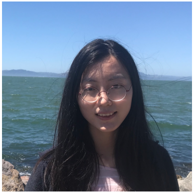
Faster Spectral Density Estimation and Sparsification in the Nuclear Norm

COLT 2024, Edmonton, Canada

Apoorv Vikram Singh
(NYU)



Yujia Jin (Stanford)

Ishani Karmarkar (Stanford)

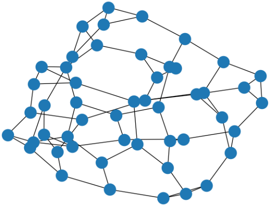
Christopher Musco (NYU)

Aaron Sidford (Stanford)

Problem Setup

- Graph G, with normalized adjacency matrix N.

$$N := D^{-1/2} A D^{-1/2}, \quad D = \mathrm{diag}(\deg(v_1), \dots, \deg(v_n))$$



- Spectrum of $G \equiv$ Eigenvalues of N:

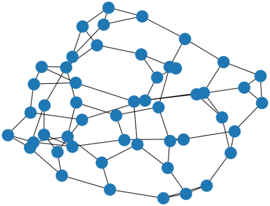
$$-1 \leq \lambda_n \leq \dots \leq \lambda_1 = 1.$$

Goal: Estimate the spectrum of G.

Problem Setup

- Graph G, with normalized adjacency matrix N.

$$N := D^{-1/2} A D^{-1/2}, \quad D = \mathrm{diag}(\mathsf{deg}(v_1), \ldots, \mathsf{deg}(v_n))$$



compute in $\lesssim \; n^{2.371}$

- Spectrum of $G \equiv$ Eigenvalues of N.

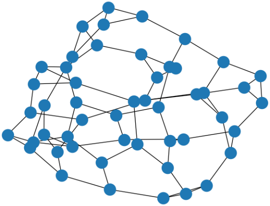
$$-1 \leq \lambda_n \leq \cdots \leq \lambda_1 = 1.$$

Goal: Estimate the spectrum of G.

Problem Setup

- Graph G, with normalized adjacency matrix N.

$$N := D^{-1/2} A D^{-1/2}, \quad D = \mathrm{diag}(\deg(v_1), \ldots, \deg(v_n))$$



- Spectrum of $G \equiv$ Eigenvalues of N:

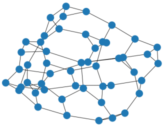
sublinear time: $o(n^2)$

$$-1 \le \lambda_n \le \cdots \le \lambda_1 = 1.$$

Goal: Estimate the spectrum of G.

Goal

(Graph G, normalized adjacency matrix N)



Graph G

Eigenvalues of N

$$-1 \leq \lambda_n \leq \cdots \leq \lambda_1 = 1.$$

Goal: Output $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$ such that

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

Representing Spectrum as a Distribution

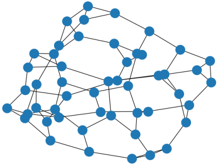
Eigenvalues of N: $-1 \leq \lambda_n \leq \cdots \leq \lambda_1 = 1$

This is a purely algorithmic problem. Faster algorithms use statistical techniques to solve it.

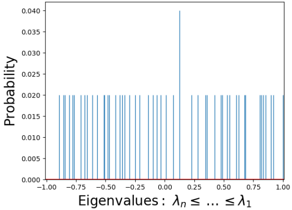
Representing Spectrum as a Distribution

Eigenvalues of N: $-1 \leq \lambda_n \leq \cdots \leq \lambda_1 = 1$

This is a purely algorithmic problem. Faster algorithms use statistical techniques to solve it.



Graph G



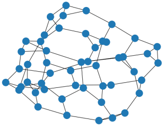
Spectral Density p (pmf)

$p \equiv \mathrm{SpectralDensity}(G)$:

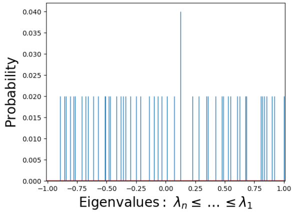
$$p = \mathrm{Unif}(\{\lambda_n, \ldots, \lambda_1\})$$

Spectral Density Estimation (SDE)

Eigenvalues of N: $-1 \leq \lambda_n \leq \cdots \leq \lambda_1 = 1$



Graph G



$[p = \mathrm{Unif}(\{\lambda_n, \ldots, \lambda_1\})]$

Spectral Density p (pmf)

Recall Goal: Output $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$ such that $\sum_{i=1}^{n} \left|\lambda_i - \tilde{\lambda}_i\right| \leq n\varepsilon$.

