Spectral graph matching and regularized quadratic relaxations

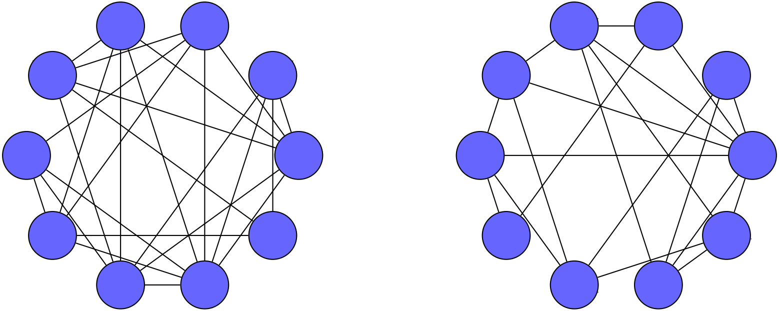
Jiaming Xu

The Fuqua School of Business
Duke University

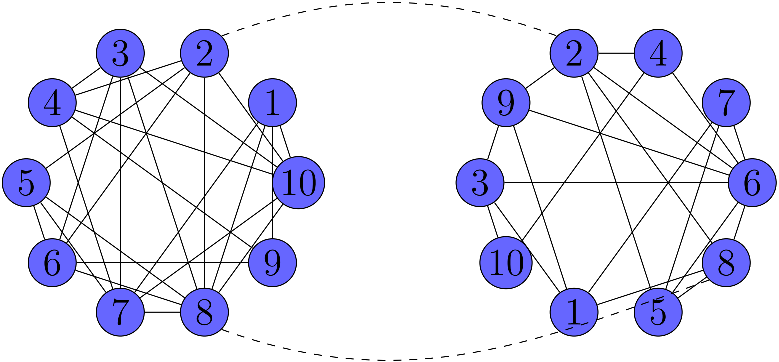
Joint work with
Zhou Fan, Cheng Mao, Yihong Wu
Department of Statistics and Data Science, Yale

Workshop on Graphical models, Exchangeable models and Graphons
August 20, 2019

Graph matching (network alignment)

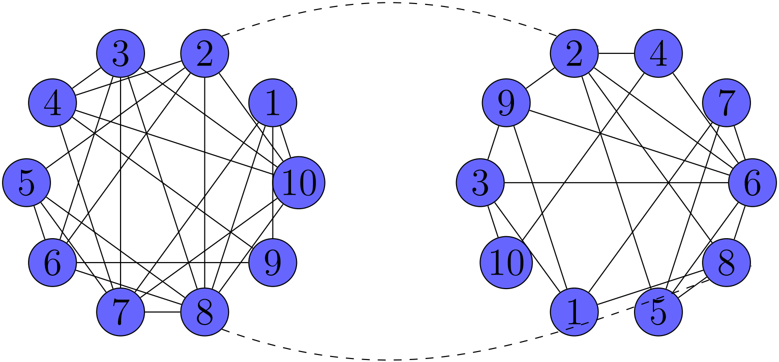


Goal: find a mapping between two node sets that maximally aligns the edges (i.e. minimizes # of adjacency disagreements)



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

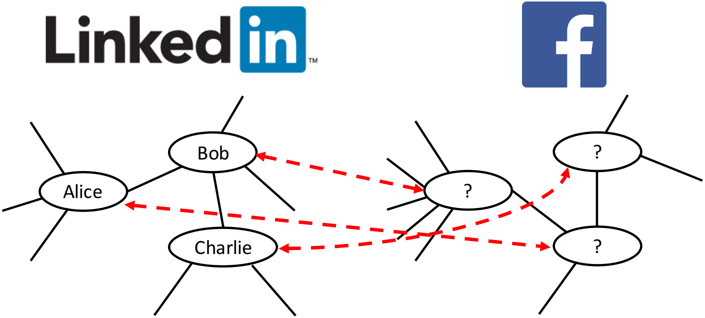
Graph matching (network alignment)



