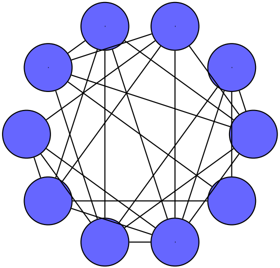
S&DS 684 Lecture 13: Random Graph Matching: spectral methods and convex relaxation

Yihong Wu

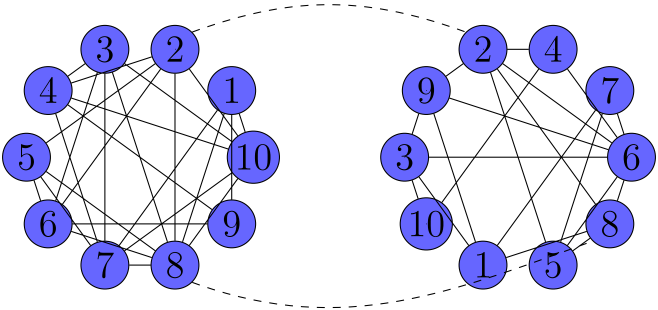
Department of Statistics and Data Science
Yale University

Apr 25, 2023

Graph matching (network alignment)

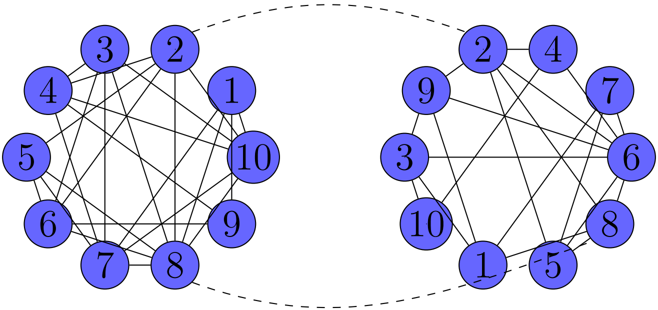


Graph matching (network alignment)



Goal: find a mapping between two node sets that maximally aligns the edges (i.e. minimizes # of adjacency disagreements)

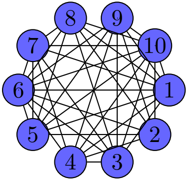
Graph matching (network alignment)



Goal: find a mapping between two node sets that maximally aligns the edges (i.e. minimizes # of adjacency disagreements)

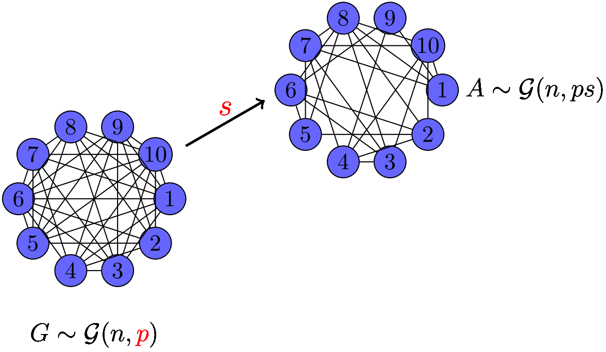
$$\text{Quadratic Assignment Problem (QAP)}: \quad \max_{\Pi \in S_n} \langle A, \Pi B \Pi^\top \rangle$$

Correlated Erdős-Rényi graphs model $\mathcal{G}(n, p, s)$

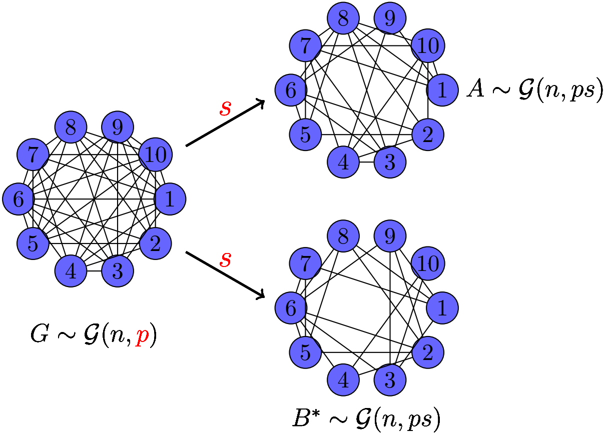


$G \sim \mathcal{G}(n, p)$

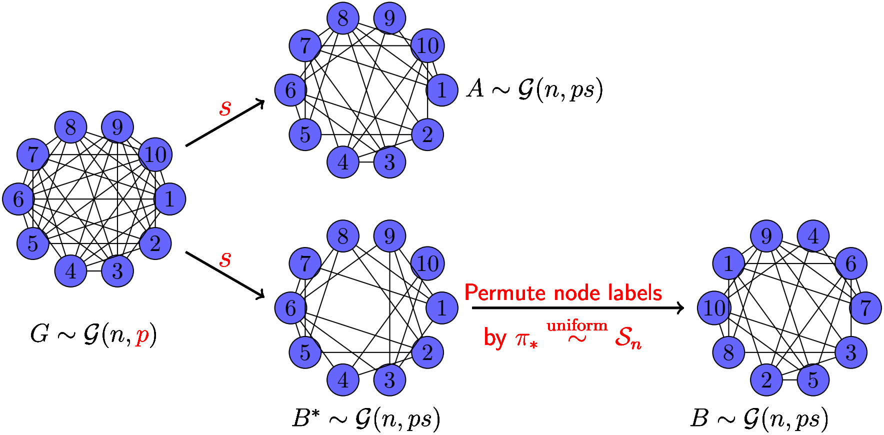
Correlated Erdős-Rényi graphs model $\mathcal{G}(n, p, s)$



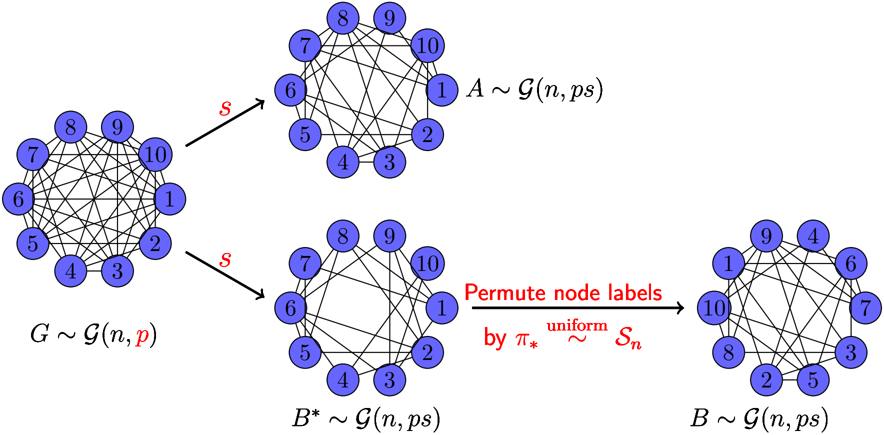
Correlated Erdős-Rényi graphs model $\mathcal{G}(n, p, s)$



Correlated Erdős-Rényi graphs model $\mathcal{G}(n, p, s)$



Correlated Erdős-Rényi graphs model $\mathcal{G}(n, p, s)$



- $(A_{\pi_*(i)\pi_*(j)}, B_{ij})$ are iid pairs of correlated $\mathrm{Bern}(ps)$
- A and B differ by a fraction $\delta \triangleq 1 - s$ of edges, under the latent π_*
- Key parameter nps^2: average degree of intersection graph $A \wedge B^*$;

Information-theoretic limits

p: edge probability $\delta = 1 - s$: fraction of errors (differed edges)

Theorem (Cullina-Kiyavash '18, Wu-Xu-Yu' 21)

For $p = o(1)$, exact recovery of π^* is information-theoretically possible if and only if

$$nps^2 - \log n \to +\infty$$

Interpretation: Intersection graph $A \wedge B^* \sim \mathcal{G}(n, ps^2)$ is connected

Information-theoretic limits

p: edge probability $\delta = 1 - s$: fraction of errors (differed edges)

Theorem (Cullina-Kiyavash '18, Wu-Xu-Yu' 21)

For $p = o(1)$, exact recovery of π^* is information-theoretically possible if and only if

$$nps^2 - \log n \to +\infty$$

Interpretation: Intersection graph $A \wedge B^* \sim \mathcal{G}(n, ps^2)$ is connected

Computationally:

- Noiseless $s = 1 (\delta = 0)$: Random graph isomorphism. Optimal condition is attained in linear-time [Bollobás '82, Czajka-Pandurangan '08]
- Noisy case $s < 1 (\delta > 0)$: little is known for efficient algorithms until recently

Main result

p: edge probability $\quad\quad \delta = 1 - s$: fraction of errors (differed edges)

Theorem (Fan-Mao-W-Xu '19)

Exact recovery is achieved efficiently by **a spectral method** *whp if*

$$np \gtrsim (\log n)^C \quad \textit{and} \quad \delta \lesssim (\log n)^{-C}$$

for some absolute constant C.

Main result

p: edge probability $\qquad \delta = 1 - s$: fraction of errors (differed edges)

Theorem (Fan-Mao-W-Xu '19)

Exact recovery is achieved efficiently by **a spectral method** *whp if*

$$np \gtrsim (\log n)^C \quad \textit{and} \quad \delta \lesssim (\log n)^{-C}$$

for some absolute constant C*.*

- Classical spectral methods require $\delta \leq n^{-C}$ [Ganassali-Lelarge-Massoulié '19]

Main result

p: edge probability $\delta = 1 - s$: fraction of errors (differed edges)

Theorem (Fan-Mao-W-Xu '19)

Exact recovery is achieved efficiently by **a spectral method** *whp if*

$$np \gtrsim (\log n)^C \quad \textit{and} \quad \delta \lesssim (\log n)^{-C}$$

for some absolute constant C.

