Low-degree Hardness for Detection in Correlated Erdős-Rényi Graphs

Zhangsong Li

School of Mathematical Sciences
Peking University

January 22, 2024

Based on a joint work with J. Ding (PKU) and H. Du (MIT)

Application: network de-anonymization

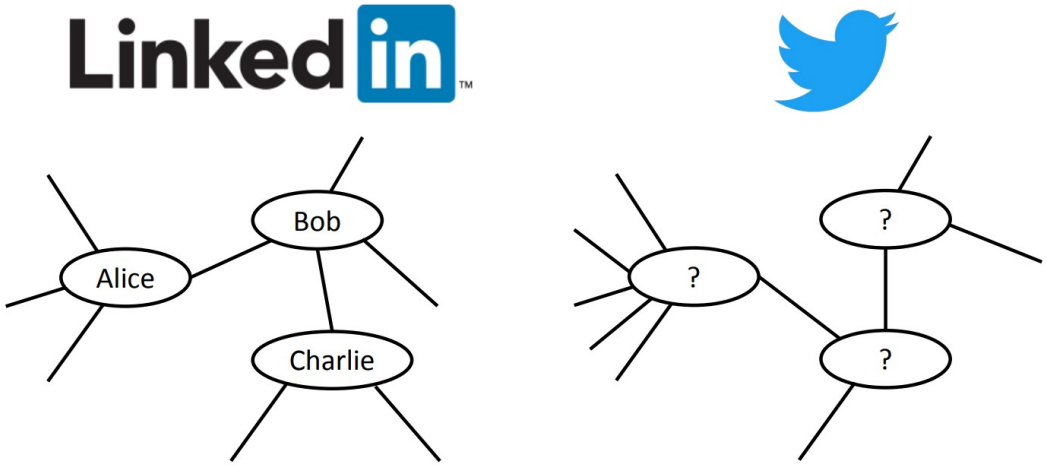


Figure 1: Picture courtesy of R.Srikant

Application: network de-anonymization

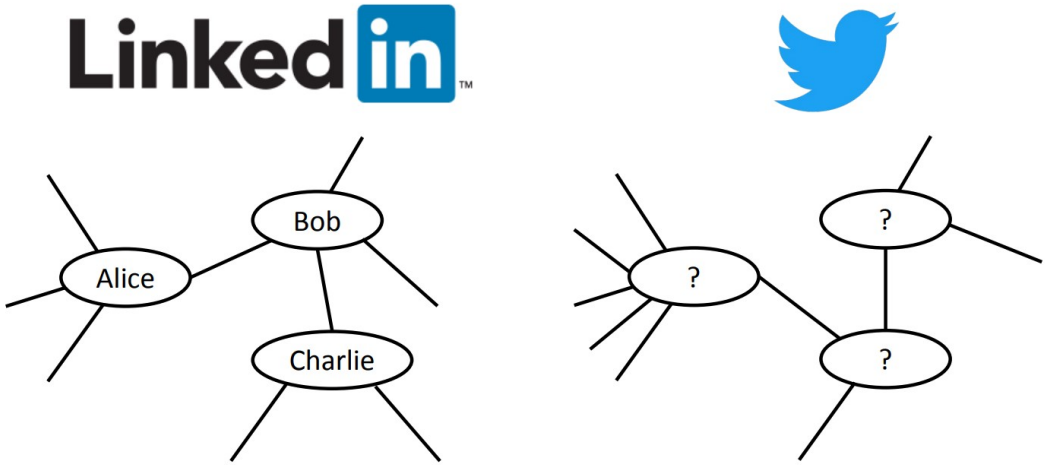


Figure 1: Picture courtesy of R.Srikant

- Successfully de-anonymize Netflix by matching it to IMDB. [Narayanan-Shmatikov '08]

Application: network de-anonymization

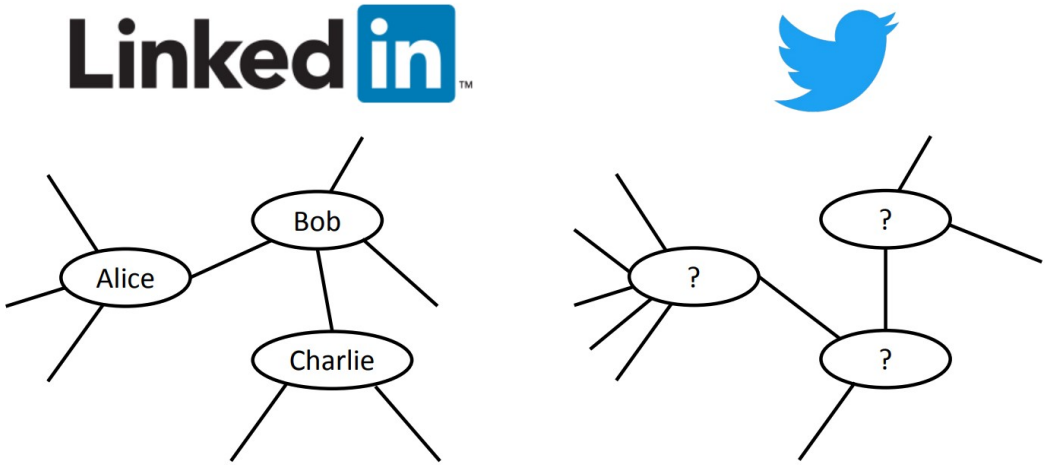
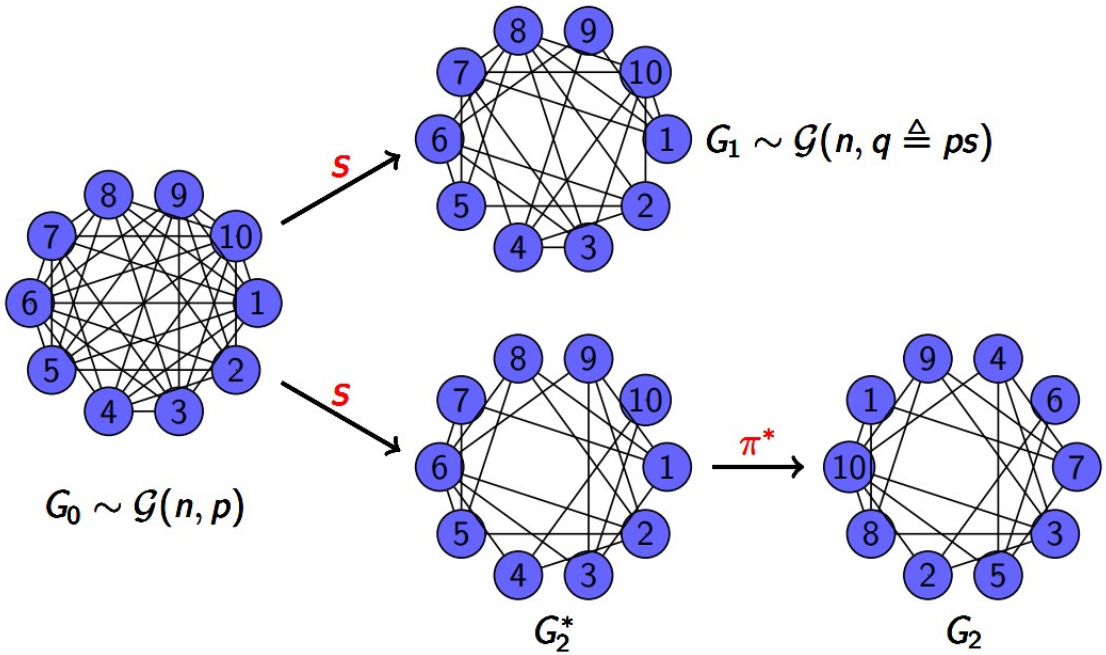