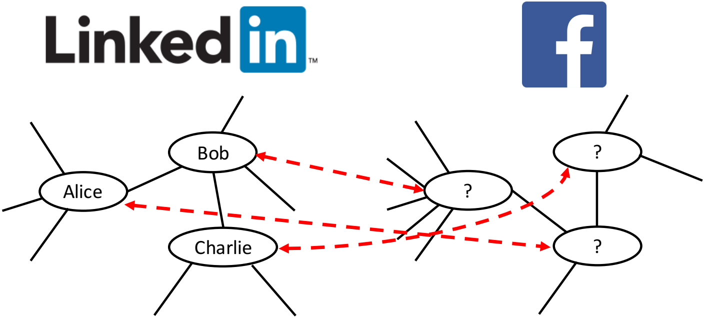
Goal: find a mapping between two node sets that maximally aligns the edges (i.e. minimizes # of adjacency disagreements)

Noiseless case: reduce to graph isomoprhism

Application 1: Network de-anonymization

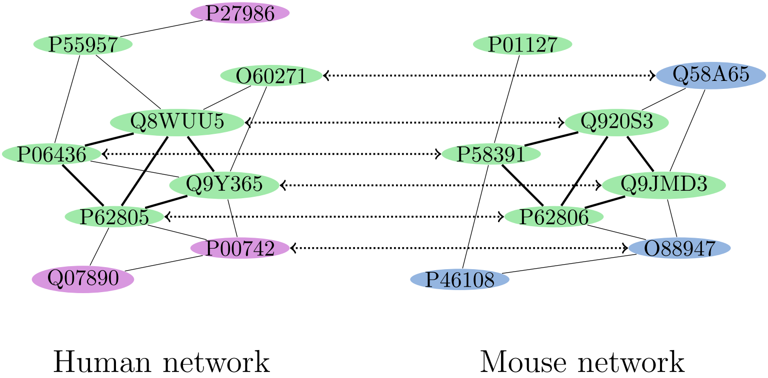


Application 1: Network de-anonymization



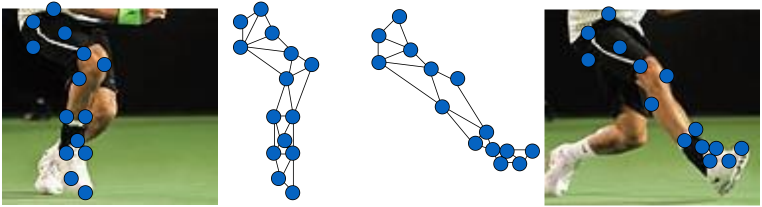
- Successfully de-anonymize Netflix by matching it to IMDB [Narayanan-Shmatikov '08]
- Correctly identify 30.8% of node pairings between Twitter and Flickr [Narayanan-Shmatikov '09]

Application 2: Protein-protein interaction network



[Kazemi-Hassani-Grossglauser-Modarres '16]

Ontology: Discover proteins with similar functions across different species based on interaction network topology

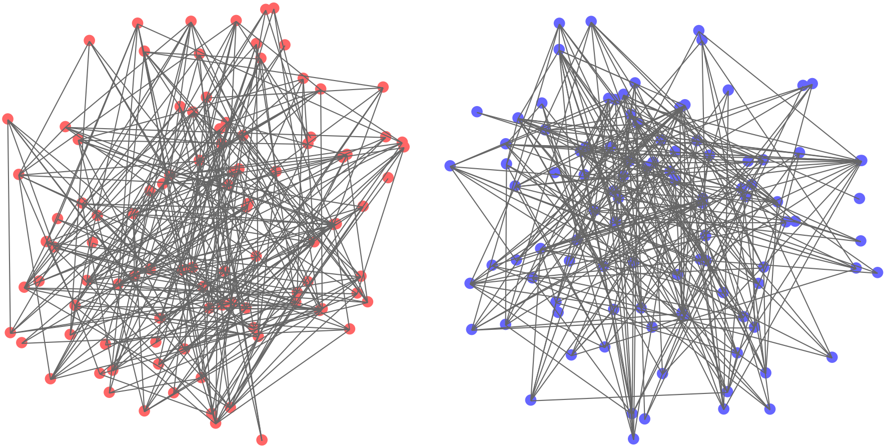


objects → graphs (features → nodes, distances → edges)
match objects by matching graphs

A fundamental problem in computer vision with applications in 3D reconstruction, object tracking, shape matching, image classification, autonomous driving, ...

