Faster Spectral Density Estimation and Sparsification in the Nuclear Norm

Yujia Jin
Stanford University
yujia@stanford.edu

Ishani Karmarkar
Stanford University
ishanik@stanford.edu

Christopher Musco
New York University
cmusco@nyu.edu

Aaron Sidford
Stanford University
sidford@stanford.edu

Apoorv Vikram Singh
New York University
apoorv.singh@nyu.edu

Abstract

In this paper we consider the problem of estimating the spectral density of the normalized adjacency matrix of an n-node undirected graph. We provide an algorithm that, with $O(n\varepsilon^{-2})$ queries to a degree and neighbor oracle and in $O(n\varepsilon^{-3})$ time, estimates the spectrum up to ε accuracy in the Wasserstein-1 metric. This improves on previous state-of-the-art methods, including an $O(n\varepsilon^{-7})$ time algorithm from [Braverman et al., STOC 2022] and, for sufficiently small ε, a $2^{O(\varepsilon^{-1})}$ time method from [Cohen-Steiner et al., KDD 2018]. To achieve this result, we introduce a new notion of graph sparsification, which we call nuclear sparsification. We provide a deterministic, $O(n\varepsilon^{-2})$-time and $O(n\varepsilon^{-2})$-query algorithm for computing $O(n\varepsilon^{-2})$-sparse nuclear sparsifiers. We show that this bound is optimal in both its sparsity and query complexity, and we further separate our results from the related notion of additive spectral sparsification.

Contents

1 Introduction

In this paper, we study the fundamental problem of estimating the spectrum of an (undirected) graph. This problem has received significant attention due applications in visualizing, classifying, and understanding large networks [FDBV06; EG17; CKSV18; DBB19; CTU21; BKM22]. Concretely, we consider a standard version of the problem based on the graph's normalized adjacency matrix:

Problem 1.1 (Spectral Density Estimation)**.** Given an n-node undirected graph $G = (V, E, w)$ with edge weights $w \in \mathbb{R}^E_{>0}$, output eigenvalue estimates $\widehat{\lambda}_1 \leq \cdots \leq \widehat{\lambda}_n$ such that

$$\frac{1}{n} \sum_{i \in [n]} |\widehat{\lambda}_i - \lambda_i| \leq \varepsilon, \tag{1}$$

where $\lambda_1 \leq \cdots \leq \lambda_n$ are the true eigenvalues of the normalized adjacency matrix of G. Note that the metric above is equivalent to requiring that the Wasserstein-1 distance between the uniform distribution on $\lambda_1, \ldots, \lambda_n$ and the uniform distribution on $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_n$ is less than ε.

Excitingly, at least for unweighted graphs, it has been shown that spectral density estimation can be solved in time that is *sublinear* in the size of G's representation, which can be as large as $\Omega(n^2)$. Concretely, given a "neighbor" oracle that when queried with $a \in V$ and $i \in \mathbb{Z}_{>0}$ outputs the degree of a and the i-th vertex incident to a (if there is one) in $O(1)$ time, there are randomized spectral density estimation algorithms running in $2^{O(\varepsilon^{-1})}$ [CKSV18] and $O(n\varepsilon^{-7})$ time [BKM22].[1] Recent work also studies query complexity lower bounds in various models [JMSS23]. However, large gaps remain: the best query lower bound for spectral density estimation in this model is just $\Omega(1/\varepsilon^2)$.

Motivated by these advances and the importance of sublinear-time graph algorithms in general [CS10], we ask whether it is possible to improve on existing spectral density estimation algorithms, or to generalize them to broader settings, e.g., weighted graphs. In addition, we ask whether it is possible to efficiently obtain new sparse approximations, i.e., *sparsifications*, of G that facilitate more efficient spectral density estimation.

In this paper, we provide an affirmative answer to each of these questions. We describe a randomized $O(n\varepsilon^{-3})$-time algorithm for solving the spectral density estimation problem in unweighted graphs. Interestingly, this algorithm only makes $O(n\varepsilon^{-2})$ *deterministic queries* to the neighbor oracle and computes a *nuclear norm sparsifier*, a new notion of sparsifier that we introduce below. Additionally, we obtain an algorithm with the same complexity for weighted graphs provided that, when queried with $a \in V$ and $i \in \mathbb{Z}_{>0}$, the oracle outputs the weighted-degree of a, the weight of the i-th largest edge incident to a (with ties broken arbitrarily), and the endpoint of this edge. Furthermore, we show that these complexities are obtainable even in a weaker *random walk model* for accessing G.

1.1 A Sparsification Approach to Spectral Density Estimation

To obtain our main results, we consider a natural *two-stage* approach to approximating the spectral density of graphs. In particular, we take advantage of the fact that there exist *linear* time methods for spectral density estimation of general matrices whose computational cost is dominated by $O(\varepsilon^{-1})$ matrix-vector multiplications with the target matrix. We obtain our results by applying such methods to a *sparsified* approximation to G's normalized adjacency matrix, which is constructed in a deterministic way. This approach differs substantially from previous methods, which obtain sublinear time by speeding up matrix-vector multiplications via random sampling [BKM22].

[1] The runtime of the method from [CKSV18] is independent of n; so instead of outputting a list of n eigenvalues, it returns a list of $O(1/\varepsilon)$ distinct eigenvalue magnitudes and corresponding multiplicities.

Concretely, linear time methods have emerged from a long line of work in computational chemistry, applied mathematics, and, recently, computer science [WWAF06; LSY16]. In [BKM22] and [CTU21] it was proven that common algorithms like the kernel polynomial, moment matching, and stochastic Lanczos quadrature methods, can all provably approximate the spectral density of *any* symmetric matrix A up to ε-accuracy in Wasserstein distance using roughly $O(\varepsilon^{-1})$ matrix-vector multiplications with A.[2] Matrix-vector multiplication dominates the runtime of these methods, so they run in time $O(\text{nnz}(A)\varepsilon^{-1})$, where $\text{nnz}(A)$ denotes the number of non-zeros in the matrix A.

It follows that we can solve the spectral density estimation problem in sublinear time as long as we can produce, in sublinear time, a *sparse* approximation, M, to any normalized adjacency matrix, N_G, that preserves the spectrum of N_G in the sense of (1). Consequently, in this paper, we study the following natural question: What notions of graph approximation (and ultimately, sparsification) are sufficient for preserving the spectrum in the sense of (1) *and* can be obtained in sublinear time? In addressing this question, our goal is to decouple the question of how many queries are needed to approximate N_G from the computational question of how to efficiently approximate eigenvalues.

We make progress on this central question by introducing a new notion of *ε-additive nuclear approximation* and presenting algorithms that obtain near-optimal sparsity and query complexity for producing such approximations in sublinear time. Formally, we define:

Definition 1.2 (ε-additive nuclear approximation and sparsification)**.** $M \in \mathbb{R}^{n\times n}$ is an ε-additive *nuclear approximation* to an n-node graph G with normalized adjacency matrix N_G if $\|N_G - M\|_* \leq \varepsilon n$. If M is also s-sparse for some $s = o(n^2)$, we say that M is an ε-additive *nuclear sparsifier* of G.[3]

Above, $\|N_G - M\|_*$ denotes the nuclear norm of $N_G - M$, i.e., the sum of the matrix's singular values. Importantly, nuclear approximation is sufficient for obtaining accurate spectral density estimates. In particular, let $W_1(A, B)$ denote the Wasserstein-1 distance between the probability distributions p and q induced by the real-valued eigenvalues of the two symmetric matrices as in (1), i.e., $W_1(A, B) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} |\widehat{\lambda}_i - \lambda_i|$ where $\lambda_1 \leq \cdots \leq \lambda_n$ are the eigenvalues of A and $\widehat{\lambda}_1 \leq \cdots \leq \widehat{\lambda}_n$ are the eigenvalues of B. A short proof establishes (see Appendix B) that:

Lemma 1.3. *For symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, if $\|A - B\|_* \leq n\varepsilon$, then $W_1(A, B) \leq \varepsilon$.*[4]

In particular, if $\|M - N_G\|_* \leq \varepsilon n/2$ and we find eigenvalues $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_n$ that are $\varepsilon/2$ close in Wasserstein distance to the eigenvalues of M (e.g. using the moment matching method of [BKM22]), then by triangle inequality, those eigenvalues will satisfy the requirement of Problem 1.1.

1.2 Sublinear-time sublinear-size sparsifiers

Our main algorithmic result is that nuclear sparsifies with $O(n\varepsilon^{-2})$ non-zero entries can be computed in $O(n\varepsilon^{-2})$ time. Formally, we assume access to the graph G via the following oracle:

Definition 1.4 (Adjacency Query Model)**.** We say we have *adjacency query access* to a weighted graph $G = (V, E, w)$ if the vertices V are known and there is an $O(1)$ time procedure, $\mathsf{GetNeighbor}(a, i)$

[2] For example, the method from [BKM22] uses $\min(\varepsilon^{-1}, \varepsilon^{-2}\log^4(\varepsilon^{-1})/n$ matrix-vector multiplications, which is $O(\varepsilon^{-1})$ for sufficiently large n.

[3] Note that M itself is not required to be the adjacency matrix of a graph and our algorithms will in general return matrices that are not. However, in all cases, we show that one can easily modify our algorithm to ensure that the output is indeed a normalized adjacency matrix if this is required.

[4] When A and B are diagonal with entries sorted in decreasing order, Lemma 1.3 can be strengthened to an if and only if. In this sense, nuclear approximation is a natural strengthening of approximation in the Wasserstein-1 distance.

that when queried with any $a \in V$ and $i \in \mathbb{Z}_{>0}$ outputs $\deg_G(a)$ as well as the i-th largest edge $\{a, b\}$ and $w_{\{a,b\}}$ (with ties broken arbitrarily) if there is one and $\emptyset$ otherwise.

Such queries are easily supported by standard data structures for storing graphs. For example, an adajency list where each node's neighbors are stored in an array supports adajency queries for unweighted graphs. For weighted graphs, we just need those arrays to be sorted by edge weight.

In Section 3, we provide an algorithm that uses a sublinear number of GetNeighbor queries to construct a sublinear-sized nuclear sparsifier. The method works even for weighted graphs.

Theorem 1.5 (Sublinear time nuclear sparsification)**.** *For any $\varepsilon \in (0, 1)$, there is a* deterministic *algorithm (Algorithm 1) that runs in $O(n\varepsilon^{-2})$ time in the Adjacency Query Model (Definition 1.4) which, for any undirected weighted graph G, returns an $O(n\varepsilon^{-2})$-sparse ε-additive nuclear sparsifier of G in $O(n\varepsilon^{-2})$ time.*

We obtain Theorem 1.5 using a greedy algorithm that deterministically removes edges from G based on their weight, and the degree of their end-points. The fact that such an approach works is surprising in the context of prior work on graph sparisfication. For example, efficient methods for spectral sparsification have all been based on random sampling, which inherently introduces $\log(n)$ factors [SS11; Lee13]. The fact that randomness is not need in our setting highlights the strong qualitative difference between our new notion of nuclear approximation and prior work.

By combining Thoerem 1.5 with the existing linear time spectral density estimation algorithm from [BKM22], we immediately obtain the following theorem:

Theorem 1.6 (Spectral Density Estimation)**.** *There is an algorithm that runs in $O(n\varepsilon^{-3})$ time in the Adjacency Query Model (Definition 1.4) which, with probability $9/10$, solves the spectral density estimation problem (Problem 1.1) for any weighted, undirected graph G.*

As discussed earlier, for sufficiently small ε, Theorem 1.6 gives the fastest known result for spectral density estimation of graphs, improving on the $O(n\varepsilon^{-7})$ time method from [BKM22]. The result is incomparable to the the $2^{O(\varepsilon^{-1})}$ algorithm from [CKSV18], which is based on computing moments of G's spectral density via random walks. That method has *no dependence* on n, but a prohibitive exponential dependence on ε.[5] Notably, our method is also the first to apply to arbitrary weighted graphs; prior work work assumes that G is unweighted.

Beyond a faster algorithm, our two-stage approach to spectral density estimation also motivates new questions about graph sparsification that may be of independent interest. For example, nuclear approximation is a natural relaxation of more stringent notions of approximation studied in the literature, such as spectral sparsification [ST11; SS11] or probabilistic spectral sparsification [Lee13]. It is natural to ask if our current bounds for nuclear approximation are tight, both in terms of sparsity and sample complexity. We address this question below in Section 1.3. Additionally, we can ask whether this relaxed sense of approximation is necessary, or if one can obtain comparable sublinear algorithms for the stronger notions of spectral approximation. Unfortunately, we show barriers against this possibility, motivating our results on nuclear sparsification (see Section 1.4.)

1.3 Lower Bounds for Nuclear Sparsification

As mentioned above, given that nuclear approximation is a natural relaxation of widely studied graph sparsification metrics, it is desirable to fully understand the complexity of the problem. We

[5] Initial evidence in the form of lower bounds in restricted models suggests that it may not be impossible to improve the ε dependence in [CKSV18] to sub-exponential while maintaining no dependence on n [JMSS23].

complement our algorithmic result from Theorem 3.1 with nearly matching sparsity and query lower bounds. First, we show that the sparsity of our sparsifiers is near-optimal:

Theorem 1.7 (Nuclear Norm Sparsity Lower Bound)**.** *There exists constant $\varepsilon_0 \in (0,1)$ such that for any $\varepsilon \in (0,\varepsilon_0)$, and $n \geq c_1 n\varepsilon^2$, there is a graph G with vertex size n and normalized adjacency matrix N_G, such that for any matrix M with $\text{nnz}(M) \leq n\varepsilon^{-2} \cdot (c_2 \log^2 \varepsilon^{-1})^{-1}$, we have $\|N_G - M\|_* > n\varepsilon$, for some constants c_1, c_2.*

To understand our proof of Theorem 1.7, consider the extreme case when $\varepsilon = 1/\sqrt{cn}$ for a small constant c. This is the smallest value of ε for which our upper bound from Theorem 3.1 gives a non-trivial result (i.e. a matrix with less than n^2 entries). Using the probabilistic method, we construct $2^{O(n^2)}$ graphs whose adjacency matrices are ε far away from each other in the nuclear norm. The number of $\Omega(n^2/\log^2 \varepsilon^{-1})$-sparse matrices that can meaningfully approximate these matrices is at most $2^{O(n^2/\log n)}$, so there must be many adjacency matrices that cannot be approximated by any of the sparse matrices. We prove the result for general ε via a reduction takes many disjoint copies of one of these graphs that is hard for the $\varepsilon = 1/\sqrt{cn}$ case.

