^{c_2/\varepsilon^{1/6}}$$

for an appropriate constant $c_2 < c_1 c/3$, this completes the proof. ■

Acknowledgments

This work is supported in part by NSFC Grant 62272431 and Quantum Science and Technology - National Science and Technology Major Project (Grant No. 2021ZD0302901).

References

Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time. *Foundations of Computational Mathematics*, 23(6):1959–2047, 2023.

Roland Bauerschmidt, Jiaoyang Huang, and Horng-Tzer Yau. Local kesten–mckay law for random regular graphs. *Communications in Mathematical Physics*, 369(2):523–636, 2019.

Rajarshi Bhattacharjee, Rajesh Jayaram, Cameron Musco, Christopher Musco, and Archan Ray. Improved spectral density estimation via explicit and implicit deflation. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2025.

Vladimir Braverman, Aditya Krishnan, and Christopher Musco. Sublinear time spectral density estimation. In *Proceedings of the 54th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1144–1157, 2022.

Tyler Chen, Thomas Trogdon, and Shashanka Ubaru. Analysis of stochastic lanczos quadrature for spectrum approximation. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021.

David Cohen-Steiner, Weihao Kong, Christian Sohler, and Gregory Valiant. Approximating the spectrum of a graph. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1263–1271, 2018.

Kun Dong, Austin R Benson, and David Bindel. Network density of states. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1152–1161, 2019.

Nicole Eikmeier and David F. Gleich. Revisiting power-law distributions in spectra of real world networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, page 817–826, 2017.

Illés J Farkas, Imre Derényi, Albert-László Barabási, and Tamas Vicsek. Spectra of "real-world" graphs: Beyond the semicircle law. *Physical Review E*, 64(2):026704, 2001.

Grzegorz Gluch, Michael Kapralov, Silvio Lattanzi, Aida Mousavifar, and Christian Sohler. Spectral clustering oracles in sublinear time. In *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1598–1617, 2021.

Oded Goldreich and Dana Ron. *On Testing Expansion in Bounded-Degree Graphs*, pages 68–75. Springer Berlin Heidelberg, 2011.

Jiaoyang Huang, Theo McKenzie, and Horng-Tzer Yau. Optimal eigenvalue rigidity of random regular graphs, 2024. URL https://arxiv.org/abs/2405.12161.

Yujia Jin, Christopher Musco, Aaron Sidford, and Apoorv Vikram Singh. Moments, random walks, and limits for spectrum approximation. In *Proceedings of the 36th Annual Conference on Computational Learning Theory (COLT)*, pages 5373–5394, 2023.

Yujia Jin, Ishani Karmarkar, Christopher Musco, Aaron Sidford, and Apoorv Vikram Singh. Faster spectral density estimation and sparsification in the nuclear norm. In *Proceedings of the 37th Annual Conference on Computational Learning Theory (COLT)*, 2024.

Ruipeng Li, Yuanzhe Xi, Lucas Erlandson, and Yousef Saad. The eigenvalues slicing library (evsl): Algorithms, implementation, and software. *SIAM Journal on Scientific Computing*, 2019.

Lin Lin, Yousef Saad, and Chao Yang. Approximating spectral densities of large matrices. *SIAM Review*, 2016.

Michael Mahoney and Charles Martin. Traditional and heavy tailed self regularization in neural network models. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 2019.

Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

Dean Moldovan, Misa Andelkovic, and Francois Peeters. pybinding v0.9.4: a python package for tight-binding calculations. *Zenodo*, 2020.

Cameron Musco, Christopher Musco, Lucas Rosenblatt, and Apoorv Vikram Singh. Sharper bounds for chebyshev moment matching, with applications. In *The Thirty Eighth Annual Conference on Learning Theory*, pages 4309–4358. PMLR, 2025.

Victor M. Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Application*, 6(1):405–431, March 2019. ISSN 2326-831X.

Beresford N Parlett. *The symmetric eigenvalue problem*. SIAM, 1998.

Pan Peng. Robust clustering oracle and local reconstructor of cluster structure of graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2953–2972. SIAM, 2020.

Ronitt Rubinfeld and Asaf Shapira. Sublinear time algorithms. *SIAM Journal on Discrete Mathematics*, 25(4):1562–1588, 2011.

Filippo Santambrogio. Optimal transport for applied mathematicians. 2015.

R.N. Silver and H. Röder. Densities of states of mega-dimensional hamiltonian matrices. *International Journal of Modern Physics C*, 5(4), 1994.

John Skilling. *The Eigenvalues of Mega-dimensional Matrices*, pages 455–466. Springer Netherlands, 1989.

Daniel Spielman. *Spectral and Algebraic Graph Theory*. Unpublished Manuscript, 2019.