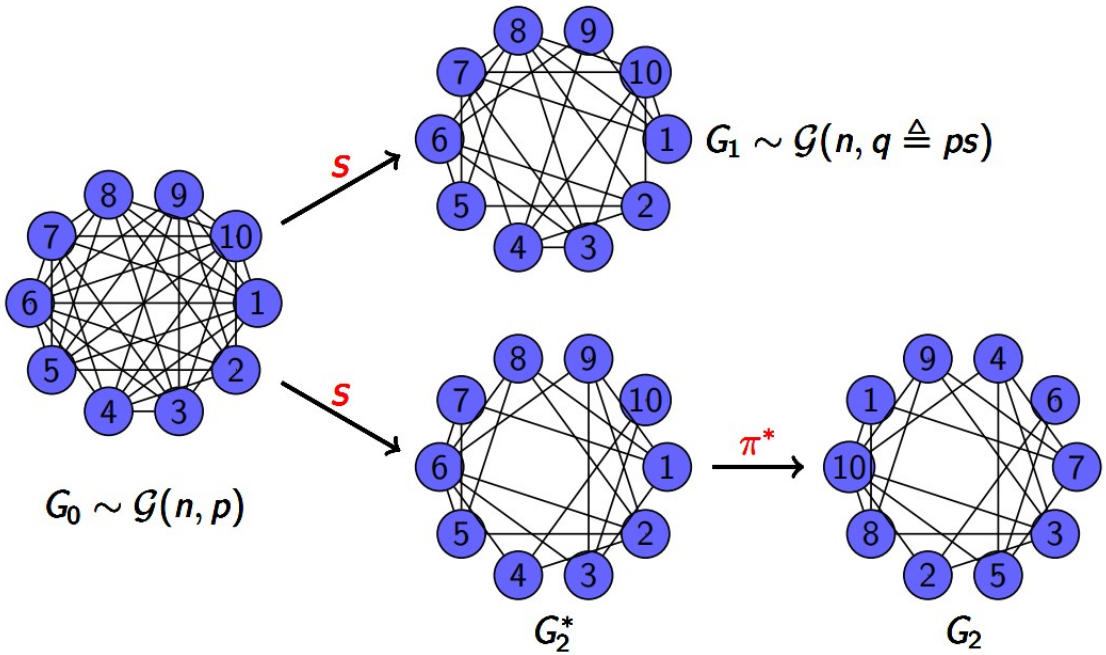
Figure 1: Picture courtesy of R.Srikant

- Successfully de-anonymize Netflix by matching it to IMDB. [Narayanan-Shmatikov '08]
- Correctly identified 30.8% of node mappings between Twitter and Flickr. [Narayanan-Shmatikov '09]

An idealized model: correlated Erdős-Rényi graphs

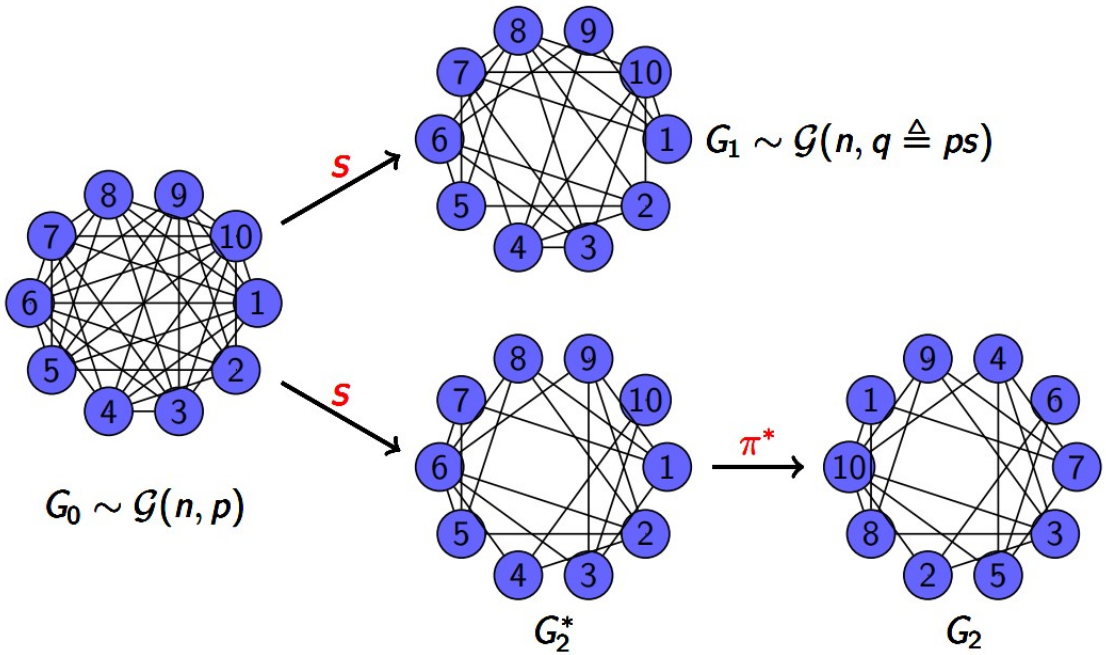


An idealized model: correlated Erdős-Rényi graphs



- G_1 and G_2 are Erdős-Rényi graphs with edge-density $q = ps$, and edge-correlation $\rho = \frac{s(1-p)}{1-ps}$.