Two key challenges

- Statistical: two graphs may not be the same
- Computational: # of possible node mappings is $n!$ ($100! \approx 10^{158}$)



Beyond computational intractability

- NP-hard for matching two general graphs
- However, real networks are not arbitrary and have latent structures

Beyond computational intractability

- NP-hard for matching two general graphs
- However, real networks are not arbitrary and have latent structures

Focus of this talk

Statistical models for graph matching: correlated random graphs

Beyond computational intractability

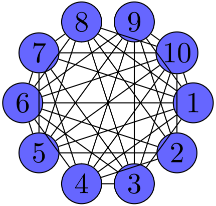
- NP-hard for matching two general graphs
- However, real networks are not arbitrary and have latent structures

Focus of this talk

Statistical models for graph matching: correlated random graphs

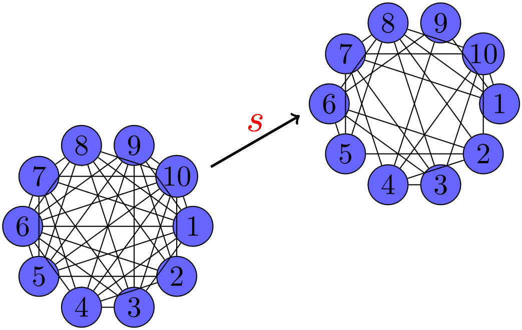
- Focus on correlated Erdős-Rényi graphs model [Pedarsani-Grossglauser '11]
- Performance of our algorithm is proven to be universal

Correlated Erdős-Rényi graphs model



$$G_0 \sim \mathcal{G}(n, p)$$

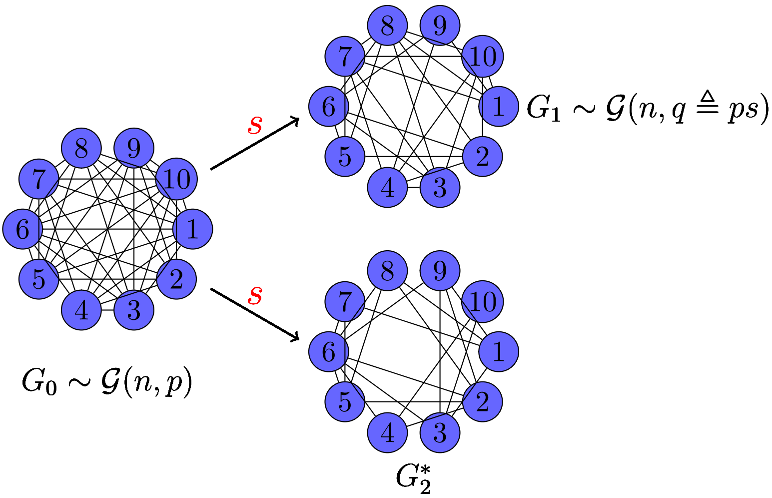
Correlated Erdős-Rényi graphs model



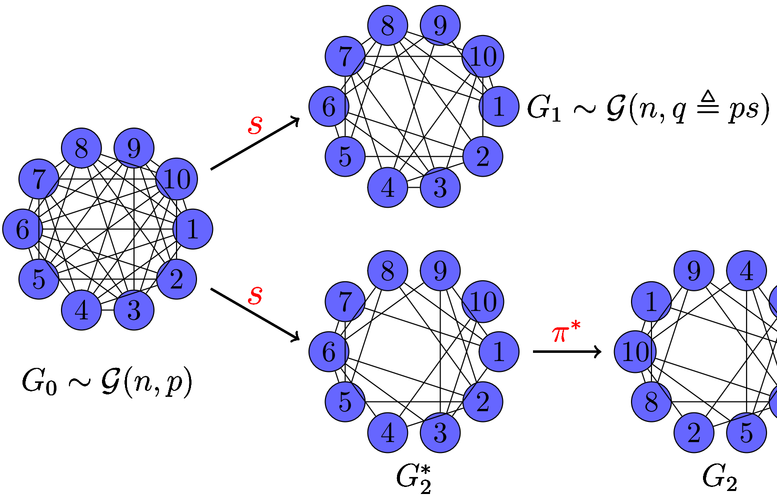
$G_1 \sim \mathcal{G}(n, q \triangleq ps)$

