# 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 a randomized 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 $\varepsilon$ 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 $\varepsilon$, 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 an $O(n\varepsilon^{-2})$-query and $O(n\varepsilon^{-2})$-time 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. Of independent interest, we show that our sparsification method also yields the first *deterministic* algorithm for spectral density estimation that scales linearly with $n$ (sublinear in the representation size of the graph).

## Contents

# 1 Introduction

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 [FDBV01; 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 (SDE))**.** Given an undirected graph $G = (V, E, w)$ with $n$ nodes and positive edge weights $w \in \mathbb{R}^E_{>0}$, return eigenvalue estimates $\widehat{\lambda}_1 \leq \cdots \leq \widehat{\lambda}_n$ such that

$$\frac{1}{n} \sum_{i \in \{1,\ldots,n\}} |\widehat{\lambda}_i - \lambda_i(N_G)| \leq \varepsilon, \tag{1}$$

where $\lambda_1(N_G) \leq \cdots \leq \lambda_n(N_G)$ are the eigenvalues of $G$'s normalized adjacency matrix, $N_G \in \mathbb{R}^{V \times V}$. Note that (1) 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 $\varepsilon$.

Outside of applications in network science, SDE has seen recent interest from the machine learning community due to its connections to learning distributions based on moment estimates [KV17; JMSS23] and learning graph structure from random walk traces [CKSV18]. The general version of the problem (involving arbitrary symmetric matrices) also has applications in deep learning [GKX19; PSG18]. Additionally, algorithms for alternative (often stronger) notations of spectral approximation, like relative-error eigenvalue histograms, or point-wise eigenvalue estimates, have been widely studied in recent literature for applications like approximating matrix norms and spectral sums [Mus+18; Bha+23; SW23; Bha+24].

Excitingly, at least for unweighted graphs, it has been shown that Problem 1.1 can be solved in time that is *sublinear* in the size of $G$'s representation, which can be as large as $\Omega(n^2)$. 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 SDE 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 neighbor oracles is just $\Omega(1/\varepsilon^2)$ [JMSS23].

Our work is motivated by these advances, the general importance of sublinear-time graph algorithms [CRT05; Czu+03; CS10; ORRR12; BKM14; ERS18; MOP01], and progress on sublinear time algorithms for other spectral problems like expander testing [CS07; NS10; GR11] and spectral clustering [Glu+21; Mou21]. We ask: Is it possible to improve on existing SDE algorithms? Is possible to efficiently obtain new sparse approximations, i.e., *sparsifications*, of $G$ that facilitate more efficient spectral density estimation? Finally, all previous sublinear time SDE algorithms make critical use randomness. Are there sublinear time *deterministic* algorithms?

In this paper, we provide an affirmative answer to each of these questions. For unweighted graphs, we provide a randomized $O(n\varepsilon^{-3})$-time algorithm for solving Problem 1.1 and a deterministic $n \cdot 2^{\tilde{O}(\varepsilon^{-1})}$-time algorithm. This is the fastest known randomized method, improving on the $O(n\varepsilon^{-7})$-time method from Braverman et al. [BKM22] and on the $2^{O(\varepsilon^{-1})}$-time method from Cohen-Steiner et al. [CKSV18] for small values of $\varepsilon$. To the best of our knowledge, our deterministic method is the first to achieve a sub-quadratic dependence on the number of nodes, $n$. Moreover, we obtain an algorithm with the same complexity for weighted graphs provided that, when queried with $a \in V$

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

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. Finally, 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

Our results follow a natural *two-stage* approach to approximating the spectral density of graphs. First, in sublinear time, we construct a *sparse matrix* that approximates $N_G \in \mathbb{R}^{V \times V}$, the normalized adjacency matrix of $G$, and has $o(n^2)$ entries. Second, we apply and adapt existing spectral density esitmation methods to efficiently approximate the spectrum of the sparse approximation. This approach decouples the problem of how to approximate $N_G$ using a small number of queries from the computational problem of how to efficiently approximate eigenvalues.

Concretely, our approach motivates the following natural question: What notions of graph approximation (ultimately, sparsification) are sufficient for preserving the spectrum in the sense of (1) *and* can be obtained in sublinear time? We make progress on this problem by introducing a new notion of *$\varepsilon$-additive nuclear approximation* and presenting algorithms that obtain near-optimal sparsity and query complexity for producing such approximations. Formally, we define:

**Definition 1.2** (Additive Nuclear Approximation and Sparsification)**.** $M \in \mathbb{R}^{V \times V}$ is an $\varepsilon$-additive *nuclear approximation* to $n$-vertex graph $G = (V, E, w)$ if $\|N_G - M\|_* \leq \varepsilon n$. $M$ is an $\varepsilon$-additive *nuclear sparsifier* of $G$ if it is also $s$-sparse, i.e., has at most $s$ non-zero entries, for $s = o(n^2)$.[2]

In Definition 1.2, $\|N_G - M\|_*$ denotes the nuclear norm, i.e., the sum of the singular values of $N_G - M$. Importantly, additive nuclear approximation is sufficient for solving the SDE problem. For symmetric matrices, $A, B \in \mathbb{R}^{n \times n}$, let $W_1(A, B) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} |\lambda_i(A) - \lambda_i(B)|$ denote Wasserstein-1 distance between the probability distributions $p$ and $q$ induced by the real-valued eigenvalues of $A$ and of $B$. A short proof establishes (see Appendix A) that:

**Fact 1.3.** *For symmetric matrices $A, B \in \mathbb{R}^{n \times n}$, if $\|A - B\|_* \leq n\varepsilon$, then $W_1(A, B) \leq \varepsilon$.*[3]

As a result, 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 those of $M$, then by triangle inequality, those eigenvalues solve Problem 1.1.

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

**Definition 1.4** (Adjacency query model)**.** We say we have *adjacency query access* to a weighted graph $G = (V, E, w)$ if $V$ is known and there is an $O(1)$ time procedure, $\mathsf{GetNeighbor}(a, i)$, that when queried with any $a \in V$ and $i \in \mathbb{Z}_{>0}$ outputs $\deg_G(a)$ and, if there is one, the $i$-th largest edge $\{a, b\}$ and $w_{\{a,b\}}$ (with ties broken arbitrarily). If the $a$ has $< i$ edges, the oracle returns $\emptyset$.

Adjacency queries are easily supported by standard graph data structures. For example, an adjacency list where each node's neighbors are stored in an array supports adjacency queries for unweighted graphs. To support weighted graphs, these arrays just need to be sorted by edge weight.

Within the query model established, we prove the following main result on constructing sublinear size nuclear sparsifiers in Section 2. Our method works even for weighted graphs.

[2]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 modify our algorithm to ensure that the output is indeed a normalized adjacency matrix. See Lemma A.2 in Appendix A for details.

[3]When $A$ and $B$ are diagonal with monotonically decreasing diagonal entries then $\|A - B\|_* = n \cdot W_1(A, B)$. In this sense, nuclear approximation is a natural strengthening of approximation in the Wasserstein-1 distance.

**Theorem 1.5** (Sublinear Time Nuclear Sparsification)**.** *There is a* deterministic *method (Algorithm 1) that, for any $\varepsilon \in (0,1)$, returns a $2n\varepsilon^{-2}$-sparse $\varepsilon$-additive nuclear sparsifier for any undirected weighted graph $G$, and runs in $O(n\varepsilon^{-2})$ time in the adjacency query model (Definition 1.4).*

We obtain Theorem 1.5 using a greedy approach that deterministically adds edges from $G$ to the sparsifier based on their weight and end-point degrees. Importantly, we leverage that this approach is deterministic to obtain the first *deterministic sublinear time SDE algorithms*. Before discussing these SDE algorithms, in Section 1.2 we highlight that deterministic sublinear time methods like Theorem 1.5 are *not possible* for stronger and more well-studied notions of graph sparsification.

## 1.2 Comparison to Prior Work on Sparsification

A significant line of work studies *spectral graph sparsification*, which generalizes cut sparsification [BK96], and has applications in linear system solving, combinatorial graph algorithms, and beyond [ST11; SS11; BSS12; Kap+17; LS18]. Concretely, a spectral sparsifier is defined as:

**Definition 1.6** (Spectral Sparsifier, [ST11])**.** Given $\varepsilon > 0$ and a graph $G$ with (unnormalized) Laplacian matrix $L$, a symmetric matrix $\widetilde{L} \in \mathbb{R}^{n\times n}$ is an $\varepsilon$-spectral sparsifier of $L$ if, for all $x \in \mathbb{R}^n$, $(1-\varepsilon)x^\top Lx \leq x^\top \widetilde{L}x \leq (1+\varepsilon)x^\top Lx$.

This definition involves the Laplacian (see Section 1.6), whereas we focus on the (normalized) adjacency matrix. Nevertheless, it can be checked that the normalized adjacency spectrum of an $\varepsilon$-spectral sparsifier is $\varepsilon$-close to that of $G$ in Wasserstein distance. However, it is impossible to obtain $\varepsilon$-spectral sparsifiers using $o(n^2)$ adjacency queries (see Theorem 11 of [Lee13] with $\delta = 1/n$). This lower bound motivates us to consider notions of spectral sparsification that are weaker than Definition 1.6, but stronger than nuclear sparsification (Definition 1.2). In particular, we introduce the following notion of *additive error* spectral sparsification, which strictly strengthens Definition 1.2:

**Definition 1.7** (Additive Spectral Sparsifier)**.** A symmetric matrix $M \in \mathbb{R}^{V\times V}$ is an $\varepsilon$-additive spectral sparsifier of $G$ if $\|M - \widetilde{N}_G\|_2 \leq \varepsilon$.