Equivalently: Output q such that Wasserstein-1 distance:

$$W_1(p, q) \leq \varepsilon.$$

Learning Distribution p From Moments

$[X \sim p = \mathrm{Unif}(\{\lambda_n, \ldots, \lambda_1\})]$

Learn p by estimating its moments.

$$m_j = \mathbb{E}[X^j] = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^j$$
$$= \frac{1}{n} \mathrm{Tr}(N^j).$$

Idea:

- Estimate $m_1, \ldots, m_{1/\varepsilon}$ to high accuracy.
- Find a distribution q that has moments close to these estimates.

Existing Results

Graph G with spectral density p

Assume G is stored as adjacency list.

$\underline{\mathcal{O}(1)\text{-time}}$: Access degree of node &
Identity of its i-th neighbour, for any i.

Existing Results

Graph G with spectral density p

Assume G is stored as adjacency list.

$\underline{\mathcal{O}(1)\text{-time}}$: Access degree of node &
Identity of its i-th neighbour, for any i.

- [Braverman et al. (2022)] Output q: $W_1(p, q) \leq \varepsilon$ in time $\mathcal{O}\left(n\varepsilon^{-7}\right)$.
- [Cohen-Steiner et al. (2018)] Output q: $W_1(p, q) \leq \varepsilon$ in time $2^{\mathcal{O}(1/\varepsilon)}$.

Existing Results

Graph G with spectral density p

Assume G is stored as adjacency list.

$\underline{\mathcal{O}(1)\text{-time}}$: Access degree of node &
Identity of its i-th neighbour, for any i.

- [Braverman et al. (2022)] Output q: $W_1(p,q) \leq \varepsilon$ in time $\mathcal{O}\left(n\varepsilon^{-7}\right)$.
- [Cohen-Steiner et al. (2018)] Output q: $W_1(p,q) \leq \varepsilon$ in time $2^{\mathcal{O}(1/\varepsilon)}$.

Our Results

- Output q: $W_1(p,q) \leq \varepsilon$ whp in time $\mathcal{O}\left(n\varepsilon^{-3}\right)$.
- Deterministic $\mathcal{O}\left(n/\varepsilon^2\right)$ queries to graph.
- *Simple algorithm via "Nuclear Norm Sparsification"*

Existing Results

Graph G with spectral density p

Assume G is stored as adjacency list.

$\underline{\mathcal{O}(1)\text{-time}}$: Access degree of node &
Identity of its i-th neighbour, for any i.

- [Braverman et al. (2022)] Output q: $W_1(p,q) \le \varepsilon$ in time $\mathcal{O}\left(n\varepsilon^{-7}\right)$.
- [Cohen-Steiner et al. (2018)] Output q: $W_1(p,q) \le \varepsilon$ in time $2^{\mathcal{O}(1/\varepsilon)}$.

Our Results

- Output q: $W_1(p,q) \le \varepsilon$ whp in time $\mathcal{O}\left(n\varepsilon^{-3}\right)$.
- Deterministic $\mathcal{O}\left(n/\varepsilon^2\right)$ queries to graph.
- *Simple algorithm via "Nuclear Norm Sparsification"*

Also works for weighted graphs.
$\mathcal{O}\left(n/\varepsilon^2\right)$ query complexity in random-sampling model.

Our Goal

[Eigenvalues of $N : \lambda_n \leq \cdots \leq \lambda_1$]

[BKM'22] Return q such that $W_1(p, q) \leq \varepsilon$ whp in time $\mathcal{O}\left(\mathrm{nnz}(N) \cdot \frac{1}{\varepsilon}\right)$.

Idea: Construct <u>sparse</u> approximation to N.