$G_0 \sim \mathcal{G}(n, p)$

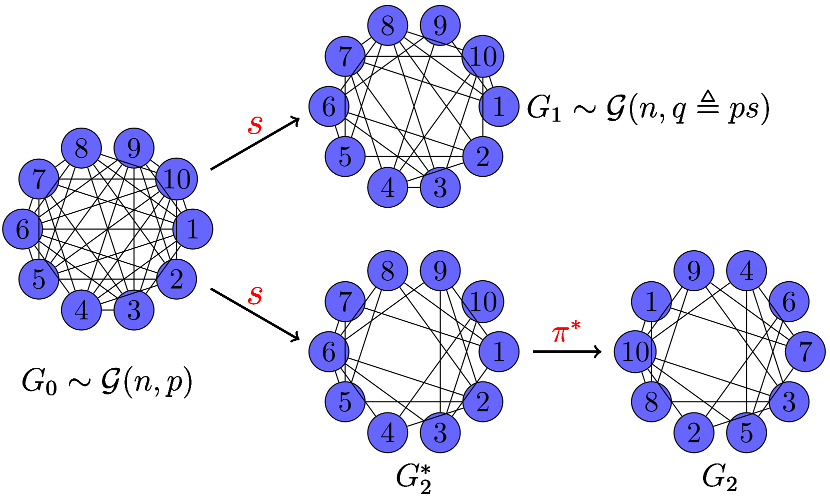
Correlated Erdős-Rényi graphs model



Correlated Erdős-Rényi graphs model



Correlated Erdős-Rényi graphs model



G_1 and G_2 differ by a fraction $\delta \triangleq 1 - s$ of edges, under the correct node mapping

Prior results: Information-theoretic limits

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

Theorem (Cullina-Kiyavash '18)

For $q < 1/2$, exact recovery of π^* is information-theoretically possible if and only if

$$nqs - \log n \to +\infty$$

Interpretation: Intersection graph $G_1 \wedge G_2^* \sim \mathcal{G}(n, qs)$ is connected

Prior results: Information-theoretic limits

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

Theorem (Cullina-Kiyavash '18)

For $q < 1/2$, exact recovery of π^* is information-theoretically possible if and only if

$$nqs - \log n \to +\infty$$

Interpretation: Intersection graph $G_1 \wedge G_2^* \sim \mathcal{G}(n, qs)$ is connected

Computationally:

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

Main result

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

Theorem (Fan-Mao-Wu-X. '19)

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

$$nq \gtrsim (\log n)^{48+\epsilon} \quad \textit{and} \quad \delta \lesssim (\log n)^{-(8+\epsilon)}$$

Main result

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

Theorem (Fan-Mao-Wu-X. '19)

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

$$nq \gtrsim (\log n)^{48+\epsilon} \quad \textit{and} \quad \delta \lesssim (\log n)^{-(8+\epsilon)}$$

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$

Main result

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

Theorem (Fan-Mao-Wu-X. '19)

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

$$nq \gtrsim (\log n)^{48+\epsilon} \quad \textit{and} \quad \delta \lesssim (\log n)^{-(8+\epsilon)}$$

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Previous spectral methods require $\delta \leq n^{-\Omega(1)}$

Main result

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

Theorem (Fan-Mao-Wu-X. '19)

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

$$nq \gtrsim (\log n)^{48+\epsilon} \quad \textit{and} \quad \delta \lesssim (\log n)^{-(8+\epsilon)}$$

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Previous spectral methods require $\delta \leq n^{-\Omega(1)}$
- Match the best known guarantee for polynomial-time algorithms [Ding-Ma-Wu-X.'18] (using degree profile)

Main result

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

Theorem (Fan-Mao-Wu-X. '19)