Appendix A. Deferred proofs from Section 2 and Section 4

A.1. Proof of Fact 9

By definition, $\rho_{A_G}(x) = \frac{1}{n}\sum_{i=1}^{n}\delta(x-\lambda_i)$ and $\rho_{A_H}(x) = \frac{1}{m}\sum_{i=1}^{m}\delta(x-\mu_i)$. So the following holds for any $z \in \mathbb{R}$:

$$\begin{aligned}
\rho_{A_G} * \rho_{A_H}(z) &= \int_{\mathbb{R}} \rho_{A_G}(x)\cdot\rho_{A_H}(z-x)dx \\
&= \int_{\mathbb{R}} \left(\frac{1}{n}\sum_{i=1}^{n}\delta(x-\lambda_i)\right)\cdot\left(\frac{1}{m}\sum_{j=1}^{m}\delta(z-x-\mu_j)\right)dx \\
&= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\int_{\mathbb{R}}\delta(x-\lambda_i)\delta(z-x-\mu_j)dx \\
&= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\delta(z-\lambda_i-\mu_j).
\end{aligned}$$

By Fact 7, the last equality is exactly the spectral density function $\rho_{A_{G\square H}}(z)$. ■

A.2. Proof of Lemma 11

We will make use of the following eigenvalue concentration result.

Lemma 25 (Theorem 1.1 in (Huang et al., 2024)) *Fix d as constant and n sufficiently large, there is a constant $\alpha_d \geq 1$, depending only on d, such that with probability $1-n^{-\alpha_d}$, the eigenvalues $\{\lambda_i\}_{1\leq i\leq n}$ of the adjacency matrix of a random d-regular graph satisfy*

$$|\lambda_i - \gamma_i| \leq \sqrt{d-1}\cdot n^{-2/3+o(1)}(\min\{i, n-i+1\})^{-1/3}$$

for every $2 \leq i \leq n$, where γ_i satisfies $\int_{\gamma_i}^{2\sqrt{d-1}}\rho_d(x)dx = \frac{i-1/2}{n-1}$ for $2 \leq i \leq n$.

Remark 26 *Compared to the original statement in (Huang et al., 2024), we formulate the result for the adjacency matrix rather than their normalized matrix $A/\sqrt{d-1}$.*

Next we use Lemma 25 to prove Lemma 11. For notational convenience, let $\rho_\lambda = \rho_{A_H} = \frac{1}{n}\sum_{i=1}^{n}\delta(x-\lambda_i)$. Define

$$\rho_\gamma = \frac{1}{n}\Big(\delta(x-\lambda_1) + \sum_{i=2}^{n}\delta(x-\gamma_i)\Big).$$

Let F_λ, F_γ, and F_d denote the CDFs of ρ_λ, ρ_γ, and ρ_d, respectively. By definition,

$$F_d(\gamma_i) = \int_{-2\sqrt{d-1}}^{\gamma_i}\rho_d(x)dx = 1 - \frac{i-1/2}{n-1}.$$

$$F_\gamma(\gamma_i) = \frac{1}{n}|\text{number of eigenvalues } \leq \gamma_i| = 1 - \frac{i-1}{n}.$$

Thus, $|F_d(\gamma_i) - F_\gamma(\gamma_i)| \le 1/n$ for $2 \le i \le n$. Additionally, for $x \in (\gamma_i, \gamma_{i-1})$, we have $F_\gamma(x) = F_\gamma(\gamma_i)$ and $F_d(\gamma_i) \le F_d(x) \le F_d(\gamma_{i-1}) = F_d(\gamma_i) + \frac{1}{n}$. Thus,

$$|F_d(x) - F_\gamma(x)| \le |F_d(\gamma_i) - F_\gamma(\gamma_i)| + \frac{1}{n} \le \frac{2}{n},$$

which holds for all $x \in [-2\sqrt{d-1}, 2\sqrt{d-1}]$. By the definition of W_1 distance,

$$W_1(\rho_d, \rho_\gamma) = \int_{\mathbb{R}} |F_d(x) - F_\gamma(x)| dx \le 4\sqrt{d-1} \cdot \frac{2}{n} + \frac{d}{n}.$$

Next, by Lemma 25,

$$W_1(\rho_\gamma, \rho_\lambda) = \frac{1}{n}\sum_{i=2}^{n} |\lambda_i - \gamma_i| \le \frac{\sqrt{d-1}}{n} n^{-2/3+o(1)} \cdot 2 \sum_{i=2}^{\lceil n/2 \rceil} i^{-1/3} \le 2\sqrt{d-1} \cdot n^{-1+o(1)},$$

holding with probability $1 - n^{-\alpha}$.

Combining the two bounds gives

$$W_1(\rho_d, \rho_\lambda) \le W_1(\rho_d, \rho_\gamma) + W_1(\rho_\gamma, \rho_\lambda) \le 3\sqrt{d-1} \cdot n^{-1+o(1)},$$

which completes the proof. ■

A.3. Proof of Lemma 16

By Fact 9, we have

$$\rho_{A_{G_1}} = \rho_{A_H} * \rho_{A_{C_1}}; \rho_{A_{G_2}} = \rho_{A_H} * \rho_{A_{C_2}}.$$

Thus,

$$W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}}) = W_1(\rho_{A_H} * \rho_{A_{C_1}}, \rho_{A_H} * \rho_{A_{C_2}})$$

We first control the effect of replacing ρ_{A_H} by its limit ρ_d. By Lemma 11 and Fact 9, we have

$$W_1(\rho_{A_H} * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_1}}) \le W_1(\rho_{A_H}, \rho_d) \le n^{-0.99}; \tag{1}$$

$$W_1(\rho_{A_H} * \rho_{A_{C_2}}, \rho_d * \rho_{A_{C_2}}) \le W_1(\rho_{A_H}, \rho_d) \le n^{-0.99}. \tag{2}$$

Here, the first inequalities in (1) and (2) follow from (Santambrogio, 2015, Lemma 5.2)[2].