For unweighted graphs, Lee [Lee13] gives a randomized algorithm that computes an $\varepsilon$-additive spectral sparsifier with $O(n\varepsilon^{-2}\log n)$ entries in $O(n\varepsilon^{-3})$-time in the adjacency query model.[4] We show in Section 6 that an $O(n\varepsilon^{-2}\log n)$-sparse $\varepsilon$-additive spectral sparsifier can also be obtained for all weighted graphs in $O(n\varepsilon^{-2}\log n)$ time using standard random sampling methods.[5] Given this result and the sublinear time randomized algorithm in Lee [Lee13], it is natural to ask whether we can obtain a *deterministic* algorithm for additive spectral approximation with a linear dependence on $n$, as we do for nuclear sparsification in Theorem 1.5. Interestingly, we prove that this is impossible:

**Theorem 1.8.** *Any deterministic algorithm requires $\Omega(n^2)$ Adjacency Queries (Definition 1.4) to compute a $\frac{1}{4}$-additive spectral sparsifier (Definition 1.7), even for unweighted graphs.*

Theorem 1.8 highlights a strong separation between our new notation of nuclear sparsification, for which we have a sublinear time deterministic algorithm, and stronger notions of sparsification.

[4] Formally, Lee [Lee13] considers a slightly stronger guarantee that he calls "probilistic spectral sparsification." Their guarantee implies our Definition 1.7 for unweighted graphs.

[5] In both cases, the $O(n\varepsilon^{-2}\log n)$ sparsity can be improved to $O(n\varepsilon^{-2})$ at the cost of additional $\log(n)\mathsf{poly}(1/\varepsilon)$ runtime factors using near-linear time $\varepsilon$-spectral sparsification methods [JRT23; LS18].

### 1.3 Applications to Sublinear Time Spectral Density Estimation

Our main application of our result on nuclear sparsification (Theorem 1.5) is faster deterministic and randomized sublinear time SDE algorithms for graphs. In the randomized setting, a long line of work in computational chemistry, applied math, and, recently, computer science [Ski89; SR94; WWAF06; LSY16] studies *linear time* methods for spectral density estimation. In Chen et al. [CTU21] and Braverman et al. [BKM22] it was proven that common randomized algorithms like the stochastic Lanczos quadrature method and moment matching can approximate the spectral density of *any* symmetric matrix $A$ up to $\varepsilon$-accuracy in Wasserstein distance using roughly $O(\varepsilon^{-1})$ matrix-vector multiplications with $A$.[6] By simply applying such methods to our nuclear norm sparsifier we obtain the following result (proven in Section 2):

**Theorem 1.9** (Randomized Sublinear Time SDE)**.** *There is a randomized algorithm that, with probability* 99/100*, solves the SDE problem (Problem 1.1) in* $O(n\varepsilon^{-3})$ *time in the adjacency query model (Definition 1.4).*

Theorem 1.9 directly improves on an $O(n\varepsilon^{-7})$ time method from Braverman et al. [BKM22], which only applied to unweighted graphs, and takes a different approach from ours. That work leverages randomized methods for approximating matrix-vector products instead of sparsification. For sufficiently small $\varepsilon$, we also improve on the $2^{O(\varepsilon^{-1})}$ time method from Cohen-Steiner et al. [CKSV18], which has no dependence on $n$.[7]

In addition to a faster randomized algorithm, in Section 2 we show how to use our deterministic nuclear norm sparsifiers to obtain the first sublinear time deterministic SDE method for graphs:

**Theorem 1.10** (Deterministic Sublinear Time SDE)**.** *There is a deterministic algorithm that solves the SDE problem (Problem 1.1) in* $n \cdot 2^{O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})}$ *time in the adjacency query model (Definition 1.4).*

We prove Theorem 1.10 by showing that it is possible to exactly compute eigenvalue moments of the form $\mathrm{Tr}\left(M^i\right)$ for our particular nuclear norm sparsifier $M$. We then show that these moments can be used to obtain an $\varepsilon$-accurate approximation to the spectral density. Crucially, we leverage that fact that our nuclear norm sparsifiers are not just sparse, i.e, have a small number of non-zero entries, but that every vertex has bounded combinatorial degree, i.e., number of neighbors. Previously, the best deterministic algorithm for spectral density estimation would have been via a full eigendecomposition of $N_G$, which takes time $O(n^\omega)$.

### 1.4 Lower Bounds for Nuclear Sparsification

Given that nuclear approximation is a natural relaxation of widely studied graph sparsification notions like spectral sparsification, it is desirable to fully understand the complexity of the problem. We complement our algorithmic result from Theorem 1.5 with nearly matching sparsity and query lower bounds. First, we show that the sparsity of our sparsifiers is near-optimal:

**Theorem 1.11** (Sparsity Lower Bound)**.** *For any* $\varepsilon \leq \varepsilon_0$*, where* $\varepsilon_0 \in (0,1)$ *is a fixed constant, there is a graph* $G$ *with normalized adjacency matrix* $N_G$*, such that any matrix* $M$ *satisfying* $\|N_G - M\|_* \leq n\varepsilon$ *must have* $\Omega(n\varepsilon^{-2}/\log^2\varepsilon^{-1})$ *non-zero entries.*

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

[7] Initial evidence (lower bounds in restricted models) suggests that it may not be impossible to improve the $\varepsilon$ dependence in Cohen-Steiner et al. [CKSV18] to sub-exponential while maintaining no dependence on $n$ [JMSS23].

We prove Theorem 1.11 by considering the extreme case when $\varepsilon = 1/\sqrt{cn}$ for a small constant $c$. This is the smallest value of $\varepsilon$ for which Theorem 1.5 gives a non-trivial result (i.e., a matrix with less than $n^2$ entries). Using the probabilistic method, we show that there are $2^{O(n^2)}$ graphs whose adjacency matrices are all $\varepsilon$-far in the nuclear norm. A pigeonhole argument is then used to show that not all of these matrices can be well approximated by $O(n^2/\log^2 n)$-sparse matrices. We prove the result for general $\varepsilon$ via a reduction to the $\varepsilon = 1/\sqrt{cn}$ case.

We also show that any algorithm for constructing an $\varepsilon$-additive nuclear sparsifier must, in the worst case, make $\widetilde{\Omega}(n\varepsilon^{-2})$ GetNeighbor queries, matching Theorem 1.5 up to log factors:

**Theorem 1.12** (Query Lower Bound)**.** *For any $\varepsilon \leq \varepsilon_0$, where $\varepsilon_0 \in (0,1)$ is a fixed constant, any algorithm working in the adjacency query model (Definition 1.4) requires $\Omega(n\varepsilon^{-2}\log^{-2}(\varepsilon^{-1}))$* GetNeighbor *queries to return an $\varepsilon$-additive nuclear sparsifier with probability $3/4$.*

Notably, Theorem 1.12 even applies to randomized algorithms. It is interesting to ask if the same query complexity is optimal for spectral density estimation itself (Problem 1.1). Currently, the best query lower bound is just $\Omega(1/\varepsilon^2)$, which is due to a recent result of Jin et al. [JMSS23]. That work also proves a lower bound of $\Omega(\exp(\varepsilon^{-1}))$ under a weaker random walk query model (see Section 1.5). However, 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.

### 1.5 Random Walk Query Model

Finally, motivated by the SDE algorithms of Braverman et al. [BKM22] and Cohen-Steiner et al. [CKSV18], we consider a weaker graph access model than the adjacency query model. Those algorithms only assume access to $G$ via random walks, i.e., that in $O(k)$ time we can sample a random walk $v_0, v_1, ..., v_k$, where $v_0$ is a uniformly random node in $G$ and $v_i$ is chosen from the neighbors of $v_{i-1}$ with probability proportional to edge weight. Interestingly, we show that it is possible to construct nuclear norm sparsifiers in an even weaker model where only one-step walks are allowed:

**Definition 1.13** (One-step Random Walk Query Model)**.** We say we have *one-step random walk query access* to a graph $G = (V, E, w)$ if there is an $O(1)$ time procedure, RandomNeighbor, that returns a uniformly random vertex $a \in V$ and either $\emptyset$ if $a$ has no neighbors, or an edge $\{a, b\}$ selected with probability proportional to its weight, along with the degree of $a$ and $b$.

While our algorithm discussed in Section 1.1 are not directly implementable in this more restrictive query model, in Section 6 we present an alternative, randomized algorithm that achieves identical query- and time-complexity in the model:

**Theorem 1.14.** *There is an algorithm (Algorithm 2) that, for any $\varepsilon \in (0,1)$, returns with probability $2/3$ an $O(n\varepsilon^{-2})$-sparse $\varepsilon$-additive nuclear sparsifier for any undirected weighted graph $G$ using just $O(n\varepsilon^{-2})$ queries in the one-step random walk query model (Definition 1.13).*

We also obtain an algorithm for $\varepsilon$-additive spectral sparsification in the one-step random walk query model. We obtain the following via a natural application of matrix concentration. Specifically, we directly use an algorithm and theorem from Cohen et al. [Coh+17].