An idealized model: correlated Erdős-Rényi graphs



- G_1 and G_2 are Erdős-Rényi graphs with edge-density $q = ps$, and edge-correlation $\rho = \frac{s(1-p)}{1-ps}$.
- Two basic problems regarding this model: (1) the detection problem, i.e., testing $\mathbb{P}$ against the law of two independent Erdős-Rényi graphs on with edge density q; (2) the matching problem, i.e., recovering the latent matching π_* from the adjacency matrices (A, B) of (G_1, G_2).

Quick overview on information threshold

Quick overview on information threshold

- [Wu-Xu-Yu '20,21]: progress based analyzing MLE (the maximum common graph).

Quick overview on information threshold

- [Wu-Xu-Yu '20,21]: progress based analyzing MLE (the maximum common graph).
- Results: determine the exact information threshold for exact recovery; Determine the information threshold for partial-recovery and detection in the dense region ($p = n^{o(1)}$) exactly and in the non-dense region ($p = n^{-c+o(1)}$ where $0 < c < 1$) up to constants.

Quick overview on information threshold

- [Wu-Xu-Yu '20,21]: progress based analyzing MLE (the maximum common graph).
- Results: determine the exact information threshold for exact recovery; Determine the information threshold for partial-recovery and detection in the dense region ($p = n^{o(1)}$) exactly and in the non-dense region ($p = n^{-c+o(1)}$ where $0 < c < 1$) up to constants.
- [Ding-Du '23a,23b]: determine the exact information threshold for detection and partial-recovery in the non-dense region via a modified statistics based on densest subgraphs.

Efficient algorithms

- Progressively improved algorithms have been obtained by commmunity (e.g. [Dai-Cullina-Kiyavash-Grossglauser '18], [Barak-Chou-Lei-Schramm-Sheng '19], [Ding-Ma-Wu-Xu '21], [Mao-Rudelson-Tikhomirov '21], etc.)

Efficient algorithms

- Progressively improved algorithms have been obtained by commmunity (e.g. [Dai-Cullina-Kiyavash-Grossglauser '18], [Barak-Chou-Lei-Schramm-Sheng '19], [Ding-Ma-Wu-Xu '21], [Mao-Rudelson-Tikhomirov '21], etc.)
- The state-of-the-art algorithm:

Efficient algorithms

- Progressively improved algorithms have been obtained by commmunity (e.g. [Dai-Cullina-Kiyavash-Grossglauser '18], [Barak-Chou-Lei-Schramm-Sheng '19], [Ding-Ma-Wu-Xu '21], [Mao-Rudelson-Tikhomirov '21], etc.)
- The state-of-the-art algorithm:
  - [Mao-Wu-Xu-Yu '21+,23]: polynomial time algorithm for detection/matching when $q > \frac{\log n}{n}$ and correlation $\rho > \sqrt{\alpha}$ where $\alpha \approx 0.338$ is the Otter's constant, based on counting carefully curated family of rooted trees.

Efficient algorithms

- Progressively improved algorithms have been obtained by commmunity (e.g. [Dai-Cullina-Kiyavash-Grossglauser '18], [Barak-Chou-Lei-Schramm-Sheng '19], [Ding-Ma-Wu-Xu '21], [Mao-Rudelson-Tikhomirov '21], etc.)
- The state-of-the-art algorithm:
  - [Mao-Wu-Xu-Yu '21+,23]: polynomial time algorithm for detection/matching when $q > \frac{\log n}{n}$ and correlation $\rho > \sqrt{\alpha}$ where $\alpha \approx 0.338$ is the Otter's constant, based on counting carefully curated family of rooted trees.
  - [Ding-L. '22+,23+]: polynomial time iterative algorithm for matching when $p \geq n^{-1+\delta}$ and correlation ρ non-vanishing.

Information-computation gap

regime	Info, Detection	Info, ExaMatch	Alg
$q = n^{o(1)}$	$\rho^2 > \frac{\log n}{nq \log q^{-1}}$	$\rho^2 > \frac{\log n}{nq \log q^{-1}}$	$\rho = \Omega(1)$
$q = n^{-1+c+o(1)}$	$\rho^2 > \frac{\lambda_*}{nq}$	$\rho^2 > \frac{\log n}{nq}$	$\rho = \Omega(1)$
$q = n^{-1+o(1)}$	$\rho^2 > \left(\frac{1}{nq} \wedge u_0\right)$	$\rho^2 > \frac{\log n}{nq}$	$\rho^2 > \alpha$

- λ_* (resp. u_0) are constants that can be determined in [DD23a] (resp. [WXY23]); α is the Otter's constant.

Information-computation gap

regime	Info, Detection	Info, ExaMatch	Alg
$q = n^{o(1)}$	$\rho^2 > \frac{\log n}{nq \log q^{-1}}$	$\rho^2 > \frac{\log n}{nq \log q^{-1}}$	$\rho = \Omega(1)$
$q = n^{-1+c+o(1)}$	$\rho^2 > \frac{\lambda_*}{nq}$	$\rho^2 > \frac{\log n}{nq}$	$\rho = \Omega(1)$
$q = n^{-1+o(1)}$	$\rho^2 > \left(\frac{1}{nq} \wedge u_0\right)$	$\rho^2 > \frac{\log n}{nq}$	$\rho^2 > \alpha$

- λ_* (resp. u_0) are constants that can be determined in [DD23a] (resp. [WXY23]); α is the Otter's constant.
- Information-computation gaps: a major challenge in many random combinatorial optimization problems.

Information-computation gap

regime	Info, Detection	Info, ExaMatch	Alg
$q = n^{o(1)}$	$\rho^2 > \frac{\log n}{nq \log q^{-1}}$	$\rho^2 > \frac{\log n}{nq \log q^{-1}}$	$\rho = \Omega(1)$
$q = n^{-1+c+o(1)}$	$\rho^2 > \frac{\lambda_*}{nq}$	$\rho^2 > \frac{\log n}{nq}$	$\rho = \Omega(1)$
$q = n^{-1+o(1)}$	$\rho^2 > \left(\frac{1}{nq} \wedge u_0\right)$	$\rho^2 > \frac{\log n}{nq}$	$\rho^2 > \alpha$

- λ_* (resp. u_0) are constants that can be determined in [DD23a] (resp. [WXY23]); α is the Otter's constant.
- Information-computation gaps: a major challenge in many random combinatorial optimization problems.
- Question: how to give lower bounds on computational complexity for problems with random input?