By the triangle inequality, we have that

$$\begin{aligned} W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}) \le W_1(\rho_d * \rho_{A_{C_1}}, \rho_{A_H} * \rho_{A_{C_1}}) &+ W_1(\rho_{A_H} * \rho_{A_{C_1}}, \rho_{A_H} * \rho_{A_{C_2}}) \\ &+ W_1(\rho_{A_H} * \rho_{A_{C_2}}, \rho_d * \rho_{A_{C_2}}). \end{aligned}$$

Rearranging, we have

$$\begin{aligned} & W_1(\rho_{A_H} * \rho_{A_{C_1}}, \rho_{A_H} * \rho_{A_{C_2}}) \\ \ge & W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}) - W_1(\rho_d * \rho_{A_{C_1}}, \rho_{A_H} * \rho_{A_{C_1}}) - W_1(\rho_{A_H} * \rho_{A_{C_2}}, \rho_d * \rho_{A_{C_2}}) \\ \ge & W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}) - O\left(n^{-0.99}\right). \end{aligned}$$

■

2. Note that Item 1 of (Santambrogio, 2015, Lemma 5.2) does not require the regularizing kernel assumption; only Item 2 relies on that assumption.

A.4. Proof of Lemma 17

Let $F_d(x) = \int_{-\infty}^{x} \rho_d(s)ds$ be the CDF of ρ_d. In the following, we let

$$\{\lambda_i\}_{0\le i\le 2\ell-1} = \left\{2\cdot\cos\left(\frac{2\pi i}{\ell}\right)\right\}_{0\le i\le 2\ell-1}, \{r_i\}_{0\le i\le 2\ell-1} = \left\{2\cdot\cos\left(\frac{\pi i}{\ell}\right)\right\}_{0\le i\le 2\ell-1}$$

be the eigenvalues of A_{C_1} and of A_{C_2}, respectively. By the definition of convolution, for any $x\in\mathbb{R}$,

$$\rho_d * \rho_{A_{C_1}}(x) = \frac{1}{2\ell}\sum_{i=0}^{2\ell-1}\rho_d(x-\lambda_i)$$

$$\rho_d * \rho_{A_{C_2}}(x) = \frac{1}{2\ell}\sum_{i=0}^{2\ell-1}\rho_d(x-r_i)$$

Let F_1, F_2 be the CDFs of $\rho_d*\rho_{A_{C_1}}$ and $\rho_d*\rho_{A_{C_2}}$ respectively. By definition, $F_1(x) = \frac{1}{2\ell}\sum_{i=0}^{2\ell-1}F_d(x-\lambda_i)$ and $F_2(x) = \frac{1}{2\ell}\sum_{i=0}^{2\ell-1}F_d(x-r_i)$ for any $x\in\mathbb{R}$. Therefore,

$$\begin{aligned} W_1(\rho_d*\rho_{A_{C_1}}, \rho_d*\rho_{A_{C_2}}) &= \int_{\mathbb{R}} |F_1(t)-F_2(t)|\,dt \\ &= \int_{\mathbb{R}} \frac{1}{2\ell}\left|\sum_{i=0}^{2\ell-1} F_d(t-\lambda_i) - F_d(t-r_i)\right| dt. \end{aligned} \tag{3}$$

Next, we use the fact that $F_d(x) = 0$ for all $x \le -2\sqrt{d-1}$. Consider

$$t \le -2\sqrt{d-1} - 2\cos\left(\frac{\pi}{\ell}\right).$$

For any r_i, this implies

$$t - r_i \le -2\sqrt{d-1} - 2\cos\left(\frac{\pi}{\ell}\right) - r_i.$$

In particular, if $r_i \ge -2\cos\left(\frac{\pi}{\ell}\right)$, then $t - r_i \le -2\sqrt{d-1}$, and hence $F_d(t-r_i) = 0$. An identical argument shows that $F_d(t-\lambda_i) = 0$ whenever $\lambda_i \ge -2\cos\left(\frac{\pi}{\ell}\right)$.

Without loss of generality, assume that ℓ is even. Using the identities

$$\lambda_i = 2\cos\left(\frac{2\pi i}{\ell}\right), \qquad r_i = 2\cos\left(\frac{\pi i}{\ell}\right),$$

the condition $\lambda_i, r_i < -2\cos\left(\frac{\pi}{\ell}\right)$ occurs only for

$$\lambda_{\ell/2} = \lambda_{3\ell/2} = r_\ell = 2\cos(\pi) = -2.$$

(If ℓ is odd, the condition $\lambda_i, r_i < -2\cos\left(\frac{\pi}{\ell}\right)$ occurs only for $r_\ell = -2$; the argument below remains unchanged.)