**Theorem 1.15.** *There is an algorithm (Algorithm 3) that, for any $\varepsilon \in (0,1)$, returns with probability $2/3$ an $O(n\varepsilon^{-2}\log n)$-sparse $\varepsilon$-additive spectral sparsifier for any undirected weighted graph $G$ using just $O(n\varepsilon^{-2}\log n)$ queries in the one-step random walk query model (Definition 1.13).*

Theorem 1.15 achieves the stronger notion of $\varepsilon$-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. We show that this $\log n$ factor is unavoidable in the following sense:

**Theorem 1.16.** *For a fixed constant $\varepsilon \in (0, 1/8)$ any algorithm requires $\Omega(n \log n)$ one-step random walk model queries to output an $\varepsilon$-additive spectral sparsifier with probability $> c$.*

Theorem 1.16 is proved via a coupon collector argument. Consider a graph of size $2n$ with $n$ isolated pairs of nodes, each of which is either connected by an edge or not. To obtain a constant factor additive spectral sparsifier, we must observe one node in each pair, which takes $\Omega(n \log n)$ samples via the standard coupon collector lower bound. Note that this lower bound also applies to the stronger $k$-step random walk model studied in Braverman et al. [BKM22] and Cohen-Steiner et al. [CKSV18], since for such a graph, no additional information is gained from a longer walk.

### 1.6 Paper Organization and Notation

**Paper Organization.** Section 2 presents our nuclear sparsification and SDE algorithms in the adjacency query model. Section 3 and Section 4 presents our sparsity and query lower bounds for nuclear sparsification. Section 5 presents a lower bound against deterministic algorithms for $\varepsilon$-additive spectral sparsification. Section 6 covers our results in the more restrictive one-step random walk model. Some proof details are deferred to Appendix A.

**Graph Notation.** In this paper, we consider undirected graphs $G = (V, E, w)$ with positive edge weights $w \in \mathbb{R}^E$. $A_G \in \mathbb{R}^{V \times V}_{\geq 0}$ denotes the adjacency matrix of $G$, i.e., $[A_G]_{v,v'} \stackrel{\text{def}}{=} w_e$ if $e = \{v, v'\} \in E$ and is 0 otherwise. $\deg(v) \stackrel{\text{def}}{=} \sum_{e=\{v,v'\} \in E} w_e$ denotes the weighted degree of vertex $v$, and $D_G \in \mathbb{R}^{V \times V}_{\geq 0}$ is the diagonal degree matrix of $G$, where $D_{G v,v} := \deg(v)$ for all $v \in V$. Throughout, we let $N_G \stackrel{\text{def}}{=} D_G^{-1/2} A_G D_G^{-1/2}$ denote the normalized adjacency matrix of $G$.

**Vectors and Matrix Notation.** For positive integers $n$, we let $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. For a vector $v \in \mathbb{R}^n$, $\|v\| \stackrel{\text{def}}{=} \sqrt{\sum_{i \in [n]} v_i^2}$ denotes the Euclidean norm. For a matrix $A \in \mathbb{R}^{n \times n}$, the spectral norm is defined as $\|A\|_2 \stackrel{\text{def}}{=} \max_{v : v \in \mathbb{R}^n} \|Av\|/\|v\|$, which equals $\max_i |\lambda_i(A)|$ when $A$ is symmetric. For symmetric $A$, the nuclear norm is equal to $\|A\|_* \stackrel{\text{def}}{=} \sum_{i=1}^n |\lambda_i(A)|$ and the Frobenius norm is $\|A\|_F \stackrel{\text{def}}{=} \sqrt{\sum_{i,j} A(i,j)^2}$. We also define the matrix inner-product as $\langle A, X \rangle \stackrel{\text{def}}{=} \sum_{i,j} A(i,j) X(i,j)$ for any matrices $A, X \in \mathbb{R}^{m \times n}$. We use Loewner order notation $A \succcurlyeq 0$ or $0 \preccurlyeq A$ to denote that a symmetric $A$ is positive semidefinite (PSD), i.e., that $A$ has non-negative eigenvalues. $A \succcurlyeq B$ indicates that $A - B$ is PSD.

## 2 Additive Nuclear Sparsifiers

In this section, we present our main algorithm (Algorithm 1) for constructing $O(n\varepsilon^{-2})$-sparse $\varepsilon$-additive nuclear norm sparsifiers (Theorem 1.5). We first prove a structural result, which suggests a natural greedy sparsification procedure. Specifically, we show that the normalized adjacency matrix can be sparsified by eliminating any edge with small weight compared to the degree of its endpoints.

**Theorem 2.1.** *Let $G = (V, E, w)$ be a weighted graph with adjacency matrix $A_G \in \mathbb{R}^{n\times n}$, degree matrix $D_G$, and normalized adjacency matrix $N_G$. For $\varepsilon \in (0, 1)$ define $E' \subseteq E$ as:*

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

*Let $G'$ be obtained by removing all edges from $G$ that are not in $E'$, and let $A_{G'} \in \mathbb{R}^{n\times n}$ denote the adjacency matrix of $G'$. Let $\widetilde{N} \stackrel{\text{def}}{=} D_G^{-1/2} A_{G'} D_G^{-1/2}$. Then $\|N_G - \widetilde{N}\|_* \leq \varepsilon n$. Moreover, $\widetilde{N}$ has at most $2\varepsilon^{-2}$ non-zeros in each row (resp. column) and $\|\widetilde{N}\|_2 \leq 1$.*

*Proof.* Let $D_{G'}$ denote the degree matrix of $A_{G'}$. Since $D_{G'}$ is entrywise smaller than $D_G$, and since $D_{G'} - A_{G'}$ is a graph Laplacian and thus positive semidefinite, we have the relationship:

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

Our claim that $\|\widetilde{N}\|_2 \leq 1$ then follows by multiplying by $D_G^{-1/2}$ on the left and right:

$$-I = -D_G^{-1/2} D_G D_G^{-1/2} \preccurlyeq D_G^{-1/2} A_{G'} D_G^{-1/2} = \widetilde{N} \preccurlyeq D_G^{-1/2} D_G D_G^{-1/2} = I.$$

Next, observe that for each $v \in V$ there are at most $2\varepsilon^{-2}$ nodes $v'$ for which $w_{\{v,v'\}} \geq \frac{\varepsilon^2}{2}\deg(v)$. Only edges for which $w_{\{v,v'\}} \geq \frac{\varepsilon^2}{2}\deg(v)$ are contained in $E$. Consequently, $A_{G'}$ and $\widetilde{N}$ have at most $2\varepsilon^{-2}$ non-zeros per row/column, as claimed

Finally, we bound the error in the nuclear norm by first bounding the Frobenius norm error. Without loss of generality, assume that $V = \{1, \ldots, n\}$ and the vertices are ordered such that $\deg(i) \geq \deg(j)$ for $i < j$.

$$\begin{aligned}
\left\|N_G - \widetilde{N}\right\|_F^2 = 2 \sum_{\{i,j\}\in E\setminus E'} \left(\frac{w_{\{i,j\}}}{\sqrt{\deg(i)\deg(j)}}\right)^2 &= 2\sum_i \sum_{j:\{i,j\}\in E\setminus E', j<i} \frac{w_{\{i,j\}}^2}{\deg(i)\deg(j)} \\
&\leq 2\sum_i \sum_{j:\{i,j\}\in E\setminus E', j<i} \frac{w_{\{i,j\}}}{\deg(i)} \frac{\varepsilon^2}{2} \\
&= \varepsilon^2 \sum_i \frac{1}{\deg(i)} \sum_{j:\{i,j\}\in E\setminus E', j<i} w_{\{i,j\}} \leq \varepsilon^2 n.
\end{aligned}$$

Since $\|A\|_* \leq \sqrt{n}\,\|A\|_{\mathrm{F}}$ for any matrix $A$ (Fact A.1), we conclude that $\left\|N_G - \widetilde{N}\right\|_* \leq \varepsilon n$, as desired. ∎

Next, we show that Theorem 2.1 yields an efficient nuclear norm sparsification algorithm (Algorithm 1), whose analysis yields our first main result, Theorem 1.5.

*Proof of Theorem 1.5.* We first claim that Algorithm 1 returns exactly the matrix $\tilde{N}$ described in Theorem 2.1. In particular, it suffices to show that the while loop is executed for all $\{v, v'\} \in E'$, where $E'$ is as in Theorem 2.1. To see why this is the case, observe that, because edges are processed in decreasing order of weight, when executing the for loop for vertex $v$, the inner while loop certainly gets executed for all $\{v, v'\}$ such that $w_{\{v,v'\}} \geq \frac{\varepsilon^2}{2}\deg(v)$. Because $\max\{\deg(v'), \deg(v)\} \geq \deg(v)$, the loop executes for a superset of the edges in $E'$ that have $v$ as an endpoint.

**Algorithm 1:** Additive nuclear norm approximation

**Input:** Graph $G = (V, E, w)$ supporting GetNeighbor queries as in Definition 1.4, accuracy $\varepsilon$.

**Output:** Nuclear norm sparsified $\widetilde{N}$ for $G$.

Initialize $\widetilde{N} = 0$, $c = 1$
**for** $v \in V$ **do**
  $v' \leftarrow \mathsf{GetNeighbor}(v, c)$
  **while** $v' \neq \emptyset$ *and* $w_{\{v,v'\}} \geq \frac{\varepsilon^2}{2}\deg(v)$ **do**
    **if** $w_{\{v,v'\}} \geq \frac{\varepsilon^2}{2}\deg(v')$ **then**
      Set $\widetilde{N}(v, v') = \widetilde{N}(v', v) = w_{\{v,v'\}}/(\sqrt{\deg(v)\deg(v')})$
    $c \leftarrow c + 1$, $v' \leftarrow \mathsf{GetNeighbor}(v, c)$
**Return:** $\widetilde{N}$

Next, we bound the complexity. Every execution of the while loop run in $O(1)$ time because it requires just two GetNeighbor queries: one involving $v$, and a second to obtain the degree of $v'$. Since there are at most $2\varepsilon^{-2}$ edges connected to $v$ with weight $\geq \frac{\varepsilon^2}{2}\deg(v)$, the loop executes at most $O(\varepsilon^{-2})$ time per vertex in the graph. We conclude a total runtime of $O(n\varepsilon^{-2})$, as desired. ■

Note that the nuclear sparsifier $\widetilde{N}$ returned by Algorithm 1 is *not* guaranteed to be the normalized adjacency matrix of any undirected graph. As access to a graphical sparsifier may be desirable, we show how to modify the procedure above to obtain a graphical nuclear sparsifier in Lemma A.2.