We also show that any algorithm for constructing an ε-additive nuclear sparsifier must, in the worst case, make $\widetilde{\Omega}(n\varepsilon^{-2})$ GetNeighbor queries.

Theorem 1.8 (Corollary of Theorem 4.9)**.** *There exists constant $\varepsilon_0 \in (0,1)$ such that for any $\varepsilon \in (0,\varepsilon_0)$, any algorithm working in the Adjacency Query Model (Definition 1.4) requires $\Omega(n\varepsilon^{-2} \cdot \log^{-2}(\varepsilon^{-1}))$ GetNeighbor queries to return an ε-additive nuclear sparsifier with probability $2/3$.*

We prove Theorem 1.8 by showing there exists a distribution over graphs such that any deterministic algorithm with query complexity $o(n\varepsilon^{-2} \cdot \log^{-2}(\varepsilon^{-1}))$ cannot distinguish between an Erdős Rényi graph G_1 and another graph G_2 that is also marginally distributed as an Erdős Rényi graph, but is a complement of G_1 except on the queries considered by the algorithm. G_1 and G_2 look the same to the algorithm but are far in nuclear norm. The theorem then follows from Yao's minimax principle.

It is interesting to ask if the same query complexity remains optimal for spectral density estimation itself (Problem 1.1). Notably, there has been recent progress on this problem by Jin et al. [JMSS23], who show a lower bound of $\Omega(1/\varepsilon^2)$ and a lower bound of $\Omega(\exp(\varepsilon^{-1}))$ under a weaker random walk query model (see Section 1.5). However, depending on the regimes of n and ε, large gaps still remain in understanding the optimal query complexity and running times for spectral density estimation. Addressing these gaps is an exciting open direction for future work.

1.4 Additive Error Spectral Sparsifiers

In Section 1.1, we motivated nuclear sparsifiers for their use in computing accurate spectral densities estimates. However, there are other notions of approximation in the literature that one could consider. For example there is a long line of work on spectral sparsification [LS18; BSST13; ST11],

Definition 1.9 (ε-spectral sparsifier)**.** Given $\varepsilon > 0$ and a graph G with Laplacian L, a symmetric matrix $\widetilde{L} \in \mathbb{R}^{n \times n}$ is an ε-spectral sparsifier of L if for all $x \in \mathbb{R}^n$, $(1-\varepsilon)x^\top Lx \leq x^\top \widetilde{L}x \leq (1+\varepsilon)x^\top Lx$.

The above definition involves the Laplacian whereas we focus on the adjacency matrix, and the normalized adjacency specifically. However, it is not hard to check that, at least if L is unweighted, then an ε-spectral sparsifier will have a normalized adjacency spectrum that is ε-close to that of G in Wasserstein distance. However in [Lee13], it was shown that it is impossible to obtain spectral sparsifiers using $o(n^2)$ queries (see Theorem 11 of [Lee13] with $\delta = 1/n$). Addressing this challenge, [Lee13] defined the following weaker notion of probabilistic spectral sparsification.

Definition 1.10 ((ε, δ)-probabilistic spectral sparsifier)**.** Given $\varepsilon, \delta > 0$ and a graph G with Laplacian L, a symmetric matrix $\widetilde{L} \in \mathbb{R}^{n \times n}$ is an (ε, δ)-probabilistic spectral sparsifier of L if, for all $x \in \mathbb{R}^n$, $(1-\varepsilon)x^\top Lx - \delta x^\top x \leq x^\top \widetilde{L} x \leq (1+\varepsilon)x^\top Lx + \delta x^\top x$.

[Lee13] provides an algorithm which runs in $O(n\varepsilon^{-2}\delta^{-1})$-time and obtains an $O(n \log n \varepsilon^{-2})$-sparse (ε, δ)-probabilistic spectral sparsifier of an unweighted graph. This weaker notion of spectral sparsification is sufficient for obtaining nuclear sparsification of the normalized adjacency matrix. For weighted graphs, we introduce a closely related notion, which we call ε-additive spectral sparsification.

Definition 1.11 (ε-additive spectral sparsifier)**.** A symmetric matrix $M \in \mathbb{R}^{n \times n}$ is an ε-additive spectral sparsifier of G if and only if $\left\| M - \widetilde{N}_G \right\|_2 \leq \varepsilon$.

For a unweighted graphs, an ε-additive spectral sparsifier can be obtained by computing an $(\varepsilon, \varepsilon)$ probabilistic spectral sparsifier in the sense of [Lee13].

Given the long line of research on spectral sparsification (see Section 1.6 for discussion of related work in this area), it is natural to ask whether we can obtain a similar sparsity and query complexity as our nuclear sparsification for spectral sparsification in the sense of Definition 1.11 using a *deterministic* algorithm. Interestingly, we obtain the following result that highlights a separation between what is possible with nuclear sparsification versus spectral sparsification.

Theorem 1.12 (Corollary of Corollary 5.2)**.** *For a small constant ε and large n, any deterministic algorithm working with degree queries and weight-ordered neighbor queries that takes in graph G and computes an ε-additive spectral sparsifier has query complexity $\Omega(n^2)$.*

We prove Theorem 1.12 by constructing two d-regular graphs (for some value of d) that have the same output of the queries made by the deterministic algorithm, but differ everywhere else while maintaining degree regularity. It is an interesting open problem to construct or prove a lower bound against any *randomized* algorithm working with degree queries and weight-ordered neighbor queries that runs in $O(n \log n)$-time for constant ε. However, as we discuss in the next Section 1.5, we *do* rule out the existence of such an algorithm in the more restrictive random walk access model.

1.5 Random Walk Query Model

The prior spectral density estimation algorithms [BKM22; KV17] work in a slightly different model than the adjacency query model which we call the *random walk model* (see Definition 1.13). In this model, one knows the vertices V and there is an $O(1)$ time procedure to either (1) query a new node chosen uniformly at random along with a random neighbor of this node (chosen with probability proportional to the edge weight) along with the degrees of the edge endpoints; or (2) visit a random neighbor of the current node (chosen with probability proportional to the edge weight) along with the degrees of the edge endpoints.

Definition 1.13 (Random Walk Query Model)**.** We say we have *random walk query access* to a weighted graph $G = (V, E, w)$ if the vertices V are known and there is an $O(1)$ time procedure, RandomNeighbor that when queried selects a $a \in V$ uniformly at random and then samples a random edge $\{a, b\}$ with probability proportional to its weight and outputs that edge, and the degree of its endpoints, or $\emptyset$ if a has no neighbors.

We call random $v_0, v_1, ..., v_k$ a length k random walk from v_0 if each v_i is chosen probability proportional to weight. We call the model of Definition 1.13 the random walk query model as it naturally allows for the simulation of a 1-step random walk on G from a random initial vertex:

first, we sample a random vertex $v_0 \in V$ and then visit a neighbor of v_0 chosen randomly with probability proportional to the edge weight. Note that previous work [BKM22] actually considers a *stronger* query model, in which one can simulate a k-step random walk on the graph for any $k \in \mathbb{Z}_{>1}$. That is, in the k-step random walk model of [BKM22], one can sample a $v_0 \in V$ uniformly at random, and then, for each $i \in [k]$ one can sample v_i from the neighbors of v_{i-1} with probability proportional to the edge weight from v_i to v_{i-1}. Consequently, our algorithms in the random walk model (Definition 1.13) automatically apply in the setting of [BKM22] because we only require the ability to simulate 1-step random walks.

This model is more restrictive than the adjacency query model. Although our algorithm discussed in Section 1.2 does not work in this model, in Section 6 we present an alternative, randomized algorithm that achieves identical query- and time-complexity in the random-walk graph query model using entry-wise sampling and matrix concentration results. For nuclear sparsification, we obtain the following result using a vector variance bound.

Theorem 1.14. *Given access to a graph G under the random walk model (Definition 1.13) and some $\varepsilon \in (0,1)$, for $T \geq 2n\varepsilon^{-2}$, with probability $2/3$, $X = \mathsf{NuclearSparsify}(G, \varepsilon, n, T)$ (Algorithm 3) is an $O(n\varepsilon^{-2})$-sparse ε-additive nuclear sparsifier. Moreover, X is computed using only $O(n\varepsilon^{-2})$ queries to the random walk model.*

Interestingly, in this random walk model, we can also obtain an algorithm for ε-additive spectral sparsification. We obtain the following Theorem 1.15 by appealing to an entry-wise sampling result of [Coh+17]. Theorem 1.15 improves over the algorithm for ε-additive spectral approximation on unweighted graphs implied by [Lee13], which requires $O(n\varepsilon^{-3})$ queries to achieve the same sparsity of $O(n\log(n)\varepsilon^{-2})$ for ε-additive spectral approximation.

Theorem 1.15. *Given access to a graph G under the random walk model (Definition 1.13) and some $\varepsilon \in (0,1)$, for $T \geq 256n\varepsilon^{-2}\log(3n)$, with probability $2/3$, $X = \mathsf{SpectralAdditiveSparsify}(G, \varepsilon, n, T)$ (Algorithm 4) is an ε-additive $O(n\varepsilon^{-2}\log(n))$-sparse ε-additive spectral sparsifier for an unweighted graph G. Moreover, X is computed using $O(n\varepsilon^{-2}\log(n))$ queries to the random walk model.*

Note that the result of Theorem 1.15 achieves the stronger notion of ε-additive spectral sparsification, at the cost of an extra factor of $\log(n)$ in the sparsity, query complexity, and runtime as compared to Theorem 1.14, which obtains only a nuclear sparsifier. We show that this $\log(n)$ factor is unavoidable in the following sense:

Lemma 1.16. *Let $0 < \varepsilon < 1/8$ be a constant. Given access to a graph G under the random walk model (Definition 1.13), any algorithm that uses $o(n\log n)$ oracle queries cannot output a spectral sparsifier with a constant probability.*

Lemma 1.16 follows from the famous coupon collector's problem in probability theory. We prove Lemma 1.16 by observing that spectral sparsification is a stringent condition that sometimes needs to query all the edges of a sparse graph. The lower bound applies to the k-step random walk model since the instance we construct has many disconnected components, and therefore 1-step random walk is as informative as a k-step random walk.

1.6 Additional Related Work

Sublinear time graph algorithms. Estimating graph properties in sublinear time $o(n^2)$ is central for several algorithmic applications of graph theory. Positive results along this line have been shown for a variety of graph properties, e.g. estimating the average degree [ERS17], the

number of connected components [BKM14], the number of k-cliques and clusters [ERS18; MAC20; Ede+22; MOP01; BLHK16], the size of minimum spanning tree [CRT05; Czu+03], largest cut, maximum matching and minimum vertex cover [ORRR12], etc. Beyond these combinatorial results and aforemention work on sublinear-time density estimation there is a long line of work on sublinear time graph algorithms for various additional spectral graph properties. For example, there are sublinear time algorithms for spectral clustering [Glu+21; Mou21] and expander testing [PY23].

Graph sparsification Graph sparsification is incredibly well-studied. Given a graph on n nodes and m edges, [BK96] showed a ε-cut sparsifier with $O(n \log n/\varepsilon^2)$ edges that can be constructed in time $O(n^2 \log^3 n)$. Inspired by their work [ST11] introduced and proved $(1+\varepsilon)$-spectral sparsifiers with $O(n \log^c n/\varepsilon^2)$ (for some large value of c) edges can be constructed in time $\tilde{O}(m)$. The sparsity was improved by [BSST13], where they showed a $(1+\varepsilon)$-spectral sparsifier with $O(n/\varepsilon^2)$ edges in $O(mn^3/\varepsilon^2)$ time. While this is optimal in the sparsity, the run-time is not useful in applications. To improve upon this [SS11] showed a slightly worse sparsifier with $O(n \log n/\varepsilon^2)$ edges but with a near linear running time of $\tilde{O}(m)$. Spectral sparsification cannot fundamentally run in sublinear time [Lee13]. To design sublinear time algorithms, [Lee13] introduced the notion of (ε, δ) probabilistic spectral sparsifier and gave an algorithm that produces a sparsifier with $O(n \log n/\varepsilon^2)$ edges in time $O(n\varepsilon^{-2}\delta^{-1})$.

Spectrum estimation. It is known that the eigenvalues of a symmetric matrix can be computed accurately in time $O(n^\omega)$, where $\omega \approx 2.4$ is the matrix multiplication exponent [DDH07]. This problem was of interest in the quantum chemistry and condensed matter physics community, where methods for spectral density estimation were introduced that run in time $o(n^\omega)$ [Ski89; Wan94; SR94]. Most spectrum estimation methods rely on moment matching and stochastic trace estimation. This approach has been applied successfully for a wide variety of matrices [LSY16; WWAF06]. However, theoretical guarantees have only been proven recently. For any Hermitian matrix, it can be shown that these approaches provably run in linear time, i.e., given a matrix, one can solve the spectral density estimation problem with just $\mathsf{poly}(1/\varepsilon)$ matrix-vector multiplications with the matrix [CTU21; BKM22]. For graphs, [CKSV18] used random walks to estimate the moments of spectral density distribution, and obtain a $2^{O(\varepsilon^{-1})}$ time algorithm to estimate the spectral density of the graph. Prior work has also considered estimating properties of the spectrum. For instance, [Mus+18] show a $\tilde{O}(n^{2.18}\varepsilon^{-3})$ time to compute the nuclear norm of a matrix.