Our Goal

[Eigenvalues of $N : \lambda_n \leq \cdots \leq \lambda_1$]

[BKM'22] Return q such that $W_1(p,q) \leq \varepsilon$ whp in time $\mathcal{O}\left(\mathrm{nnz}(N) \cdot \frac{1}{\varepsilon}\right)$.

Idea: Construct sparse approximation to N.

Aim:

1. **Sparse**: Find a sparse matrix $\tilde{N}$.
2. **Close Eigenvalues**: $\tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_n$ of $\tilde{N}$ such that

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

3. **Fast**: Construct $\tilde{N}$ fast.

Sparsification Ideas

[Eigenvalues of N: $\lambda_n \leq \cdots \leq \lambda_1$. Eigenvalues of $\tilde{N}$: $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$]

Recall aim: Sparse, close eigenvalues, and fast to construct

Idea 1: Spectral Sparsification: $(1-\varepsilon)N \preccurlyeq \tilde{N} \preccurlyeq (1+\varepsilon)N$

- $\implies \tilde{\lambda}_i \in (1 \pm \varepsilon)\lambda_i$.

1. Achieves sparsity $o(n^2)$.
2. Eigenvalues are close: $\sum_{i=1}^{n} \left|\lambda_i - \tilde{\lambda}_i\right| \leq \varepsilon n$.
3. [Lee'13] Needs $\Omega(n^2)$ queries to G.

Sparsification Ideas

[Eigenvalues of N: $\lambda_n \leq \cdots \leq \lambda_1$. Eigenvalues of $\tilde{N}$: $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$]

Recall aim: Sparse, close eigenvalues, and fast to construct

Idea 1: Spectral Sparsification: $(1-\varepsilon)N \preccurlyeq \tilde{N} \preccurlyeq (1+\varepsilon)N$

- $\implies \tilde{\lambda}_i \in (1 \pm \varepsilon)\lambda_i$.

1. Achieves sparsity $o(n^2)$.
2. Eigenvalues are close: $\sum_{i=1}^{n} \left|\lambda_i - \tilde{\lambda}_i\right| \leq \varepsilon n$.
3. [Lee'13] Needs $\Omega(n^2)$ queries to G.

Idea 2: Additive Spectral Sparsification: $\left\|N - \tilde{N}\right\|_2 \leq \varepsilon$

- $\implies \tilde{\lambda}_i \in \lambda_i \pm \varepsilon$.

1. Achieves sparsity $o(n^2)$.
2. Eigenvalues are close: $\sum_{i=1}^{n} \left|\lambda_i - \tilde{\lambda}_i\right| \leq \varepsilon n$.
3. Randomized procedure to construct $\tilde{N}$ needs $\mathcal{O}\left(n \log n/\varepsilon^2\right)$ queries to G.

- Deterministic procedure requires $\Omega(n^2)$ queries to G.
- Randomized procedure needs $\Omega(n \log n)$ queries to G.

Our Idea: Nuclear Norm Sparsification

[Eigenvalues of N: $\lambda_n \leq \cdots \leq \lambda_1$. Eigenvalues of $\tilde{N}$: $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$]

$$\boxed{\|N\|_* = |\lambda_n| + \cdots + |\lambda_1|.}$$

Our Idea: Nuclear Norm Sparsification

[Eigenvalues of N: $\lambda_n \leq \cdots \leq \lambda_1$. Eigenvalues of $\tilde{N}$: $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$]

$$\|N\|_* = |\lambda_n| + \cdots + |\lambda_1|.$$

$\tilde{N}$ is a nuclear norm sparsifier if

$$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n, \text{ and } \mathrm{nnz}(\tilde{N}) = o(n^2).$$

Our Idea: Nuclear Norm Sparsification

[Eigenvalues of N: $\lambda_n \leq \cdots \leq \lambda_1$. Eigenvalues of $\tilde{N}$: $\tilde{\lambda}_n \leq \cdots \leq \tilde{\lambda}_1$]

$$\|N\|_* = |\lambda_n| + \cdots + |\lambda_1|.$$

$\tilde{N}$ is a nuclear norm sparsifier if

$$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n, \text{ and } \operatorname{nnz}(\tilde{N}) = o(n^2).$$

By Weyl's Inequality,

$$\left\|N - \tilde{N}\right\|_* \geq \sum_{i=1}^{n} \left|\lambda_i - \tilde{\lambda}_i\right|.$$

$$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n \implies \sum_{i=1}^{n} \left|\lambda_i - \tilde{\lambda}_i\right| \leq \varepsilon n.$$

Nuclear Norm Sparsification

$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$, and $\tilde{N}$ is sparse

Sparsify N Under adjacency list access to G.