## 2.1 Applications to Spectral Density Estimation

An important application of our nuclear norm approximation method from Algorithm 1 is to develop faster algorithms for graph spectral density estimation. In particular, from Fact 1.3, if we compute $\widetilde{N}$ such that $\|N - \widetilde{N}\|_* \leq \frac{\varepsilon}{2}n$, then computing an $\frac{\varepsilon}{2}$ accurate SDE for $\widetilde{N}$ immediately yields an $\varepsilon$ accurate SDE for $N_G$. If we compute the SDE for $\widetilde{N}$ using the existing linear time method from Braverman et al. [BKM22], which runs in roughly $O(\mathrm{nnz}(\widetilde{N})\varepsilon^{-1})$ time, then we immediately obtain our Theorem 1.9, i.e., that there is an $O(n\varepsilon^{-3})$ time randomized algorithm for approximating the spectral density of any weighted graph $G$.

**Theorem 1.9** (Randomized Sublinear Time SDE)**.** *There is a randomized algorithm that, with probability* 99/100*, solves the SDE problem (Problem 1.1) in* $O(n\varepsilon^{-3})$ *time in the adjacency query model (Definition 1.4).*

*Proof.* Theorem 1.4 in Braverman et al. [BKM22] combined with the discretization procedure described in the appendix of that paper yields an algorithm that solves Problem 1.1 for any matrix $M$. The method succeeds with probability 0.99 and runs in time $O\left(\mathrm{nnz}(M)\varepsilon^{-1} \cdot (1 + \varepsilon^{-2}n^{-2}\log^3(1/\varepsilon)\right)$.

Suppose that $n \geq \varepsilon^{-2}\log^3(1/\varepsilon)$ so that $\varepsilon^{-2}n^{-2}\log^3(1/\varepsilon) \leq 1$. For such values of $n$, we obtain Theorem 1.9 by simply computing an $\frac{\varepsilon}{2}$-additive nuclear sparsifier $\widetilde{N}$ for $N_G$ using Theorem 1.5 in $O(n\varepsilon^{-2})$ time, then running the method of Braverman et al. [BKM22] with error $\frac{\varepsilon}{2}$ on $\widetilde{N}$. By Fact 1.3 and triangle inequality, the eigenvalues returned are guaranteed to approximate those of $N_G$ to error at most $\frac{\varepsilon}{2} + \frac{\varepsilon}{2}$ in Wasserstein distance. Moreover, since $\widetilde{N}$ is guaranteed to have at most $O(n\varepsilon^{-2})$ entries, the total runtime in $O(n\varepsilon^{-3})$ as required.

Alternatively, suppose that that $n < \varepsilon^{-2}\log^3(1/\varepsilon)$. In this case, we can simply use a direct eigendecomposition method to compute the eigenvalues of $N_G$. Doing so takes $\tilde{O}(n^\omega) \leq O(n^{2.372})$ [BGKS20], where $\omega$ is current exponent of fast matrix multiplication [DWZ23]. We have that $\tilde{O}(n^\omega) \leq O(n\varepsilon^{-2\cdot 1.372}) \leq O(n\varepsilon^{-3})$, which yields the theorem. ■

We can also use Algorithm 1 to obtain the first *deterministic* sublinear time method for spectral density estimation. In particular, we claim:

**Theorem 1.10** (Deterministic Sublinear Time SDE)**.** *There is a deterministic algorithm that solves the SDE problem (Problem 1.1) in $n \cdot 2^{O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})}$ time in the adjacency query model (Definition 1.4).*

*Proof.* We prove this result by again appealing to a prior method from Braverman et al. [BKM22]. Like the algorithm in Cohen-Steiner et al. [CKSV18], the main algorithm in Braverman et al. [BKM22] is based on *moment matching*. In particular, it was shown by Braverman et al. [BKM22] that if two distributions $p, q$ have the same first $\frac{c}{\varepsilon}$ (uncentered) moments for a fixed constant $c$, then $W_1(p, q) \leq \varepsilon$.[8] So, given moments of an unknown distribution $p$, we can find a distribution approximating $p$ in Wasserstein distance by simply returning any distribution that (approximately) matches those moments. Braverman et al. [BKM22] shows (in their Lemma 3.4) that such a distribution can be found in $\mathsf{poly}(1/\varepsilon)$ time using their Algorithm 1. Furthermore, by their Theorem B.1, this distribution can be converted to a uniform distribution over a set of $n$ approximate eigenvalues in $O(n + 1/\varepsilon)$ time.

So, to prove Theorem 1.10, it suffices to show that we can compute the first $\frac{c}{2\varepsilon}$ moments of the eigenvalues of $\widetilde{N}$ in $n \cdot 2^{O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})}$ time, where $\widetilde{N}$ is the result of Algorithm 1 run on $N_G$ with accuracy parameter $\frac{\varepsilon}{2}$. Doing so would allow us to compute an $\frac{\varepsilon}{2}$-approximate SDE for $\widetilde{N}$ using the moment matching method of Braverman et al. [BKM22], which by triangle inequality, would be an $\varepsilon$-approximate SDE (in Wasserstein distance) for the eigenvalue distribution of $N_G$.

To see why this is possible, observe that the $j^{\text{th}}$ moment of $\widetilde{N}$'s eigenvalue distribution equals: $\frac{1}{n}\sum_{i=1}^n \tilde{\lambda}_i^j = \mathrm{Tr}\left(\widetilde{N}^j\right)$, where $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ are $\widetilde{N}$'s eigenvalues. As $\widetilde{N}$ is symmetric with at most $8\varepsilon^{-2}$ non-zeros per row/column (guaranteed by Theorem 2.1), for any positive integer $k$, $\widetilde{N}^k$ has at most $(8\varepsilon^{-2})^k$ non-zeros per row/column. Thus, given $\widetilde{N}^k$, the matrix $\widetilde{N}^{k+1} = \widetilde{N}^k\widetilde{N}$ can be computed in $n(8\varepsilon^{-2})^{k+1} = n \cdot 2^{O(k\log\frac{1}{\varepsilon})}$ time. So, we conclude that $\widetilde{N}^2, \widetilde{N}^3, \ldots, \widetilde{N}^j$ can be computed (and traces evaluated exactly) in $n \cdot 2^{O(j\log\frac{1}{\varepsilon})}$ time. As we need $j = O(1/\varepsilon)$ moments (i.e., traces) to obtain an $\frac{\varepsilon}{2}$ accurate SDE for $\widetilde{N}$, the claimed runtime of $n \cdot 2^{O(\frac{1}{\varepsilon}\log\frac{1}{\varepsilon})}$ follows. ■

# 3 Lower Bounds for Nuclear Sparsification

In this section, we show that our sparsification result of Theorem 1.5 is tight in terms of sparsity. In particular, we prove the sparsity lower bound of Theorem 1.11.

## 3.1 Sparsity Lower Bound

We start by considering the case when $\varepsilon \approx 1/\sqrt{n}$. We prove that, in this regime, a random Erdős-Rényi graph is unsparsifiable. I.e., a nuclear norm approximation requires $\Omega(n^2)$ entries with high probability. This shows that the dependence of $\varepsilon^{-2}$ in Theorem 1.5 cannot be improve for all

[8] By Lemma 3.1 in Braverman et al. [BKM22], $c \leq 36$, although it is likely that this upper bound is loose.

values of $\varepsilon$. We then extend the lower bound to show tightness for larger values of $\varepsilon$ via a repetition argument.

The lower bound for the case when $\varepsilon \approx 1/\sqrt{n}$ proceeds via a counting argument. We first show that all normalized adjacency matrices in a sample of roughly $2^{O(n^2)}$ independent random graphs are *far* in nuclear norm. Approximating this sample thus requires a large number of *different* sparsifiers. If we restrict to $s = o(n^2)$-sparse matrices, there would simply not be enough to approximate all graphs in our sample. Indeed, there are only roughly $2^s$ matrices with sparsity $s$ that are sufficiently far in nuclear norm, so less than the required $2^{O(n^2)}$. We leverage the following (simplified) matrix concentration result to show that two Erdős-Rényi random graphs tend to be far apart in nuclear norm.