Therefore, we have

$$\int_{\mathbb{R}} \frac{1}{2\ell}\left|\sum_{i=0}^{2\ell-1} F_d(t-\lambda_i) - F_d(t-r_i)\right| dt$$

$$
\begin{aligned}
&\geq \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell}\left|\sum_{i=0}^{2\ell-1} F_d(t-\lambda_i) - F_d(t-r_i)\right| dt \\
&= \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell}\left|F_d(t-\lambda_{\ell/2}) + F_d(t-\lambda_{3\ell/2}) - F_d(t-r_l)\right| dt \\
&= \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} F_d(t+2) dt \\
&= \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} \int_{-2\sqrt{d-1}}^{t+2} \rho_d(x) dx dt. \qquad (4)
\end{aligned}
$$

We now use Lemma 31, which states that for $d \geq 7$ the density $\rho_d(t)$ is concave. Let $\alpha \in (0,1)$. By Jensen's inequality,

$$
\begin{aligned}
\rho_d\Big((1-\alpha)(-2\sqrt{d-1}) + \alpha(t+2)\Big) &\geq (1-\alpha)\rho_d(-2\sqrt{d-1}) + \alpha\rho_d(t+2) \\
&= \alpha\rho_d(t+2),
\end{aligned}
$$

where the equality uses the fact that $\rho_d(x) = 0$ for all $x \leq -2\sqrt{d-1}$.

Let

$$s = (1-\alpha)(-2\sqrt{d-1}) + \alpha(t+2).$$

Solving for α yields

$$\alpha = \frac{s + 2\sqrt{d-1}}{2\sqrt{d-1} + t + 2}.$$

Substituting this expression back gives

$$\rho_d(s) \;\geq\; \frac{s + 2\sqrt{d-1}}{2\sqrt{d-1} + t + 2} \cdot \rho_d(t+2).$$

Thus, we have

$$
\begin{aligned}
\int_{-2\sqrt{d-1}}^{t+2} \rho_d(s) ds &\geq \int_{-2\sqrt{d-1}}^{t+2} \frac{s + 2\sqrt{d-1}}{2\sqrt{d-1} + t + 2} \rho_d(t+2) ds \\
&= \frac{(t+2)^2 - 4(d-1)}{2\left(2\sqrt{d-1} + t + 2\right)} \cdot \rho_d(t+2) + \frac{2\sqrt{d-1} \cdot (t + 2 + 2\sqrt{d-1})}{2\sqrt{d-1} + t + 2} \cdot \rho_d(t+2) \\
&= \frac{(t+2)^2 + 4\sqrt{d-1}(t+2) + 4(d-1)}{2\left(2\sqrt{d-1} + t + 2\right)} \cdot \rho_d(t+2). \qquad (5)
\end{aligned}
$$

Therefore, by Eq. (3), (4) and (5), we have

$$W_1(\rho_d * \rho_{A_{C_1}}, \rho_d * \rho_{A_{C_2}}) \geq \int_{-2\sqrt{d-1}-2}^{-2\sqrt{d-1}-2\cos\left(\frac{\pi}{\ell}\right)} \frac{1}{2\ell} \frac{(t+2)^2 + 4\sqrt{d-1}(t+2) + 4(d-1)}{2\left(2\sqrt{d-1} + t + 2\right)} \cdot \rho_d(t+2) dt.$$

■

A.5. Proof of Lemma 18

Notice that

$$\rho_d(x) = \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)} \geq \frac{1}{2\pi d}\sqrt{4(d-1)-x^2} \quad \text{for } x \in \left(-2\sqrt{d-1},\, -1.99\sqrt{d-1}\right).$$

Set $\lambda = x + 2\sqrt{d-1}$, so that $\lambda \in (0, 0.01\sqrt{d-1})$. Then

$$\begin{aligned}\rho_d(x) &\geq \frac{1}{2\pi d}\sqrt{4(d-1)-x^2} = \frac{1}{2\pi d}\sqrt{4(d-1)-(\lambda-2\sqrt{d-1})^2} \\ &= \frac{1}{2\pi d}\sqrt{4\sqrt{d-1}\,\lambda-\lambda^2} \geq \frac{1}{2\pi d}\sqrt{3.99\sqrt{d-1}\,\lambda} \geq \frac{\sqrt{\lambda}}{2\pi d^{3/4}}.\end{aligned}$$

This completes the proof. ■

Appendix B. Lower-bounding the probabilities of good events in Section 5

We restate the good events defined in Section 5 for convenience, to complete the proof of Lemma 22.