- Classical spectral methods require $\delta \leq n^{-C}$ [Ganassali-Lelarge-Massoulié '19]
- Holds for general correlated Wigner model, e.g., Gaussian model

$$(A_{\pi_*(i)\pi_*(j)}, B_{ij}) \overset{\text{iid}}{\sim} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Outline

1. A new spectral algorithm
2. Analysis
3. Concluding remarks

Spectral methods and applications

Estimate hidden structure using leading eigenvectors of data matrix A

- Planted clique [Alon-Krivelevich-Sudakov '98]
- Planted partition/Stochastic block model [Mcsherry '98] [Massoulié '13] [Bordenave-Lelarge-Massoulié '15]
- Clustering [von-Luxburg-Bousquet-Belkin '05]
- Graphon estimation [Chatterjee '15]
- Matrix completion [Keshavan-Montanari-Oh '09]
- Ranking [Negahban-Oh-Shah '17]

Spectral methods and applications

Estimate hidden structure using leading eigenvectors of data matrix A

- Planted clique [Alon-Krivelevich-Sudakov '98]
- Planted partition/Stochastic block model [Mcsherry '98] [Massoulié '13] [Bordenave-Lelarge-Massoulié '15]
- Clustering [von-Luxburg-Bousquet-Belkin '05]
- Graphon estimation [Chatterjee '15]
- Matrix completion [Keshavan-Montanari-Oh '09]
- Ranking [Negahban-Oh-Shah '17]

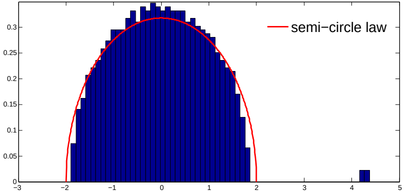
Underlying structure: A is approximately low-rank with large eigen-gap

Analysis of classical spectral method

<u>Example</u>: SBM with 3 communities:

$$A = \begin{bmatrix} p & q & q \\ q & p & q \\ q & q & p \end{bmatrix} + A - \mathbb{E}[A]$$

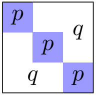
“signal” “noise”



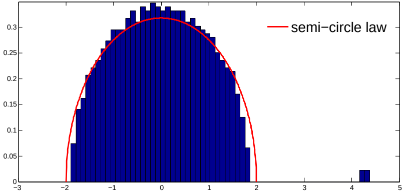
- Perturbation bound (Davis-Kahan and variants – Chap 3): Top eigenvectors of $A \approx$ those of $\mathbb{E}[A]$, if eigengap $\gtrsim \|A - \mathbb{E}[A]\|_2$

Analysis of classical spectral method

Example: SBM with 3 communities:

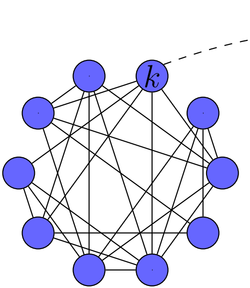
$A \quad =$  $+ \; A - \mathbb{E}[A]$

"signal" "noise"



- Perturbation bound (Davis-Kahan and variants – Chap 3): Top eigenvectors of $A \approx$ those of $\mathbb{E}\left[A\right]$, if eigengap $\gtrsim \|A - \mathbb{E}\left[A\right]\|_2$
- However, in graph matching: signal $=$ adjacency matrix of Erdős-Rényi graph, which has full rank and vanishing eigengap
- Need to rethink spectral methods for graph matching

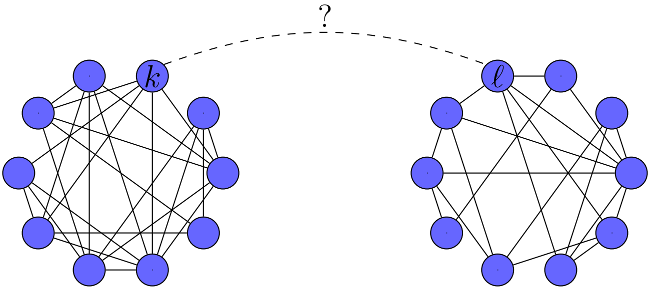
Spectral graph matching paradigm



$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^\top \qquad B = \sum_{j=1}^{n} \mu_j v_j v_j^\top$$

$$\lambda_1 \geq \cdots \geq \lambda_n \qquad \mu_1 \geq \cdots \geq \mu_n$$

Spectral graph matching paradigm



$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^\top \qquad B = \sum_{j=1}^{n} \mu_j v_j v_j^\top$$

$$\lambda_1 \geq \cdots \geq \lambda_n \qquad \mu_1 \geq \cdots \geq \mu_n$$

1. Construct a similarity matrix X based on (λ_i, u_i) and (μ_j, v_j)
2. Project X to permutation by linear assignment: $\widehat{\Pi} \in \arg\max \langle X, \Pi \rangle$

Failure of previous spectral methods

- Low-rank methods: Aligning the leading eigenvectors

$$X = s_1 u_1 v_1^\top, \qquad s_1 \in \{\pm 1\}$$

Failure of previous spectral methods

- Low-rank methods: Aligning the leading eigenvectors

$$X = s_1 u_1 v_1^\top, \qquad s_1 \in \{\pm 1\}$$

Similar ideas used in IsoRank [Singh-Xu-Berger '08] and EigenAlign [Feizi-Quon-Mendoza-Medard-Kellis-Jadbabaie '19]

Failure of previous spectral methods

- Low-rank methods: Aligning the leading eigenvectors

$$X = s_1 u_1 v_1^\top, \qquad s_1 \in \{\pm 1\}$$

  Similar ideas used in IsoRank [Singh-Xu-Berger '08] and EigenAlign [Feizi-Quon-Mendoza-Medard-Kellis-Jadbabaie '19]

- Full-rank methods: [Umeyama '88]

$$X = \sum_{i=1}^{n} s_i u_i v_i^\top, \qquad s_i \in \{\pm 1\}$$

Failure of previous spectral methods

- Low-rank methods: Aligning the leading eigenvectors

$$X = s_1 u_1 v_1^\top, \qquad s_1 \in \{\pm 1\}$$

Similar ideas used in IsoRank [Singh-Xu-Berger '08] and EigenAlign [Feizi-Quon-Mendoza-Medard-Kellis-Jadbabaie '19]

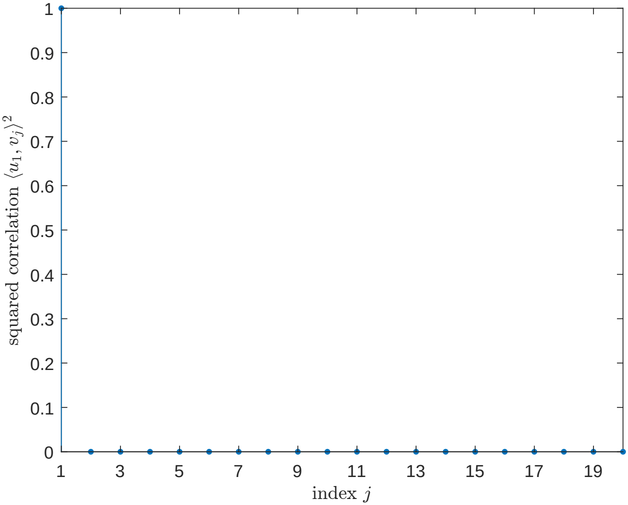
- Full-rank methods: [Umeyama '88]

$$X = \sum_{i=1}^{n} s_i u_i v_i^\top, \qquad s_i \in \{\pm 1\}$$

- All perform well with no noise, but are extremely fragile with noise
- A and B have full rank and vanishing eigengaps $\Rightarrow$ decorrelation of u_i and v_i when $\delta = n^{-c}$ [Chatterjee '14, Bourgade-Yau '17, Benigni '17, Ganassali-Lelarge-Massoulié '19]

Eigenvector correlation

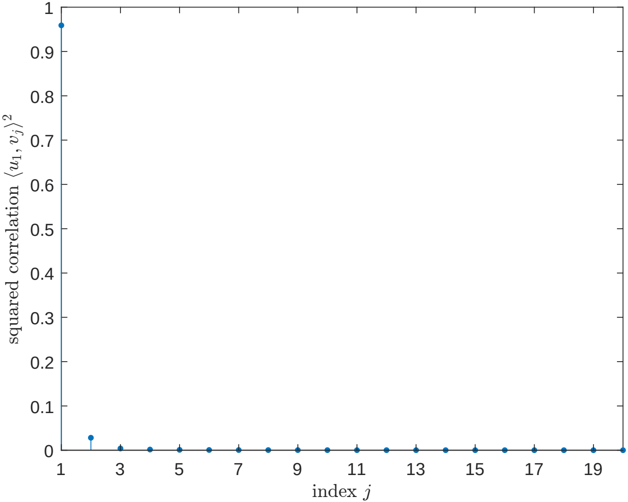
Isomorphic Erdős-Rényi graphs: 500 vertices, edge probability $\frac{1}{2}$



$\langle u_1, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

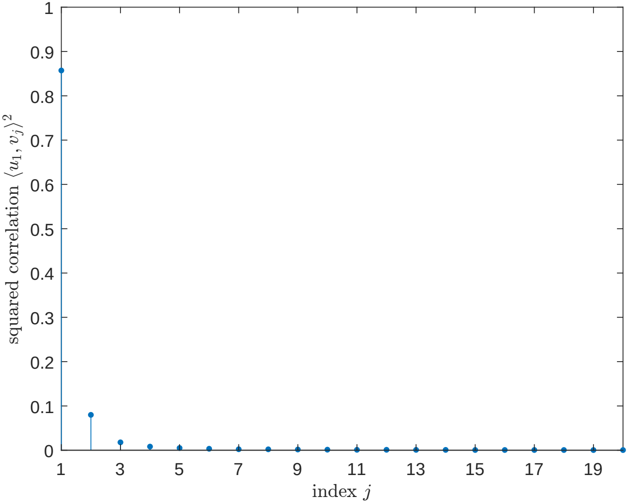
Erdős-Rényi graphs with $\delta = 0.1\%$ differed edges