Low-degree polynomial method

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.
  - Based on multivariate $f : \mathbb{R}^N \to \mathbb{R}^M$ with degree $\leq D$.

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.
  - Based on multivariate $f : \mathbb{R}^N \to \mathbb{R}^M$ with degree $\leq D$.
  - Usually low-degree means $D = O(\log N)$.

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.
  - Based on multivariate $f : \mathbb{R}^N \to \mathbb{R}^M$ with degree $\leq D$.
  - Usually low-degree means $D = O(\log N)$.
  - In some cases (e.g. [Montanari-Wein' 22+]), the "optimal" algorithm is captured by a degree-$O(1)$ polynomial.

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.
  - Based on multivariate $f : \mathbb{R}^N \to \mathbb{R}^M$ with degree $\leq D$.
  - Usually low-degree means $D = O(\log N)$.
  - In some cases (e.g. [Montanari-Wein' 22+]), the "optimal" algorithm is captured by a degree-$O(1)$ polynomial.
- Some low-degree algorithms:

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.
  - Based on multivariate $f : \mathbb{R}^N \to \mathbb{R}^M$ with degree $\leq D$.
  - Usually low-degree means $D = O(\log N)$.
  - In some cases (e.g. [Montanari-Wein' 22+]), the "optimal" algorithm is captured by a degree-$O(1)$ polynomial.
- Some low-degree algorithms:
  - Spectral methods (power iteration)

Low-degree polynomial method

- The low-degree polynomial method, originated from sum-of-squares literature, provides a framework for predicting and explaining computational hardness in average-case.
- It studies a restricted class of algorithms: low-degree polynomial algorithms.
  - Based on multivariate $f : \mathbb{R}^N \to \mathbb{R}^M$ with degree $\leq D$.
  - Usually low-degree means $D = O(\log N)$.
  - In some cases (e.g. [Montanari-Wein' 22+]), the "optimal" algorithm is captured by a degree-$O(1)$ polynomial.
- Some low-degree algorithms:
  - Spectral methods (power iteration)
  - Approximate message passing (AMP)

Optimality of low-degree polynomials

Optimality of low-degree polynomials

- Low-degree polynomials seem to be optimal for many problems! E.g., for planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, etc. It is the case that

Optimality of low-degree polynomials

- Low-degree polynomials seem to be optimal for many problems! E.g., for planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);

Optimality of low-degree polynomials

- Low-degree polynomials seem to be optimal for many problems! E.g., for planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.

Optimality of low-degree polynomials

- Low-degree polynomials seem to be optimal for many problems! E.g., for planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.
- [Hopkins'18] "Low-degree conjecture" (informal): for "natural" problems, the failure of degree-D polynomials implies the failure of all algorithms with running time $e^{\widetilde{\Theta}(D)}$.

Our Results

Theorem (Ding-Du-L., informal)

There is evidence suggesting that algorithms based on polynomials of degree $O(\rho^{-1})$ fail for detection in the correlated Erdős-Rényi graph model.

Furthermore, if q, ρ satisfies $nq = n^{o(1)}$ and $\rho^2 < \alpha - \varepsilon$ for some arbitrary constant $\varepsilon > 0$ (where $\alpha \approx 0.338$ denotes the Otter's constant), then there is evidence suggesting that algorithms based on polynomials of degree D fail for detection as long as

$$\log D = o\left(\frac{\log n}{\log nq} \wedge \sqrt{\log n}\right) .$$

Our Results

Theorem (Ding-Du-L., informal)

There is evidence suggesting that algorithms based on polynomials of degree $O(\rho^{-1})$ fail for detection in the correlated Erdős-Rényi graph model.

Furthermore, if q, ρ satisfies $nq = n^{o(1)}$ and $\rho^2 < \alpha - \varepsilon$ for some arbitrary constant $\varepsilon > 0$ (where $\alpha \approx 0.338$ denotes the Otter's constant), then there is evidence suggesting that algorithms based on polynomials of degree D fail for detection as long as

$$\log D = o\left(\frac{\log n}{\log nq} \wedge \sqrt{\log n}\right) .$$

- Also suggest that the (exact) graph matching problem is computationally hard in the aforementioned regimes.

Our Results

Theorem (Ding-Du-L., informal)

There is evidence suggesting that algorithms based on polynomials of degree $O(\rho^{-1})$ fail for detection in the correlated Erdős-Rényi graph model.

Furthermore, if q, ρ satisfies $nq = n^{o(1)}$ and $\rho^2 < \alpha - \varepsilon$ for some arbitrary constant $\varepsilon > 0$ (where $\alpha \approx 0.338$ denotes the Otter's constant), then there is evidence suggesting that algorithms based on polynomials of degree D fail for detection as long as

$$\log D = o\left(\frac{\log n}{\log nq} \wedge \sqrt{\log n}\right) .$$

- Also suggest that the (exact) graph matching problem is computationally hard in the aforementioned regimes.
- In particular, suggest that the algorithms in [MWXY21+,23] [DL22+,23+] have nearly reached the limit of efficient algorithms.

Comparison between algorithms and hardness

	Algorithms	Hardness
Detection dense regime	$\rho = \Omega(1); D = O(1)$ [DL23+]	$\rho = o(1); D = O(\rho^{-1})$
Detection sparse regime	$\rho > \sqrt{\alpha}; D = O(1)$ [MWXY21+]	$\rho < \sqrt{\alpha}; \log D = o(d(n,q))$
ExaMatch dense regime	$\rho = \Omega(1); D = O(1)$ [DL23+]	$\rho = o(1); D = O(\rho^{-1})$
ExaMatch sparse regime	$\rho > \sqrt{\alpha}; D = O(\log n)$ [MWXY23]	$\rho < \sqrt{\alpha}; \log D = o(d(n,q))$

- $d(n,q) = \frac{\log n}{\log nq} \wedge \sqrt{\log n}$.
- $\alpha \approx 0.338$ is the Otter's constant.