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

$$nq \gtrsim (\log n)^{48+\epsilon} \quad \textit{and} \quad \delta \lesssim (\log n)^{-(8+\epsilon)}$$

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Previous spectral methods require $\delta \leq n^{-\Omega(1)}$
- Match the best known guarantee for polynomial-time algorithms [Ding-Ma-Wu-X.'18] (using degree profile)
- $n^{\log n}$ time algorithm [Barak-Chou-Lei-Schramm-Sheng '18] for $nq \geq n^{\epsilon}$ and constant δ (using rare small subgraphs)

Main result

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

Theorem (Fan-Mao-Wu-X. '19)

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

$$nq \gtrsim (\log n)^{48+\epsilon} \qquad \textit{and} \qquad \delta \lesssim (\log n)^{-(8+\epsilon)}$$

- Dense graphs $q = \Theta(1)$: improvement to $\delta \lesssim (\log n)^{-(4+\epsilon)}$
- Previous spectral methods require $\delta \leq n^{-\Omega(1)}$
- Match the best known guarantee for polynomial-time algorithms [Ding-Ma-Wu-X.'18] (using degree profile)
- $n^{\log n}$ time algorithm [Barak-Chou-Lei-Schramm-Sheng '18] for $nq \geq n^{\epsilon}$ and constant δ (using rare small subgraphs)
- Polynomial-time recovery for constant δ is open

Outline

1. A new spectral algorithm
2. Analysis of our algorithm
3. Experimental results
4. 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

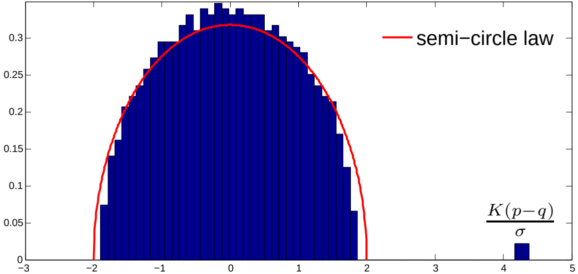
- Planted clique [Alon-Krivelevich-Sudakov '98]
- Planted partition/Stochastic block model [Mcsherry '98] [Massoulié '13] [Bordenave-Lelarge-Massoulié '15]
- Clustering [von-Luxburg-Bousquet-Belkin '05]
- Graphon estimation [Chatterjee '15]
- Matrix completion [Keshavan-Montanari-Oh '09]
- Ranking [Negahban-Oh-Shah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

Common rationale: A is approximately low-rank with large eigen-gap

Analyzing spectral methods: an example

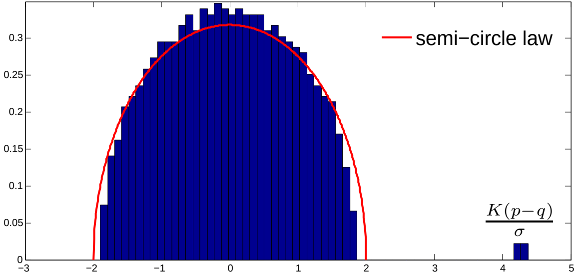
Planted partition: $A = \begin{bmatrix} p & & q \\ & p & \\ q & & p \end{bmatrix} + A - \mathbb{E}[A]$

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



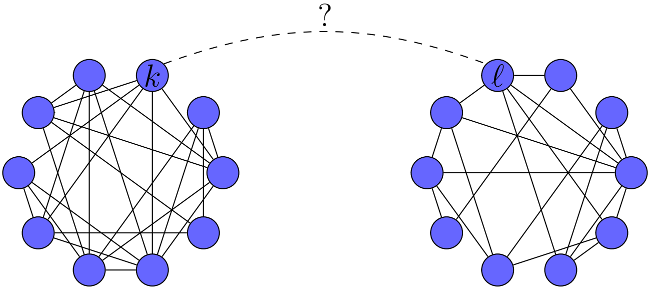
Analyzing spectral methods: an example

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



- Davis-Kahan and variants: Leading eigenvectors of $A \approx$ those of $\mathbb{E}[A]$, if eigen-gap $\gtrsim \|A - \mathbb{E}[A]\|_2$
- However, adjacency matrix of Erdős-Rényi graph has full rank and vanishing eigen-gaps