Sketching approaches to spectrum approximation. Guarantees for spectrum approximation of a general matrix have also been studied. Given a symmetric matrix $A \in \mathbb{R}^{n\times n}$, [SW23] show that one can estimate all the eigenvalues of A to within $\varepsilon \|A\|_F$ with constant probability, with just $O(1/\varepsilon^2)$ matrix-vector multiplies with A. [Bha+23] show an algorithm that (when the entries of A are bounded) can compute all the eigenvalues of A with constant probability to additive accuracy εn which only reads $\tilde{O}(\log^3 n/\varepsilon^5)$ entries of A. We note that these results do not assume a graph structure and are more general. In the graph case, the guarantees are not comparable to spectral density estimation because they approximate *all* the eigenvalues to additive accuracy $\varepsilon \|A\|_F$, but the average error is $\varepsilon n \|A\|_F$, whereas the average error required by the spectral density estimation problem is $\varepsilon n \|A\|_2$.

1.7 Paper organization

We discuss notation in Section 2 and then present technical results. Section 3 presents our nuclear sparsification algorithm in the neighbor oracle model. Section 4 presents our sparsity and query

lower bounds for nuclear sparsification. Section 5 presents a lower bound against deterministic algorithms for ε-additive spectral sparsification. Section 6 covers our results in the more restrictive random walk model. Omitted details are deferred to the Appendix (A and B.)

2 Notation

We summarize notation used throughout the paper below.

Graphs. Given a graph $G = (V, E, w)$, we use $A_G \in \mathbb{R}^{V \times V}_{\geq 0}$ to denote the weighted adjacency matrix of G where $A_{Gv,v'} = w_e$ if $e = \{v, v'\} \in E$ and $A_{Gv,v'} = 0$ otherwise. We use $\deg(v) \stackrel{\text{def}}{=} \sum_{v':e=\{v,v'\}\in E} w_e$ to denote the degree of vertex v, and $D_G \in \mathbb{R}^{V \times V}_{\geq 0}$ to denote the diagonal degree matrix of G where D_G is diagonal with $D_{Gv,v} \stackrel{\text{def}}{=} \deg(v)$ for all $v \in V$. We also use $\text{cdeg}(v) \stackrel{\text{def}}{=} |\{\{v', v\} \in E\}|$ to denote the combinatorial degree of vertex $v \in V$. We let $N_G \in \mathbb{R}^{V \times V}$ denote the normalized adjacency matrix of G, i.e. $N_G \stackrel{\text{def}}{=} D_G^{-1/2} A_G D_G^{-1/2}$. We let $L_G \stackrel{\text{def}}{=} D_G^{-1/2}(D_G - A_G)D_G^{-1/2} = I - N_G$ to denote the normalized Laplacian matrix. We may drop the subscripts G when the graph is clear from context.

Vectors and Matrix Norms. For a vector $v \in \mathbb{R}^n$, we use $\|v\| = \sqrt{\sum_{i\in[n]} v_i^2}$ to denote the standard Euclidean norm. For any symmetric square matrix $A \in \mathbb{R}^{n\times n}$, the spectral norm is defined as $\|A\|_2 = \|A\|_{\text{op}} = \max_{v:v\in\mathbb{R}^n} \|Av\|/\|v\| = \max_i |\lambda_i(A)|$; it's Frobenius norm is defined as $\|A\|_{\text{F}} \stackrel{\text{def}}{=} \sqrt{\sum_{i,j} A(i,j)^2}$; and its nuclear norm is defined as $\|A\|_* \stackrel{\text{def}}{=} \sum_{i=1}^n |\lambda_i(A)|$. We also define the inner-product in the matrix space to be $\langle A, X\rangle = \sum_{i,j} A(i,j)X(i,j)$ for any matrices $A, X \in \mathbb{R}^{m\times n}$.

3 Additive Nuclear Sparsifiers

In this section, we present our main algorithm for additive nuclear sparsification (Definition 1.2). We work with the graph access model in Definition 1.4 for weighted graphs and design a deterministic algorithm that constructs a $O(n\varepsilon^{-2})$-sparse ε-additive nuclear sparsifier $\widetilde{N}$ for the normalized adjacency matrix N of a graph.

Given a weighted graph $G = (V, E, w)$ with weighted adjacency matrix A_G and degree matrix D_G, the following theorem provides a natural scheme to construct sparse matrices that a close to N_G in nuclear norm. In particular, it shows it suffices to only consider edges with a large ratio of weight to the degrees of the endpoints.

Theorem 3.1. *Given weighted graph $G = (V, E, w)$ with weighted adjacency matrix $A_G \in \mathbb{R}^{n\times n}$ and degree matrix D_G and accuracy parameter $\varepsilon \in (0, 1)$. Define $E' \subseteq E$ as*

$$E' \stackrel{\text{def}}{=} \left\{ e = \{v, v'\} \in E \;\middle|\; w_e \geq \varepsilon^2 \cdot \min\left\{\deg(v), \deg(v')\right\} \right\}.$$

Further, let G' denote the subgraph induced by E', and let $A_{G'} \in \mathbb{R}^{n\times n}$ denote the weighted adjacency matrix of G'. Let $\widetilde{N} \stackrel{\text{def}}{=} D_G^{-1/2} A_{G'} D_G^{-1/2}$. Then $\left\|\widetilde{N}\right\|_2 \leq 1$, $\text{nnz}(\widetilde{N}) \leq 2n\varepsilon^{-2}$, and $\left\|N_G - \widetilde{N}\right\|_* \leq \varepsilon n$.

Proof. Let $D_{G'}$ denote the degree matrix of $A_{G'}$. As $A_{E'}$ is the subgraph induced by $E' \subseteq E$,

$$-D_G \preccurlyeq -D_{G'} \preccurlyeq A_{G'} \preccurlyeq D_{G'} \preccurlyeq D_G\,,$$

Algorithm 1: Sparse nuclear norm approximation

Input: Graph $G = (V = [n], E, w)$, under query model as per Definition 1.4, degree matrix D, accuracy ε
Output: Matrix $\widetilde{N}$
Initialize $\widetilde{N} = 0$
for $v \in V$ **do**
  Query the next neighbor v' of v from oracle model
  while $w_{\{v,v'\}} \geq \varepsilon^2 \deg(v)$ **do**
    Set $\widetilde{N}(v, v') = \widetilde{N}(v', v) = w_{\{v,v'\}}/(\sqrt{\deg(v)\deg(v')})$
    Query the next neighbor v' of v from oracle model
Return: $\widetilde{N}$

which implies that $\left\|\widetilde{N}\right\|_2 \leq 1$. Note that for each $v \in V$ there are at most ε^{-2} different v' for which $w_{\{v,v'\}} \geq \varepsilon^2 \deg(v)$. Consequently, the total number of edges in E' is bounded by $|E'| \leq n\varepsilon^{-2}$, and thus $\mathrm{nnz}(\widetilde{N}) \leq 2n\varepsilon^{-2}$ since every edge contributes at most 2 to the sparsity of $\widetilde{N}$. To bound the error in the nuclear norm, we first bound the error in the Frobenius norm as follows.

$$\begin{aligned}\left\|N_G - \widetilde{N}\right\|_F^2 &= \sum_{e=\{v,v'\}\in E\setminus E'} 2\left(\frac{w_e}{\sqrt{\deg(v)\deg(v')}}\right)^2 = \sum_{v\in V}\sum_{v'\in V: e=\{v,v'\}\in E\setminus E'} \frac{w_e^2}{\deg(v)\deg(v')} \\ &\leq \sum_{v\in V}\sum_{v'\in V: e=\{v,v'\}\in E\setminus E'} \varepsilon^2 \frac{w_e}{\deg(v)} = \varepsilon^2 n\,.\end{aligned}$$

Consequently, by the norm inequality in Fact A.1, we have

$$\left\|N_G - \widetilde{N}\right\|_* \leq \sqrt{n}\left\|N_G - \widetilde{N}\right\|_{\mathrm{F}} \leq \varepsilon n\,. \qquad \blacksquare$$

The nuclear sparsifier guaranteed in Theorem 3.1 is *not* guaranteed to be the normalized adjacency matrix of any undirected graph. As access to a graphical nuclear norm approximation may be desirable in certain applications, we show how to modify the procedure above to obtain a graphical nuclear sparsifier in Lemma B.1.

We show how to easily turn the procedure analyzed in Theorem 3.1 into an algorithm by working with the oracle access model in Definition 1.4, in Algorithm 1.

Theorem 1.5 (Sublinear time nuclear sparsification)**.** *For any $\varepsilon \in (0, 1)$, there is a* deterministic *algorithm (Algorithm 1) that runs in $O(n\varepsilon^{-2})$ time in the Adjacency Query Model (Definition 1.4) which, for any undirected weighted graph G, returns an $O(n\varepsilon^{-2})$-sparse ε-additive nuclear sparsifier of G in $O(n\varepsilon^{-2})$ time.*

Proof. Let $\widetilde{N}$ denote the output of Algorithm 1. We first show the following properties of $\widetilde{N}$: $\left\|\widetilde{N}\right\|_2 \leq 1$, $\mathrm{nnz}(\widetilde{N}) \leq 2n\varepsilon^{-2}$, and $\left\|N_G - \widetilde{N}\right\|_* \leq \varepsilon n$. To do so, it suffices to show such $\widetilde{N}$ that Algorithm 1 outputs is exactly the same as the defined $\widetilde{N}$ in Theorem 3.1. To avoid confusion we use $\widetilde{N}'$ to denote the matrix defined in Theorem 3.1. By definition we know for any $\widetilde{N}(v, v') \neq 0$, $w_{\{v,v'\}} \geq \varepsilon^2 \min\{\deg(v), \deg(v')\}$ and thus $\widetilde{N}'(v, v') = \widetilde{N}'(v', v) = \widetilde{N}(v, v') = \widetilde{N}(v', v)$. Now for any

Algorithm 2: Spectral Density Estimation for Graphs

1 **Input:** Graph $G = (V = [n], E, w)$, under query model as per Definition 1.4, accuracy ε, probability δ
2 **Output:** Spectral density estimate q of G
3 $\widetilde{N} \leftarrow$ Algorithm 1 on G with target accuracy $\varepsilon/2$
4 $q \leftarrow$ Approximate Chebyshev Moment Matching [BKM22, Algorithm 1, 2] on $\widetilde{N}$
5 **Return:** q

$\widetilde{N}'(v, v') \neq 0$, without loss of generality we can assume $\deg(v) \geq \deg(v')$ and $w_{\{v,v'\}} \geq \varepsilon^2 \deg(v')$. we can see such an edge $\{v', v\}$ must be queried when querying edges at vertex v' in descending weights and thus will be added to $\widetilde{N}$. This shows we also have $\widetilde{N}'(v, v') = \widetilde{N}'(v', v) = \widetilde{N}(v, v') = \widetilde{N}(v', v)$. Altogether this shows $\widetilde{N} = \widetilde{N}'$, and the properties of $\widetilde{N}$ follow from Theorem 3.1.

Next we bound the query complexity. Similar to the proof of Theorem 3.1, we can show for each $v \in V$, we at most query $\varepsilon^{-2} + 1$ edges at vertex v until seeing some $w_{\{v,v'\}} < \varepsilon^2 \deg(v)$. Summing this up for all $v \in V$ and noting the vertex size is n concludes the query complexity. ■

3.1 Spectral Density Estimation via Nuclear Norm Sparsification

From Lemma 1.3, we know that to approximate the spectral density N up to additive error ε, it suffices to find a matrix $\widetilde{N}$ such that $\left\|N - \widetilde{N}\right\|_* \leq \varepsilon n$. Thus, an important application of Algorithm 1 is to estimate spectral density of the normalized adjacency matrix N, as shown in Algorithm 2. We state the guarantee of Algorithm 2 below.

Theorem 1.6 (Spectral Density Estimation)**.** *There is an algorithm that runs in $O(n\varepsilon^{-3})$ time in the Adjacency Query Model (Definition 1.4) which, with probability 9/10, solves the spectral density estimation problem (Problem 1.1) for any weighted, undirected graph G.*

Proof. Set $\delta = 0.1$. Note that we can assume that $|V| = n \geq \varepsilon^{-2} \log^2(1/(\varepsilon\delta)) \log^2(1/\varepsilon)$. If otherwise, then we can compute the entire eigen-decomposition to high accuracy in time roughly $\tilde{O}(n^\omega)$ [BGKS20] which is less than $O(n\varepsilon^{-3})$, where $\omega < 2.371866$ is the exponent of the matrix multiplication time [DWZ22].

Correctness. Let p denote the spectral density of G. For the normalized adjacency matrix $N_G = D_G^{-1/2} A_G D_G^{-1/2}$ of graph G, from Theorem 3.1, we know that the $\widetilde{N}$ computed in Line 3 of Algorithm 2 satisfies $\left\|N_G - \widetilde{N}\right\|_* \leq \varepsilon n/2$. Let $\tilde{p}$ denote the spectral density of $\widetilde{N}$, we thus have $W_1(p, \tilde{p}) \leq \varepsilon/2$ following from Lemma 1.3.

From Theorem 1.4 of Braverman et al. [BKM22], we know that if $|V| = n \geq \varepsilon^{-2} \log^2(1/(\varepsilon\delta)) \log^2(1/\varepsilon)$ then the Approximate Chebyshev Moment Matching Algorithm [BKM22, Algorithm 1 with Algorithm 2] in Line 4 of our Algorithm 2 returns spectral density q such that $W_1(\tilde{p}, q) \leq \varepsilon/2$ with probability $1 - \delta$. Note that q can be discretized to output a list of eigenvalues.