$\langle u_1, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

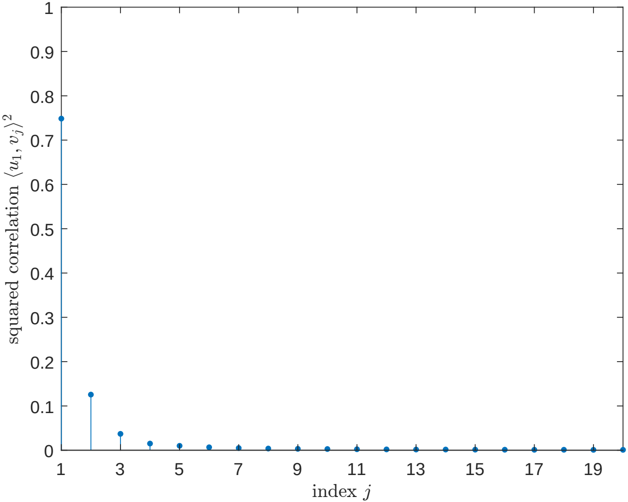
Erdős-Rényi graphs with $\delta = 0.5\%$ differed edges



$\langle u_1, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

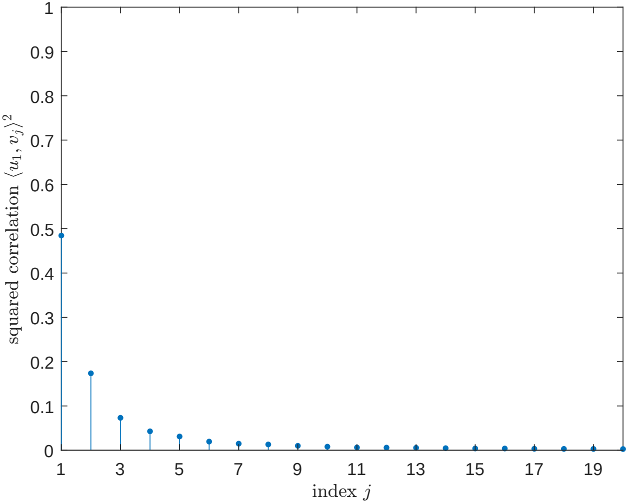
Erdős-Rényi graphs with $\delta = 1\%$ differed edges



$\langle u_1, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

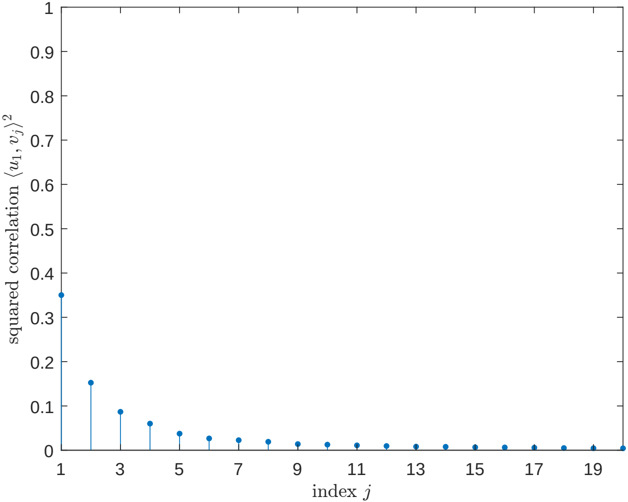
Erdős-Rényi graphs with $\delta = 3\%$ differed edges



$\langle u_1, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

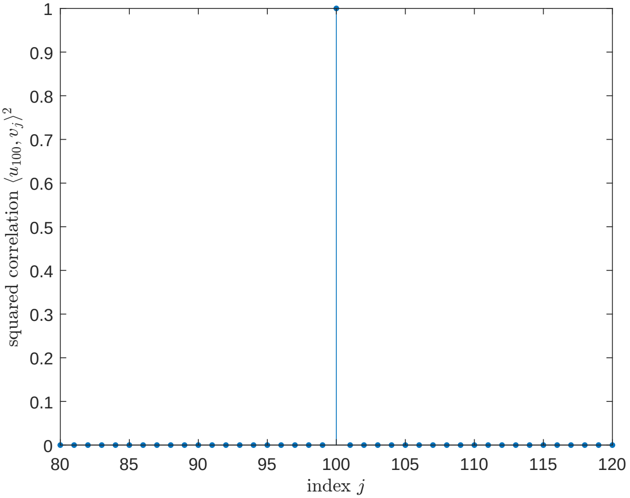
Erdős-Rényi graphs with $\delta = 5\%$ differed edges



$\langle u_1, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

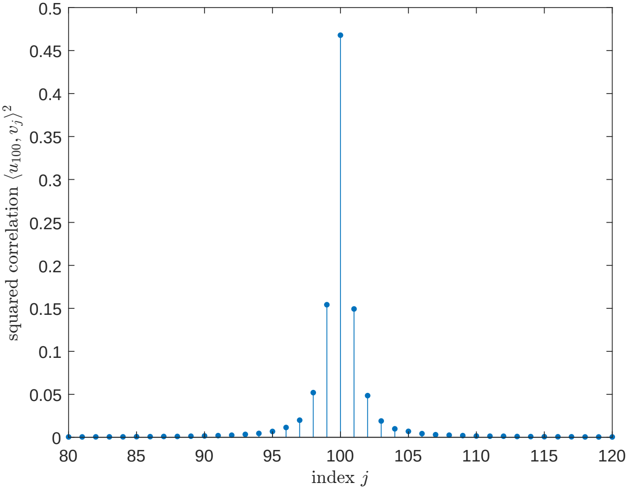
Isomorphic Erdős-Rényi graphs: 500 vertices, edge probability $\frac{1}{2}$



$\langle u_{100}, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

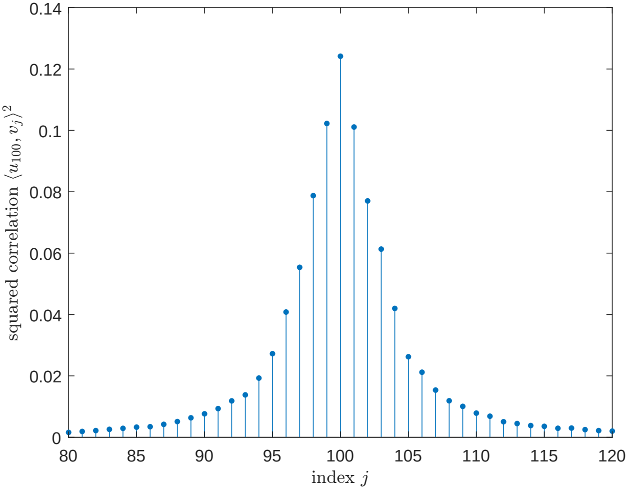
Erdős-Rényi graphs with $\delta = 0.1\%$ differed edges



$\langle u_{100}, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

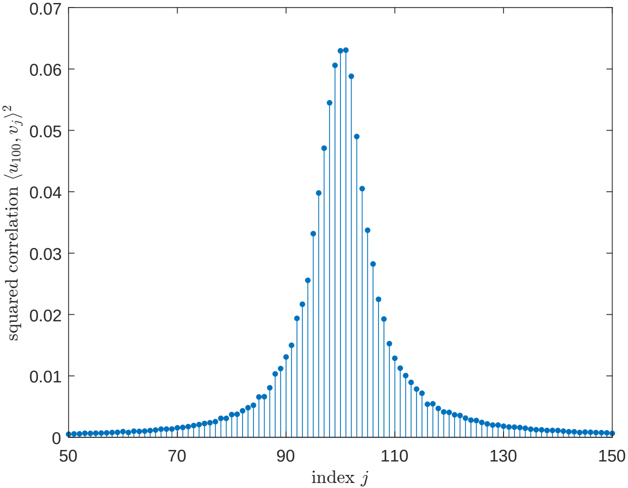
Erdős-Rényi graphs with $\delta = 0.5\%$ differed edges



$\langle u_{100}, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

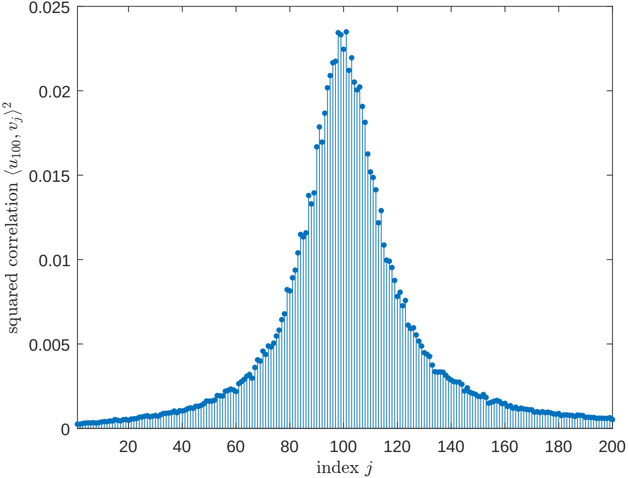
Erdős-Rényi graphs with $\delta = 1\%$ differed edges



$\langle u_{100}, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

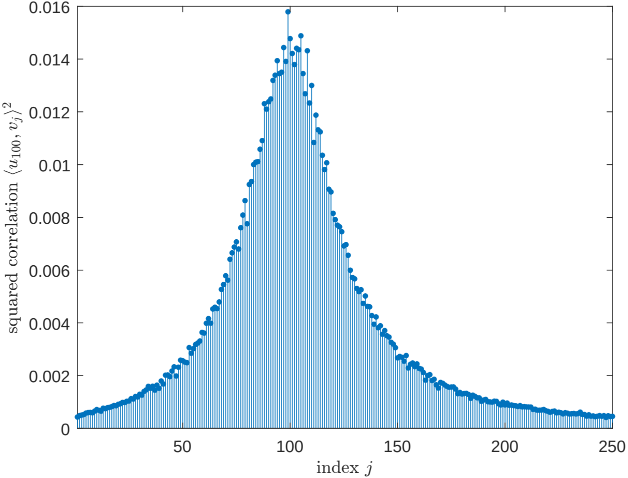
Erdős-Rényi graphs with $\delta = 3\%$ differed edges