Spectral graph matching paradigm



$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^\top$$

$$\lambda_1 \geq \cdots \geq \lambda_n$$

$$B = \sum_{j=1}^{n} \mu_j v_j v_j^\top$$

$$\mu_1 \geq \cdots \geq \mu_n$$

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

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

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

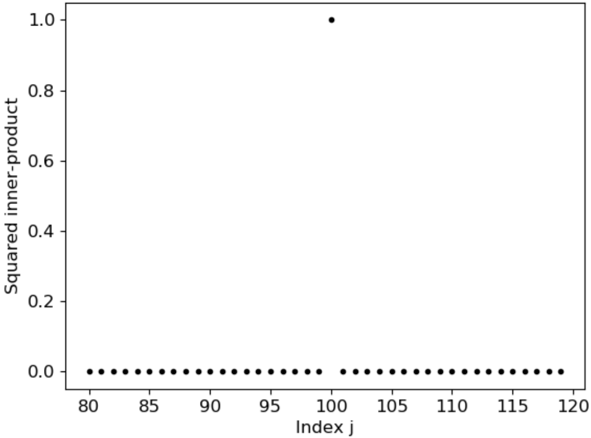
- Full-rank methods: [Umeyama '88]

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

- All perform well with no noise, but are extremely fragile with noise
- A and B have full rank and vanishing eigen-gaps $\Rightarrow$ decorrelation of u_i and v_i

Eigenvector correlation decay

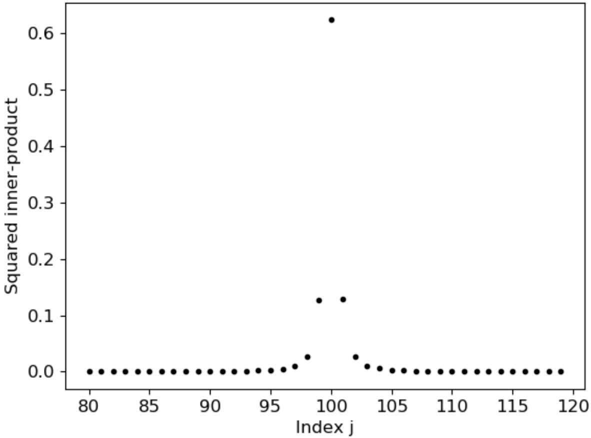
Isomorphic Erdős-Rényi graphs: 500 vertices, edge probability $\frac{1}{2}$



$\langle u_{100}, v_j \rangle^2$ for $j \in \{80, \ldots, 120\}$, averaged across 1000 simulations

Eigenvector correlation decay

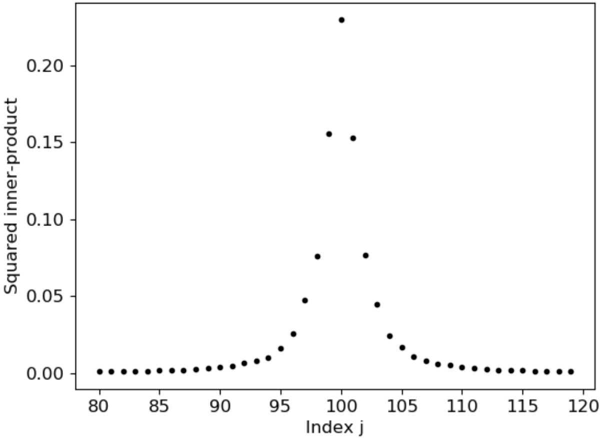
Erdős-Rényi graphs with $\delta = 0.1\%$ differed edges



$\langle u_{100}, v_j \rangle^2$ for $j \in \{80, \ldots, 120\}$, averaged across 1000 simulations