**Fact 3.1** ([GZ00], Theorem 1.1 with $f(x) = |x|$)**.** *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 random symmetric matrix with $X_{ij} = C_{ij}\omega_{ij}$, where $\omega_{ij} = \omega_{ji}$, and $\{\omega_{ij}, 1 \le i \le j \le n\}$ are independent random variables supported on $[-1, 1]$. Then, for any $\delta > 16\sqrt{\pi}/n$,*

$$\mathbb{P}\left\{ |\|X\|_* - \mathbb{E}[\|X\|_*]| \ge n^{1.5}\delta \right\} \le 4 \cdot \exp\left\{ -\frac{n^2(\delta - 16\sqrt{\pi}/n)^2}{64} \right\}.$$

To apply Fact 3.1, we first bound the expected value of the nuclear norm distance between two Erdős-Rényi $G(n, 1/2)$ random graphs:

**Lemma 3.2.** *For $n \ge 10$, consider a symmetric random matrix $X \in \{-1, 0, 1\}^{n\times n}$ with zeros on the diagonal, $X_{ji} = X_{ij}$, and*

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

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

*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 $\mathbb{P}\left\{\|X\|_{\mathrm{F}}^2 \le n^2/9\right\} \le \exp(-n^2/32)$. 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, $\|X\|_* \ge n^{1.5}/45$ and $\mathbb{E}\,\|X\|_* \ge (1 - \exp(-n/5)) \cdot n^{1.5}/45 \ge n^{1.5}/100$. Moreover, we have that $\|X\|_* \le \sqrt{n}\,\|X\|_{\mathrm{F}} \le n^{1.5}$, where the last step follows because $X$ is a $\{-1, 0, 1\}$ matrix, ∎

In the following theorem, we consider the case of $\varepsilon \approx 1/\sqrt{n}$ and show that the (unnormalized) adjacency matrix of an Erdős-Rényi graph cannot be approximated well in the nuclear norm by sparse matrices. We state the theorem for the unnormalized adjacency matrix as it serves as a building block for a larger regime of $\varepsilon$, proved in Theorem 1.11 (which is stated for normalized adjacency matrix).

**Theorem 3.3.** *Let $A$ denote the adjacency matrix of an Erdős-Rényi random graph $G(n, 1/2)$, for $n \ge 3000$. With probability greater than $5/8$, there is no matrix $B'$ with $\mathrm{nnz}(B') \le n^2/(16\cdot 10^6 \cdot \log^2 n)$ such that $\|A - B'\|_* \le n^{1.5}/500$.*

*Proof.* Let $A_1$ and $A_2$ be drawn from $G(n, 1/2)$. Then $A_1 - A_2$ follows the distribution of the matrix $X$ in Lemma 3.2. Therefore, we have that $\mathbb{E}\,\|A_1 - A_2\|_* \geq n^{1.5}/100$. From Fact 3.1, 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 3.2 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 $A_1, \ldots, A_m$ be the adjacency matrices of $m$ independent $G(n, 1/2)$ graphs. By union bound, $\mathbb{P}\left\{\exists i \neq j : \|A_i - A_j\|_* \leq \frac{n^{1.5}}{200}\right\} \leq 4\binom{m}{2}\cdot\exp\left(\frac{-n^2}{4\cdot 10^6}\right)$. Setting $m = \exp\left(n^2/(16\cdot 10^6)\right)$, we get from the above equation that

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

The above equation implies that $m = e^{\Omega(n^2)}$ random $G(n, 1/2)$ graphs will have adjacency matrices that are all more than $n^{1.5}/200$ apart in nuclear norm from each other. Let $\mathcal{A}$ denote the set of matrices $A_1, \ldots, A_m$.

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 $\forall i, j \in [n]$, $B_{ij} \in \{-n^2, -(n^2-\varepsilon), \ldots, (n^2-\varepsilon), n^2\}$. Let $c = 16\cdot 10^6$, and $\varepsilon = 2/(n^2)$. Bounding the size of $\mathcal{S}$, we get

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

Therefore, for any sparse matrix $B'$ with $\mathrm{nnz}(B') \leq n^2/(c\log^2 n)$, and $|B'_{ij}| \leq n^2$, 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 $\|B - B'\|_* \leq 2/\sqrt{n}$. We note that it suffices to consider $|B'_{ij}| \leq n^2$, else $B'$ would not be able to approximate any matrix in $\mathcal{A}$ in nuclear norm due to Fact A.1. Next, we will show that any sparse matrix $B'$ cannot be a good approximation (in the sense of the statement of the theorem) to most matrices in $\mathcal{A}$.

Recall that $m = \exp\left(n^2/c\right)$ and size of set $|\mathcal{S}| \leq \exp\left(\frac{7n^2}{c\log n}\right)$. By counting argument, there are at least $m - |\mathcal{S}|$ random Erdős-Rényi graph adjacency matrices for which no matrix in $\mathcal{S}$ is a good approximation. Formally, since $|\mathcal{A}| = m$, there exists $m - |\mathcal{S}|$ random Erdős-Rényi graph adjacency matrices $A_{i_1}, \ldots, A_{i_{m-|\mathcal{S}|}} \in \mathcal{A}$ such that for any $B \in \mathcal{S}$, $t \in [m - |\mathcal{S}|]$, $\|A_{i_t} - B\|_* \geq n^{1.5}/400$, where the last inequality follows from triangle inequality.

For $n \geq 3000$, $m - |\mathcal{S}| > (3/4)\cdot\exp(n^2/c)$. Therefore, by triangle inequality, 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 any sparse matrix $B'$ with $\mathrm{nnz}(B') \leq n^2/(c\log^2 n)$, i.e., for $t \in [m - |\mathcal{S}|]$, $\|A_{i_t} - B'\|_* \geq n^{1.5}/400 - 2/\sqrt{n} > n^{1.5}/500$, for any matrix $B'$ with $\mathrm{nnz}(B') \leq n^2/(c\log^2 n)$. ∎

Our lower bound construction follows from considering $n/b$ such $b$-by-$b$ dimensional matrices $A$ satisfying the properties in Theorem 3.3, 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 3.4.** *Let $n, k, b \in \mathbb{Z}_{\geq 0}$ be such that $k = \lfloor n/b \rfloor \geq 1$. 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 $\mathrm{nnz}(B') \leq \left(c'b^2/\log^2 b\right)$, $\|E_i - B'\|_* \geq c''b^{1.5}$. Then, for any matrix $B$ with $\mathrm{nnz}(B) \leq (k/2) \cdot \left(c'b^2/\log^2 b\right)$, we have that $\|E - B\|_* \geq c''n\sqrt{b}/4$.*

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

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

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 $\left\|\widetilde{X}\right\|_2 \leq 1$, and by the dual formulation of nuclear norm (Fact A.1), we get that

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

Using Proposition 3.4, we get the following lower bound on the sparsifiability of normalized adjacency matrices in the nuclear norm by setting $b = O(1/\varepsilon^2)$ and noting that (from Theorem 3.3) Erdős-Rényi random graphs satisfy the properties of the matrices $E_1, \ldots E_k$ in Proposition 3.4.

**Theorem 1.11** (Sparsity Lower Bound)**.** *For any $\varepsilon \leq \varepsilon_0$, where $\varepsilon_0 \in (0, 1)$ is a fixed constant, there is a graph $G$ with normalized adjacency matrix $N_G$, such that any matrix $M$ satisfying $\|N_G - M\|_* \leq n\varepsilon$ must have $\Omega(n\varepsilon^{-2}/\log^2 \varepsilon^{-1})$ non-zero entries.*

*Proof.* The theorem follows from Proposition 3.4. Let $G_1, \ldots, G_k$ denote the "unsparsifiable" Erdős Rényi random graphs of size $b \geq 3000$ from Theorem 3.3. Let $A_1, \ldots, A_k$ denote the adjacency 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 3.4, with $E_i = A_i$, $i \in [k]$. This is true because Theorem 3.3 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 3.4 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, 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 ensure that $n/b \geq 1$, which implies that $n \geq 16\varepsilon^2/(c'')^2$. 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 $B$, by $M = D_G^{-1/2}BD_G^{-1/2}$. Hence, by Proposition 3.4, 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 Nuclear Sparsification Query Complexity Lower Bound

Next, we prove that our algorithm from Section 2 also achieves the optimal query complexity for nuclear approximation, up to logarithmic factors in $\varepsilon^{-1}$. Interestingly, our lower bounds apply in an *even more general* query model which supports edge queries in addition to neighbor queries.

**Definition 4.1** (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 $\mathsf{True}$ if $\{u, v\} \in E$ and $\mathsf{False}$ 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 4.2.** *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 $\|A\|_* \geq \langle A, \widetilde{X}\rangle = \langle B, X\rangle = \|B\|_*$. $\blacksquare$

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

**Lemma 4.3.** *Let $G = (V, E)$ be an Erdős-Rényi $G(n, 1/2)$ 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 by Theorem 3.3, 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

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

Next, we show that any algorithm that makes $o(n^2/\log^2(n))$ queries to the generalized adjacency query model (Definition 4.1) is unable to distinguish between a random Erdős-Rényi graph and the graph which is the complement restricted to the edges which were not queried.

**Lemma 4.4.** *There exists an $\varepsilon_0 \in (0, 1)$ such that for any $\varepsilon \in (0, \varepsilon_0)$ and any $n_0 \geq \varepsilon^{-2}$, there exists a distribution $q$ over graphs on $n \geq n_0$ nodes such that the following holds. 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 4.1) such that $|Q_a(G)| \leq n\varepsilon^{-2}/(256 \cdot 10^6 \log^2(n))$ and outputs some $X_a(G) \in \mathbb{R}^{n \times n}$. Then*

$$\sum_{S \in \mathcal{S}} \mathbb{P}\left\{\|X_a(G) - N_G\|_* > n\varepsilon | Q_a(G) = S\right\} \mathbb{P}\left\{Q_a(G) = S\right\} \geq 17/32,$$

*where $\mathcal{S} = \left\{E' \subset E : |E'| \leq \frac{162}{(256000\varepsilon)^2} \cdot \frac{1}{(512 \cdot 10^6 \log^2(\varepsilon^{-1} \cdot 9\sqrt{2}/256000)}\right\}$.*

*Proof.* For $\varepsilon_0$ sufficiently small, we can set $b$ such that $b \stackrel{\text{def}}{=} (9\sqrt{2}/(256000\varepsilon))^2 \geq 3000$. Now, set $k = \lceil n_0/2b \rceil$. Note that for $n_0 \geq \varepsilon^{-2}$, we have $k \geq n_0(256000\varepsilon)^2/162 \geq 300$.