$\langle u_{100}, v_j \rangle^2$, averaged across 1000 simulations

Eigenvector correlation

Erdős-Rényi graphs with $\delta = 5\%$ differed edges



$\langle u_{100}, v_j \rangle^2$, averaged across 1000 simulations

A new spectral method: GRAMPA

GRAph Matching by Pairwise eigen-Alignments:

$$X = \sum_{i,j=1}^{n} \underbrace{K\left(\frac{\lambda_i - \mu_j}{\eta}\right)}_{\text{spectral weights}} \times \underbrace{u_i^\top \mathbf{J} v_j \cdot u_i v_j^\top}_{\text{"Alignment" between } u_i \text{ and } v_j}$$

where $\eta =$ bandwidth parameter, $\mathbf{J} =$ all-one matrix

A new spectral method: GRAMPA

GRAph Matching by Pairwise eigen-Alignments:

$$X = \sum_{i,j=1}^{n} \underbrace{\frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2}}_{\text{spectral weights}} \times \underbrace{u_i^\top \mathbf{J} v_j \cdot u_i v_j^\top}_{\text{"Alignment" between } u_i \text{ and } v_j}$$

where $\eta =$ bandwidth parameter, $\mathbf{J} =$ all-one matrix

A new spectral method: GRAMPA

GRAph Matching by Pairwise eigen-Alignments:

$$X = \sum_{i,j=1}^{n} \underbrace{\frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2}}_{\text{spectral weights}} \times \underbrace{u_i^\top \mathbf{J} v_j \cdot u_i v_j^\top}_{\text{"Alignment" between } u_i \text{ and } v_j}$$

where $\eta =$ bandwidth parameter, $\mathbf{J} =$ all-one matrix

- All pairs matter:
  - Spectral weight penalizes pairs whose eigenvalues are far apart
  - Cauchy weight kernel is inspired by the eigenvector correlation decay [Bourgade-Yau '17], [Benigni '17]:

$$n \cdot \mathbb{E}\left[\langle u_i, v_j \rangle^2\right] \approx \frac{\delta}{(\lambda_i - \mu_j)^2 + C\delta^2}$$

  (See previous plot for this Cauchy form.)

- GRAMPA is invariant to the choices of signs for u_i and v_j

GRAMPA as regularized QP relaxation

- Graph matching as a quadratic assignment problem (QAP):

$$\arg\max_{\Pi \in S_n} \langle A, \Pi B \Pi^\top \rangle = \arg\min_{\Pi \in S_n} \|A - \Pi B \Pi^\top\|_F^2$$

GRAMPA as regularized QP relaxation

- Graph matching as a quadratic assignment problem (QAP):

$$\arg\max_{\Pi \in S_n} \langle A, \Pi B \Pi^\top \rangle = \arg\min_{\Pi \in S_n} \|A - \Pi B \Pi^\top\|_F^2 = \arg\min_{\Pi \in S_n} \|A\Pi - \Pi B\|_F^2$$

GRAMPA as regularized QP relaxation

- Graph matching as a quadratic assignment problem (QAP):

$$\arg\max_{\Pi \in S_n} \langle A, \Pi B \Pi^\top \rangle = \arg\min_{\Pi \in S_n} \|A - \Pi B \Pi^\top\|_F^2 = \arg\min_{\Pi \in S_n} \|A\Pi - \Pi B\|_F^2$$

- A popular quadratic programming relaxation [Zaslavskiy-Bach-Vert '09], [Aflalo-Bronstein-Kimmel '15], [Lyzinski-Fishkind-Fiori-Vogelstein-Priebe-Sapiro '15]

$$\arg\min_{X \geq 0:\ X\mathbf{1}=\mathbf{1},\ X^\top\mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 \qquad \text{(QP-DS)}$$

GRAMPA as regularized QP relaxation

- Graph matching as a quadratic assignment problem (QAP):

$$\arg\max_{\Pi \in S_n} \langle A, \Pi B \Pi^\top \rangle = \arg\min_{\Pi \in S_n} \|A - \Pi B \Pi^\top\|_F^2 = \arg\min_{\Pi \in S_n} \|A\Pi - \Pi B\|_F^2$$

- A popular quadratic programming relaxation [Zaslavskiy-Bach-Vert '09], [Aflalo-Bronstein-Kimmel '15], [Lyzinski-Fishkind-Fiori-Vogelstein-Priebe-Sapiro '15]

$$\arg\min_{X \geq 0:\, X\mathbf{1}=\mathbf{1},\, X^\top\mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 \qquad \text{(QP-DS)}$$

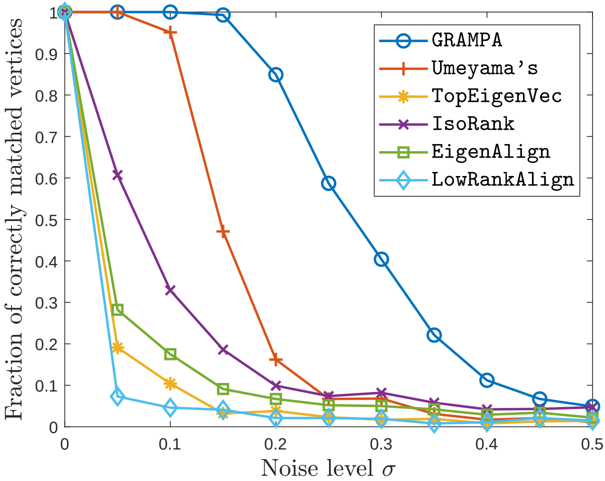
- The GRAMPA similarity matrix X is (a multiple of)

$$\arg\min_{X:\, \mathbf{1}^\top X\mathbf{1}=n} \|AX - XB\|_F^2 + \eta^2\|X\|_F^2$$

This further relaxes the DS constraint and adds a ridge regularizer

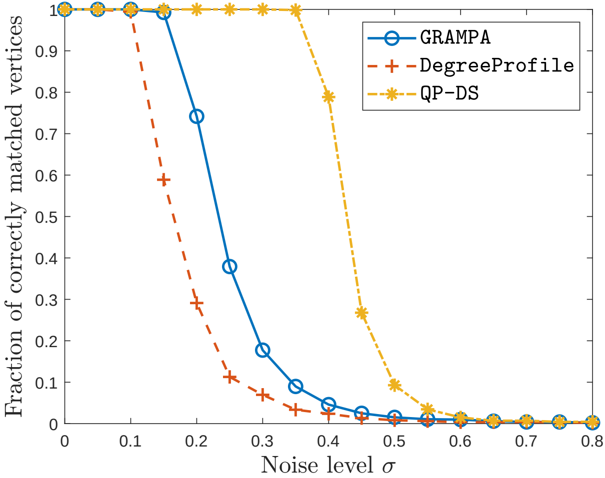
Experiments

Spectral algorithms on Erdős-Rényi graphs



$n = 100$ vertices, edge probability $\frac{1}{2}$, noise level $\sigma = \sqrt{\delta}$

Competitive methods on Erdős-Rényi graphs



$n = 500$ vertices, edge probability $\frac{1}{2}$, noise level $\sigma = \sqrt{\delta}$

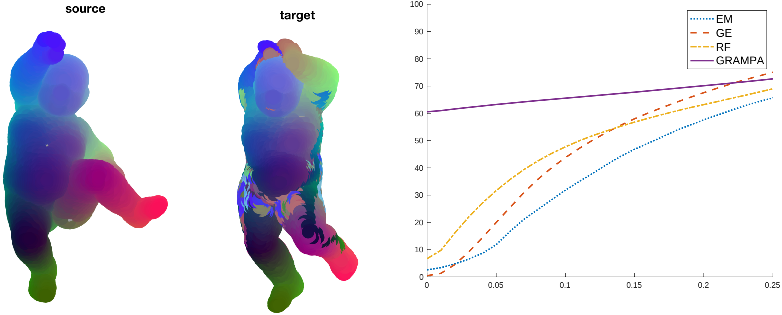
GRAMPA is 100–1000x faster than QP-DS and scalable to larger networks

Computer vision dataset SHREC'16

A realistic and challenging noisy dataset of 3D deformable shapes [Lähner et al '16]

- Triangulated mesh graphs: ∼10K vertices, ∼20K triangular faces.
- Very sparse graphs: average degree 3
- Noise from both local perturbations and large topological changes

Performance of GRAMPA



Bonus:

- Unsupervised (does not use training data)
- Does not use 3D coordinate info

Analysis of GRAMPA

Diagonal dominance in population version

Question: Is X "close" to true permutation matrix Π^*?

Diagonal dominance in population version

Question: Is X "close" to true permutation matrix Π^*?

Consider the "population version" of the regularized QP:

$$X \propto \arg \min_{X:\ \mathbf{1}^\top X \mathbf{1} = n} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

Diagonal dominance in population version

Question: Is X "close" to true permutation matrix Π^*?

Consider the "population version" of the regularized QP:

$$X_{\mathsf{pop}} = \arg \min_{X:\ \mathbf{1}^\top X \mathbf{1} = n} \mathbb{E}\left[\|AX - XB\|_F^2\right] + \eta^2 \|X\|_F^2$$

Diagonal dominance in population version

Question: Is X "close" to true permutation matrix Π^*?

Consider the "population version" of the regularized QP:

$$X_{\mathsf{pop}} = \arg \min_{X:\, \mathbf{1}^\top X \mathbf{1} = n} \mathbb{E}\left[\|AX - XB\|_F^2\right] + \eta^2 \|X\|_F^2$$

Assume $\Pi^* = \mathbf{I}$ and $A \leftarrow \frac{A - \mathbb{E}[A]}{\sqrt{nq(1-q)}}$ and $B \leftarrow \frac{B - \mathbb{E}[B]}{\sqrt{nq(1-q)}}$:

$$X_{\mathsf{pop}} = \epsilon \mathbf{I} + (1-\epsilon)\frac{\mathbf{J}}{n}, \qquad \epsilon \approx \frac{2(1-\delta)}{n(2\delta + \eta^2)}$$

- X_{pop} is close to $\frac{\mathbf{J}}{n}$ (center of the Birkhoff polytope)

Diagonal dominance in population version

Question: Is X "close" to true permutation matrix Π^*?