Therefore, by triangle inequality of the Wasserstein-1 distance, we get that $W_1(p, q) \leq W_1(p, \tilde{p}) + W_1(\tilde{p}, q) \leq \varepsilon$.

Runtime. Theorem 3.1 shows that the matrix $\widetilde{N}$ constructed in Line 3 of Algorithm 2 has $\mathrm{nnz}(\widetilde{N}) \leq 8n\varepsilon^{-2}$. Under the access model in Definition 1.4 and by Theorem 1.5, it takes time $O(n\varepsilon^{-2})$ to construct such matrix $\widetilde{N}$. From Braverman et al. [BKM22, Theorem 1.4], Line 3 of

Algorithm 2 takes time $O(n\varepsilon^{-2} \cdot \varepsilon^{-1})$, as the algorithm takes $O(\varepsilon^{-1})$ of matrix-vector products with matrix $\widetilde{N}$. Therefore, the total running time of Algorithm 2 is $O(n\varepsilon^{-3})$. ■

4 Lower Bounds for Nuclear Sparsification

In this section, we present the lower bounds on sparsity and query complexity for constructing nuclear norm approximation , in Section 4.1 and Section 4.2 respectively, showing the optimality of our algorithm and upper bounds in Section 3.

4.1 Sparsity Lower Bound

To prove our lower bound on the sparsity of a nuclear sparsifier, we first show a lower bound on the sparsity of a nuclear sparsifier for the adjacency matrix of an unweighted, unnormalized graph. In particular, we show that for n sufficiently large, with constant probability, a random Erdős-Rényi graph is unsparsifiable. We then take many disjoint copies of the graph constructed to make it work for a larger range of $\varepsilon \in (0, 1)$.

In order to prove our lower bound for unweighted graphs, we use the following (simplified) matrix concentration result of Guionnet and Zeitouni [GZ00] to show that random $G_{n,1/2}$ graphs tend to be far apart in nuclear norm.

Theorem 4.1 ([GZ00, Theorem 1.1])**.** *Let $C \in \mathbb{R}^{n\times n}$ be a fixed symmetric matrix, whose entries are uniformly bounded by 1. Let $X \in \mathbb{R}^{n\times n}$ be a symmetric random matrix, such that $X_{ij} = C_{ij}\omega_{ij}$, where $\omega_{ij} = \omega_{ji}$, and $\{\omega_{ij}, 1 \le i \le j \le n\}$ are independent random variables with laws $\{P_{ij}, 1 \le i \le j \le n\}$. Assume that $\{P_{ij}, 1 \le i \le j \le n\}$ are supported in $[-1, 1]$. Then, for any $\delta > \delta_0(n) \stackrel{\text{def}}{=} 16\sqrt{\pi}/n$,*

$$\mathbb{P}\left\{\left|\mathrm{Tr}\left(|X|\right) - \mathbb{E}[\mathrm{Tr}\left(|X|\right)]\right| \ge n^{1.5}\delta\right\} \le 4 \cdot \exp\left\{-\frac{n^2(\delta - \delta_0(n))^2}{64}\right\}.$$

In Lemma 4.2, we bound the expected value of this nuclear norm distance between $G_{n,1/2}$ graphs.

Lemma 4.2. *Let $n \in \mathbb{Z}_{\ge 0}$ such that $n \ge 500$. Consider a symmetric random matrix $X \in \{-1, 0, 1\}^{n\times n}$ with zeros on the diagonal and iid entries for $i > j$ according to the distribution*

$$X_{ij} \overset{iid}{\sim} \begin{cases} -1, & \text{with probability } 1/4 \\ 0, & \text{with probability } 1/2 \\ 1, & \text{with probability } 1/4 \end{cases}, \quad \text{for } i > j\,. \tag{2}$$

Then, $n^{1.5} \ge \mathbb{E}\,\|X\|_* \ge n^{1.5}/100$.

Proof. From Corollary 3.9 of Bandeira and Handel [BH16], we know that $\mathbb{P}\{\|X\|_2 > 5\sqrt{n}\} \le \exp(-n/4)$. Moreover, by Chernoff bound, we get that $\mathbb{P}\{\|X\|_{\mathrm{F}} \le n/3\} \le \exp(-n^2/32)$. Conditioning on these two events, we see that (for $n \ge 10$) with probability $1 - \exp(-n/5)$.

$$n^2/9 \le \|X\|_F^2 \le \|X\|_2 \|X\|_* \le 5\sqrt{n}\,\|X\|_*\,,$$

so, $\|X\|_* \ge n^{1.5}/45$. Therefore, for n sufficiently large, $\mathbb{E}\,\|X\|_* \ge n^{1.5}/45 \cdot (1 - \exp(-n/2)) \ge n^{1.5}/100$. Moreover, since X is a $\{-1, 0, 1\}$ matrix, we get that $\|X\|_* \le \sqrt{n}\,\|X\|_{\mathrm{F}} \le n^{1.5}$. ■

Proposition 4.3. *Let A denote the adjacency matrix of an Erdős-Rényi random graph $G(n, 1/2)$, for $n \geq 3000 \in \mathbb{Z}_{\geq 0}$. With probability greater than $5/8$, for any sparse matrix B' with $\mathrm{nnz}(B') \leq n^2/(16 \cdot 10^6 \cdot \log^2 n)$ we have that $\|A\|_* \geq n^{1.5}/400$ and $\|A - B'\|_* \geq n^{1.5}/500$.*

Proof. Let A_1 and A_2 be two Erdős-Rényi random graph $G(n, 1/2)$. The matrix $A_1 - A_2$ follows the distribution of the matrix X in Lemma 4.2. Therefore, we have that $\mathbb{E}\,\|A_1 - A_2\|_* \geq n^{1.5}/100$. From Theorem 4.1, we get that

$$\mathbb{P}\left\{\left|\|A_1 - A_2\|_* - \mathbb{E}\,\|A_1 - A_2\|_*\right| \geq \delta n^{1.5}\right\} \leq 4\exp\left(\frac{-n^2(\delta - 30/n)^2}{64}\right).$$

Using the fact that $n^{1.5}/100 \leq \mathbb{E}\,\|A_1 - A_2\|_* \leq n^{1.5}$ from Lemma 4.2 and choosing $\delta = 1/200$ in the above bound, we get that for $n \geq 3000$

$$\mathbb{P}\left\{\|A_1 - A_2\|_* \leq \frac{n^{1.5}}{200}\right\} \leq 4\exp\left(\frac{-n^2}{4 \cdot 10^6}\right).$$

Let $A_1, A_2, \ldots, A_m$ denote adjacency matrices of m independent Erdős-Rényi random graph $G(n, 1/2)$. By union bound, we get that for any $i \neq j \in [m]$,

$$\mathbb{P}\left\{\|A_i - A_j\|_* \leq \frac{n^{1.5}}{200}\right\} \leq 4\binom{m}{2} \cdot \exp\left(\frac{-n^2}{4 \cdot 10^6}\right).$$

Choosing $m = \exp\left(n^2/(16 \cdot 10^6)\right)$, we get from the above equation that

$$\mathbb{P}\left\{\|A_i - A_j\|_* \leq \frac{n^{1.5}}{200}\right\} \leq 2\exp\left(\frac{-n^2}{8 \cdot 10^6}\right).$$

The above equation implies that there exists more than m random matrices such that with high probability all of them are at a distance more than $n^{1.5}/200$ in nuclear norm from each other. Let $\mathcal{A}' \stackrel{\text{def}}{=} \{A_i, i \in [m]\}$ denote the set of m such matrices. By triangle inequality, we get that at least $m - 1$ matrices $A_j \in \mathcal{A}$ such that $\|A_j\|_* \geq n^{1.5}/400$. Let $\mathcal{A}$ be the set of such matrices, i.e., $\mathcal{A} \stackrel{\text{def}}{=} \{A_j \in \mathcal{A}' \ : \ \|A_j\|_* \geq n^{1.5}/400\}$. We have that $|\mathcal{A}| \geq m - 1$.

Let $\mathcal{S}$ denote the set of all $n \times n$ "discretized-sparse" matrices with the following properties: For any $B \in \mathcal{S}$, $\mathrm{nnz}(B) \leq n^2/(16 \cdot 10^6 \cdot \log^2 n)$ and $\forall i, j \in [n]$, $B_{ij} \in \{-n^2, -(n^2 - \varepsilon), \ldots, (n^2 - \varepsilon), n^2\}$. Note that the entries of B cannot be larger than n^2, else it would not be able to approximate the matrices in $\mathcal{A}$ in nuclear norm due to Fact A.1. Let $c = 16 \cdot 10^6$, and $\varepsilon = 2/(n^2)$. We get that

$$\begin{aligned}
|\mathcal{S}| &\leq \binom{n^2}{\frac{n^2}{c\log^2 n}} \cdot \left(\frac{2n^2}{\varepsilon}\right)^{\left(\frac{n^2}{c\log^2 n}\right)} \\
&\leq (ce\log^2 n)^{\left(\frac{n^2}{c\log^2 n}\right)} \cdot n^{\left(\frac{4n^2}{c\log^2 n}\right)} \\
&= \exp\left(\frac{n^2\log c}{c\log^2 n} + \frac{n^2}{c\log^2 n} + \frac{2n^2\log\log n}{c\log^2 n} + \frac{4n^2\log n}{c\log^2 n}\right) \\
&\leq \exp\left(\frac{7n^2}{c\log n}\right).
\end{aligned}$$

Therefore, for any sparse matrix B' with $\mathrm{nnz}(B') \le n^2/(c\log^2 n)$, we get by the relation between Frobenius norm and nuclear norm Fact A.1, and the choice of $\varepsilon = 2/n^2$ that there exists a matrix $B \in \mathcal{S}$ such that $\|B - B'\|_* \le 2/\sqrt{n}$.

Recalling $m = \exp(n^2/c)$ and size of set $|\mathcal{S}| \le \exp\left(\frac{7n^2}{c\log n}\right)$, we get by counting argument that there are $m - 1 - |\mathcal{S}|$ random Erdős-Rényi high-energy adjacency matrices such that no matrix in $\mathcal{S}$ gives a good approximation. Formally, since $\mathcal{A} \subseteq \mathcal{A}'$ and $|\mathcal{A}| = m - 1$, we get that there exists $m-1-|\mathcal{S}|$ random Erdős-Rényi adjacency matrices $A_{i_1}, \ldots, A_{i_{m-1-|\mathcal{S}|}} \in \mathcal{A}$ such that for any $B \in \mathcal{S}$, $t \in [m - 1 - |\mathcal{S}|]$, $\|A_{i_t} - B\|_* \ge n^{1.5}/400$, where the last inequality follows from triangle inequality.

For $n \ge 3000$, we get that $m - 1 - |\mathcal{S}| > (3/4)\cdot\exp(n^2/c)$. Therefore, by triangle inequality, for any sparse matrix B' with $\mathrm{nnz}(B') \le n^2/(c\log^2 n)$, we get that there exists more than $(3/4)\cdot\exp(n^2/c)$ random Erdős-Rényi matrices in $\mathcal{A}$ with nuclear norm greater than $n^{1.5}/400$ that cannot be approximated by B', i.e., for any matrix B' with $\mathrm{nnz}(B') \le n^2/(c\log^2 n)$, $t \in [m - 1 - |\mathcal{S}|]$, $\|A_{i_t} - B'\|_* \ge n^{1.5}/400 - 2/\sqrt{n} > n^{1.5}/500$. ■

Our lower bound construction follows from considering n/b such b-by-b dimensional matrices A satisfying the properties in Proposition 4.3, and tiling them up into a block-diagonal matrix. Next we show a general statement on the lower bound of the nuclear norm tiled "unsparsifiable" matrices.

Proposition 4.4. *Let $n, k, b \in \mathbb{Z}_{\ge 0}$ be such that $k = \lfloor n/b \rfloor \ge 1$. Let $E \in \mathbb{R}^{n\times n}$ be a block diagonal matrix $E = \mathrm{diag}(E_1, E_2, \ldots, E_k, I_{(n-kb)})$, where $E_1, \ldots, E_k \in \mathbb{R}^{b\times b}$ and $I_{(n-kb)}$ is an identity matrix of size $(n - kb)$. Let E_i, for each $i \in [k]$, be such that for any sparse matrix B' with $\mathrm{nnz}(B') \le \left(c'b^2/\log^2 b\right)$, $\|E_i - B'\|_* \ge c''b^{1.5}$. Then, for any matrix B with $\mathrm{nnz}(B) \le (k/2)\cdot\left(c'b^2/\log^2 b\right)$, we have that*

$$\|E - B\|_* \ge \left(\frac{c''n\sqrt{b}}{4}\right).$$

Proof. Let B be any $n \times n$ matrix with the sparsity condition satisfied by M, i.e., $\mathrm{nnz}(B) \le (k/2)\cdot\left(c'b^2/\log^2 b\right)$, where. Define B_i as the submatrix of B with entries corresponding to E_i. Let $\mathcal{I} \stackrel{\text{def}}{=} \{i \in [k] : \mathrm{nnz}(B_i) \le c'b^2/\log^2 b\}$ be the index set for "not too dense" blocks B_i. Thus we have $|\mathcal{I}| \ge \lfloor k/2 \rfloor$. This is because if $|\mathcal{I}| < \lfloor k/2 \rfloor$, then we have $\mathrm{nnz}(B) \ge \sum_{i\in[k]\setminus\mathcal{I}} \mathrm{nnz}(B_i) > (k/2)\cdot\left(c'b^2/\log^2 b\right)$, leading to a contradiction.