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:
  - Null model $(A, B) \sim \mathbb{Q}$ (i.e., two independent $\mathbf{G}(n, q)$)

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:
  - Null model $(A, B) \sim \mathbb{Q}$ (i.e., two independent $\mathbf{G}(n, q)$)
  - Planted model $(A, B) \sim \mathbb{P}$ (i.e., two correlated graphs $\mathbf{G}(n, p, s)$)

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:
  - Null model $(A, B) \sim \mathbb{Q}$ (i.e., two independent $\mathbf{G}(n, q)$)
  - Planted model $(A, B) \sim \mathbb{P}$ (i.e., two correlated graphs $\mathbf{G}(n, p, s)$)
- Look for a degree-D polynomial $f : \mathbb{R}^{n*n} \otimes \mathbb{R}^{n*n} \to \mathbb{R}$ such that f is "large" under $\mathbb{P}$ and "small" under $\mathbb{Q}$.

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:
  - Null model $(A, B) \sim \mathbb{Q}$ (i.e., two independent $\mathbf{G}(n, q)$)
  - Planted model $(A, B) \sim \mathbb{P}$ (i.e., two correlated graphs $\mathbf{G}(n, p, s)$)
- Look for a degree-D polynomial $f : \mathbb{R}^{n*n} \otimes \mathbb{R}^{n*n} \to \mathbb{R}$ such that f is "large" under $\mathbb{P}$ and "small" under $\mathbb{Q}$.
- Key items: signal-to-noise ratio

$$\mathrm{SNR}_{\leq D} = \max_{\deg(f) \leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \frac{\text{mean in } \mathbb{P}}{\text{fluctuation in } \mathbb{Q}}$$

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:
  - Null model $(A, B) \sim \mathbb{Q}$ (i.e., two independent $\mathbf{G}(n, q)$)
  - Planted model $(A, B) \sim \mathbb{P}$ (i.e., two correlated graphs $\mathbf{G}(n, p, s)$)
- Look for a degree-D polynomial $f : \mathbb{R}^{n*n} \otimes \mathbb{R}^{n*n} \to \mathbb{R}$ such that f is "large" under $\mathbb{P}$ and "small" under $\mathbb{Q}$.
- Key items: signal-to-noise ratio

$$\mathrm{SNR}_{\leq D} = \max_{\deg(f) \leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \frac{\text{mean in } \mathbb{P}}{\text{fluctuation in } \mathbb{Q}}$$

  - $\mathrm{SNR}_{\leq D} \to \infty$: degree-$D$ polynomials succeed;

Low-degree framework for detection

- Goal: hypothesis test with error probability $o(1)$ between:
  - Null model $(A, B) \sim \mathbb{Q}$ (i.e., two independent $\mathbf{G}(n, q)$)
  - Planted model $(A, B) \sim \mathbb{P}$ (i.e., two correlated graphs $\mathbf{G}(n, p, s)$)
- Look for a degree-D polynomial $f : \mathbb{R}^{n*n} \otimes \mathbb{R}^{n*n} \to \mathbb{R}$ such that f is "large" under $\mathbb{P}$ and "small" under $\mathbb{Q}$.
- Key items: signal-to-noise ratio

$$\mathrm{SNR}_{\leq D} = \max_{\deg(f) \leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \frac{\text{mean in } \mathbb{P}}{\text{fluctuation in } \mathbb{Q}}$$

  - $\mathrm{SNR}_{\leq D} \to \infty$: degree-$D$ polynomials succeed;
  - $\mathrm{SNR}_{\leq D} = O(1)$: degree-$D$ polynomials fail.

Low-degree framework for detection

$$\mathrm{SNR}_{\leq D} = \max_{\deg(f)\leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$$

Low-degree framework for detection

$$\mathrm{SNR}_{\le D} = \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$$

$$= \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \langle f, g\rangle = \mathbb{E}_{\mathbb{Q}}[f\cdot g]$$

Low-degree framework for detection

$$\mathrm{SNR}_{\le D} = \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$$

$$= \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \langle f, g\rangle = \mathbb{E}_{\mathbb{Q}}[f\cdot g]$$

$$= \max_{\deg(f)\le D} \frac{\langle L, f\rangle}{\|f\|} \qquad \|f\| = \sqrt{\langle f, f\rangle}$$

Low-degree framework for detection

$$\mathrm{SNR}_{\le D} = \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$$

$$= \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \langle f, g\rangle = \mathbb{E}_{\mathbb{Q}}[f\cdot g]$$

$$= \max_{\deg(f)\le D} \frac{\langle L, f\rangle}{\|f\|} \qquad \|f\| = \sqrt{\langle f, f\rangle}$$

$$= \|L^{\le D}\|$$

Maximizer: $f = L^{\le D}$:= projection of L onto degree-D subspace.

Low-degree framework for detection

$$\mathrm{SNR}_{\le D} = \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \mathrm{d}\mathbb{P}/\mathrm{d}\mathbb{Q}$$

$$= \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \langle f, g\rangle = \mathbb{E}_{\mathbb{Q}}[f\cdot g]$$

$$= \max_{\deg(f)\le D} \frac{\langle L, f\rangle}{\|f\|} \qquad \|f\| = \sqrt{\langle f, f\rangle}$$

$$= \|L^{\le D}\|$$

Maximizer: $f = L^{\le D} :=$ projection of L onto degree-D subspace.

- Since $\mathbb{Q}$ has independent entries, we can directly calculate the projection using orthogonal polynomials w.r.t. $\mathbb{Q}$.

Warm up: the proof in dense regime

- For two graphs S_1, S_2 with no isolated vertices, define the polynomial ϕ_{S_1,S_2} by

$$\phi_{S_1,S_2}\left(\{A_{i,j}\},\{B_{i,j}\}\right) = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j},$$

where $\overline{A}_{i,j} = \frac{A_{i,j}-q}{\sqrt{q(1-q)}}, \overline{B}_{i,j} = \frac{B_{i,j}-q}{\sqrt{q(1-q)}}$.

Warm up: the proof in dense regime

- For two graphs S_1, S_2 with no isolated vertices, define the polynomial ϕ_{S_1,S_2} by

$$\phi_{S_1,S_2}\left(\{A_{i,j}\},\{B_{i,j}\}\right) = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j},$$

  where $\overline{A}_{i,j} = \frac{A_{i,j}-q}{\sqrt{q(1-q)}}, \overline{B}_{i,j} = \frac{B_{i,j}-q}{\sqrt{q(1-q)}}$.