Event 1: For all $k \in [m]$, H_k^1 and H_k^2 remain trees.
Event 2: For all $k \in [m]$ and depth $x \neq y$, $H_{k,x}^1 \cap H_{k,y}^1 = \emptyset$ and $H_{k,x}^2 \cap H_{k,y}^2 = \emptyset$ if $\ell \mid (x-y)$.
Event 3: For all $j \neq k$, $H_j^1 \cap H_k^1 = \emptyset$ and $H_j^2 \cap H_k^2 = \emptyset$.

To this end, we apply the following lemma for random d-regular graph.

Lemma 27 (Proposition 4.1 in Bauerschmidt et al. (2019)) *Let $\delta > 0$ and $\omega \geq 1$ be an integer. There is $0 < c < \delta/(2\omega+2)$ such that, if $L \leq c \log_{d-1} n$, then the following holds for a uniformly chosen random d-regular G on vertex set $[n]$ for sufficiently large n, with probability at least $1 - o(n^{-\omega+\delta})$.*

- *Most L-neighborhoods are trees:*

$$|\{v \in [n] \mid \textit{the subgraph } N_L(v,G) \textit{ contains a cycle}\}| \leq n^{\delta}.$$

Here, $N_L(v,G)$ is the subgraph of G induced by L-hop neighborhood rooted at v.

Bound on Event 1. Consider the i-th projected walk on H, say $W_i = (v_0, v_1, \cdots, v_T)$.

Claim 28 $\Pr[W_i \textit{ contains a cycle}] \leq \frac{T^2}{n^c}$, *where $c > 0$ is a small universal constant.*

Proof We define $S := \{v \mid N_L(v,H) \text{ is not a } d\text{-ray tree}\}$, for $L = 0.01 \cdot \log n / \log d$. Lemma 27 tells that $|S| \leq n^{0.1}$ w.p. $\geq 1 - o(\frac{1}{n})$ over randomness of H, by setting $\delta = 0.1$ and $\omega = 2$. We conditioned on H s.t. $|S| \leq n^{0.1}$.

Define event $\mathcal{E} := \{W_i \text{ hits at least one vertex in } S\}$. Since $\Pr[v_i = u] = \frac{1}{n}$ for every i and $u \in V_H$, a direct argument using Markov's inequality tells that

$$\Pr[\mathcal{E}] \leq T \cdot n^{-0.9},$$

hence we have

$$\Pr[W_i \text{ contains a cycle}]$$

$$\leq \Pr[W_i \text{ contains a cycle of length} \leq L] + \Pr[W_i \text{ only contains cycle of length} \geq L+1]$$
$$\leq \Pr[\mathcal{E}] + \Pr[W_i \text{ only contains cycle of length} \geq L+1]$$
$$\leq T \cdot n^{-0.9} + \Pr[W_i \text{ only contains cycle of length} \geq L+1].$$

To bound the latter term, we bound the probability that the $(t+1)$-th step finishes a cycle. Here we only care for $t \in [L, T-1]$, since any cycle is of length $\geq L+1$.

We define $C_t = \{v_0, v_1, \ldots, v_{t-L}\}$ to be vertices visited during the first $t-L$ steps of W_i. We denote by F_{t+1} the event that the $(t+1)$-th step of W_i finishes a cycle of length $\geq L+1$, which occurs only when $v_{t+1} \in C_t$. Let $\mathbf{1}_{F_{t+1}}$ be the indicator of F_{t+1}, then for any fixed C_t, we have

$$\begin{aligned}
\mathbf{E}[\mathbf{1}_{F_{t+1}} \mid C_t] &\leq \sum_{u \in C_t} \Pr[v_{t+1} = u \mid C_t] \\
&\leq \sum_{u \in C_t} e_{v_{t-L}} \cdot P^{L+1} \cdot e_u \\
&\leq \sum_{u \in C_t} (\frac{1}{n} + \lambda^{L+1}) \\
&\leq T \cdot n^{-c}
\end{aligned}$$

The last inequality follows from the fact that H is a (near) Ramanujan graph (i.e. $\lambda = \max\{\lambda_2, |\lambda_n|\} = \Theta(1/\sqrt{d})$) and $L = \Theta(\log_d n)$. Furthermore, we have

$$\mathbf{E}[\mathbf{1}_{F_{t+1}}] = \mathop{\mathbf{E}}_{C_t}[\mathbf{E}[\mathbf{1}_{F_{t+1}} \mid C_t]] \leq T \cdot n^{-c}.$$

Hence

$$\begin{aligned}
&\Pr[W_i \text{ only contains cycle of length} \geq L+1] \\
\leq &\sum_{t=L}^{T-1} \mathbf{E}[F_{t+1}] \\
\leq &T^2 \cdot n^{-c}.
\end{aligned}$$

This finishes the proof. ■

By a union bound over all m projected walks, we have

$$\Pr[\text{Event 1}] \geq 1 - \frac{mT^2}{n^c}.$$

Bound on Event 2. We consider the T-step lazy random walk on H. We denote by v_i the vertex visited at i-th step, and x_i the depth. Here, the notion *depth* is defined analogously in the coupling process. We define event $\mathcal{E}_k$ characterizing the following bad event:

$$\mathcal{E}_k\text{: } \exists s < t \text{ s.t. } x_t - x_s = k\ell \text{ and } v_s = v_t.$$

In other words, the event that we visit the same node after traversing a cycle of length ℓ for the k-th time. Note that $1 \leq |k| \leq \lfloor \frac{T}{\ell} \rfloor$.