By the assumption on the matrices $E_1, \ldots, E_k$, we get that for each $i \in \mathcal{I}$, by the dual-norm formulation of the nuclear norm (Fact A.1), there exists X_i with $\|X_i\|_2 \le 1$ such that $\|E_i - B_i\|_* = \langle E_i - B_i, X_i\rangle \ge c''b^{1.5}$.

Let $\widetilde{X} \in \mathbb{R}^{n\times n}$ be a block diagonal matrix $\widetilde{X} = \mathrm{diag}(\widetilde{X}_1, \ldots \widetilde{X}_k, \mathbf{0}_{(n-kb)})$, where $\widetilde{X}_i$ (for $i \in [k]$) is a matrix of size $b \times b$, and $\mathbf{0}_{(n-kb)}$ is the all-zeros matrix of size $(n - kb) \times (n - kb)$. For all $i \in \mathcal{I}$, we set $\widetilde{X}_i = X_i$, and for all $i \notin \mathcal{I}$, we set $\widetilde{X}_i = \mathbf{0}_{b\times b}$. By construction, we have that $\left\|\widetilde{X}\right\|_2 \le 1$, and by the dual formulation of nuclear norm (Fact A.1), we get that

$$\|E - B\|_* \ge \left\langle E - B, \widetilde{X}\right\rangle \ge \sum_{i\in\mathcal{I}} \langle E_i - B_i, X_i\rangle \ge c''b^{1.5}\cdot\lfloor k/2\rfloor \ge c''b^{1.5}\cdot\frac{n}{4b} = \frac{c''n\sqrt{b}}{4}.$$ ■

Using the proposition above, we get a lower bound on the sparsifiability of normalized adjacency matrices in the nuclear norm.

Theorem 1.7 (Nuclear Norm Sparsity Lower Bound)**.** *There exists constant $\varepsilon_0 \in (0,1)$ such that for any $\varepsilon \in (0, \varepsilon_0)$, and $n \geq c_1 n\varepsilon^2$, there is a graph G with vertex size n and normalized adjacency matrix N_G, such that for any matrix M with $\mathrm{nnz}(M) \leq n\varepsilon^{-2} \cdot (c_2 \log^2 \varepsilon^{-1})^{-1}$, we have $\|N_G - M\|_* > n\varepsilon$, for some constants c_1, c_2.*

Proof. The theorem follows from Proposition 4.4. Let $G_1, \ldots, G_k$ denote the "unsparsifiable" Erdős Rényi random graphs of size $b \geq 3000$ from Proposition 4.3. Let $A_1, \ldots, A_k$ denote the adjacecny matrix of $G_1, \ldots, G_k$ respectively. We define the graph G to be a graph of size n which is a disjoint union of $G_1, \ldots, G_k$ and $(n - kb)$ disjoint vertices with a self-loop, where $k = \lfloor n/b \rfloor \geq 1$. Let A_G denote the adjacency matrix of G. We get that $A_G = \mathrm{diag}(A_1, A_2, \ldots, A_k, I_{(n-kb)})$, where $I_{(n-kb)}$ is an identity matrix of size $(n - kb)$. Note that the matrix A satisfies the conditions of the matrix E from Proposition 4.4, with $E_i = A_i$, $i \in [k]$. This is true because Proposition 4.3 guarantees that, for each $i \in [k]$, any matrix B' with $\mathrm{nnz}(B') \leq c' b^2 / \log^2 b$, $\|A_i - B'\|_* \geq c'' b^{1.5}$, where $c' = 1/(16 \cdot 10^6)$ and $c'' = 1/500$.

From Proposition 4.4 and the dual norm characterization of the nuclear norm, we get that for any matrix B with $\mathrm{nnz}(B) \leq \frac{n}{2b}(c' b^2 / \log^2 b)$, there exists a matrix $\widetilde{X}$ with $\left\|\widetilde{X}\right\|_2 \leq 1$ such that

$$\|A_G - B\|_* = \left\langle A_G - B, \widetilde{X} \right\rangle \geq \frac{c'' n\sqrt{b}}{4}\,.$$

Since the graphs $G_1, \ldots, G_k$ are Erdős-Rényi random graphs of size b, the degree matrix D_G of graph G satisfies $D_G \preccurlyeq \mathrm{diag}(b \cdot \mathbf{I}, \cdots, b \cdot \mathbf{I}, 1, \cdots, 1)$ where each $\mathbf{I}$ is a b-by-b identity matrix, and we have that

$$\left\|D_G^{-1/2}(A_G - B)D_G^{-1/2}\right\|_* \geq \left\langle D_G^{-1/2}(A_G - B)D_G^{-1/2}, \widetilde{X} \right\rangle = \left\langle (A_G - B), D_G^{-1/2}\widetilde{X}D_G^{-1/2} \right\rangle \geq \frac{c'' n\sqrt{b}}{4b}\,.$$

where in the last inequality we used the definition of the inner product and the fact that all the degrees are less than b.

Setting $\varepsilon_0 = c''/(4\sqrt{3000})$, and $b = (c'')^2/(16\varepsilon^2)$ ensures that $b \geq 3000$ for all $\varepsilon \leq \varepsilon_0$. We also need to esure that $n/b \geq 1$, which implies that $n \geq 16n\varepsilon^2/(c'')^2$. From the equation above and the value of b, we get that for all $\varepsilon \leq \varepsilon_0$,

$$\left\|D_G^{-1/2}(A_G - B)D_G^{-1/2}\right\|_* = \left\|N_G - D_G^{-1/2}BD_G^{-1/2}\right\|_* \geq n\varepsilon\,.$$

Note that $\mathrm{nnz}(B) = \mathrm{nnz}(D_G^{-1/2}BD_G^{-1/2})$, since D_G is a diagonal matrix. Since B was an arbitrary matrix with just sparsity constraint, we can represent any arbitrary matrix M with the same sparsity as B, as $M = D_G^{-1/2}BD_G^{-1/2}$. Hence, we get that for any matrix M with

$$\mathrm{nnz}(M) \leq \left(\frac{c'(c'')^2}{64 \log^2\left(\frac{c''/4}{\varepsilon}\right)}\right) \cdot \frac{n}{\varepsilon^2}, \qquad \|N_G - M\|_* \geq n\varepsilon\,. \qquad \blacksquare$$

4.2 Query Complexity Lower Bound

Next, we prove that our algorithm from Section 3 also achieves the optimal query complexity for nuclear approximation, up to logarithmic factors in ε^{-1}. Interestingly, our lower bounds apply in an *even more general* query model which supports edge queries in addition to neighbor queries.

Definition 4.5 (Generalized Adjacency Query Model)**.** We say we have *generalized adjacency query access* to a weighted graph $G = (V, E, w)$ if in addition to the access under the adjacency query model (Definition 1.4), there is an $O(1)$ time procedure, $\mathsf{GetEdge}(u, v)$ that when queried with any $u, v \in V$ returns True if $\{u, v\} \in E$ and False otherwise.

To begin, we state a lemma that allows us to lower bound the nuclear norm of a block-diagonal matrix in terms of the nuclear norm of its blocks.

Lemma 4.6. *Let $A \in \mathbb{R}^{2n \times 2n}$ be of the form $\begin{pmatrix} B & C \\ D & E \end{pmatrix}$, for $B, C, D, E \in \mathbb{R}^{n \times n}$. Then, $\|A\|_* \geq \|B\|_*$.*

Proof. By the dual characterization of the nuclear norm, there exists a matrix $X \in \mathbb{R}^{n \times n}$ with $\|X\|_2 = 1$ such that $\langle B, X \rangle = \|B\|_*$. Consider $\left\|\widetilde{X}\right\|_2 = \begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{2n \times 2n}$. Then, $\left\|\widetilde{X}\right\|_2 = 1$, and $\|A\|_* \geq \langle A, \widetilde{X} \rangle = \langle B, X \rangle = \|B\|_*$. ■

Next, we show that the difference matrix between an Erdos-Renyi matrix and its complement is unsparsifiable in the nuclear norm.

Lemma 4.7. *Let $G = (V, E)$ be an Erdos-Renyi $G(n, 1/2)$ graph. Let $G' = (V, E')$ denote the complement of G. Then, for $n \geq 4 \cdot 10^6$, with probability $5/8$,*

$$\left\|(A_G - A_{G'}) - B'\right\|_* \geq n^{1.5}/1000.$$

for any B' with $\mathrm{nnz}(B)' \leq n^2/(16 \cdot 10^6 \log^2(n))$.

Proof. Let $Z \in \mathbb{R}^{n \times n}$ be defined as $Z \stackrel{\text{def}}{=} 1 - I$ where 1 denotes the all-ones matrix. Then,

$$\|A_G - A_{G'}\| * = \|A_G - (Z - A_G)\|_* = \|2A_G - Z\|_* .$$

Consider any B' with $\mathrm{nnz}(B)' \leq n^2/(16 \cdot 10^6 \log^2(n))$. Then, with probability $5/8$, $\|2A_G - B'\|_* \geq n^{1.5}/500$. Thus, for $n \geq 4 \cdot 10^6$, by triangle inequality we have that

$$\left\|(2A_G - Z) - B'\right\|_* \geq \left\|2A_G - B'\right\|_* - \|Z\|_* \geq n^{1.5}/500 - 2(n-1) \geq (1/1000) \cdot n^{1.5}.$$ ■

Next, we show that any algorithm that makes $o(n^2/\log^2(n))$ queries to the generalized graph access model (Definition 4.5) is unable to distinguish between a random Erods-Renyi graph and the graph which is the complement restricted to the edges which were not queried.

Lemma 4.8. *Let a be any algorithm such that on any n-node unweighted graph $G = (V, E)$, a makes $Q_a(G)$ queries to the generalized adjacency query model (Definition 4.5) such that $|Q_a(G)| \leq n\varepsilon^{-2}/(64 \cdot 10^6 \log^2(n))$ and outputs $X_a(G) \in \mathbb{R}^{n \times n}$. There exists an $\varepsilon_0 \in (0, 1)$ such that for any $\varepsilon \in (0, \varepsilon_0)$ and $n_0 \geq 4 \cdot 10^6$, there exists a distribution q over $n \geq n_0$ nodes such that the following holds:*

$$\sum_{S \in \mathcal{S}} \mathbb{P}\left\{\|X_a(G) - N_E\|_* > n\varepsilon | Q_a(G) = S\right\} \mathbb{P}\left\{Q_a(G) = S\right\} \geq 9/16,$$

where $\mathcal{S} = \{E' \subset E : |E'| \leq n\varepsilon^{-2}/(64 \cdot 10^6 \log^2(n))\}$.

Proof. For ε_0 sufficiently small, we can set b such that $b \stackrel{\text{def}}{=} 1/(16 \cdot 4000(^2\varepsilon^2)) \geq 3000$. Now, set $k = \lceil n_0/2b \rceil$. Note that for $n_0 \geq 4 \cdot 10^6$, we have $k \geq 150$.

Constructing q: We will construct the distribution q over graphs on $n \stackrel{\text{def}}{=} 2kb$ nodes as follows. Let $V = \bigsqcup_{r\in[k]} V^{r,1} \sqcup V^{r,2}$, where $|V^{r,1}| = |V^{r,2}| = b$ for each $r \in [k]$. We use $v_j^{r,1}$ and $v_j^{r,2}$ to denote the j-th vertex in $V^{r,1}$ and $V^{r,2}$, respectively.

For each $i, j \in [b]$ and $r \in [k]$, we draw $X_{i,j,r} \sim \text{Ber}(1/2)$. If $X_{i,j,r} = 1$, we add the edges $\{v_i^{r,1}, v_j^{r,1}\}, \{v_i^{r,2}, v_j^{r,2}\}$ to E. Meanwhile, if $X_{i,j,r} = 0$, we add the edges $\{v_i^{r,1}, v_j^{r,2}\}, \{v_i^{r,2}, v_j^{r,1}\}$ to E. Each $G = (V, E) \sim q$ now has n vertices and is $d \stackrel{\text{def}}{=}: (b-1)$-regular. Consequently, the degree oracle does not provide any information for graphs drawn from q.

Constructing the complement with respect to $Q_a(G)$: With each realization $G \sim q$, we shall associate two graphs $G' = (V, E')$ and $\bar{G} = (V, \bar{E})$ defined as follows. For each $r \in [k]$,

- If $X_{i,j,r} = 1$, we add the edge $\{v_i^{r,1}, v_j^{r,2}\}, \{v_i^{r,2}, v_j^{r,1}\}$ to E', and if $X_{i,j,r} = 0$, we add the edge $\{v_i^{r,1}, v_j^{r,1}\}, \{v_i^{r,2}, v_j^{r,2}\}$ to E'.
- Construct $\bar{E}$ by modifying edges in E' as follows. Initialize $\bar{E} = E'$. Now, if any of $\{v_i^{r,1}, v_j^{r,2}\}$, $\{v_i^{r,2}, v_j^{r,1}\}$, $\{v_i^{r,1}, v_j^{r,1}\}$, or $\{v_i^{r,2}, v_j^{r,2}\}$ are in $Q_a(G)$, then delete any edges of the form $\{v_i^{r,1}, v_j^{r,1}\}$ $\{v_i^{r,2}, v_j^{r,2}\}$ and replace them with $\{v_i^{r,1}, v_j^{r,2}\}$ and $\{v_i^{r,2}, v_j^{r,1}\}$; likewise delete any edges $\{v_i^{r,1}, v_j^{r,2}\}$ and $\{v_i^{r,2}, v_j^{r,1}\}$ and replace them with $\{v_i^{r,1}, v_j^{r,1}\}$ and $\{v_i^{r,2}, v_j^{r,2}\}$.