- $\{\phi_{S_1,S_2} : |E(S_1)| + |E(S_2)| \leq D\}$ constitutes a standard orthogonal basis.

Warm up: the proof in dense regime

- For two graphs S_1, S_2 with no isolated vertices, define the polynomial ϕ_{S_1,S_2} by

$$\phi_{S_1,S_2}\left(\{A_{i,j}\},\{B_{i,j}\}\right) = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j},$$

where $\overline{A}_{i,j} = \frac{A_{i,j}-q}{\sqrt{q(1-q)}}, \overline{B}_{i,j} = \frac{B_{i,j}-q}{\sqrt{q(1-q)}}$.

- $\{\phi_{S_1,S_2} : |E(S_1)| + |E(S_2)| \le D\}$ constitutes a standard orthogonal basis.
- It can be directly calculated that

$$\left\|L_{\le D}\right\|^2 = \sum_{|E(S_1)|+|E(S_2)|\le D} \left(\mathbb{E}_{\mathbb{P}}[\phi_{S_1,S_2}]\right)^2$$

and

$$\mathbb{E}_{\mathbb{P}}[\phi_{S_1,S_2}] = \begin{cases} \rho^{|E(S_1)|} \cdot \frac{\mathrm{Aut}(S_1)}{n(n-1)\cdots(n-|V(S_1)|+1)}, & \text{if } S_1 \cong S_2, \\ 0, & \text{otherwise.} \end{cases}$$

Warm up: the proof in dense regime

The problem reduces to controlling

$$\sum_{S_1 \cong S_2, |E(S_1)|+|E(S_2)| \leq D} \rho^{2|E(S_1)|} \cdot \frac{\mathrm{Aut}(S_1)^2}{[n(n-1)\cdots(n-|V(S_1)|+1)]^2}$$

Warm up: the proof in dense regime

The problem reduces to controlling

$$\sum_{S_1 \cong S_2, |E(S_1)|+|E(S_2)| \le D} \rho^{2|E(S_1)|} \cdot \frac{\mathrm{Aut}(S_1)^2}{[n(n-1)\cdots(n-|V(S_1)|+1)]^2}$$

$$= \sum_{|E(\mathbf{H})| \le D/2} \rho^{2|E(\mathbf{H})|} \cdot \frac{\mathrm{Aut}(\mathbf{H})^2 \cdot \#\{S_1 \cong S_2 \cong \mathbf{H}\}}{[n(n-1)\cdots(n-|V(\mathbf{H})|+1)]^2}$$

Warm up: the proof in dense regime

The problem reduces to controlling

$$\sum_{S_1 \cong S_2, |E(S_1)|+|E(S_2)| \le D} \rho^{2|E(S_1)|} \cdot \frac{\mathrm{Aut}(S_1)^2}{[n(n-1)\cdots(n-|V(S_1)|+1)]^2}$$

$$= \sum_{|E(\mathbf{H})| \le D/2} \rho^{2|E(\mathbf{H})|} \cdot \frac{\mathrm{Aut}(\mathbf{H})^2 \cdot \#\{S_1 \cong S_2 \cong \mathbf{H}\}}{[n(n-1)\cdots(n-|V(\mathbf{H})|+1)]^2}$$

$$= \sum_{|E(\mathbf{H})| \le D/2} \rho^{2|E(\mathbf{H})|}$$

Warm up: the proof in dense regime

The problem reduces to controlling

$$\sum_{S_1 \cong S_2, |E(S_1)|+|E(S_2)| \le D} \rho^{2|E(S_1)|} \cdot \frac{\mathrm{Aut}(S_1)^2}{[n(n-1)\cdots(n-|V(S_1)|+1)]^2}$$

$$= \sum_{|E(\mathbf{H})| \le D/2} \rho^{2|E(\mathbf{H})|} \cdot \frac{\mathrm{Aut}(\mathbf{H})^2 \cdot \#\{S_1 \cong S_2 \cong \mathbf{H}\}}{[n(n-1)\cdots(n-|V(\mathbf{H})|+1)]^2}$$

$$= \sum_{|E(\mathbf{H})| \le D/2} \rho^{2|E(\mathbf{H})|}$$

- It is $O(1)$ provided that $D = O(\rho^{-1})$, using the fact that

$$\#\{|E(\mathbf{H})| = k\} \le \sum_{\ell=1}^{2k} \#\{|E(\mathbf{H})| = k, |V(\mathbf{H})| = \ell\} \le \sum_{\ell=1}^{2k} \binom{\ell(\ell-1)/2}{k}$$

The proof in sparse regime

- Same proof as dense regime?

The proof in sparse regime

- Same proof as dense regime?
- Unfortunately, we still have

$$\mathrm{SNR}_{\leq D} = \|L_{\leq D}\|^2 = \sum_{|E(\mathbf{H})| \leq D/2} \rho^{2|E(\mathbf{H})|}$$

which blows-up for general $\rho < \sqrt{\alpha}$.

The proof in sparse regime

- Same proof as dense regime?
- Unfortunately, we still have

$$\mathrm{SNR}_{\leq D} = \|L_{\leq D}\|^2 = \sum_{|E(\mathbf{H})| \leq D/2} \rho^{2|E(\mathbf{H})|}$$

  which blows-up for general $\rho < \sqrt{\alpha}$.
- Observation: the main contribution of $\|L_{\leq D}\|$ comes from counting "very dense" subgraphs;

The proof in sparse regime

- Same proof as dense regime?
- Unfortunately, we still have

$$\mathrm{SNR}_{\leq D} = \|L_{\leq D}\|^2 = \sum_{|E(\mathbf{H})| \leq D/2} \rho^{2|E(\mathbf{H})|}$$

  which blows-up for general $\rho < \sqrt{\alpha}$.
- Observation: the main contribution of $\|L_{\leq D}\|$ comes from counting "very dense" subgraphs;
- However, since $q = n^{-1+o(1)}$, with high probability all subgraphs (with at most D edges) of (G_1, G_2) should have "low" edge-density.

The proof in sparse regime

- Same proof as dense regime?
- Unfortunately, we still have

$$\mathrm{SNR}_{\leq D} = \|L_{\leq D}\|^2 = \sum_{|E(\mathbf{H})| \leq D/2} \rho^{2|E(\mathbf{H})|}$$

  which blows-up for general $\rho < \sqrt{\alpha}$.
- Observation: the main contribution of $\|L_{\leq D}\|$ comes from counting "very dense" subgraphs;
- However, since $q = n^{-1+o(1)}$, with high probability all subgraphs (with at most D edges) of (G_1, G_2) should have "low" edge-density.
- This inspired us to work with a truncated version of $\mathbb{P}$ rather than $\mathbb{P}$ itself.