**Constructing $q$:** Note that $n \stackrel{\text{def}}{=} 2kb \geq n_0$. We will construct the distribution $q$ over graphs on $n$ 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 $i \neq j \in [b]$ and $r \in [k]$, we draw $X_{i,j,r} \sim \mathrm{Ber}(1/2)$. 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 in the adjacency query model (Definition 4.1) 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 $\{v_i^{r,1}, v_j^{r,2}\}$ or $\{v_i^{r,2}, v_j^{r,1}\}$ 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 if any edges of $\{v_i^{r,1}, v_j^{r,1}\}$ or $\{v_i^{r,2}, v_j^{r,2}\}$ are in $Q_a(G)$, 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 $\bar{G} \in \mathsf{supp}(q)$ where $\mathsf{supp}$ denote the support of the distribution, 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.

In the remainder of the proof, for $S \subset V$, we use $G[S]$ to denote the vertex-induced subgraph of $G$ corresponding to $S$.

**Indistinguishably between $G$ and $\bar{G}$:** Observe that $A_{G[V^{r,1}]} \sim G(b, 1/2)$ for each $r \in [k]$. Let $\mathcal{T}_{c'',c'}$ denote the set of all graphs $H$ on $n_H$ nodes such that

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

Proposition 3.3 ensures that if $H$ is an Erdős-Rényi $G(n, 1/2)$ graph, then $H \in \mathcal{T}_{1/500,1/(16\cdot 10^6)}$ with probability at least $5/8$. Moreover, Lemma 4.2 ensures that if $G[V^{r,1}] \in \mathcal{T}_{1/500,1/(16\cdot 10^6)}$ for some $r \in [k]$, then $G[V^{r,1} \cup V^{r,2}] \in \mathcal{T}_{1/500,1/(64\cdot 10^6)}$ as well (using the fact that $\left|V^{r,1} \cup V^{r,2}\right|^2 = 4\left|V^{r,1}\right|^2$.)

Let $\mathcal{H} \stackrel{\text{def}}{=} \left\{G \in \mathsf{supp}(q) : G[V^{r,1} \cup V^{r,2}] \in \mathcal{T}_{1/500,1/(64\cdot 10^6)} \text{ for more than } 9k/16 \text{ values of } r \in [k]\right\}$. By the independence of the $k$ components in $G$ and Chernoff bound, we have that whenever $k \geq 300$, $\mathbb{P}_{G=(V,E)\sim q}\{G \in \mathcal{H}\} \geq 1 - \exp(-5/8k \cdot 1/200) \geq 1 - \exp(-5k/(1600)) \geq 17/32$.

Consequently, at least a $17/32$ fraction of the graphs in $\mathsf{supp}(q)$ are also 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

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

So, assume $G \in \mathcal{H}$, and without loss of generality, assume that $G[V^{r,1} \cup V^{r,2}] \in \mathcal{T}_{1/500,1/(64\cdot 10^6)}$ for all $r \in [\lceil 9/16 \cdot k \rceil]$. $Q_a(G) = Q_a(\bar{G})$ by construction, so it remains to prove (3). The $r$-th block of $N_G - N_{\bar{G}}$ has the following form for $r \in [\lceil 9/16 \cdot k \rceil]$:

$$N_{G[V^{r,1}\cup V^{r,2}]} - N_{\bar{G}[V^{r,1}\cup V^{r,2}]} = \frac{1}{b-1}\left(\left(A_{G[V^{r,1}\cup V^{r,2}]} - A_{G'[V^{r,1}\cup V^{r,2}]}\right) - P_r\right),$$

where $P_r \stackrel{\text{def}}{=} (A_{\bar{G}[V^{r,1}\cup V^{r,2}]} - A_{G'[V^{r,1}\cup V^{r,2}]})$ is the edge perturbation matrix in the $r$-th block corresponding to the queries in $Q_a(G)$. Now, Lemma 4.3 guarantees that for $r \in [\lceil 9/16 \cdot k \rceil]$,

$$\left(A_{G[V^{r,1}\cup V^{r,2}]} - A_{G'[V^{r,1}\cup V^{r,2}]}\right) \in \mathcal{T}_{1/1000,1/(64\cdot 10^6)}.$$

Let $P$ be the block diagonal matrix containing $P_r$ in the $r$-th block. Since $Q_a(G)$ contains at most $nb \cdot 1/(256 \cdot \log^2(b))$ queries,

$$\mathrm{nnz}(P) \leq 4 \cdot |Q_a(G)| \leq nb \cdot 1/(64 \cdot 10^6 \cdot \log^2(b)).$$

This is because each edge in $Q_a(G)$ corresponds to two edge changes (one edge deleted from and one edge added) to $\bar{G}$ relative to $G'$ and correspondingly, corresponds to four entries modified in the adjacency matrix of $\bar{G}$ relative to $G'$. Consequently, by Proposition 3.4 and Lemma 4.2, it follows that

$$\left\| A_{G\left[\bigcup_{r\in[\lceil 9k/16\rceil]} V^{r,1}\cup V^{r,2}\right]} - A_{G'\left[\bigcup_{r\in[\lceil 9k/16\rceil]} V^{r,1}\cup V^{r,2}\right]} - P' \right\|_* \geq \frac{1}{1000} \cdot 9/16 \cdot n \cdot \sqrt{b} \cdot \frac{1}{4}.$$

Since each $G[V^{r,1} \cup V^{r,2}]$ forms its own connected component and each $G[V^{r,1}$ is drawn from $G(b, 1/2)$, the maximum degree in $G'$ and $\bar{G}$ is at most $2b$. Thus,

$$\|N_G - N_{\bar{G}}\|_* \geq \frac{1}{1000} \cdot \frac{1}{4\sqrt{2}} \cdot \frac{9}{16} \cdot \frac{n}{\sqrt{b}} = \frac{9\sqrt{2}}{128000} \frac{n}{\sqrt{b}} \geq 2n\varepsilon. \qquad \blacksquare$$

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

**Theorem 1.12** (Query Lower Bound)**.** *For any $\varepsilon \leq \varepsilon_0$, where $\varepsilon_0 \in (0, 1)$ is a fixed constant, any algorithm working in the adjacency query model (Definition 1.4) requires $\Omega(n\varepsilon^{-2} \log^{-2}(\varepsilon^{-1}))$* GetNeighbor *queries to return an $\varepsilon$-additive nuclear sparsifier with probability $3/4$.*

*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 generalized adjacency query 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 $c(a, G = (V, E)) \stackrel{\text{def}}{=} \mathbb{1}\{\|X_a(G) - N_E\|_* > n\varepsilon\}$ (where $\mathbb{1}\{\cdot\}$ denotes the indicator function). 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 $q$ such that for any $S \in \mathcal{S}$ and any $a \in A$, the following hold

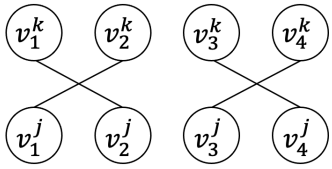
$$\sum_{S\in\mathcal{S}} \mathbb{P}\left\{\|X_a(G) - N_G\|_* > n\varepsilon | Q_a(G) = S\right\} \mathbb{P}\left\{Q_a(G) = S\right\} \geq 17/32 \cdot 1/2 \geq 1/4.$$

Lemma 4.4 guarantees this distribution $q$, concluding the proof. $\blacksquare$

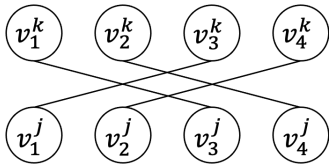
## 5 Separation from Additive Spectral Sparsifiers

In this section, we discuss the query lower bounds for any deterministic algorithm that produces $\varepsilon$-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 generalized adjacency query model for required by any deterministic $\varepsilon$ additive spectral norm approximation algorithm on undirected graphs.




(a) If any of the four depicted edges are queried, then we add all four edges to $E$, and $E$ will not contain any other edges between these eight vertices.




(b) If any of the four depicted edges are queried, then we add all four edges to $E$, and $E$ will not contain any other edges between these eight vertices.

Figure 1: Illustration of construction of $E$ in the proof of Theorem 5.1.

**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 4.1). 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 $\left| v^\top \bar{L}_G v - X \right| > \frac{1}{8} v^\top v$.*

*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 $Q$ be the set of edges queried 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 $\{v_1^k, v_3^j\}, \{v_1^j, v_3^k\}, \{v_2^k, v_4^j\}, \{v_2^j, v_4^k\} \in E$.