In other words, within each component of b nodes in G, G' is the complement of G. Meanwhile, within each component of b nodes in G, $\bar{G}$ is the complement of G on all edges except those revealed by the queries in $Q_a(G)$ (while on the edges revealed by the queries, $\bar{G}$ matches G). Each $\bar{G} \in \mathsf{supp}(q)$ corresponds to a unique $G \in \mathsf{supp}(q)$, and q assigns equal mass to G and $\bar{G}$. Moreover, $\bar{G}$ is also d-regular.

Indistinguishably between G and $\bar{G}$: For $S \subset V$, we use $G[S]$ to denote the vertex-induced subgraph of G corresponding to S. Observe that $A_{G[V^{r,1}]} \sim G(b, 1/2)$ for each $r \in [k]$, where Let $\mathcal{T}_{c',c''}$ denote the set of all graphs H such that

$$\|A_H - B'\|_* \geq c''b^{1.5}, \text{ for all } B' \text{ with } \text{nnz}(B)' \leq c'b^2/\log^2(b).$$

Proposition 4.3 ensures that if H is an Erdos-Renyi $G(n, 1/2)$ graph, then $H \in \mathcal{T}_{1/500,1/(16\cdot 10^6)}$ with probability at least $5/8$. Moreover, Lemma 4.6 ensures that if $G[V^{r,1}] \in \mathcal{T}_{1/500,1/(16\cdot 10^6)}$ for some $r \in [k]$, then $G[V^{r,1} \cup V^{r,2}] \in \mathcal{T}_{1/500,1/(16\cdot 10^6)}$ as well.

Let $\mathcal{H} \stackrel{\text{def}}{=} \left\{G \in \mathsf{supp}(q) : G[V^{r,1} \cup V^{r,2}] \in \mathcal{T}_{1/500,1/(16\cdot 10^6)} \text{ for at least } 9k/16 \text{ values of } r \in [k]\right\}$. By the independence of the k components in G and Chernoff bound, we have that $\mathbb{P}_{G=(V,E)\sim q}\{G \in \mathcal{H}\} \geq 1 - \exp(-1/100 \cdot 5k/8) = 1 - \exp(-5n/(1600b)) \geq 9/16$ whenever $k \geq 150$.

Consequently, at least $9/16$'s of the graphs in $\mathsf{supp}(q)$ are in $\mathcal{H}$. So, to prove the lemma, it suffices to show that for every $G \in \mathcal{H}$, $Q_a(G) = Q_a(\bar{G})$ and

$$\max\left\{\|N_{\bar{G}} - X_a(G)\|_*, \|N_G - X_a(G)\|_*\right\} \geq \frac{1}{2}\|N_{\bar{G}} - N_G\|_* \geq 2n\varepsilon. \tag{3}$$

So, assume $G \in \mathcal{H}$, and without loss of generality, assume that $G[V^{r,1} \cup V^{r,2}] \in \mathcal{T}_{1/500,1/(16\cdot 10^6)}$ for all $r \in [9/16 \cdot k]$. $Q_a(G) = Q_a(\bar{G})$ by construction, so it remains to prove (3). In the following, we

use $X^{a:b,a:b}$ to denote the block of X corresponding to rows and columns. $\{a, a+1, ..., b-1, b\}$. The r-th block of $N_G - N_{\bar{G}}$ has the following form for $r \in [9/16 \cdot k]$:

$$(N_G - N_{\bar{G}})^{rb+1:r(b+1)} = \frac{1}{b-1}\left((A_G - A_{G'})^{rb+1:r(b+1)} - P'_r\right),$$

where P'_r and $\tilde{P}'_r \stackrel{\text{def}}{=} \frac{1}{b-1}P'_r$ is the edge perturbation matrix in the r-th block corresponding to the queries in $Q_a(G)$. Now, Lemma 4.7 guarantees that for $r \in [9/16 \cdot k]$,

$$(A_G - A_{G'})^{rb+1:r(b+1)} \in \mathcal{T}_{1/1000,1/(16\cdot 10^6)}.$$

Let P' be the block diagonal matrix containing P'_r in the r-th block. Since $Q_a(G)$ contains at most $n\varepsilon^{-2}/(64\cdot 10^6 \log^2(n))$ queries, P' is $n\varepsilon^{-2}/(16\cdot 10^6 \log^2(n))$-sparse. Consequently, by Proposition 4.4, it follows that

$$\left\|A_G - A_{G'} - P'\right\|_* \geq 9n\sqrt{b}/(16 \cdot 4000),$$

and

$$\|N_G - N_{\bar{G}}\|_* \geq \frac{9n\sqrt{b}}{16 \cdot 4000(b-1)} \geq \frac{9n}{16 \cdot 4000\sqrt{b}} \geq 2n\varepsilon. \qquad \blacksquare$$

Finally, we can prove our lower bound on the query complexity for nuclear sparsification using Yao's min-max principle.

Theorem 4.9. *There exists a constant $\varepsilon_0 > 0$, $n_0 \geq 4 \cdot 10^6$ such that for any $\varepsilon \in (0, \varepsilon_0)$ and $n \geq n_0$, there exists $G = (V, E)$ on n nodes such that any randomized algorithm which calls the generalized graph access model (Definition 4.5) at most $n\varepsilon^{-2}/(64 \cdot 10^6 \log^2(n))$ times and outputs $X \in \mathbb{R}^{n\times n}$ has the following property: $\mathbb{P}\left[\|N_G - X\|_* > \varepsilon n\right] > 9/16$.*

Proof. Let $\mathcal{A}$ be any set of algorithms such that for any $a \in \mathcal{A}$ and n node unweighted graph $G = (V, E)$ on n nodes, a makes $Q_a(G)$ queries to the graph access model such that $|Q_a(G)| = n\varepsilon^{-2}/(64\cdot 10^6 \log^2(n))$ and outputs $X_a(G) \in \mathbb{R}^{n\times n}$. Let $c(a, G = (V, E)) \stackrel{\text{def}}{=} 1[\|X_a(G) - N_E\|_* > n\varepsilon]$. Let $\mathcal{G}$ denote the set of all n node undirected unweighted graphs. By Yao's minmax principle, we have that for any distribution p over $\mathcal{A}$ and distribution q over $\mathcal{G}$,

$$\begin{aligned}\max_{G\in\mathcal{G}} \mathbb{P}_{a\sim p}\left\{\|X_a(G) - N_G\|_* > n\varepsilon\right\} &= \max_{G\in\mathcal{G}} \mathop{\mathbb{E}}_{a\sim p}[c(a, G)] \geq \min_{a\in\mathcal{A}} \mathop{\mathbb{E}}_{G\sim q}[c(a, G)] \\ &= \min_{a\in\mathcal{A}} \mathbb{P}_{G=(V,E)\sim q}\left\{\|X_a(G) - N_G\|_* > n\varepsilon\right\}. \qquad (4)\end{aligned}$$

As any randomized algorithm can be formulated as a distribution over deterministic algorithms, to prove the lemma it suffices to lower bound (4) for a particular choice of q. In the remainder of the proof, for conciseness, we may drop the subscripts on the probabilities where it is clear from context.

Now, let $\mathcal{S} = \{S \subset V \times V : |S| < |Q_a(G)|\}$. For any $a \in \mathcal{A}$,

$$\mathbb{P}\left\{\|X_a(G) - N_G\|_* > n\varepsilon\right\} = \sum_{S\in\mathcal{S}} \mathbb{P}\left\{\|X_a(G) - N_G\|_* > n\varepsilon | Q_a(G) = S\right\} \mathbb{P}\left\{Q_a(G) = S\right\},$$

where the probabilities are taken with respect to the random choice of $G = (V, E) \sim q$. Consequently, to prove the lemma, it suffices to show that there exists a distribution q and $\mathcal{H} \subset \mathsf{supp}(q)$, such that for any $S \in \mathcal{S}$ and any $a \in A$, the following hold

$$\sum_{S\in\mathcal{S}} \mathbb{P}\left\{\|X_a(G) - N_G\|_* > n\varepsilon | Q_a(G) = S\right\} \mathbb{P}\left\{Q_a(G) = S\right\} \geq 9/16.$$

Lemma 4.8 guarantees this distribution q, concluding the proof. $\blacksquare$

5 Separation from Additive Spectral Sparsifiers

In this section, we discuss the query lower bounds for any deterministic algorithm that produces ε-additive spectral sparsifiers. The result shows that a separation between nuclear sparsifiers and additive spectral sparsifiers, explaining one of the reasons that this paper consider nuclear sparsifiers for spectral density estimation.

Concretely, here we prove an $\Omega(n^2)$ lower bound on the number of queries to the access model for undirected graphs required by any deterministic ε additive spectral norm approximation algorithm on undirected graphs.

Theorem 5.1. *Let X be the output of any deterministic algorithm which, given an input graph on n nodes, makes at most $\frac{n^2}{128}$ queries to the generalized adjacency query model (Definition 4.5). Then, for any $m \geq 1$ and $n \geq 4m$, there exists an undirected graph $G = (V, E)$ on n nodes along with a vector $v \in \mathbb{R}^V$, such that $\left|v^\top L_E v - X\right| > \frac{1}{8} v^\top v$.*

Proof. Set $V = V_1 \sqcup V_2 \sqcup V_3 \sqcup V_4$ where $|V_i| = m$. We use v_i^r to denote the r-th vertex in V_i.

Let Q denote the set of edges queries by the algorithm. Let E be constructed as follows. If any of $\{v_1^k, v_2^j\}, \{v_1^j, v_2^k\}, \{v_3^k, v_4^j\}$ or $\{v_3^j, v_4^k\} \in Q$ then we include $\{v_1^k, v_2^j\}, \{v_1^j, v_2^k\}, \{v_3^k, v_4^j\}, \{v_3^j, v_4^k\} \in E$.

On the other hand, if any of $\{v_1^k, v_3^j\}, \{v_1^j, v_3^k\}, \{v_2^k, v_4^j\}$ or $\{v_2^j, v_4^k\} \in Q$, then we instead include $\{v_1^k, v_3^j\}, \{v_1^j, v_3^k\}, \{v_2^k, v_4^j\}, \{v_2^j, v_4^k\} \in E$.

Now, let $E_1 = \{\{v_1^k, v_2^j\}, \{v_3^k, v_4^j\} : \{v_i^k, v_\ell^j\} \notin E \text{ for any } i, \ell \in [4]\}$ and $E_2 = \{\{v_1^k, v_3^j\}, \{v_2^k, v_4^j\} : \{v_i^k, v_\ell^j\} \notin E \text{ for any } i, \ell \in [4]\}$. Let $G_1 = (V, E_1 \cup E)$ and $G_2 = (V, E_2 \cup E)$. Note that G_1 and G_2 are both m-regular graphs on $4m$ nodes. Let L_1 be the normalized Laplacian of G_1 and L_2 be the normalized Laplacian of G_2.

Take $v \in \mathbb{R}^V$ to be the vector which is 1 on $V_1 \cup V_2$ and 0 on $V_3 \cup V_4$. We have that for $i \in [2]$,

$$v^\top L_i v = \frac{1}{m} \left|\partial_{G_i}(V_1 \cup V_2)\right|, \qquad v^\top v = 2m,$$

where $\partial_{G_i}(S) \stackrel{\text{def}}{=} \{\{u, v\} \in E_i : u \in S, u \notin S\}$. As each edge in Q fixes at most four edges in E, we have

$$\left|v^\top L_1 v - v^\top L_2 v\right| \geq \frac{1}{m}\left(m^2 - 4\left|Q\right|\right) = m - \frac{4\left|Q\right|}{m} \geq \frac{m}{2} = \frac{1}{4} v^\top v,$$

whenever $4\left|Q\right|/m \leq m/2$, i.e., $|Q| \leq m^2/8 = n^2/128$. So, either $\left|v^\top L_1 v - X\right| \geq \frac{1}{8} v^\top v$ or $\left|v^\top L_2 v - X\right| \geq \frac{1}{8} v^\top v$. ∎

Corollary 5.2. *Any deterministic algorithm for computing an ε additive spectral norm approximation to an undirected graph requires at least $\frac{n^2}{32}$ queries to the generalized adjacency query model (Definition 4.5).*

Proof. Note that given $f : \mathbb{R}^n \mapsto \mathbb{R}$, an ε additive spectral norm approximation to an undirected graph $G = (V, E)$, we can immediately compute a function $g : \mathbb{R}^n \mapsto \mathbb{R}$ defined by $g(x) = x^\top x - f(x)$ such that

$$\left|g(x) - x^\top L_G x\right| = \left|x^\top x - f(x) - x^\top (I - A_G) x\right| = \left|f(x) - x^\top A_G x\right| \leq \varepsilon \left\|x\right\|_2^2.$$

Algorithm 3: $\mathsf{NuclearSparsify}(G, \varepsilon, n, T)$

Input: Graph $G = (V = [n], E, w)$, under the random walk access model (Definition 1.13.)
Input: Accuracy ε
Input: Number of queries $T \in \mathbb{Z}_{\geq 0}$
Output: X, a sparsifier of N_G
Initialize $X \leftarrow 0 \in \mathbb{R}^{n \times n}$
for $t = 1, ..., T$ **do**
  $(i, j, \deg(i), \deg(j)) \leftarrow \mathsf{RandomNeighbor}$
  $X_{i,j} \leftarrow X_{i,j} + 1/(2T) \cdot n \cdot \sqrt{\deg(i)/\deg(j)}$
  $X_{j,i} \leftarrow X_{j,i} + 1/(2T) \cdot n \cdot \sqrt{\deg(i)/\deg(j)}$
Return: X

Note that given f, computing g requires no additional queries to the unweighted graph access oracle. Consequently, the result follows by taking any $\varepsilon < 1/4$ and applying Lemma 5.1. We also note that since the graph we construct is unweighted, the theorem holds for the weighted oracle. ■

6 Sparsification in the Random Walk Model

In this section, we show how our notion of nuclear sparsification (and indeed, ε-additive spectral sparsification) can also be achieved in the random walk model (Definition 1.13 considered by [BKM22].