Eigenvector correlation decay

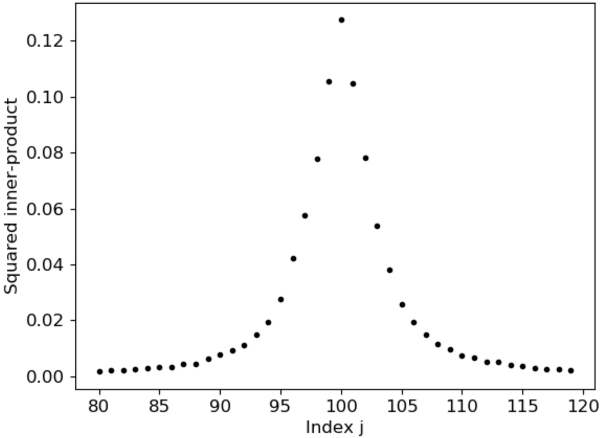
Erdős-Rényi graphs with $\delta = 0.5\%$ differed edges



$\langle u_{100}, v_j \rangle^2$ for $j \in \{80, \dots, 120\}$, averaged across 1000 simulations

Eigenvector correlation decay

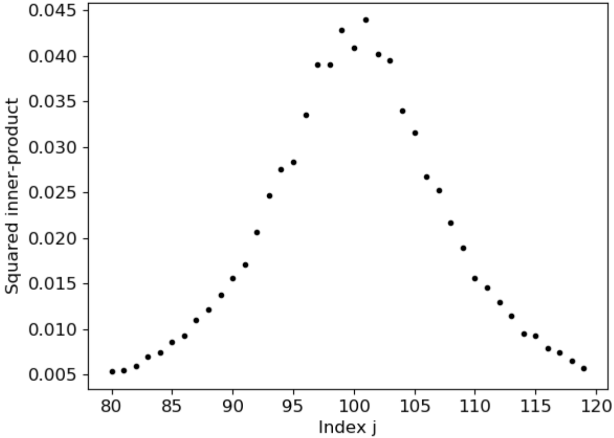
Erdős-Rényi graphs with $\delta = 1\%$ differed edges



$\langle u_{100}, v_j \rangle^2$ for $j \in \{80, \ldots, 120\}$, averaged across 1000 simulations

Eigenvector correlation decay

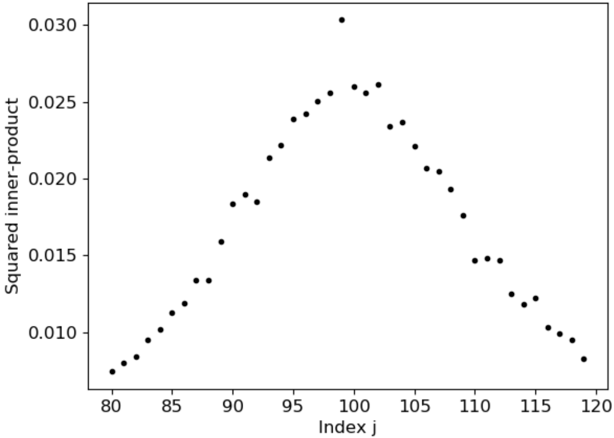
Erdős-Rényi graphs with $\delta = 3\%$ differed edges



$\langle u_{100}, v_j \rangle^2$ for $j \in \{80, \ldots, 120\}$, averaged across 1000 simulations

Eigenvector correlation decay

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



$\langle u_{100}, v_j \rangle^2$ for $j \in \{80, \ldots, 120\}$, averaged across 1000 simulations

A new spectral method: GRAMPA

GRAph Matching by Pairwise eigen-Alignments:

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

Here η is a regularization parameter, and $\mathbf{J}$ the all-one matrix

A new spectral method: GRAMPA

GRAph Matching by Pairwise eigen-Alignments:

$$\boxed{X = \sum_{i,j=1}^{n} \underbrace{\frac{\eta}{(\lambda_i - \mu_j)^2 + \eta^2}}_{\text{Cauchy kernel applied to } \lambda_i \text{ and } \mu_j} \times \underbrace{u_i u_i^\top \mathbf{J} v_j v_j^\top}_{\text{``Alignment'' between } u_i \text{ and } v_j}}$$