Claim 29 $\Pr[\mathcal{E}_k] \leq O(\frac{T^2}{n})$ *if* $\ell \geq 10 \log_d n$.

Proof

$$\begin{aligned}
\Pr[\mathcal{E}_k] &\leq \sum_{\substack{s<t \\ s,t\in[T]}} \Pr[v_s = v_t, x_t - x_s = k\ell] \\
&\leq \sum_{u\in V_H} \sum_{\substack{s<t \\ s,t\in[T]}} \Pr[v_s = v_t = u, x_t - x_s = k\ell] \\
&= \sum_{u\in V_H} \sum_{\substack{s<t \\ s,t\in[T]}} \Pr[v_t = u, x_t - x_s = k\ell \mid v_s = u] \cdot \Pr[v_s = u] \\
&= \frac{1}{n} \sum_{u\in V_H} \sum_{\substack{s<t \\ s,t\in[T]}} \Pr[v_t = u, x_t - x_s = k\ell \mid v_s = u] \\
&\leq \frac{1}{n} \sum_{u\in V_H} \sum_{\substack{t-s\geq|k|\ell \\ s,t\in[T]}} \Pr[v_t = u \mid v_s = u] \\
&= \frac{1}{n} \sum_{u\in V_H} \sum_{\substack{t-s\geq|k|\ell \\ s,t\in[T]}} e_u^T P^{t-s} e_u
\end{aligned}$$

The fourth equality is due to the fact that we are in the non-adaptive random walk setting and thus the node we visit at the s step is uniformly random. As discussed above, $e_u^T P^{t-s} e_u \leq \frac{1}{n} + \lambda^{t-s}$, hence

$$\begin{aligned}
\Pr[\mathcal{E}_k] &\leq \frac{1}{n} \sum_{u\in V_H} \sum_{\substack{t-s\geq|k|\ell \\ s,t\in[T]}} e_u^T P^{t-s} e_u \\
&\leq \frac{1}{n} \sum_{u\in V_H} \sum_{\substack{t-s\geq|k|\ell \\ s,t\in[T]}} (\frac{1}{n} + \lambda^{t-s}) \\
&\leq \sum_{\substack{t-s\geq|k|\ell \\ s,t\in[T]}} (\frac{1}{n} + \lambda^{t-s}) \\
&\leq \frac{T^2}{n} + T \cdot \sum_{T\geq t'\geq|k|\ell} \lambda^{t'} \\
&\leq \frac{T^2}{n} + O(T\lambda^{|k|\ell}) \\
&\leq \frac{2T^2}{n}.
\end{aligned}$$

The last equation follows from $\lambda \approx 1/\sqrt{d}$, $\ell \geq 10 \log_d n$ and $k \geq 1$. ■

A union bound over all k and m independent walks gives us the following

$$\Pr[\text{Event 2}] \geq 1 - \frac{2mT^3}{n}.$$

Bound on Event 3. Let S be the set of vertices on H visited by first k projected walks $W_1, W_2, \cdots, W_k$. We consider the probability that the $(k+1)$-th projected walk W_{k+1} visits S.

Claim 30 $\Pr[W_{k+1}$ *visits at least one vertex in* $S] \leq \frac{T}{n} \cdot |S|$.

Proof Since W_{k+1} starts from a stationary distribution, we have $\Pr[v_i = u] = \frac{1}{n}$ for every i and $u \in V_H$, where v_i is the vertex visited at the i-th step of W_{k+1}. Hence, the expected number of visits to S is $\frac{|S|}{n}|T|$. A straightforward argument using Markov's inequality finishes the proof. ■

Since $|S| \leq mT$, we have

$$\Pr[\text{Event 3}] \geq 1 - m \cdot \frac{T}{n} \cdot mT = 1 - \frac{m^2T^2}{n},$$

by a union bound over all m projected walks.

Putting together. **Proof [of Lemma 22]** By a union bound over Event 1, 2 and 3, we see that the output of coupling process remains same with probability at least

$$1 - \frac{mT^2}{n^c} - \frac{2mT^3}{n} - \frac{m^2T^2}{n} \geq 1 - \frac{m^2T^3}{n^c}.$$

This finishes the proof. ■

Appendix C. Deferred proofs from Section 6

C.1. Proof of Lemma 24

Suppose, for contradiction, that an algorithm A succeeds with probability at least $3/4$ on both G_1 and G_2. We construct a distinguisher D between $\mathcal{D}_{G_1}$ and $\mathcal{D}_{G_2}$. Given a transcript $\mathcal{S}$, run A on $\mathcal{S}$, obtaining $\widehat{\rho}$. Output 1 if

$$W_1(\widehat{\rho}, \rho_{N_{G_1}}) < W_1(\widehat{\rho}, \rho_{N_{G_2}}),$$

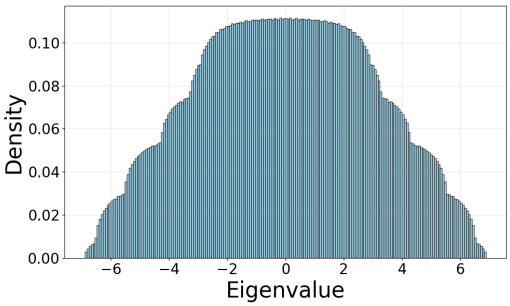
and output 2 otherwise.