In Section 6.1 we present sparsification algorithms in this model. In Section 6.2 we present a lower bound that separates the query complexity obtainable for spectral sparsification what is obtainable with nuclear sparsification.

6.1 Algorithms for sparsification in the random walk model

In this section, we present algorithms to build an ε-additive nuclear sparsifier and an ε-additive spectral sparsifier under the random walk access model (Definition 1.13). In Theorem 1.14, we show that by taking $T = O(n\varepsilon^{-2})$ in Algorithm 3, with constant probability we can obtain a nuclear sparsifier. In Theorem 1.15, we show that by taking $T = O(n\varepsilon^{-2}\log(n))$, we can obtain an ε-additive spectral sparsifier.

Theorem 1.14. *Given access to a graph G under the random walk model (Definition 1.13) and some $\varepsilon \in (0, 1)$, for $T \geq 2n\varepsilon^{-2}$, with probability $2/3$, $X = \mathsf{NuclearSparsify}(G, \varepsilon, n, T)$ (Algorithm 3) is an $O(n\varepsilon^{-2})$-sparse ε-additive nuclear sparsifier. Moreover, X is computed using only $O(n\varepsilon^{-2})$ queries to the random walk model.*

Proof. Let $i^{(t)} \sim \mathsf{Uniform}([n])$ and $j^{(t)} = j$ with probability w_{ij} for each $j \in [n]$, independently for each $t \in [T]$. Let $A^{(t)} \in \mathbb{R}^{V \times V}$ be the matrix where

$$A^{(t)}_{i^{(t)}, j^{(t)}} = n \cdot \sqrt{\deg(i^{(t)})/\deg(j^{(t)})}.$$

and all other entries are 0. For all $i, j \in V$

$$\mathbb{E}[A^{(t)}]_{i,j} = \left[\frac{1}{n} \cdot \frac{w_{i,j}}{\deg(i)}\right] \cdot \left[n \cdot \sqrt{\deg(i)/\deg(j)}\right] = \frac{w_{i,j}}{\sqrt{\deg(i) \cdot \deg(j)}} = N_{i,j}\,.$$

Consequently, as N is symmetric, $\mathbb{E}\,\frac{1}{2}(A^{(t)}+A^{(t)\top})=N$ and $\mathbb{E}\,\frac{1}{2T}\sum_{t\in[T]}A^{(t)}+A^{(t)\top}=N$. Using the independence of $A^{(t)}$ for $t\in[T]$, and linearity of expectation, we get that

$$\begin{aligned}
\mathbb{E}\left\|\left(\frac{1}{2T}\sum_{t\in[T]}A^{(t)}+A^{(t)\top}\right)-N\right\|_F^2 &= \mathbb{E}\left\|\frac{1}{2T}\left(\sum_{t\in[T]}A^{(t)}+A^{(t)\top}-2N\right)\right\|_F^2 \\
&= \frac{1}{4T^2}\,\mathbb{E}\,\mathrm{tr}\left(\left(\sum_{t\in[T]}A^{(t)}+A^{(t)\top}-2N\right)^\top\left(\sum_{t\in[T]}A^{(t)}+A^{(t)\top}-2N\right)\right) \\
&= \frac{1}{4T^2}\mathrm{tr}\left(\sum_{t\in[T]}\mathbb{E}\left(\left(A^{(t)}+A^{(t)\top}-2N\right)^\top\left(A^{(t)}+A^{(t)\top}-2N\right)\right)\right) \\
&= \frac{1}{4T}\,\mathbb{E}\left(\left\|A^{(t)}+A^{(t)\top}-2N\right\|_F^2\right)=\frac{1}{T}\,\mathbb{E}\left(\left\|\frac{1}{2}\left(A^{(t)}+A^{(t)\top}\right)-N\right\|_F^2\right).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{E}\left(\left\|\frac{1}{2}\left(A^{(1)}+A^{(1)\top}\right)-N\right\|_F^2\right) &\le \mathbb{E}\left(\left\|\frac{1}{2}\left(A^{(1)}+A^{(1)\top}\right)\right\|_F^2\right) \\
&= \sum_{i\in[n]}\sum_{j\in[n]}\frac{w_{i,j}}{n\deg(i)}\left[n\cdot\sqrt{\deg(i)/\deg(j)}\right]^2=n^2.
\end{aligned}$$

Note that X has the same distribution as $\frac{1}{2}\left(A^{(1)}+A^{(1)\top}\right)$. So, by Markov's inequality for the second moment,

$$\mathbb{P}\left\{\|X-N\|_*\ge n\varepsilon\right\}\le\mathbb{P}\left\{\|X-N\|_F\ge\sqrt{n}\varepsilon\right\}\le\frac{(1/T)n^2}{n\varepsilon^2},$$

so, picking $T=3n\varepsilon^{-2}$ suffices to succeed with probability $2/3$. ■

By combining the sparsification procedure in Algorithm 3 with the Chebyshev moment-matching algorithm of [BKM22], we obtain a $O(n\varepsilon^{-3})$ algorithm for spectral density estimation in this random walk query model (Definition 1.13).

In the following Theorem, we note that by using a result of [Coh+17] we can also achieve the stronger notion of ε-additive spectral sparsification in the random walk access model (Definition 1.13.) However, the cost of doing so is an additional $O(\log(n))$-factor in the number of queries.

Theorem 6.1 (Theorem 3.9 (Simplified), Cohen et al. [Coh+17])**.** *Let $A\in\mathbb{R}^{n\times n}$ be a symmetric matrix where no row or column is all zeros. The $\varepsilon,p\in(0,1)$. Let $r=A\mathbf{1}$ where $\mathbf{1}$ is the all ones vector. Let E_{ij} denote a matrix whose (i,j)-th entry is 1 and rest of the entries are zero, and $\mathcal{D}$ be a distribution over $\mathbb{R}^{n\times n}$ such that $X\sim\mathcal{D}$ takes value*

$$X=\left(\frac{A_{ij}}{p_{ij}}\right)E_{ij}\ \text{with probability}\ p_{ij}=\frac{A_{ij}}{s}\left[\frac{1}{r_i}+\frac{1}{r_j}\right]\forall A_{ij}\neq 0\,.$$

If $A_i,\ldots,A_k$ are sampled independently from $\mathcal{D}$ for $T\ge 128\frac{2n}{\varepsilon^2}\log\frac{2n}{p}$, $R=\mathrm{diag}(r)$ then the average $\tilde{A}\overset{\mathrm{def}}{=}\frac{1}{T}\sum_{t\in[T]}A_t$ satisfies

$$\mathbb{P}\left\{\left\|R^{-1/2}(\tilde{A}-A)R^{-1/2}\right\|_2\ge\varepsilon\right\}\le p\,.$$

Algorithm 4: SpectralAdditiveSparsify(G, ε, n, T)

Input: Graph $G = (V = [n], E, w)$, under the random walk access model (Definition 1.13.)
Input: Accuracy ε
Input: Number of queries $T \in \mathbb{Z}_{\geq 0}$
Output: X, a sparsifier of N_G
Initialize $X \leftarrow 0 \in \mathbb{R}^{n \times n}$
for $t = 1, ..., T$ **do**
  $(i, j, \deg(i), \deg(j)) \leftarrow$ RandomNeighbor
  $Z \leftarrow$ Bernoulli$(1/2)$
  if $Z = 1$ **then**
    $X_{i,j} \leftarrow X_{i,j} + \frac{2n}{\sqrt{\deg(i)}\sqrt{\deg(j)}} \left(\frac{1}{\deg(i)} + \frac{1}{\deg(j)}\right)^{-1}$
  else
    $X_{j,i} \leftarrow X_{j,i} + \frac{2n}{\sqrt{\deg(i)}\sqrt{\deg(j)}} \left(\frac{1}{\deg(i)} + \frac{1}{\deg(j)}\right)^{-1}$
Return: X

Theorem 1.15. *Given access to a graph G under the random walk model (Definition 1.13) and some $\varepsilon \in (0, 1)$, for $T \geq 256 n \varepsilon^{-2} \log(3n)$, with probability $2/3$, $X =$ SpectralAdditiveSparsify(G, ε, n, T) (Algorithm 4) is an ε-additive $O(n\varepsilon^{-2}\log(n))$-sparse ε-additive spectral sparsifier for an unweighted graph G. Moreover, X is computed using $O(n\varepsilon^{-2}\log(n))$ queries to the random walk model.*

Proof. Let $i^{(t)} \sim$ Uniform$([n])$ and $j^{(t)} = j$ with probability w_{ij} for each $j \in [n]$, independently for each $t \in [T]$. Let $A^{(t)} \in \mathbb{R}^{V \times V}$ be the matrix where with probability $1/2$, either

$$A^{(t)}_{i^{(t)}, j^{(t)}} = \frac{2n}{\sqrt{\deg(i^{(t)})}\sqrt{\deg(j^{(t)})}} \left(\frac{1}{\deg(i^{(t)})} + \frac{1}{\deg(j^{(t)})}\right)^{-1}.$$

or else

$$A^{(t)}_{j^{(t)}, i^{(t)}} = \frac{2n}{\sqrt{\deg(i^{(t)})}\sqrt{\deg(j^{(t)})}} \left(\frac{1}{\deg(i^{(t)})} + \frac{1}{\deg(j^{(t)})}\right)^{-1}.$$

All other entries are 0. Note that then,

$$A_{i,j} = \frac{2n}{\sqrt{\deg(i)}\sqrt{\deg(j)}} \left(\frac{1}{\deg(i)} + \frac{1}{\deg(j)}\right)^{-1},$$

with probability $\frac{1}{2n} \cdot \left(\frac{w_{ij}}{\deg(i)} + \frac{w_{ij}}{\deg(j)}\right)$. By Theorem 6.1, setting $A = N_G$, $A_i = A^{(t)}$, $p = 2/3$, the result follows. ∎

6.2 Separation between spectral and nuclear sparsification

In this section, we show that there exists a constant c such that any algorithm for constructing an c-additive spectral sparsifier for some must make at least $\Omega(n \log(n))$ queries to the random walk graph access oracle. Our result relies on the following (slightly modified) classical result on the sample complexity for the coupon collector problem.

Lemma 6.2 (Coupon Collector)**.** *Suppose there is a universe of $2n$ labeled balls in which there are n labeled black balls and n labeled red balls. At each trial, a call is sampled uniformly at random from the universe, with replacement. Let T be a random variable which denotes the number of trials needed to see all the labels of the n red balls. Then,*

$$\mathbb{P}\left[T \leq 2n \log n - 2cn\right] \leq \exp(-c)\,.$$

In the following lemma, we obtain our lower bound for the query complexity for ε-additive spectral sparsification by reducing to the coupon collector problem.

Lemma 1.16. *Let $0 < \varepsilon < 1/8$ be a constant. Given access to a graph G under the random walk model (Definition 1.13), any algorithm that uses $o(n \log n)$ oracle queries cannot output a spectral sparsifier with a constant probability.*

Proof. Consider a bi-partite graph $G = (V, E)$, where $V = A \sqcup B$, such that $|A| = |B| = 2n$. Let $a_1, \ldots, a_{2n}$ denote the labels of the vertices in A, and $b_1, \ldots, b_{2n}$ denote the vertices of B. Suppose the graph G has n edges of the form (a_i, b_i) for $i \in [2n]$. Suppose the algorithm knows the labels of the vertices and that it is a graph with n edges. The labels of the edges are unknown to the algorithm. The graph has $3n$ connected components, where each connected component is either a singleton vertex or an edge. The algorithm can go from one connected component to another connected component using only the RandomNeighbor query. For $i \in [2n]$, when the RandomNeighbor query sees a vertex a_i (or b_i) it can figure out if there is an edge (a_i, b_i).

Let A denote the adjacency matrix of G. Let $\mathcal{I} \stackrel{\text{def}}{=} \{i : (a_i, b_i) \in E\}$. The RandomNeighbor query either returns a vertex a_i or b_i for $i \in [2n] \setminus \mathcal{I}$, or a vertex a_i or b_i for $i \in \mathcal{I}$. The RandomNeighbor query corresponds to random sampling from the universe in the coupon collector problem. The output a_i or b_i for $i \in \mathcal{I}$ corresponds to the labeled red ball in the coupon collector problem. The output a_i or b_i for $i \notin \mathcal{I}$ corresponds to a black ball in the coupon collector problem. Suppose there exists an algorithm that takes $T \leq 2n \log n - 2\log(2/c)n$ RandomNeighbor queries and outputs a spectral sparsifier $\widetilde{A}$ of A such that $\left\|A - \widetilde{A}\right\|_2 \leq \varepsilon$ with constant probability c. Then, we will show that we can solve the coupon collector problem in $T \leq 2n \log n - 2\log(2/c)n$ queries with constant probability c, which leads to a contradiction.

This follows because we can figure out the labels of edges in the graph G from $\widetilde{A}$ by taking $x \in \mathbb{R}^{4n}$ indexed by a_i and b_i, for $i \in [2n]$. We take x to be all 0 vector except at two coordinates where it is 1, i.e., $x_{a_i} = x_{b_i} = 1$, and $x_k = 0$ for $k \neq a_i, b_i$. Then, we get that if $3/4 \leq x^\top \widetilde{A} x \leq 5/4$ then $(a_i, b_i) \in E$ and if $-1/4 \leq x^\top \widetilde{A} x \leq 2/4$, then $(a_i, b_i) \notin E$.

Therefore, any algorithm that takes $T \leq n \log n - \log(2/c)n$ oracle queries, fails to produce a spectral sparsifier, with probability greater than c.