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0
- Given a graph $H = H(V, E)$, define

$$\Phi(H) = n^{|V(H)|} q^{|E(H)|} \geq \mathbb{E}[\text{number of } H \text{ in } \mathcal{G}(n, q)]$$

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0
- Given a graph $H = H(V, E)$, define

$$\Phi(H) = n^{|V(H)|} q^{|E(H)|} \geq \mathbb{E}[\text{number of } H \text{ in } \mathcal{G}(n, q)]$$

- H is said to be bad if $\Phi(H) < (\log n)^{-1}$. Furthermore, we say a graph is admissible if it contains no bad subgraph, and we say it is inadimissible otherwise. (H admissible $\Longrightarrow |E(H)| \leq (1 + o(1))|V(H)|$)

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0
- Given a graph $H = H(V, E)$, define

$$\Phi(H) = n^{|V(H)|} q^{|E(H)|} \geq \mathbb{E}[\text{number of } H \text{ in } \mathcal{G}(n, q)]$$

- H is said to be bad if $\Phi(H) < (\log n)^{-1}$. Furthermore, we say a graph is admissible if it contains no bad subgraph, and we say it is inadimissible otherwise. (H admissible $\Longrightarrow |E(H)| \leq (1 + o(1))|V(H)|$)
- Denote $\mathcal{E}$ for the event that G_0 does not contain any bad subgraph with no more than D^2 vertices and $\mathbb{P}' = \mathbb{P}(\cdot \mid \mathcal{E})$.

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0
- Given a graph $H = H(V, E)$, define

$$\Phi(H) = n^{|V(H)|} q^{|E(H)|} \geq \mathbb{E}[\text{number of } H \text{ in } \mathcal{G}(n, q)]$$

- H is said to be bad if $\Phi(H) < (\log n)^{-1}$. Furthermore, we say a graph is admissible if it contains no bad subgraph, and we say it is inadimissible otherwise. (H admissible $\Longrightarrow |E(H)| \leq (1 + o(1))|V(H)|$)
- Denote $\mathcal{E}$ for the event that G_0 does not contain any bad subgraph with no more than D^2 vertices and $\mathbb{P}' = \mathbb{P}(\cdot \mid \mathcal{E})$.
- Our goal is to prove the following:

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0
- Given a graph $H = H(V, E)$, define

$$\Phi(H) = n^{|V(H)|} q^{|E(H)|} \geq \mathbb{E}[\text{number of } H \text{ in } \mathcal{G}(n, q)]$$

- H is said to be bad if $\Phi(H) < (\log n)^{-1}$. Furthermore, we say a graph is admissible if it contains no bad subgraph, and we say it is inadimissible otherwise. (H admissible $\Longrightarrow |E(H)| \leq (1 + o(1))|V(H)|$)
- Denote $\mathcal{E}$ for the event that G_0 does not contain any bad subgraph with no more than D^2 vertices and $\mathbb{P}' = \mathbb{P}(\cdot \mid \mathcal{E})$.
- Our goal is to prove the following:
  - $\mathrm{TV}(\mathbb{P}, \mathbb{P}') = o(1)$ (first moment method).

The proof in sparse regime: truncation

- Recall that (G_1, G_2) are subsampled from the parent graph G_0
- Given a graph $H = H(V, E)$, define

$$\Phi(H) = n^{|V(H)|} q^{|E(H)|} \geq \mathbb{E}[\text{number of } H \text{ in } \mathcal{G}(n, q)]$$

- H is said to be bad if $\Phi(H) < (\log n)^{-1}$. Furthermore, we say a graph is admissible if it contains no bad subgraph, and we say it is inadimissible otherwise. (H admissible $\Longrightarrow |E(H)| \leq (1 + o(1))|V(H)|$)
- Denote $\mathcal{E}$ for the event that G_0 does not contain any bad subgraph with no more than D^2 vertices and $\mathbb{P}' = \mathbb{P}(\cdot \mid \mathcal{E})$.
- Our goal is to prove the following:
  - $\mathrm{TV}(\mathbb{P}, \mathbb{P}') = o(1)$ (first moment method).
  - $\mathrm{SNR}'_{\leq D} = \max_{\deg(f) \leq D} \frac{\mathbb{E}_{\mathbb{P}'}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} = O(1)$.

The proof in sparse regime (sketch)

- Define $L' = \mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{Q}$, we still have

$$(\mathrm{SNR}'_{\leq D})^2 = \|L'_{\leq D}\|^2 = \sum_{|E(S_1)|+|E(S_2)|\leq D} \left(\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]\right)^2$$

where $\phi_{S_1,S_2} = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j}$.

The proof in sparse regime (sketch)

- Define $L' = \mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{Q}$, we still have

$$(\mathrm{SNR}'_{\leq D})^2 = \|L'_{\leq D}\|^2 = \sum_{|E(S_1)|+|E(S_2)|\leq D} \left(\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]\right)^2$$

  where $\phi_{S_1,S_2} = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j}$.

- Our intuition to bound $\mathrm{SNR}'_{\leq D}$:

The proof in sparse regime (sketch)

- Define $L' = \mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{Q}$, we still have

$$(\mathrm{SNR}'_{\leq D})^2 = \|L'_{\leq D}\|^2 = \sum_{|E(S_1)|+|E(S_2)|\leq D} \left(\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]\right)^2$$

  where $\phi_{S_1,S_2} = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j}$.

- Our intuition to bound $\mathrm{SNR}'_{\leq D}$:
  1. For S_1 or S_2 inadmissible, since (G_1, G_2) do not contain inadmissible subgraph under $\mathbb{P}'$, we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]$ is negligible.

The proof in sparse regime (sketch)

- Define $L' = \mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{Q}$, we still have

$$(\mathrm{SNR}'_{\leq D})^2 = \|L'_{\leq D}\|^2 = \sum_{|E(S_1)|+|E(S_2)|\leq D} \left(\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]\right)^2$$

  where $\phi_{S_1,S_2} = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j}$.

- Our intuition to bound $\mathrm{SNR}'_{\leq D}$:
  1. For S_1 or S_2 inadmissible, since (G_1, G_2) do not contain inadmissible subgraph under $\mathbb{P}'$, we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]$ is negligible.
  2. For S_1, S_2 admissible, since we condition on a typical event we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}] \approx \mathbb{E}_{\mathbb{P}}[\phi_{S_1,S_2}]$.