Let $E_1 = \{\{v_1^k, v_2^j\}, \{v_3^k, v_4^j\} : \{v_i^k, v_\ell^j\} \notin E,\ i, \ell \in [4]\}$ and $E_2 = \{\{v_1^k, v_3^j\}, \{v_2^k, v_4^j\} : \{v_i^k, v_\ell^j\} \notin E,\ i, \ell \in [4]\}$. 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. For an illustration, see Figure 1. Let $\overline{L}_1 = I - N_{G_1}$ be the normalized Laplacian of $G_1$ and $\overline{L}_2 = I - N_{G_2}$ 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 $i \in [2]$,

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

**Algorithm 2:** NuclearSparsify$(G, \varepsilon, n, T)$

**Input:** Graph $G = (V = [n], E, w)$, under the one-step random walk query model (Definition 1.13), Accuracy $\varepsilon$, 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
    $X_{i,j} \leftarrow X_{i,j} + 1/(2T) \cdot n \cdot \sqrt{\deg(i)/\deg(j)}$
    $X_{j,i} \leftarrow X_{j,i} + 1/(2T) \cdot n \cdot \sqrt{\deg(i)/\deg(j)}$
**Return:** $X$

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

$$\left|v^\top \overline{L}_1 v - v^\top \overline{L}_2 v\right| \geq \frac{1}{m}\left(m^2 - 4\left|Q\right|\right) = m - \frac{4\left|Q\right|}{m} \geq \frac{m}{2} = \frac{1}{4} v^\top v,$$

whenever $4\left|Q\right|/m \leq m/2$, i.e., $|Q| \leq m^2/8 = n^2/128$. So, either $\left|v^\top \overline{L}_1 v - X\right| \geq \frac{1}{8} v^\top v$ or $\left|v^\top \overline{L}_2 v - X\right| \geq \frac{1}{8} v^\top v$. ■

**Theorem 1.8.** *Any deterministic algorithm requires $\Omega(n^2)$ Adjacency Queries (Definition 1.4) to compute a $\frac{1}{4}$-additive spectral sparsifier (Definition 1.7), even for unweighted graphs.*

*Proof.* Note that given $f : \mathbb{R}^n \mapsto \mathbb{R}$, an $\varepsilon$ 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

$$\left|g(x) - x^\top \bar{L}_G x\right| = \left|x^\top x - f(x) - x^\top (I - N_G) x\right| = \left|f(x) - x^\top N_G x\right| \leq \varepsilon \left\|x\right\|_2^2.$$

Note that given $f$, computing $g$ requires no additional queries to the unweighted graph access oracle. Consequently, the result follows by taking any $\varepsilon < 1/8$ and applying Lemma 5.1. As the hard instance graph in Lemma 5.1 is unweighted, the theorem holds even for unweighted graphs. ■

# 6 Sparsification in the Random Walk Model

In this section, we show how our notion of nuclear sparsification (and indeed, $\varepsilon$-additive spectral sparsification) can also be achieved in the one-step random walk model (Definition 1.13, which is a more restrictive query model than the one considered by Braverman et al. [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 achievable for spectral sparsification from what is achievable with nuclear sparsification.

## 6.1 Algorithms for sparsification in the random walk model

In this section, we present algorithms to build an $\varepsilon$-additive nuclear sparsifier and an $\varepsilon$-additive spectral sparsifier under the one-step random walk query model (Definition 1.13). In Theorem 1.14,

we show that by taking $T = O(n\varepsilon^{-2})$ in Algorithm 2, with constant probability, we can obtain a nuclear sparsifier. In Theorem 1.15, we show that by taking $T = O(n\varepsilon^{-2}\log(n))$ in Algorithm 3, we can obtain an $\varepsilon$-additive spectral sparsifier.

**Theorem 1.14.** *There is an algorithm (Algorithm 2) that, for any $\varepsilon \in (0,1)$, returns with probability $2/3$ an $O(n\varepsilon^{-2})$-sparse $\varepsilon$-additive nuclear sparsifier for any undirected weighted graph $G$ using just $O(n\varepsilon^{-2})$ queries in the one-step random walk query model (Definition 1.13).*

*Proof.* Let $i^{(t)} \sim \mathsf{Uniform}([n])$ and $j^{(t)} = j$ with probability proportional to $w_{ij}$ 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

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

Furthermore,

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

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

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

Setting $T = 3n\varepsilon^{-2}$ suffices to succeed with probability $2/3$. ∎

**Algorithm 3:** SpectralAdditiveSparsify$(G, \varepsilon, n, T)$

**Input:** Graph $G = (V = [n], E, w)$, under the one-step random walk query model (Definition 1.13), Accuracy $\varepsilon$, 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:** $X$

By combining the sparsification procedure in Algorithm 2 with the Chebyshev moment-matching algorithm of Braverman et al. [BKM22], we obtain a $O(n\varepsilon^{-3})$ algorithm for spectral density estimation in this one-step random walk query model (Definition 1.13).

In the following Theorem, we note that by using a result of Cohen et al. [Coh+17] we can also achieve the stronger notion of $\varepsilon$-additive spectral sparsification in the one-step random walk access model (Definition 1.13.)

**Theorem 6.1** (Theorem 3.9 (Simplified), [Coh+17])**.** *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix where no row or column is all zeros. Let $\varepsilon, p \in (0, 1)$. 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*

$$X = \left(\frac{A_{ij}}{p_{ij}}\right) E_{ij} \text{ with probability } p_{ij} = \frac{A_{ij}}{2n}\left[\frac{1}{r_i} + \frac{1}{r_j}\right], \ \forall A_{ij} \neq 0\,.$$

*If $A_1, \ldots, A_T$ are sampled independently from $\mathcal{D}$ for $T \geq 128\frac{2n}{\varepsilon^2}\log\frac{2n}{p}$, $R = Diag(r)$ (where $Diag(r)$ is a diagonal matrix with $Diag(r)_{ii} = r_i$, for $i \in [n]$) then the average $\tilde{A} \stackrel{\text{def}}{=} \frac{1}{T}\sum_{t\in[T]} A_t$ satisfies*

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

Using the above theorem, we show that with an additional $O(\log(n))$-factor in the number of queries, the Algorithm 3 we can obtain the stronger notion of $\varepsilon$-additive spectral sparsifier in the one-step random walk access model (Definition 1.13.)

**Theorem 1.15.** *There is an algorithm (Algorithm 3) that, for any $\varepsilon \in (0, 1)$, returns with probability $2/3$ an $O(n\varepsilon^{-2}\log n)$-sparse $\varepsilon$-additive spectral sparsifier for any undirected weighted graph $G$ using just $O(n\varepsilon^{-2}\log n)$ queries in the one-step random walk query model (Definition 1.13).*

*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

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

All other entries are 0. Note that then

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

with probability $\frac{1}{2n} \cdot \left(\frac{w_{ij}}{\deg(i)} + \frac{w_{ij}}{\deg(j)}\right)$. By Theorem 6.1, setting $A = N_G$, for $i \in [T]$ setting $A_i = A^{(t)}$, and $p = 1/3$ in Theorem 6.1 and taking $T \geq 128\frac{2n}{\varepsilon^2}\log(2n/p)$, 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 one-step random walk graph query oracle. Our result relies on the following classical result on the sample complexity for the coupon collector problem.

**Lemma 6.2** (Coupon Collector)**.** *Consider a collection of $n \geq 1$ different coupons from which coupons are being drawn independently, with equal probability, and with replacement. Let $T$ be a random variable which denotes the number of trials needed to see all the $n$ different coupons. Then, for $c > 0$*

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

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

**Theorem 1.16.** *For a fixed constant $\varepsilon \in (0, 1/8)$ any algorithm requires $\Omega(n\log n)$ one-step random walk model queries to output an $\varepsilon$-additive spectral sparsifier with probability $> c$.*

*Proof.* For each $i \in [n]$, let $X_i$ be an independent Bernoulli random variable. We construct a family of bi-partite graphs on $2n$ nodes $G_X = (V, E)$ based on $X_1, \ldots, X_n$. Let $V = A \sqcup B$ denote the vertices of $G_X$, such that $|A| = |B| = n$. Let $a_1, \ldots, a_n$ denote the vertices in $A$, and $b_1, \ldots, b_n$ denote the vertices in $B$. If $X_i = 1$ we add $(a_i, b_i)$ to $E$.

Let $A$ denote the adjacency matrix of $G_X$. 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 $i \in [n] \setminus \mathcal{I}$, or an edge $(a_i, b_i)$ for $i \in \mathcal{I}$. If the RandomNeighbor query returns a vertex $a_i$ (or $b_i$) for $i \notin \mathcal{I}$, then the algorithm knows that $(a_i, b_i) \notin E$. Therefore the RandomNeighbor query corresponds to an sampling a pair $(a_i, b_i)$ for $i \in [n]$. The pair $(a_i, b_i)$ for $i \in [n]$ corresponds to coupons in the coupon collector problem.

Suppose there exists an algorithm that takes $T \leq n\log n - \log(1/c)n$ 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 $T \leq n\log n - \log(1/c)n$ queries with constant probability $c$, which leads to a contradiction.

This follows because we can determine all the edges in the graph $G_X$, and hence all the coupons $(a_i, b_i)$ for $i \in [n]$, from $\widetilde{A}$ by taking $x \in \mathbb{R}^{2n}$ indexed by $a_i$ and $b_i$, for $i \in [n]$. We take $x$ to be the 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 $k \neq a_i$ or $b_i$. 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 $T \leq n \log n - \log(1/c)n$ oracle queries, fails to produce a spectral sparsifier, with probability greater than $c$.