Note that for this instance since the degree is at most 1, the k-step random walk query model is no more informative than a 1-step random walk model. Therefore, this lower bound holds for the k-step query model as well. ∎

A Preliminaries Appendix

Fact A.1 (Matrix Norms)**.** *For any symmetric matrix $A \in \mathbb{R}^{n \times n}$, we have*

1. *Norm Inequalities: $\|A\|_* \leq \sqrt{n}\,\|A\|_{\mathrm{F}}$, $\|A\|_{\mathrm{F}} \leq \sqrt{n}\,\|A\|_2$.*

2. Dual Characterization of Nuclear Norm: $\|A\|_* = \max_{X:\|X\|_2\le 1} \langle A, X\rangle = \max_{X:\|X\|_2=1} \langle A, X\rangle$.

B Omitted Proofs

Lemma 1.3. *For symmetric matrices $A, B \in \mathbb{R}^{n\times n}$, if $\|A-B\|_* \le n\varepsilon$, then $W_1(A,B) \le \varepsilon$.*[6]

Proof. Note that the first fact immediately follows from the norm inequality of $\frac{1}{n}\|M-N\|_* \le \|M-N\|_2$. For the second fact, we note that for sorted eigenvalue lists of $M, N, M-N$ in form $\lambda_i(M)$ and $\lambda_i(N)$, $\lambda_i(M-N)$, $i \in [n]$, $\|M-N\|_* = \sum_{i\in[n]} \lambda_i(M-N) \ge \sum_{i\in[n]} |\lambda_i(M) - \lambda_i(N)| = nW_1(M,N)$, where the inequality holds due to the Mirsky Singular Value Perturbation Inequality ([Mir60], Theorem 5). ■

Lemma B.1. *Let $G=(V,E,w)$ be a weighted graph with weighted adjacency matrix $A_G \in \mathbb{R}^{n\times n}$ and degree matrix D_G and an accuracy parameter $\varepsilon \in (0,1)$. Without loss of generality, assume that $V=[n]$ and the vertices in V are ordered such that $\deg(i) \ge deg(n)$ for all $i \in [n]$. Define G' as in Theorem 3.1, and let $\mathbf{e} \in \mathbb{R}^{n-1}$ be defined as follows: $e_i \stackrel{\text{def}}{=} \deg(i) - \sum_{j=1}^n (A_{G'})_{i,j}$. Let G'' be a graph with adjacency matrix $A_{G''} = \begin{pmatrix} (A_{G'})_{[1:n-1]} & \mathbf{e} \\ \mathbf{e}^\top & 0 \end{pmatrix}$ with degree matrix $D_{G''} = \begin{pmatrix} (D_G)_{[1:n-1]} & \mathbf{0} \\ \mathbf{0}^\top & \|\mathbf{e}\|_1 \end{pmatrix}$, where for a matrix $M \in \mathbb{R}^{n\times n}$, we denote $M_{[1:j]}$ as the leading principal submatrix of order j, for $j \in [n]$. Then, $\|N_G - N_{G''}\|_* \le 3n\varepsilon$ whenever $\varepsilon^{-2} \le n$.*

Proof. It suffices to show that $\left\|N_G - D_{G''}^{-1/2} A_{G''} D_{G''}^{-1/2}\right\|_F^2 \le 9n\varepsilon^2$. Note that the Frobenius norm of the upper left $(n-1)\times(n-1)$ block of the matrix $N_G - D_{G''}^{-1/2} A_{G''} D_{G''}^{-1/2}$ is already bounded by $n\varepsilon^2$ by Lemma 3.1. Consequently,

$$\left\|N_G - D_{G''}^{-1/2} A_{G''} D_{G''}^{-1/2}\right\|_F^2 \le n\varepsilon + 2\sum_{i=1}^{n-1} \left(\frac{w_{in}}{\sqrt{\deg(i)}\sqrt{\deg(n)}} - \frac{e_i}{\sqrt{\deg(i)}\sqrt{\|\mathbf{e}\|_1}}\right)^2.$$

To bound the second term, observe that since $e_i \in [0, \deg(i)]$, $\sum_{i=1}^{n-1} \frac{e_i^2}{\deg(i)\|\mathbf{e}\|_1} \le \frac{1}{\|\mathbf{e}\|_1} \sum_{i=1}^{n-1} e_i = 1$. Likewise, since $\deg(n) \le \deg(i)$ we have that whenever $\varepsilon^{-2} \le n$,

$$\begin{aligned} 2\sum_{i=1}^{n-1} \left(\frac{w_{in}}{\sqrt{\deg(i)}\sqrt{\deg(n)}} - \frac{e_i}{\sqrt{\deg(i)}\sqrt{\|\mathbf{e}\|_1}}\right)^2 &\le 2\sum_{i=1}^{n-1} \frac{w_{in}^2}{\deg(i)\deg(n)} + 2\sum_{i=1}^{n-1} \frac{e_i^2}{\deg(i)\,\|\mathbf{e}\|_1} \\ &\le \frac{2}{\deg(n)^2}\left(\sum_{i=1}^{n-1} w_{in}\right)^2 + 2 \le 4 \le 4\varepsilon^2 n, \end{aligned}$$

completing the proof. ■

References

[BLHK16] Olivier Bachem, Mario Lucic, S Hamed Hassani, and Andreas Krause. "Approximate k-means++ in sublinear time". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 30. 1. 2016.

[6] When A and B are diagonal with entries sorted in decreasing order, Lemma 1.3 can be strengthened to an if and only if. In this sense, nuclear approximation is a natural strengthening of approximation in the Wasserstein-1 distance.

[BH16] Afonso S. Bandeira and Ramon van Handel. “Sharp nonasymptotic bounds on the norm of random matrices with independent entries”. In: *The Annals of Probability* 44.4 (2016), pp. 2479–2506.

[BGKS20] Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. “Pseudospectral Shattering, the Sign Function, and Diagonalization in Nearly Matrix Multiplication Time”. In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. 2020, pp. 529–540.

[BSST13] Joshua Batson, Daniel A Spielman, Nikhil Srivastava, and Shang-Hua Teng. “Spectral sparsification of graphs: theory and algorithms”. In: *Communications of the ACM* 56.8 (2013), pp. 87–94.

[BK96] András A. Benczúr and David R. Karger. “Approximating S-t Minimum Cuts in Õ(N2) Time”. In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*. STOC ’96. Association for Computing Machinery, 1996, pp. 47–55.

[BKM14] Petra Berenbrink, Bruce Krayenhoff, and Frederik Mallmann-Trenn. “Estimating the number of connected components in sublinear time”. In: *Information Processing Letters* 114.11 (2014), pp. 639–642.

[Bha+23] Rajarshi Bhattacharjee, Gregory Dexter, Petros Drineas, Cameron Musco, and Archan Ray. “Sublinear Time Eigenvalue Approximation via Random Sampling”. In: *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*. Ed. by Kousha Etessami, Uriel Feige, and Gabriele Puppis. Vol. 261. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 21:1–21:18.

[BKM22] 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)*. 2022, pp. 1144–1157.

[CRT05] Bernard Chazelle, Ronitt Rubinfeld, and Luca Trevisan. “Approximating the minimum spanning tree weight in sublinear time”. In: *SIAM Journal on computing* 34.6 (2005), pp. 1370–1379.

[CTU21] 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.

[Coh+17] Michael B Cohen, Jonathan Kelner, John Peebles, Richard Peng, Anup B Rao, Aaron Sidford, and Adrian Vladu. “Almost-linear-time algorithms for markov chains and new spectral primitives for directed graphs”. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. 2017, pp. 410–419.

[CKSV18] 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)*. 2018, pp. 1263–1271.

[Czu+03] Artur Czumaj, Funda Ergün, Lance Fortnow, Avner Magen, Ilan Newman, Ronitt Rubinfeld, Christian Sohler, et al. “Sublinear-time approximation of Euclidean minimum spanning tree.” In: *SODA*. 2003, pp. 813–822.

[CS10] Artur Czumaj and Christian Sohler. “Sublinear-time Algorithms”. In: *Property Testing: Current Research and Surveys*. Ed. by Oded Goldreich. Springer Berlin Heidelberg, 2010, pp. 41–64.

[DDH07] James Demmel, Ioana Dumitriu, and Olga Holtz. “Fast linear algebra is stable”. In: *Numerische Mathematik* 108.1 (2007), pp. 59–91.

[DBB19] 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)*. 2019, pp. 1152–1161.

[DWZ22] Ran Duan, Hongxun Wu, and Renfei Zhou. "Faster matrix multiplication via asymmetric hashing". In: *arXiv preprint arXiv:2210.10173* (2022).

[Ede+22] Talya Eden, Nimrod Fiat, Orr Fischer, Fabian Kuhn, and Rotem Oshman. "Sublinear-time distributed algorithms for detecting small cliques and even cycles". In: *Distributed Computing* (2022), pp. 1–28.

[ERS17] Talya Eden, Dana Ron, and C Seshadhri. "Sublinear time estimation of degree distribution moments: The degeneracy connection". In: *44th International Colloquium on Automata, Languages, and Programming (ICALP 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik. 2017.

[ERS18] Talya Eden, Dana Ron, and C Seshadhri. "On approximating the number of k-cliques in sublinear time". In: *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*. 2018, pp. 722–734.

[EG17] 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)*. 2017, pp. 817–826.

[FDBV06] Illes J. Farkas, Imre Derenyi, A.-L. Barabási, and Tamas Vicsek. "Spectra of "real-world" graphs: Beyond the semicircle law". In: *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[Glu+21] Grzegorz Gluch, Michael Kapralov, Silvio Lattanzi, Aida Mousavifar, and Christian Sohler. "Spectral clustering oracles in sublinear time". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 1598–1617.

[GZ00] Alice Guionnet and Ofer Zeitouni. "Concentration of the Spectral Measure for Large Matrices". In: *Electronic Communications in Probability* 5.none (2000), pp. 119–136.

[JMSS23] Yujia Jin, Christopher Musco, Aaron Sidford, and Apoorv Vikram Singh. "Moments, Random Walks, and Limits for Spectrum Approximation". In: *Proceedings of Thirty Sixth Conference on Learning Theory*. Vol. 195. Proceedings of Machine Learning Research. PMLR, 2023, pp. 5373–5394.

[KV17] Weihao Kong and Gregory Valiant. "Spectrum estimation from samples". In: *The Annals of Statistics* 45.5 (2017), pp. 2218–2247.

[Lee13] Yin Tat Lee. *Probabilistic Spectral Sparsification In Sublinear Time*. 2013.

[LS18] Yin Tat Lee and He Sun. "Constructing linear-sized spectral sparsification in almost-linear time". In: *SIAM Journal on Computing* 47.6 (2018), pp. 2315–2336.

[LSY16] Lin Lin, Yousef Saad, and Chao Yang. "Approximating Spectral Densities of Large Matrices". In: *SIAM Review* 58.1 (2016), pp. 34–65.

[MAC20] Jay Mardia, Hilal Asi, and Kabir Aladin Chandrasekher. "Finding planted cliques in sublinear time". In: *arXiv preprint arXiv:2004.12002* (2020).

[Mir60] L. Mirsky. "Symmetric Gauge Functions and Unitarily Invariant Norms". In: *The Quarterly Journal of Mathematics* 11.1 (Jan. 1960), pp. 50–59.

[MOP01] Nina Mishra, Daniel Oblinger, and Leonard Pitt. "Sublinear time approximate clustering." In: *SODA*. Vol. 1. 2001, pp. 439–447.

[Mou21] Aidasadat Mousavifar. *Sublinear Algorithms for Spectral Graph Clustering*. 2021.

[Mus+18] Cameron Musco, Praneeth Netrapalli, Aaron Sidford, Shashanka Ubaru, and David P. Woodruff. "Spectrum Approximation Beyond Fast Matrix Multiplication: Algorithms and Hardness". In: *9th Innovations in Theoretical Computer Science Conference (ITCS*

2018). Ed. by Anna R. Karlin. Vol. 94. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018, 8:1–8:21.

[ORRR12] Krzysztof Onak, Dana Ron, Michal Rosen, and Ronitt Rubinfeld. "A near-optimal sublinear-time algorithm for approximating the minimum vertex cover size". In: *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*. SIAM. 2012, pp. 1123–1131.

[PY23] Pan Peng and Yuichi Yoshida. "Sublinear-Time Algorithms for Max Cut, Max E2Lin (q), and Unique Label Cover on Expanders". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2023, pp. 4936–4965.

[SR94] R.N. Silver and H. Röder. "Densities of States of Mega-Dimensional Hamiltonian Matrices". In: *International Journal of Modern Physics C* 5.4 (1994), pp. 735–753.

[Ski89] John Skilling. "The Eigenvalues of Mega-dimensional Matrices". In: *Maximum Entropy and Bayesian Methods: Cambridge, England, 1988*. Springer Netherlands, 1989, pp. 455–466.

[ST11] Daniel A Spielman and Shang-Hua Teng. "Spectral sparsification of graphs". In: *SIAM Journal on Computing* 40.4 (2011), pp. 981–1025.

[SS11] Daniel A. Spielman and Nikhil Srivastava. "Graph Sparsification by Effective Resistances". In: *SIAM Journal on Computing* 40.6 (2011). Preliminary version in the 40th Annual ACM Symposium on Theory of Computing (STOC), pp. 1913–1926.

[SW23] William Swartworth and David P. Woodruff. "Optimal Eigenvalue Approximation via Sketching". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. STOC 2023. Association for Computing Machinery, 2023, pp. 145–155.

[Wan94] Lin-Wang Wang. "Calculating the density of states and optical-absorption spectra of large quantum systems by the plane-wave moments method". In: *Phys. Rev. B* 49 (15 1994), pp. 10154–10158.

[WWAF06] Alexander Weiße, Gerhard Wellein, Andreas Alvermann, and Holger Fehske. "The kernel polynomial method". In: *Rev. Mod. Phys.* 78 (1 2006), pp. 275–306.