Our Results: $\tilde{N}$ with following properties:

1. $\mathrm{nnz}(\tilde{N}) = \mathcal{O}\left(n/\varepsilon^2\right)$.

Nuclear Norm Sparsification

$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$, and $\tilde{N}$ is sparse

Sparsify N Under adjacency list access to G.

Our Results: $\tilde{N}$ with following properties:

1. $\mathrm{nnz}(\tilde{N}) = \mathcal{O}\left(n/\varepsilon^2\right)$.
2. $\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$.

Nuclear Norm Sparsification

$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$, and $\tilde{N}$ is sparse

Sparsify N Under adjacency list access to G.

Our Results: $\tilde{N}$ with following properties:

1. $\mathrm{nnz}(\tilde{N}) = \mathcal{O}\left(n/\varepsilon^2\right)$.
2. $\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$.
3. Requires $\mathcal{O}\left(n/\varepsilon^2\right)$ time and queries.

Nuclear Norm Sparsification

$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$, and $\tilde{N}$ is sparse

Sparsify N Under adjacency list access to G.

Our Results: $\tilde{N}$ with following properties:

1. $\mathrm{nnz}(\tilde{N}) = \mathcal{O}\left(n/\varepsilon^2\right)$.
2. $\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$.
3. Requires $\mathcal{O}\left(n/\varepsilon^2\right)$ time and queries.
4. Deterministic procedure.

Nuclear Norm Sparsification

$\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$, and $\tilde{N}$ is sparse

Sparsify N Under adjacency list access to G.

Our Results: $\tilde{N}$ with following properties:

1. $\mathrm{nnz}(\tilde{N}) = \mathcal{O}\left(n/\varepsilon^2\right)$.
2. $\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$.
3. Requires $\mathcal{O}\left(n/\varepsilon^2\right)$ time and queries.
4. Deterministic procedure.
5. Almost matching lower bound on sparsification and queries.

Nuclear Norm Sparsification

$\left\| N - \tilde{N} \right\|_* \leq \varepsilon n$, and $\tilde{N}$ is sparse

Sparsify N Under adjacency list access to G.

Our Results: $\tilde{N}$ with following properties:

1. $\mathrm{nnz}(\tilde{N}) = \mathcal{O}\left(n/\varepsilon^2\right)$.
2. $\left\| N - \tilde{N} \right\|_* \leq \varepsilon n$.
3. Requires $\mathcal{O}\left(n/\varepsilon^2\right)$ time and queries.
4. Deterministic procedure.
5. Almost matching lower bound on sparsification and queries.

For spectral density estimation:

- Total Running Time [using [BKM'22]]: $\mathcal{O}\left(n/\varepsilon^2\right) + \mathrm{nnz}(\tilde{N}) \cdot \frac{1}{\varepsilon} = \mathcal{O}\left(n/\varepsilon^3\right)$.
- Total queries to G: $\mathcal{O}\left(n/\varepsilon^2\right)$.

Sparsification Algorithm

$N = D^{-1/2}AD^{-1/2}$. Note: If (v, v') is an edge, then $N_{v,v'} = 1/\sqrt{\deg(v)\deg(v')}$

$\boxed{\tilde{N} \text{ zeroes out } v\text{-th row and column of } N \text{ if } \deg(v) > 1/\varepsilon^2.}$

Sparsification Algorithm

$N = D^{-1/2} A D^{-1/2}$. Note: If (v, v') is an edge, then $N_{v,v'} = 1/\sqrt{\deg(v)\deg(v')}$

$\tilde{N}$ zeroes out v-th row and column of N if $\deg(v) > 1/\varepsilon^2$.