Consider the "population version" of the regularized QP:

$$X_{\mathsf{pop}} = \arg \min_{X:\, \mathbf{1}^\top X \mathbf{1} = n} \mathbb{E}\left[\|AX - XB\|_F^2\right] + \eta^2 \|X\|_F^2$$

Assume $\Pi^* = \mathbf{I}$ and $A \leftarrow \frac{A - \mathbb{E}[A]}{\sqrt{nq(1-q)}}$ and $B \leftarrow \frac{B - \mathbb{E}[B]}{\sqrt{nq(1-q)}}$:

$$X_{\mathsf{pop}} = \epsilon \mathbf{I} + (1-\epsilon)\frac{\mathbf{J}}{n}, \qquad \epsilon \approx \frac{2(1-\delta)}{n(2\delta + \eta^2)}$$

- X_{pop} is close to $\frac{\mathbf{J}}{n}$ (center of the Birkhoff polytope)
- Same analysis holds for tighter QP-DS

Diagonal dominance in population version

Question: Is X "close" to true permutation matrix Π^*?

Consider the "population version" of the regularized QP:

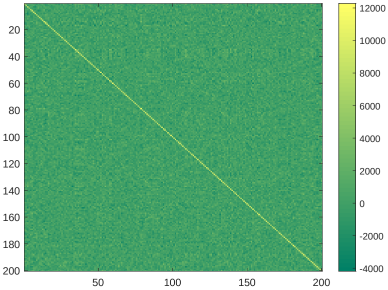
$$X_{\text{pop}} = \arg \min_{X:\, \mathbf{1}^\top X \mathbf{1} = n} \mathbb{E}\left[\|AX - XB\|_F^2\right] + \eta^2 \|X\|_F^2$$

Assume $\Pi^* = \mathbf{I}$ and $A \leftarrow \frac{A - \mathbb{E}[A]}{\sqrt{nq(1-q)}}$ and $B \leftarrow \frac{B - \mathbb{E}[B]}{\sqrt{nq(1-q)}}$:

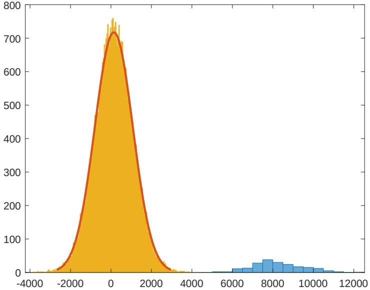
$$X_{\text{pop}} = \epsilon \mathbf{I} + (1-\epsilon)\frac{\mathbf{J}}{n}, \qquad \epsilon \approx \frac{2(1-\delta)}{n(2\delta + \eta^2)}$$

- X_{pop} is close to $\frac{\mathbf{J}}{n}$ (center of the Birkhoff polytope)
- Same analysis holds for tighter QP-DS
- X_{pop} is diagonally dominant: diagonals are $\approx \frac{2}{2\delta + \eta^2}$ times off-diagonals

Diagonal dominance of the similarity matrix



Heatmap of X



Histogram of off-diagonal (orange) and diagonal (blue) entries

When $\Pi^* = \mathbf{I}$, aim to show **diagonal dominance**

$$\min_k X_{kk} > \max_{k \neq \ell} X_{k\ell}$$

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

For Gaussian matrices

- eigenvectors are uniform:

$$u_i \sim \text{Uniform}(n\text{-sphere}) \approx \frac{1}{\sqrt{n}} \mathcal{N}(0, I_n)$$

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

For Gaussian matrices

- eigenvectors are uniform:

$$u_i \sim \text{Uniform}(n\text{-sphere}) \approx \frac{1}{\sqrt{n}} \mathcal{N}(0, I_n)$$

  - Sum: $\langle u_i, \mathbf{1} \rangle \approx 1$

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

For Gaussian matrices

- eigenvectors are uniform:

$$u_i \sim \text{Uniform}(n\text{-sphere}) \approx \frac{1}{\sqrt{n}} \mathcal{N}(0, I_n)$$

  - Sum: $\langle u_i, \mathbf{1} \rangle \approx 1$
  - kth entry: $(u_i)_k \approx \mathcal{N}(0, \frac{1}{n})$.

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

For Gaussian matrices

- eigenvectors are uniform:

$$u_i \sim \text{Uniform}(n\text{-sphere}) \approx \frac{1}{\sqrt{n}} \mathcal{N}(0, I_n)$$

  - Sum: $\langle u_i, \mathbf{1} \rangle \approx 1$
  - kth entry: $(u_i)_k \approx \mathcal{N}(0, \frac{1}{n})$.

- eigenvalues satisfies semicircle law:

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i} \approx \rho(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \cdot \mathbf{1}_{\{|x| \leq 2\}} \text{ (Wigner's semicircle law)}$$

Heuristic argument: noiseless Gaussian case

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle u_i u_j^\top$$

For Gaussian matrices

- eigenvectors are uniform:

$$u_i \sim \text{Uniform}(n\text{-sphere}) \approx \frac{1}{\sqrt{n}} \mathcal{N}(0, I_n)$$

  - Sum: $\langle u_i, \mathbf{1} \rangle \approx 1$
  - kth entry: $(u_i)_k \approx \mathcal{N}(0, \frac{1}{n})$.
- eigenvalues satisfies semicircle law:

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i} \approx \rho(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \cdot \mathbf{1}_{\{|x| \le 2\}} \text{ (Wigner's semicircle law)}$$

- eigenvalues and eigenvectors are independent

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i=1}^{n} \frac{1}{\eta} \langle u_i, \mathbf{1} \rangle^2 (u_i)_k (u_j)_\ell + \sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i=1}^{n} \frac{1}{\eta} \langle u_i, \mathbf{1} \rangle^2 (u_i)_k (u_j)_\ell + \sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

- First term is diagonally dominant:

$$\sum_{i=1}^{n} \frac{1}{\eta} \underbrace{\langle u_i, \mathbf{1} \rangle^2}_{\approx 1} \underbrace{(u_i)_k}_{\mathcal{N}(0,\frac{1}{n})} \underbrace{(u_i)_\ell}_{\mathcal{N}(0,\frac{1}{n})} \approx \begin{cases} \frac{1}{\eta} & \text{if } k = \ell \\ \frac{1}{\eta\sqrt{n}} & \text{if } k \neq \ell \end{cases}$$

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i=1}^{n} \frac{1}{\eta} \langle u_i, \mathbf{1} \rangle^2 (u_i)_k (u_j)_\ell + \sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

- First term is diagonally dominant:

$$\sum_{i=1}^{n} \frac{1}{\eta} \underbrace{\langle u_i, \mathbf{1} \rangle^2}_{\approx 1} \underbrace{(u_i)_k}_{\mathcal{N}(0,\frac{1}{n})} \underbrace{(u_i)_\ell}_{\mathcal{N}(0,\frac{1}{n})} \approx \begin{cases} \frac{1}{\eta} & \text{if } k = \ell \\ \frac{1}{\eta\sqrt{n}} & \text{if } k \neq \ell \end{cases}$$

- Second term can be viewed as perturbation:

$$\mathrm{Var}\left(\sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell \right) \approx \frac{1}{n^2} \sum_{i \neq j} \left(\frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \right)^2$$

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i=1}^{n} \frac{1}{\eta} \langle u_i, \mathbf{1} \rangle^2 (u_i)_k (u_j)_\ell + \sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

- First term is diagonally dominant:

$$\sum_{i=1}^{n} \frac{1}{\eta} \underbrace{\langle u_i, \mathbf{1} \rangle^2}_{\approx 1} \underbrace{(u_i)_k}_{\mathcal{N}(0,\frac{1}{n})} \underbrace{(u_i)_\ell}_{\mathcal{N}(0,\frac{1}{n})} \approx \begin{cases} \frac{1}{\eta} & \text{if } k = \ell \\ \frac{1}{\eta\sqrt{n}} & \text{if } k \neq \ell \end{cases}$$

- Second term can be viewed as perturbation:

$$\text{Var}\left(\sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle \left(u_i\right)_k \left(u_j\right)_\ell \right) \approx \iint \left(\frac{\eta}{(x-y)^2 + \eta^2} \right)^2 d\rho(x) d\rho($$

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i=1}^{n} \frac{1}{\eta} \langle u_i, \mathbf{1} \rangle^2 (u_i)_k (u_j)_\ell + \sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

- First term is diagonally dominant:

$$\sum_{i=1}^{n} \frac{1}{\eta} \underbrace{\langle u_i, \mathbf{1} \rangle^2}_{\approx 1} \underbrace{(u_i)_k}_{\mathcal{N}(0,\frac{1}{n})} \underbrace{(u_i)_\ell}_{\mathcal{N}(0,\frac{1}{n})} \approx \begin{cases} \frac{1}{\eta} & \text{if } k = \ell \\ \frac{1}{\eta\sqrt{n}} & \text{if } k \neq \ell \end{cases}$$

- Second term can be viewed as perturbation:

$$\text{Var}\left(\sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell \right) \approx \iint \Big(\frac{\eta}{(x-y)^2 + \eta^2} \Big)^2 d\rho(x) d\rho($$

$$\approx \frac{1}{\eta} \quad \text{(using semicircle law)}$$

Heuristic argument: noiseless Gaussian case

$$X_{k\ell} = \sum_{i=1}^{n} \frac{1}{\eta} \langle u_i, \mathbf{1} \rangle^2 (u_i)_k (u_j)_\ell + \sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell$$

- First term is diagonally dominant:

$$\sum_{i=1}^{n} \frac{1}{\eta} \underbrace{\langle u_i, \mathbf{1} \rangle^2}_{\approx 1} \underbrace{(u_i)_k}_{\mathcal{N}(0,\frac{1}{n})} \underbrace{(u_i)_\ell}_{\mathcal{N}(0,\frac{1}{n})} \approx \begin{cases} \frac{1}{\eta} & \text{if } k = \ell \\ \frac{1}{\eta\sqrt{n}} & \text{if } k \neq \ell \end{cases}$$