If $\mathcal{S} \sim \mathcal{D}_{G_1}$ and $W_1(\widehat{\rho}, \rho_{N_{G_1}}) \leq \varepsilon$, then by the triangle inequality and the assumption that $W_1(\rho_{N_{G_1}}, \rho_{N_{G_2}}) > 2\varepsilon$, we have

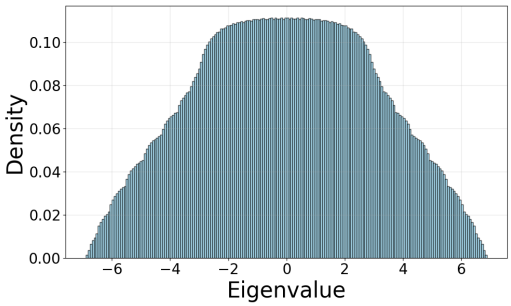
$$W_1(\widehat{\rho}, \rho_{N_{G_2}}) > \varepsilon \geq W_1(\widehat{\rho}, \rho_{N_{G_1}}),$$

so D outputs 1. Thus D is correct with probability at least $3/4$ when the input comes from $\mathcal{D}_{G_1}$. The same argument shows that D is correct with probability at least $3/4$ when the input comes from $\mathcal{D}_{G_2}$. Hence D distinguishes $\mathcal{D}_{G_1}$ and $\mathcal{D}_{G_2}$ with advantage at least $1/4$, which implies $d_{\mathrm{TV}}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2}) \geq 1/2$, a contradiction.

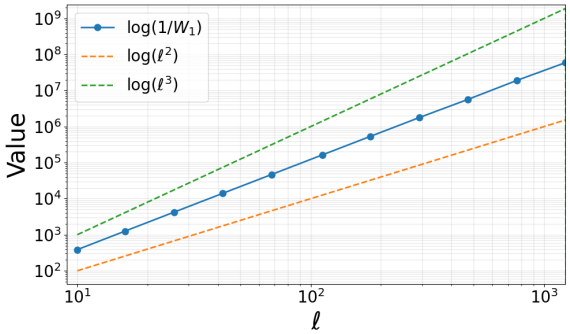
The quantitative version follows from the standard fact that the maximum distinguishing success probability between two distributions $\mathcal{D}_{G_1}, \mathcal{D}_{G_2}$ with equal prior is $1/2 + d_{\mathrm{TV}}(\mathcal{D}_{G_1}, \mathcal{D}_{G_2})/2$. ■



(*a*) Eigenvalues of A_{G_1} when $\ell = 10$



(*b*) Eigenvalues of A_{G_2} when $\ell = 10$



(*c*) $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}})$ vs. poly$\frac{1}{\ell}$ in log scales

Figure 2: Testing $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}})$

Appendix D. Numerical experiments for measuring $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}})$

Given parameters n, d, and ℓ, the eigenvalues of G_1 and G_2 can be computed exactly from our construction. In this section, we study the W_1 distance between $\rho_{A_{G_1}}$ and $\rho_{A_{G_2}}$ by varying ℓ. Specifically, we fix $n = 10{,}000$ and $d = 7$, and let ℓ range from 10 to 300. Figures 2 (a) and (b) show the eigenvalues of G_1 and G_2 for $\ell = 10$, while Figure 2 (c) plots $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}})$ against $1/\ell^2$ on a logarithmic scale as ℓ varies.

The results indicate that $W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}})$ is experimentally between $1/\ell^2$ and $1/\ell^3$, suggesting that the theoretical lower bound could ideally be improved. Consequently, the exponential lower bound for spectral density estimation could potentially be strengthened using our construction. However, due to current analytical limitations, we can only rigorously establish the weaker bound stated in Theorem 13, namely

$$W_1(\rho_{A_{G_1}}, \rho_{A_{G_2}}) = \Omega(1/\ell^6).$$

Appendix E. Concavity of $\rho_d(x)$

Lemma 31 *Let $d \geq 2$ be an integer. The Kesten-McKay density function*

$$\rho_d(x) = \frac{d\sqrt{4(d-1) - x^2}}{2\pi(d^2 - x^2)}, \quad |x| < 2\sqrt{d-1}$$

is strictly concave on its support if and only if $d \geq 7$.