Note that for family of graphs $G_X$, since the degree of the graph 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 Omitted Facts and Proofs

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

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

$$\|A\|_* = \max_{X : \|X\|_2 \leq 1} \langle A, X \rangle = \max_{X : \|X\|_2 = 1} \langle A, X \rangle .$$

## A.1 Small Nuclear Norm Error Implies Small Wasserstein Error

**Fact 1.3.** *For symmetric matrices $A, B \in \mathbb{R}^{n \times n}$, if $\|A - B\|_* \leq n\varepsilon$, then $W_1(A, B) \leq \varepsilon$.*[9]

*Proof.* Let $\lambda_{\min}$ be the smallest eigenvalue among the eigenvalues of $A$ and $B$, so that $A - \lambda_{\min} I$ and $B - \lambda_{\min} I$ are both positive semidefinite, where $I$ denotes an $n \times n$ identity matrix. The singular values of $A - \lambda_{\min} I$ and $B - \lambda_{\min} I$ equal the eigenvalues of those matrices. So, we can directly apply Mirsky's singular value perturbation inequality [Mir60, Theorem 5] to conclude that:

$$\begin{aligned} \|A - B\|_* = \|A - \lambda_{\min} I - (B - \lambda_{\min} I)\|_* &\geq \sum_{i \in [n]} |\lambda_i(A - \lambda_{\min} I) - \lambda_i(B - \lambda_{\min} I)| \\ &= \sum_{i \in [n]} |\lambda_i(A) - \lambda_i(B)| = n W_1(A, B). \end{aligned}$$ ■

## A.2 Additive Nuclear Sparsification from Additive Nuclear Approximation

**Lemma A.2.** *Let $G = (V, E, w)$ be a weighted graph with adjacency matrix $A_G \in \mathbb{R}^{n \times n}$, degree matrix $D_G$, and normalized adjacency matrix $N_G$. Without loss of generality, assume that $V = \{1, \ldots, n\}$ and the vertices are ordered so that $\deg(i) \geq deg(j)$ for $i < j$. For $\varepsilon \in (0, 1)$, let $A_{G'} \in \mathbb{R}^{n \times n}$ be any matrix such that $\|N_G - D_G^{-1/2} A_{G'} D_G^{-1/2}\|_F^2 \leq n\varepsilon^2$. As in the proof of Theorem 2.1, such a matrix can be constructed by Algorithm 1. Let $\mathbf{e} \in \mathbb{R}^{n-1}$ be a vector with $e_i = \deg(i) - \sum_{j=1}^{n-1} (A_{G'})_{i,j}$. For a matrix $M \in \mathbb{R}^{n \times n}$, let $M_{[1:j]}$ denotes the leading principal submatrix of order $j$. Let $G''$ be a graph with adjacency and degree matrix*

$$A_{G''} = \begin{pmatrix} (A_{G'})_{[1:n-1]} & \mathbf{e} \\ \mathbf{e}^\top & 0 \end{pmatrix} \qquad D_{G''} = \begin{pmatrix} (D_G)_{[1:n-1]} & \mathbf{0} \\ \mathbf{0}^\top & \|\mathbf{e}\|_1 \end{pmatrix}.$$

[9] When $A$ and $B$ are diagonal with monotonically decreasing diagonal entries then $\|A - B\|_* = n \cdot W_1(A, B)$. In this sense, nuclear approximation is a natural strengthening of approximation in the Wasserstein-1 distance.

*Then,* $\|N_G - N_{G''}\|_* \leq 3n\varepsilon$ *whenever* $\varepsilon^{-2} \leq n$.

*Proof.* As in the proof of Theorem 2.1, it suffices to show that $\left\|N_G - D_{G''}^{-1/2} A_{G''} D_{G''}^{-1/2}\right\|_F^2 \leq 9n\varepsilon^2$. 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 assumption. Consequently,

$$\begin{aligned}\left\|N_G - D_{G''}^{-1/2} A_{G''} D_{G''}^{-1/2}\right\|_F^2 &\leq n\varepsilon^2 + 2\sum_{i=1}^{n-1}\left(\frac{w_{\{i,n\}}}{\sqrt{\deg(i)}\sqrt{\deg(n)}} - \frac{e_i}{\sqrt{\deg(i)}\sqrt{\|\mathbf{e}\|_1}}\right)^2 \\ &\leq n\varepsilon^2 + 2\sum_{i=1}^{n-1}\frac{w_{\{i,n\}}^2}{\deg(i)\deg(n)} + 2\sum_{i=1}^{n-1}\frac{e_i^2}{\deg(i)\,\|\mathbf{e}\|_1}. \end{aligned} \tag{5}$$

We can bound the first term in (5) by using the fact that $\deg(n) \leq \deg(i)$:

$$\sum_{i=1}^{n-1}\frac{w_{\{i,n\}}^2}{\deg(i)\deg(n)} \leq \frac{1}{\deg(n)^2}\sum_{i=1}^{n-1} w_{\{i,n\}}^2 \leq \frac{1}{\deg(n)^2}\left(\sum_{i=1}^{n-1} w_{\{i,n\}}\right)^2 = 1.$$

We can bound the second term by observing that, since since $0 \leq e_i \leq \deg(i)$,

$$\sum_{i=1}^{n-1}\frac{e_i^2}{\deg(i)\,\|\mathbf{e}\|_1} \leq \frac{1}{\|\mathbf{e}\|_1}\sum_{i=1}^{n-1} e_i = 1.$$

Plugging into (5), we conclude that $\left\|N_G - D_{G''}^{-1/2} A_{G''} D_{G''}^{-1/2}\right\|_F^2 \leq n\varepsilon^2 + 4$, which is less than $9n\varepsilon^2$ whenever $\varepsilon^{-2} \leq n$. ∎

## References


[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: *61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2020.

[BSS12] Joshua Batson, Daniel A. Spielman, and Nikhil Srivastava. "Twice-Ramanujan sparsifiers". In: *SIAM Journal on Computing* 41.6 (2012). Preliminary version in the 41st Annual ACM Symposium on Theory of Computing (STOC), pp. 1704–1721.

[BK96] András A. Benczúr and David R. Karger. "Approximating S-T Minimum Cuts in $\tilde{O}(N^2)$ Time". In: *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC)*. 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: *Proceedings of the 50th International Colloquium on Automata, Languages and Programming (ICALP)*. 2023, 21:1–21:18.

[Bha+24] Rajarshi Bhattacharjee, Gregory Dexter, Cameron Musco, Archan Ray, Sushant Sachdeva, and David P. Woodruff. "Universal Matrix Sparsifiers and Fast Deterministic Algorithms for Linear Algebra". In: *Proceedings of the 15th Conference on Innovations in Theoretical Computer Science (ITCS)*. 2024.

[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 Symposium on Theory of Computing (STOC)*. 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, and Christian Sohler. "Sublinear-time approximation of Euclidean minimum spanning tree." In: *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2003, pp. 813–822.

[CS07] Artur Czumaj and Christian Sohler. "Testing Expansion in Bounded-Degree Graphs". In: *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2007, pp. 570–578.

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

[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.

[DWZ23] Ran Duan, Hongxun Wu, and Renfei Zhou. "Faster matrix multiplication via asymmetric hashing". In: *Proceedings of the 64th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2023, pp. 2129–2138.

[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 Symposium on Theory of Computing (STOC)*. 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.

[FDBV01] Illés J Farkas, Imre Derényi, Albert-László Barabási, and Tamas Vicsek. "Spectra of "real-world" graphs: Beyond the semicircle law". In: *Physical Review E* 64.2 (2001), p. 026704.

[GKX19] Behrooz Ghorbani, Shankar Krishnan, and Ying Xiao. "An Investigation into Neural Net Optimization via Hessian Eigenvalue Density". In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*. 2019.

[Glu+21] Grzegorz Gluch, Michael Kapralov, Silvio Lattanzi, Aida Mousavifar, and Christian Sohler. "Spectral clustering oracles in sublinear time". In: *Proceedings of the 32nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2021, pp. 1598–1617.

[GR11] Oded Goldreich and Dana Ron. "On Testing Expansion in Bounded-Degree Graphs". In: *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*. Springer Berlin Heidelberg, 2011, pp. 68–75.

[GZ00] Alice Guionnet and Ofer Zeitouni. "Concentration of the Spectral Measure for Large Matrices". In: *Electronic Communications in Probability* 5 (2000), pp. 119–136.

[JRT23] Arun Jambulapati, Victor Reis, and Kevin Tian. "Linear-Sized Sparsifiers via Near-Linear Time Discrepancy Theory". In: *arXiv preprint arXiv:2305.08434* (2023).

[JMSS23] Yujia Jin, Christopher Musco, Aaron Sidford, and Apoorv Vikram Singh. "Moments, Random Walks, and Limits for Spectrum Approximation". In: *Proceedings of the 36th Annual Conference on Computational Learning Theory (COLT)*. 2023, pp. 5373–5394.

[Kap+17] Michael Kapralov, Yin Tat Lee, Cameron Musco, Christopher Musco, and Aaron Sidford. "Single Pass Spectral Sparsification in Dynamic Streams". In: *SIAM Journal on Computing* 46.1 (2017), pp. 456–477.

[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". In: *arXiv preprint arXiv:1401.0085* (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.

[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: *Proceedings of the 12th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Vol. 1. 2001, pp. 439–447.

[Mou21] Aidasadat Mousavifar. "Sublinear Algorithms for Spectral Graph Clustering". PhD thesis. EPFL, 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: *Proceedings of the 9th Conference on Innovations in Theoretical Computer Science (ITCS)*. 2018, 8:1–8:21.

[NS10] Asaf Nachmias and Asaf Shapira. "Testing the expansion of a graph". In: *Information and Computation* 208.4 (2010), pp. 309–314.

[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 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2012, pp. 1123–1131.

[PSG18] Jeffrey Pennington, Samuel Schoenholz, and Surya Ganguli. "The emergence of spectral universality in deep networks". In: *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2018.

[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).

[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* (2011).

[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, pp. 145–155.

[WWAF06] Alexander Weiße, Gerhard Wellein, Andreas Alvermann, and Holger Fehske. "The kernel polynomial method". In: *Reviews of modern physics* 78.1 (2006), p. 275.