- Second term can be viewed as perturbation:

$$\text{Var}\left(\sum_{i \neq j} \frac{\eta}{(\lambda_i - \lambda_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle u_j, \mathbf{1} \rangle (u_i)_k (u_j)_\ell \right) \approx \iint \Big(\frac{\eta}{(x-y)^2 + \eta^2} \Big)^2 d\rho(x) d\rho($$

$$\approx \frac{1}{\eta} \text{ (using semicircle law)}$$

- Made rigorous when A, B are Gaussian since eigenvalues and eigenvectors are independent, but hard to extend to Erdős-Rényi graphs

Universality: Correlated Wigner model

Consider the standardized weighted adjacency matrices A, B where (A_{ij}, B_{ij}) are independent sub-gaussian pairs satisfying

$$\mathbb{E}[A_{ij}] = \mathbb{E}[B_{ij}] = 0, \quad \mathbb{E}[A_{ij}^2] = \mathbb{E}[B_{ij}^2] = \frac{1}{n}, \quad \mathbb{E}[A_{ij}B_{ij}] = \frac{1-\delta}{n}$$

Key proof technique: Resolvent and local laws

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i - z} u_i u_i^\top, \qquad z \in \mathbb{C} \backslash \mathbb{R}$$

Key proof technique: Resolvent and local laws

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i - z} u_i u_i^\top, \qquad z \in \mathbb{C}\backslash\mathbb{R}$$

Denote Wigner's semicircle density and its Stieltjes transform by

$$\rho(x) = \frac{1}{2\pi}\sqrt{4 - x^2} \cdot \mathbf{1}_{\{|x| \le 2\}} \quad \text{and} \quad m(z) = \int \frac{1}{x - z}\rho(x)dx = \frac{-z + \sqrt{z^2 - 4}}{2}$$

Key proof technique: Resolvent and local laws

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i - z} u_i u_i^\top, \qquad z \in \mathbb{C}\backslash\mathbb{R}$$

Denote Wigner's semicircle density and its Stieltjes transform by

$$\rho(x) = \frac{1}{2\pi}\sqrt{4 - x^2}\cdot\mathbf{1}_{\{|x|\le 2\}} \quad \text{and} \quad m(z) = \int \frac{1}{x - z}\rho(x)dx = \frac{-z + \sqrt{z^2 - 4}}{2}$$

- Classical result in RMT: empirical eigenvalue distribution $\frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i}$ converges to ρ.

Key proof technique: Resolvent and local laws

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i - z} u_i u_i^\top, \qquad z \in \mathbb{C} \backslash \mathbb{R}$$

Denote Wigner's semicircle density and its Stieltjes transform by

$$\rho(x) = \frac{1}{2\pi}\sqrt{4 - x^2} \cdot \mathbf{1}_{\{|x| \le 2\}} \quad \text{and} \quad m(z) = \int \frac{1}{x - z}\rho(x)dx = \frac{-z + \sqrt{z^2 - 4}}{2}$$

- Classical result in RMT: empirical eigenvalue distribution $\frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i}$ converges to ρ.
- So we expect

$$\frac{1}{n} \operatorname{Tr} R_A(z) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\lambda_i - z} \to m(z)$$

Key proof technique: Resolvent and local laws

Stronger results concern the entire matrix $R_A(z)$ (as opposed to just Tr), especially when $\mathrm{Im}\, z$ is small (known as **local laws**)

- $R_A(z) \approx m(z)\mathbf{I}$ entrywise [Erdos-Knowles-Yau-Yin '13]: whp

$$(R_A(z))_{ij} \approx m(z) \cdot \mathbf{1}_{\{i=j\}}$$

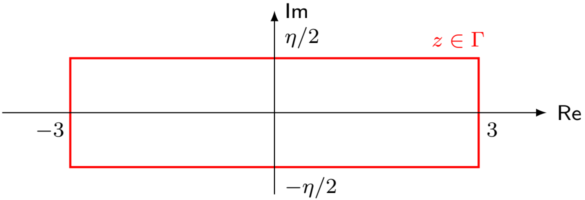
- Similarly, row sum and total sum satisfy: whp

$$\sum_j (R_A(z))_{ij} \lesssim \mathsf{polylog}(n) \qquad \sum_{i,j} (R_A(z))_{ij} \approx n \cdot m(z)$$

Universality proof step 1: Resolvent representation

Lemma

$$X \triangleq \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2} \langle u_i, \mathbf{1} \rangle \langle v_j, \mathbf{1} \rangle u_i v_j^\top$$
$$= \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma R_A(z) \mathbf{1}\mathbf{1}^\top R_B(z + \mathbf{i}\eta) dz$$



Γ encloses $\lambda_1, \ldots, \lambda_n$ but not $\mu_1 - \mathbf{i}\eta, \ldots, \mu_n - \mathbf{i}\eta$

Proof of resolvent representation

Cauchy's integral formula

Let Γ be closed curve enclosing z_0 and f is analytic inside Γ. Then

$$\frac{1}{2\pi\mathbf{i}}\oint_{\Gamma}\frac{f(z)}{z-z_0}dz = f(z_0)$$

Proof of resolvent representation

Cauchy's integral formula

Let Γ be closed curve enclosing z_0 and f is analytic inside Γ. Then

$$\frac{1}{2\pi\mathbf{i}}\oint_\Gamma \frac{f(z)}{z-z_0}dz = f(z_0)$$

Expanding

$$R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta) = \sum_{i,j=1}^{n} \frac{1}{(z-\lambda_i)(z+\mathbf{i}\eta-\mu_j)} u_i u_i^\top \mathbf{J} v_j v_j^\top$$

and

$$\begin{aligned}\frac{1}{2\pi}\operatorname{Re}\oint_\Gamma \frac{1}{(z-\lambda_i)(z+\mathbf{i}\eta-\mu_j)} &= \frac{1}{2\pi}\operatorname{Re}\frac{2\pi\mathbf{i}}{\lambda_i-\mu_j+\mathbf{i}\eta} \\ &= -\operatorname{Im}\frac{1}{\lambda_i-\mu_j+\mathbf{i}\eta} = \frac{\eta}{(\lambda_i-\mu_j)^2+\eta^2}\end{aligned}$$

Block inverse

Technical tool: Schur complement identity

Provided D is square and invertible,

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \implies M^{-1} = \begin{bmatrix} S & -SBD^{-1} \\ -D^{-1}CS & D^{-1} + D^{-1}CSBD^{-1} \end{bmatrix}$$

where $S = (A - BD^{-1}C)^{-1}$.

Utility of this result (both theoretically and algorithmically): reduce inverse to inverting smaller matrices

Block inverse

Technical tool: Schur complement identity

Provided D is square and invertible,

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \implies M^{-1} = \begin{bmatrix} S & -SBD^{-1} \\ -D^{-1}CS & D^{-1} + D^{-1}CSBD^{-1} \end{bmatrix}$$

where $S = (A - BD^{-1}C)^{-1}$.

Utility of this result (both theoretically and algorithmically): reduce inverse to inverting smaller matrices

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

- $$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_{\Gamma} \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

-
$$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$

- By the Schur-complement formula

$$\begin{aligned} R_{A,1*}(z) &= -R_{A,11}(z) \cdot a_1^\top (A^{(1)} - z\mathbf{I})^{-1} \\ &= -R_{A,11}(z) \cdot a_1^\top R_{A^{(1)}}(z) \end{aligned}$$

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

- $$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$
- By the Schur-complement formula

$$\begin{aligned} R_{A,1*}(z) &= -R_{A,11}(z) \cdot a_1^\top (A^{(1)} - z\mathbf{I})^{-1} \\ &\approx -m(z) \cdot a_1^\top R_{A^{(1)}}(z) \end{aligned}$$

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

- $$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$
- By the Schur-complement formula

$$\begin{aligned} R_{A,1*}(z) &= -R_{A,11}(z) \cdot a_1^\top (A^{(1)} - z\mathbf{I})^{-1} \\ &\approx -m(z) \cdot a_1^\top R_{A^{(1)}}(z) \end{aligned}$$

- Writing a similar expression for B, we get

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma m(z)m(z+\mathbf{i}\eta) \left[a_1^\top R_{A^{(1)}}(z)\mathbf{1}\mathbf{1}^\top R_{B^{(1)}}(z+\mathbf{i}\eta)b_1\right] dz$$

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right] \left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

- $$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$
- By the Schur-complement formula

$$\begin{aligned} R_{A,1*}(z) &= -R_{A,11}(z) \cdot a_1^\top (A^{(1)} - z\mathbf{I})^{-1} \\ &\approx -m(z) \cdot a_1^\top R_{A^{(1)}}(z) \end{aligned}$$

- Writing a similar expression for B, we get

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \underbrace{\left[\oint_\Gamma m(z)m(z+\mathbf{i}\eta)R_{A^{(1)}}(z)\mathbf{1}\mathbf{1}^\top R_{B^{(1)}}(z+\mathbf{i}\eta)dz\right]}_{M} b_1$$

Step 2: Leave-one-out relation

$$X_{11} = \frac{1}{2\pi} \operatorname{Re} \oint_\Gamma \underbrace{\left[e_1^\top R_A(z)\mathbf{1}\right]\left[\mathbf{1}^\top R_B(z+\mathbf{i}\eta)e_1\right]}_{\text{correlated 1st row sums}} dz$$

-
$$A = \begin{pmatrix} a_{11} & a_1^\top \\ a_1 & A^{(1)} \end{pmatrix} \qquad R_A(z) = \begin{pmatrix} R_{A,11} & R_{A,1*} \\ R_{A,*1} & R_{A,**} \end{pmatrix}$$

- By the Schur-complement formula

$$\begin{aligned} R_{A,1*}(z) &= -R_{A,11}(z) \cdot a_1^\top (A^{(1)} - z\mathbf{I})^{-1} \\ &\approx -m(z) \cdot a_1^\top R_{A^{(1)}}(z) \end{aligned}$$

- Writing a similar expression for B, we get

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \underbrace{\left[\oint_\Gamma m(z)m(z+\mathbf{i}\eta)R_{A^{(1)}}(z)\mathbf{1}\mathbf{1}^\top R_{B^{(1)}}(z+\mathbf{i}\eta)dz\right]}_{M} b_1$$

- (a_1, b_1) are correlated and independent of $(A^{(1)}, B^{(1)})$ and hence M

Step 2: Leave-two-out relation

Similarly, for off-diagonals

$$X_{12} \approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \left[\oint_\Gamma m(z) m(z + \mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{1}\mathbf{1}^\top R_{B^{(12)}}(z + \mathbf{i}\eta) dz \right] b_2.$$

where $A^{(12)}$ and $B^{(12)}$ are the same as A and B by deleting the first two rows and columns.