Proof Let $a^2 = 4(d-1)$, so the support of $\rho_d(x)$ is $|x| < a$. Define

$$g(x) = \sqrt{a^2 - x^2}, \quad h(x) = d^2 - x^2.$$

Then

$$\rho_d(x) = C \cdot \frac{g(x)}{h(x)}, \quad \text{where } C = \frac{d}{2\pi} > 0.$$

Since C is a positive constant, the concavity of ρ_d is equivalent to the concavity of

$$\varphi(x) = \frac{g(x)}{h(x)}.$$

We compute:

$$g'(x) = -\frac{x}{\sqrt{a^2 - x^2}}, \quad h'(x) = -2x.$$

By the quotient rule:

$$\begin{aligned}
\varphi'(x) &= \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2} \\
&= \frac{-\frac{x}{\sqrt{a^2-x^2}}(d^2 - x^2) - \sqrt{a^2 - x^2}(-2x)}{(d^2 - x^2)^2} \\
&= \frac{-x(d^2 - x^2) + 2x(a^2 - x^2)}{(d^2 - x^2)^2\sqrt{a^2 - x^2}}.
\end{aligned}$$

Simplify the numerator:

$$-x(d^2 - x^2) + 2x(a^2 - x^2) = x\left[-d^2 + x^2 + 2a^2 - 2x^2\right] = x\left[2a^2 - d^2 - x^2\right].$$

Define constant

$$B = 2a^2 - d^2 = 8(d-1) - d^2.$$

Then

$$\varphi'(x) = \frac{x\left[B - x^2\right]}{(d^2 - x^2)^2\sqrt{a^2 - x^2}}.$$

Next step we compute the second derivative. Let

$$N(x) = x(B - x^2), \quad D(x) = (d^2 - x^2)^2\sqrt{a^2 - x^2}.$$

Then $\varphi'(x) = N(x)/D(x)$. By the quotient rule:

$$\varphi''(x) = \frac{N'(x)D(x) - N(x)D'(x)}{D(x)^2}.$$

Since $D(x) > 0$ on $|x| < a$, the sign of $\varphi''(x)$ is determined by the numerator:

$$M(x) = N'(x)D(x) - N(x)D'(x).$$

We compute:

$$N'(x) = B - 3x^2.$$

To compute $D'(x)$, take logarithms:

$$\ln D(x) = 2\ln\big(d^2 - x^2\big) + \frac{1}{2}\ln\big(a^2 - x^2\big),$$

$$\frac{D'(x)}{D(x)} = 2 \cdot \frac{-2x}{d^2 - x^2} + \frac{1}{2} \cdot \frac{-2x}{a^2 - x^2} = -\frac{4x}{d^2 - x^2} - \frac{x}{a^2 - x^2}.$$

Thus

$$D'(x) = -D(x)x\left[\frac{4}{d^2 - x^2} + \frac{1}{a^2 - x^2}\right].$$

Substitute into $M(x)$:

$$\begin{aligned} M(x) &= (B - 3x^2)D(x) - x(B - x^2)\left(-D(x)x\left[\frac{4}{d^2 - x^2} + \frac{1}{a^2 - x^2}\right]\right) \\ &= D(x)\left[B - 3x^2 + x^2(B - x^2)\left(\frac{4}{d^2 - x^2} + \frac{1}{a^2 - x^2}\right)\right]. \end{aligned}$$

Therefore,

$$\varphi''(x) = \frac{M(x)}{D(x)^2} = \frac{B - 3x^2 + x^2(B - x^2)\left(\frac{4}{d^2 - x^2} + \frac{1}{a^2 - x^2}\right)}{D(x)}.$$

Since $D(x) > 0$, we have $\varphi''(x) < 0$ if and only if

$$M(x) := B - 3x^2 + x^2(B - x^2)\left(\frac{4}{d^2 - x^2} + \frac{1}{a^2 - x^2}\right) < 0.$$

For $d \geq 7$, we show that $M(x) < 0$ for all $|x| < a$. First, compute B:

$$B = -d^2 + 8d - 8.$$

For $d = 7$, $B = -1$. For $d > 7$, $B < -1$ since B is decreasing for $d > 4$. Now consider the three terms in $M(x)$:

$$\begin{aligned} T_1 &= B - 3x^2, \\ T_2 &= \frac{4x^2(B - x^2)}{d^2 - x^2}, \\ T_3 &= \frac{x^2(B - x^2)}{a^2 - x^2}. \end{aligned}$$

Since $B \leq -1 < 0$ and $x^2 \geq 0$, we have $T_2 \leq 0$ and $T_3 \leq 0$. Thus,

$$M(x) = T_1 + T_2 + T_3 \leq T_1 = B - 3x^2 < 0 \quad \text{for all } |x| < a.$$

Therefore, for $d \geq 7$, $\varphi''(x) < 0$ for all $|x| < a$. Hence $\rho_d(x)$ is strictly concave. Finally we show that when $d \leq 6$, $\rho_d(x)$ is not concave. Actually for $d \leq 6$, we have $B > 0$. In particular, at $x = 0$:

$$M(0) = B > 0,$$

so $\varphi''(0) > 0$, meaning ρ_d is convex near $x = 0$. Therefore, ρ_d cannot be concave on the entire interval. ■