The proof in sparse regime (sketch)

- Define $L' = \mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{Q}$, we still have

$$(\mathrm{SNR}'_{\le D})^2 = \|L'_{\le D}\|^2 = \sum_{|E(S_1)|+|E(S_2)|\le D} \left(\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]\right)^2$$

  where $\phi_{S_1,S_2} = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j}$.

- Our intuition to bound $\mathrm{SNR}'_{\le D}$:
  1. For S_1 or S_2 inadmissible, since (G_1, G_2) do not contain inadmissible subgraph under $\mathbb{P}'$, we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]$ is negligible.
  2. For S_1, S_2 admissible, since we condition on a typical event we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}] \approx \mathbb{E}_{\mathbb{P}}[\phi_{S_1,S_2}]$.
  3. Under previous two assumption, direct calculation yields

$$(\mathrm{SNR}'_{\le D})^2 \approx \sum_{E(\mathbf{H})\le D/2, \mathbf{H} \text{ admissible}} \rho^{2|E(\mathbf{H})|}$$

The proof in sparse regime (sketch)

- Define $L' = \mathrm{d}\mathbb{P}'/\mathrm{d}\mathbb{Q}$, we still have

$$(\mathrm{SNR}'_{\le D})^2 = \|L'_{\le D}\|^2 = \sum_{|E(S_1)|+|E(S_2)|\le D} \left(\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]\right)^2$$

  where $\phi_{S_1,S_2} = \prod_{(i,j)\in E(S_1)} \overline{A}_{i,j} \prod_{(i,j)\in E(S_2)} \overline{B}_{i,j}$.

- Our intuition to bound $\mathrm{SNR}'_{\le D}$:
  1. For S_1 or S_2 inadmissible, since (G_1, G_2) do not contain inadmissible subgraph under $\mathbb{P}'$, we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}]$ is negligible.
  2. For S_1, S_2 admissible, since we condition on a typical event we expect that $\mathbb{E}_{\mathbb{P}'}[\phi_{S_1,S_2}] \approx \mathbb{E}_{\mathbb{P}}[\phi_{S_1,S_2}]$.
  3. Under previous two assumption, direct calculation yields

$$(\mathrm{SNR}'_{\le D})^2 \approx \sum_{E(\mathbf{H})\le D/2, \mathbf{H} \text{ admissible}} \rho^{2|E(\mathbf{H})|}$$

  4. We can show the number of admissible graphs with k edges is approximately α^{-k} where α is the Otter's constant, thus $(*) = O(1)$ when $\rho < \sqrt{\alpha}$.

Future questions

- What if we consider partial matching (i.e., recover a positive fraction of π_*)?

Future questions

- What if we consider partial matching (i.e., recover a positive fraction of π_*)?
- It is "easier" than exact matching and intuitively "harder" than detection, so we expect the same result holds for partial matching.

Future questions

- What if we consider partial matching (i.e., recover a positive fraction of π_*)?
- It is "easier" than exact matching and intuitively "harder" than detection, so we expect the same result holds for partial matching.
- The partial recovery algorithm is established in [Ganassali-Massoulié-Lelarge '20+,22+]

Future questions

- What if we consider partial matching (i.e., recover a positive fraction of π_*)?
- It is "easier" than exact matching and intuitively "harder" than detection, so we expect the same result holds for partial matching.
- The partial recovery algorithm is established in [Ganassali-Massoulié-Lelarge '20+,22+]
- To prove the low-degree hardness for recovery problems, [Wein-Schramm '20] proposed the following framework: try to show that

$$\max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}|\pi_*(1)=1}[f]}{\sqrt{\mathbb{E}_{\mathbb{P}}[f^2]}} = O(1)$$

Future questions

- What if we consider partial matching (i.e., recover a positive fraction of π_*)?
- It is "easier" than exact matching and intuitively "harder" than detection, so we expect the same result holds for partial matching.
- The partial recovery algorithm is established in [Ganassali-Massoulié-Lelarge '20+,22+]
- To prove the low-degree hardness for recovery problems, [Wein-Schramm '20] proposed the following framework: try to show that

$$\max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}|\pi_*(1)=1}[f]}{\sqrt{\mathbb{E}_{\mathbb{P}}[f^2]}} = O(1)$$

- We can control $\mathbb{E}_{\mathbb{P}|\pi_*(1)=1}[f]$ in the similar manner but we don't know how to control $\mathbb{E}_{\mathbb{P}}[f^2]$.

References

1. [DD23a] J. Ding and H. Du. Detection threshold for correlated Erdős-Rényi graphs via densest subgraph. In *IEEE Transactions on Information Theory*, 69(8):5289–5298, 2023.
2. [DD23b] J. Ding and H. Du. Matching recovery threshold for correlated random graphs. In *Annals of Statistics*, 51(4):1718–1743, 2023.
3. [DDL23+] J. Ding, H. Du and Z. Li. Low-degree hardness for detection in correlated Erdős-Rényi graphs. Preprint, arXiv: 2311.15931.
4. [DL22+] J. Ding and Z. Li. A polynomial time iterative algorithm for matching Gaussian matrices with non-vanishing correlation. Preprint, arXiv:2212.13677.
5. [DL23+] J. Ding and Z. Li. A polynomial-time iterative algorithm for random graph matching with non-vanishing correlation. Preprint, arXiv:2306.00266.
6. [Hopkins18] S. Hopkins. Statistical Inference and the Sum of Squares Method. PhD thesis, Cornell University, 2018.
7. [MWXY21+] C. Mao, Y. Wu, J. Xu, and S. H. Yu. Testing network correlation efficiently via counting trees. to appear in *Annals of Statistics*.
8. [MWXY23] C. Mao, Y. Wu, J. Xu, and S. H. Yu. Random graph matching at Otter's threshold via counting chandeliers. In *STOC 2023*.
9. [WXY22] Y. Wu, J. Xu and S. H. Yu, Settling the sharp reconstruction thresholds of random graph matching. In *IEEE Transactions on Information Theory*, 68(8):5391-5417, 2022.
10. [WXY23] Y. Wu, J. Xu and S. H. Yu, Testing correlation of unlabeled random graphs. In *Annals of Applied Probability*, 33(4): 2519-2558, 2023.

Thank you!