Technical tool: concentration of bilinear forms

Recall: a_1, b_1 are correlated and a_1, b_2 are independent, s.t.

$$\mathbb{E}[a_1 b_1^\top] = \frac{1-\delta}{n}\mathbf{I}, \quad \mathbb{E}[a_1 b_2^\top] = 0.$$

So we expect, for deterministic matrix M,

$$a_1^\top M b_1 \approx \frac{1-\delta}{n}\,\mathrm{Tr}(M) \quad a_1^\top M b_2 \approx N\left(0, \frac{1}{n^2}\,\|M\|_{\mathrm{F}}^2\right)$$

Technical tool: concentration of bilinear forms

Recall: a_1, b_1 are correlated and a_1, b_2 are independent, s.t.

$$\mathbb{E}[a_1 b_1^\top] = \frac{1-\delta}{n}\mathbf{I}, \quad \mathbb{E}[a_1 b_2^\top] = 0.$$

So we expect, for deterministic matrix M,

$$a_1^\top M b_1 \approx \frac{1-\delta}{n}\operatorname{Tr}(M) \quad a_1^\top M b_2 \approx N\left(0, \frac{1}{n^2}\|M\|_{\mathrm{F}}^2\right)$$

More precisely, whp

$$\left|a_1^\top M b_1 - \frac{1-\delta}{n}\operatorname{Tr}(M)\right| \le \frac{\mathsf{polylog}(n)\,\|M\|_{\mathrm{F}}}{n} \quad \left|a_1^\top M b_2\right| \le \frac{\mathsf{polylog}(n)\,\|M\|_{\mathrm{F}}}{n}$$

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \left[\oint_\Gamma m(z) m(z + \mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z + \mathbf{i}\eta) dz \right] b_1$$

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} \underbrace{a_1^\top \left[\oint_\Gamma m(z) m(z + \mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z + \mathbf{i}\eta) dz \right] b_1}_{\approx \frac{1-\delta}{n} \operatorname{Tr}\left[\oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right]}$$

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} \underbrace{a_1^\top \left[\oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right] b_1}_{\approx \frac{1-\delta}{n} \operatorname{Tr}\left[\oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right]}$$

Off-diagonal entries:

$$X_{12} \approx \frac{1}{2\pi} \operatorname{Re} a_1^\top \left[\oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{J} R_{B^{(12)}}(z+\mathbf{i}\eta) dz \right] b_2$$

Here (a_1, b_2) are independent, so the conditional mean is 0

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

$$X_{11} \approx \frac{1}{2\pi} \operatorname{Re} \underbrace{a_1^\top \left[\oint_\Gamma m(z) m(z + \mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z + \mathbf{i}\eta) dz \right] b_1}_{\approx \frac{1-\delta}{n} \operatorname{Tr}\left[\oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right]}$$

Off-diagonal entries:

$$X_{12} \approx \frac{1}{2\pi} \operatorname{Re} \underbrace{a_1^\top \left[\oint_\Gamma m(z) m(z + \mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{J} R_{B^{(12)}}(z + \mathbf{i}\eta) dz \right] b_2}_{\lesssim \frac{\mathsf{polylog}(n)}{n} \left\| \oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{J} R_{B^{(12)}}(z+\mathbf{i}\eta) dz \right\|_F}$$

Here (a_1, b_2) are independent, so the conditional mean is 0

Step 4: Proof of diagonal dominance

- Diagonal entries:

$$X_{11} \approx \frac{1-\delta}{2\pi} \operatorname{Re} \frac{1}{n} \operatorname{Tr} \left[\oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right]$$

$$\approx \frac{1-\delta}{2\pi} \operatorname{Re} \frac{1}{\mathbf{i}\eta} \oint_\Gamma m(z) m(z+\mathbf{i}\eta) \left(m(z+\mathbf{i}\eta) - m(z) \right) dz + \frac{\sqrt{\delta}}{\eta^2}$$

$$\approx \frac{1-\delta}{\eta} + \frac{\sqrt{\delta}}{\eta^2}$$

Step 4: Proof of diagonal dominance

- Diagonal entries:

$$X_{11} \approx \frac{1-\delta}{2\pi} \operatorname{Re} \frac{1}{n} \operatorname{Tr} \left[\oint_\Gamma m(z)m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right]$$

$$\approx \frac{1-\delta}{2\pi} \operatorname{Re} \frac{1}{\mathbf{i}\eta} \oint_\Gamma m(z)m(z+\mathbf{i}\eta) \left(m(z+\mathbf{i}\eta) - m(z) \right) dz + \frac{\sqrt{\delta}}{\eta^2}$$

$$\approx \frac{1-\delta}{\eta} + \frac{\sqrt{\delta}}{\eta^2}$$

- Off-diagonal entries

$$X_{12} \lesssim \frac{\mathsf{polylog}(n)}{n} \left\| \oint_\Gamma m(z)m(z+\mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{J} R_{B^{(12)}}(z+\mathbf{i}\eta) dz \right\|_F$$

$$\lesssim \frac{\mathsf{polylog}(n)}{\sqrt{\eta}}$$

Step 4: Proof of diagonal dominance

- Diagonal entries:

$$\begin{aligned} X_{11} &\approx \frac{1-\delta}{2\pi} \operatorname{Re} \frac{1}{n} \operatorname{Tr} \left[\oint_{\Gamma} m(z) m(z+\mathbf{i}\eta) R_{A^{(1)}}(z) \mathbf{J} R_{B^{(1)}}(z+\mathbf{i}\eta) dz \right] \\ &\approx \frac{1-\delta}{2\pi} \operatorname{Re} \frac{1}{\mathbf{i}\eta} \oint_{\Gamma} m(z) m(z+\mathbf{i}\eta) \left(m(z+\mathbf{i}\eta) - m(z) \right) dz + \frac{\sqrt{\delta}}{\eta^2} \\ &\approx \frac{1-\delta}{\eta} + \frac{\sqrt{\delta}}{\eta^2} \end{aligned}$$

- Off-diagonal entries

$$\begin{aligned} X_{12} &\lesssim \frac{\mathsf{polylog}(n)}{n} \left\| \oint_{\Gamma} m(z) m(z+\mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{J} R_{B^{(12)}}(z+\mathbf{i}\eta) dz \right\|_F \\ &\lesssim \frac{\mathsf{polylog}(n)}{\sqrt{\eta}} \end{aligned}$$

- Applying this and a union bound for every $X_{k\ell}$ shows that X is diagonally dominant when

$$\sqrt{\delta} \lesssim \eta \lesssim (\log n)^{-(4+2\varepsilon)}$$

$\mathrm{Tr}(\cdot)$ calculation

Applying

- $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$
- whp, $\mathbf{1}^\top R_A(z)\mathbf{1} \approx nm(z)$, $\|A-B\| \lesssim \sqrt{\delta}$ and $\|R_A(z)\mathbf{1}\| \lesssim \sqrt{\frac{n}{\eta}}$,

we get

$$\begin{aligned}&\frac{1}{n}\operatorname{Re}\operatorname{Tr} M\\ =&\frac{1}{n}\operatorname{Re}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\operatorname{Tr}\left[R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)\right]\end{aligned}$$

$\mathrm{Tr}(\cdot)$ calculation

Applying

- $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$
- whp, $\mathbf{1}^\top R_A(z)\mathbf{1} \approx nm(z)$, $\|A-B\| \lesssim \sqrt{\delta}$ and $\|R_A(z)\mathbf{1}\| \lesssim \sqrt{\frac{n}{\eta}}$,

we get

$$\begin{aligned}&\frac{1}{n}\operatorname{Re}\operatorname{Tr} M \\ =&\frac{1}{n}\operatorname{Re}\oint_\Gamma dz\, m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1}\end{aligned}$$

$\mathrm{Tr}(\cdot)$ calculation

Applying

- $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$
- whp, $\mathbf{1}^\top R_A(z)\mathbf{1} \approx nm(z)$, $\|A-B\| \lesssim \sqrt{\delta}$ and $\|R_A(z)\mathbf{1}\| \lesssim \sqrt{\frac{n}{\eta}}$,

we get

$$
\begin{aligned}
&\frac{1}{n}\operatorname{Re}\operatorname{Tr} M\\
=&\frac{1}{n}\operatorname{Re}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1}\\
=&\frac{1}{n\eta}\operatorname{Im}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top\left(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\right)\mathbf{1}
\end{aligned}
$$

$\mathrm{Tr}(\cdot)$ calculation

Applying

- $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$
- whp, $\mathbf{1}^\top R_A(z)\mathbf{1} \approx nm(z)$, $\|A-B\| \lesssim \sqrt{\delta}$ and $\|R_A(z)\mathbf{1}\| \lesssim \sqrt{\frac{n}{\eta}}$,

we get

$$\begin{aligned}
&\frac{1}{n}\,\mathrm{Re}\,\mathrm{Tr}\,M \\
=&\frac{1}{n}\,\mathrm{Re}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1} \\
=&\frac{1}{n\eta}\,\mathrm{Im}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top\left(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\right)\mathbf{1} \\
\le&\frac{1}{\eta}\,\mathrm{Im}\oint_\Gamma dz\; m(z)m(z+\mathbf{i}\eta)(m(z+\mathbf{i}\eta) - m(z)) + O\left(\frac{\sqrt{\delta}}{\eta^2}\right)
\end{aligned}$$