Proof Sketch:

- Each row of $\tilde{N}$ has $\leq 1/\varepsilon^2$ non zero entries.
  - $\mathrm{nnz}(\tilde{N}) = n/\varepsilon^2$
- $\tilde{N}$ deletes only small entries of N.
- Prove that $\left\|N - \tilde{N}\right\|_F \leq \varepsilon\sqrt{n}$.
- Implies $\left\|N - \tilde{N}\right\|_* \leq \sqrt{n}\left\|N - \tilde{N}\right\|_F = \varepsilon n$.

Deterministic Sublinear-Time Spectral Density Estimation

Our Results: Deterministic, Sublinear Time, Spectral Density Estimation in time $n \cdot 2^{\tilde{\mathcal{O}}(\varepsilon^{-1})}$.

Deterministic Sublinear-Time Spectral Density Estimation

Our Results: Deterministic, Sublinear Time, Spectral Density Estimation in time $n \cdot 2^{\tilde{\mathcal{O}}(\varepsilon^{-1})}$.

1. Construct $\tilde{N}$: $\left\|N - \tilde{N}\right\|_* \leq \varepsilon n$.
2. Calculate moments exactly: $\frac{1}{n} \cdot \mathrm{Tr}(\tilde{N}^j)$, for $j = 1, \ldots, 1/\varepsilon$.
   - Fast to compute because $\tilde{N}$ has sparse rows/columns.
3. Find distribution q that has moments close to the moments calculated.

Sparsification Lower Bound

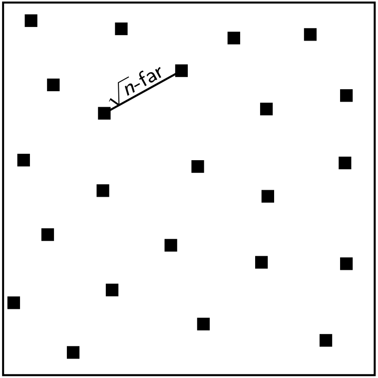
[Graph G, Normalized Adjacency Matrix N]

Our Results: Sparsity $\tilde{\Omega}(n/\varepsilon^2)$ is needed for nuclear norm sparsification.

We will sketch: for $\varepsilon = \Theta(1/\sqrt{n})$.

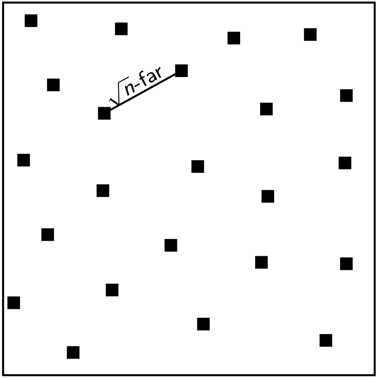
There is a graph with normalized adjacency N, such that, any $\tilde{N}$ with $\left\|N - \tilde{N}\right\|_* \leq \sqrt{n}$, must have $\mathrm{nnz}(\tilde{N}) = \Omega(n^2/\log n)$.

$\exists N : \left\| N - \tilde{N} \right\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2/\log n)$



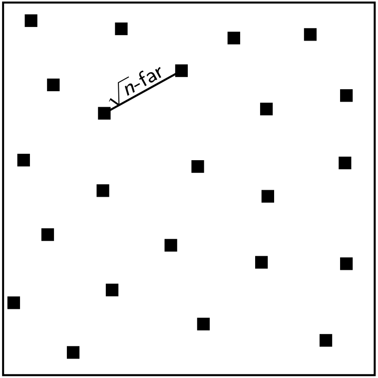
There are $\approx \exp(n^2)$ many dense graphs [▪] that are <u>all</u> $\sqrt{n}$ far in $\|\cdot\|_*$.

$\exists N : \left\| N - \tilde{N} \right\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2 / \log n)$



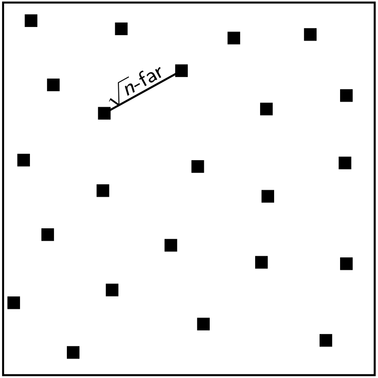
Consider the set $\mathcal{S}$ of s-sparse matrices, i.e., $\mathrm{nnz}(S) = s$.