$\mathrm{Tr}(\cdot)$ calculation

Applying

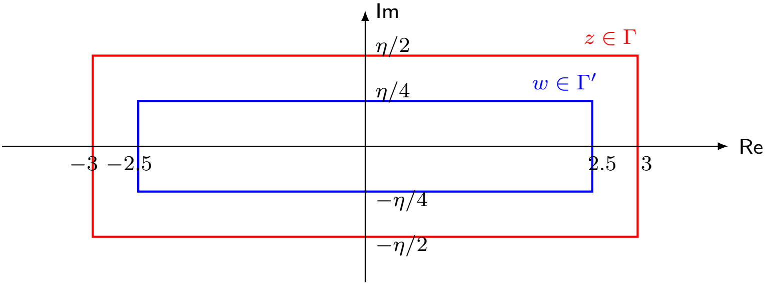
- $R_B(z+\mathbf{i}\eta)R_A(z) = \frac{1}{\mathbf{i}\eta}(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z))$
- whp, $\mathbf{1}^\top R_A(z)\mathbf{1} \approx nm(z)$, $\|A-B\| \lesssim \sqrt{\delta}$ and $\|R_A(z)\mathbf{1}\| \lesssim \sqrt{\frac{n}{\eta}}$,

we get

$$
\begin{aligned}
&\frac{1}{n}\operatorname{Re}\operatorname{Tr} M \\
=&\frac{1}{n}\operatorname{Re}\oint_\Gamma dz\, m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_A(z)\mathbf{1} \\
=&\frac{1}{n\eta}\operatorname{Im}\oint_\Gamma dz\, m(z)m(z+\mathbf{i}\eta)\mathbf{1}^\top\left(R_B(z+\mathbf{i}\eta) - R_A(z) - R_B(z+\mathbf{i}\eta)(A-B)R_A(z)\right)\mathbf{1} \\
\le&\frac{1}{\eta}\underbrace{\operatorname{Im}\oint_\Gamma dz\, m(z)m(z+\mathbf{i}\eta)(m(z+\mathbf{i}\eta)-m(z))}_{2\pi+o_\eta(1)} + O\left(\frac{\sqrt{\delta}}{\eta^2}\right)
\end{aligned}
$$

$\|\cdot\|_F$ calculation

Deforming the contour:



Then

$$M = \oint_{\Gamma} m(z)m(z+\mathbf{i}\eta)R_A(z)\mathbf{J}R_B(z+\mathbf{i}\eta)dz$$
$$= \oint_{\Gamma'} m(w)m(w+\mathbf{i}\eta)R_A(w)\mathbf{J}R_B(w+\mathbf{i}\eta)dw$$

$\|\cdot\|_{\mathrm{F}}$ calculation

Applying the facts

- $\overline{m(z)} = m(\overline{z})$, $R_A(z)^* = R_A(\overline{z})$
- $R_A(z)R_A(w) = \frac{R_A(z)-R_A(w)}{z-w}$
- $|m(z)| \lesssim 1$ and $|\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim n$

we get

$$\|M\|_{\mathrm{F}}^2 = \mathrm{Tr}(MM^*)$$
$$= \oint_\Gamma dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(\overline{w})m(\overline{w}-\mathbf{i}\eta)\,\mathrm{Tr}\left[R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(\overline{w}-\mathbf{i}\eta)\mathbf{1}\mathbf{1}^\top R_A(\overline{w})\right.$$

$\|\cdot\|_{\mathrm{F}}$ calculation

Applying the facts

- $\overline{m(z)} = m(\overline{z})$, $R_A(z)^* = R_A(\overline{z})$
- $R_A(z)R_A(w) = \frac{R_A(z)-R_A(w)}{z-w}$
- $|m(z)| \lesssim 1$ and $|\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim n$

we get

$$\begin{aligned}&\|M\|_{\mathrm{F}}^2 = \mathrm{Tr}(MM^*)\\ &= -\oint_\Gamma dz \oint_{\Gamma'} dw \; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(w-\mathbf{i}\eta)\mathbf{1}\end{aligned}$$

$\|\cdot\|_{\mathrm{F}}$ calculation

Applying the facts

- $\overline{m(z)} = m(\overline{z})$, $R_A(z)^* = R_A(\overline{z})$
- $R_A(z)R_A(w) = \frac{R_A(z)-R_A(w)}{z-w}$
- $|m(z)| \lesssim 1$ and $|\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim n$

we get

$$
\begin{aligned}
&\|M\|_{\mathrm{F}}^2 = \mathrm{Tr}(MM^*) \\
&= -\oint_\Gamma dz \oint_{\Gamma'} dw \; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(w-\mathbf{i}\eta)\mathbf{1} \\
&= -\oint_\Gamma dz \oint_{\Gamma'} dw \; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\frac{\mathbf{1}^\top(R_A(z)-R_A(w))\mathbf{1}}{z-w} \\
&\qquad\qquad \times \frac{\mathbf{1}^\top(R_B(z+\mathbf{i}\eta)-R_B(w-\mathbf{i}\eta))\mathbf{1}}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)}
\end{aligned}
$$

$\|\cdot\|_{\mathrm{F}}$ calculation

Applying the facts

- $\overline{m(z)} = m(\overline{z})$, $R_A(z)^* = R_A(\overline{z})$
- $R_A(z)R_A(w) = \frac{R_A(z)-R_A(w)}{z-w}$
- $|m(z)| \lesssim 1$ and $|\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim n$

we get

$$
\begin{aligned}
&\|M\|_{\mathrm{F}}^2 = \mathrm{Tr}(MM^*) \\
=& -\oint_\Gamma dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(w-\mathbf{i}\eta)\mathbf{1} \\
=& -\oint_\Gamma dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\frac{\mathbf{1}^\top(R_A(z)-R_A(w))\mathbf{1}}{z-w} \\
&\qquad\times \frac{\mathbf{1}^\top(R_B(z+\mathbf{i}\eta)-R_B(w-\mathbf{i}\eta))\mathbf{1}}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)} \\
\lesssim& n^2 \oint_\Gamma dz \oint_{\Gamma'} dw \frac{1}{|z-w|}\frac{1}{|z-w+2\mathbf{i}\eta|}
\end{aligned}
$$

$\|\cdot\|_{\mathrm{F}}$ calculation

Applying the facts

- $\overline{m(z)} = m(\overline{z})$, $R_A(z)^* = R_A(\overline{z})$
- $R_A(z)R_A(w) = \frac{R_A(z)-R_A(w)}{z-w}$
- $|m(z)| \lesssim 1$ and $|\mathbf{1}^\top R_A(z)\mathbf{1}| \lesssim n$

we get

$$
\begin{aligned}
&\|M\|_{\mathrm{F}}^2 = \mathrm{Tr}(MM^*) \\
=& -\oint_\Gamma dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\mathbf{1}^\top R_A(w)R_A(z)\mathbf{1}\mathbf{1}^\top R_B(z+\mathbf{i}\eta)R_B(w-\mathbf{i}\eta)\mathbf{1} \\
=& -\oint_\Gamma dz \oint_{\Gamma'} dw\; m(z)m(z+\mathbf{i}\eta)m(w)m(w-\mathbf{i}\eta)\frac{\mathbf{1}^\top (R_A(z)-R_A(w))\mathbf{1}}{z-w} \\
&\qquad\qquad\qquad\qquad \times \frac{\mathbf{1}^\top(R_B(z+\mathbf{i}\eta)-R_B(w-\mathbf{i}\eta))\mathbf{1}}{z+\mathbf{i}\eta-(w-\mathbf{i}\eta)} \\
\lesssim& n^2 \underbrace{\oint_\Gamma dz \oint_{\Gamma'} dw \frac{1}{|z-w|}\frac{1}{|z-w+2\mathbf{i}\eta|}}_{\asymp 1/\eta}
\end{aligned}
$$

Extensions

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Gaussian weighted graphs: improvement to $\delta \lesssim (\log n)^{-2}$ by direct analysis (slides 33-34)

Extensions

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Gaussian weighted graphs: improvement to $\delta \lesssim (\log n)^{-2}$ by direct analysis (slides 33-34)
- Similar result for a tighter QP relaxation (this is not automatic!)

$$\arg \max_{X:\, X\mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

  - Open problem: analyze the tigher relaxation

$$\arg \max_{X:\, X\mathbf{1}=\mathbf{1}, X^\top \mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

Extensions

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Gaussian weighted graphs: improvement to $\delta \lesssim (\log n)^{-2}$ by direct analysis (slides 33-34)
- Similar result for a tighter QP relaxation (this is not automatic!)

$$\arg \max_{X:\, X\mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

  - Open problem: analyze the tigher relaxation

$$\arg \max_{X:\, X\mathbf{1}=\mathbf{1}, X^\top\mathbf{1}=\mathbf{1}} \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2$$

- Similar results for matching bipartite graphs

Concluding remarks

- New spectral graph matching algorithm: "full-rank" spectral method

$$X = \sum_{i,j=1}^{n} \frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2} u_i u_i^\top \mathbf{J} v_j v_j^\top$$

- Efficiently matches two graphs with average degree $\geq \mathrm{polylog}(n)$ and fraction of differred edges $\leq 1/\mathrm{polylog}(n)$

Open problems

- Theoretical guarantees for QP-DS

$$\min_{X \text{ doubly stochastic}} \|AX - XB\|_F^2$$

Open problems

- Theoretical guarantees for QP-DS

$$\min_{X \text{ doubly stochastic}} \|AX - XB\|_F^2$$

- Other random graphs ensembles, e.g., geometric graphs

References

- Z. Fan, C. Mao, W, J. Xu *Spectral graph matching and regularized quadratic relaxations I: Algorithm and Gaussian analysis*, Foundations of Computational Mathematics, arxiv:1907.08880.
- Z. Fan, C. Mao, W, J. Xu *Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality*, Foundations of Computational Mathematics, arxiv:1907.08883.