$\exists N : \left\| N - \tilde{N} \right\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2 / \log n)$



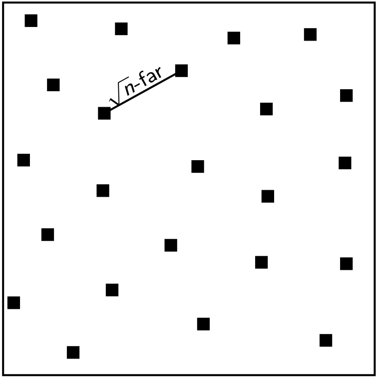
~~Consider the set $\mathcal{S}$ of s-sparse matrices, i.e., $\mathrm{nnz}(S) = s$.~~
Construct a net $\mathcal{N}_\delta[\ \circ\ ]$ of size $\exp\left(\mathcal{O}\left(s \log n\right)\right)$ that approximates $\mathcal{S}$ very well.

$\exists N : \left\|N - \tilde{N}\right\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2/\log n)$



Construct a net $\mathcal{N}_\delta[\circ]$ of size $\exp\left(\mathcal{O}\left(s \log n\right)\right)$ that approximates $\mathcal{S}$ very well.
If $s = o(n^2/\log n)$, then size of net $|\mathcal{N}_\delta[\circ]| = \exp(o(n^2))$.

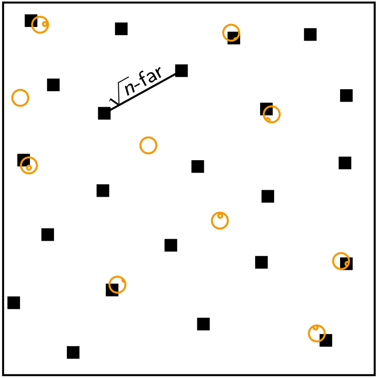
$\exists N : \left\| N - \tilde{N} \right\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2 / \log n)$



There are $\approx \exp(n^2)$ many dense graphs [■] that are <u>all</u> $\sqrt{n}$ far in $\|\cdot\|_*$.

Essentially there are only $\exp(o(n^2))$ sparse matrices [○].

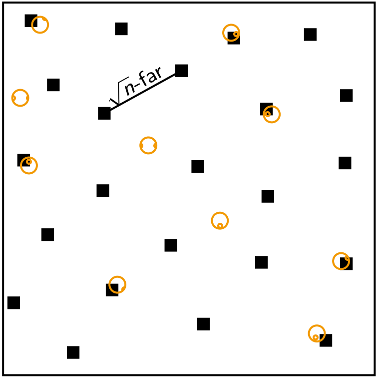
$\exists N : \|N - \tilde{N}\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2/\log n)$



There are $\approx \exp(n^2)$ many dense graphs [■] that are <u>all</u> $\sqrt{n}$ far in $\|\cdot\|_*$.

Essentially there are only $\exp(o(n^2))$ sparse matrices [○].

$\exists N : \left\| N - \tilde{N} \right\|_* \leq \sqrt{n}$ implies $\mathrm{nnz}(\tilde{N}) = \Omega(n^2/\log n)$



Can extend it to a larger range of ε.

Open Problems

- Can we further improve the running time?
  - We achieve a running time of $\mathcal{O}\left(n/\varepsilon^3\right)$.
  - [CKSV'18] achieve a running time of $2^{\mathcal{O}(1/\varepsilon)}$.
  - Something in between? Sublinear in n and polynomial in $1/\varepsilon$.
- Lower bound for spectral density estimation under adjacency list access to G and stronger access models?

Summary of Our Results

SDE: Spectral Density Estimation

1. Sublinear time SDE (randomized) in $\mathcal{O}\left(n/\varepsilon^3\right)$ time.
2. Small nuclear norm implies small W_1 distance.
3. Deterministic $\mathcal{O}\left(n/\varepsilon^2\right)$ time algorithm for nuclear norm sparsification.
4. Achieve optimal sparsity and query complexity, up to $\log n$ factors.
5. Deterministic, sublinear time SDE in $n2^{\tilde{\mathcal{O}}(1/\varepsilon)}$.