Here η is a regularization parameter, and $\mathbf{J}$ the all-one matrix

- All pairs matter: Cauchy weight kernel is inspired by the eigenvector correlation decay [Bourgade-Yau '17], [Benigni '17]:

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

- Unlike previous spectral methods, GRAMPA is invariant to the choices of signs for u_i and v_j

GRAMPA as regularized quadratic programming relaxation

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

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

GRAMPA as regularized quadratic programming relaxation

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

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

- 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 quadratic programming relaxation

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

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

- 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

Outline

1. A new spectral algorithm
2. Analysis of our algorithm
   - Diagonal dominance in "population version"
   - Universality proof via resolvent representation and local laws
3. Experimental results
4. Concluding remarks

Diagonal dominance in population version

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

Consider the "popoulation version" of the regularized QP:

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

Diagonal dominance in population version

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

Consider the "popoulation version" of the regularized QP:

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

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

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

Diagonal dominance in population version

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

Consider the "popoulation version" of the regularized QP:

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

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

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

- $\overline{X}$ 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 "popoulation version" of the regularized QP:

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

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

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

- $\overline{X}$ is close to $\frac{\mathbf{J}}{n}$ (center of the Birkhoff polytope)
- Same analysis holds for tighter QP-DS, suggesting X is not a permutation (shown by [Lyzinski-Fishkind-Fiori-Vogelstein-Priebe-Sapiro '15])

Diagonal dominance in population version

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

Consider the "popoulation version" of the regularized QP:

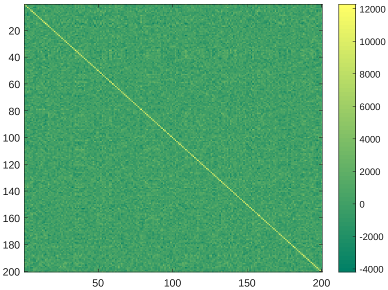
$$\overline{X} = \arg \min_{X:\, \mathbf{1}^\top X \mathbf{1} = n} \mathbb{E}\left[\|AX - XB\|_F^2\right] + \eta^2 \|X\|_F^2$$

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

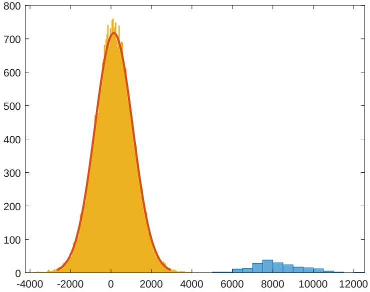
$$\overline{X} = \epsilon \mathbf{I} + (1-\epsilon)\frac{\mathbf{J}}{n}, \qquad \epsilon \approx \frac{2(1-\delta)}{n(2\delta + \eta^2)}$$

- $\overline{X}$ is close to $\frac{\mathbf{J}}{n}$ (center of the Birkhoff polytope)
- Same analysis holds for tighter QP-DS, suggesting X is not a permutation (shown by [Lyzinski-Fishkind-Fiori-Vogelstein-Priebe-Sapiro '15])
- $\overline{X}$ is diagonally dominant: diagonals are $\frac{2(1-\delta)}{2\delta+\eta^2}$ times off-diagonals

Diagonal dominance of the GRAMPA similarity matrix

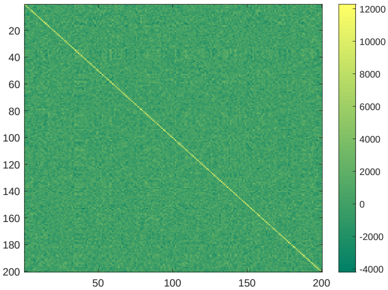


Heatmap of X

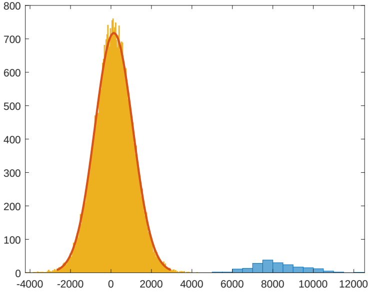


Histogram of diagonal (orange) and off-diagonal (blue) entries

Diagonal dominance of the GRAMPA similarity matrix



Heatmap of X



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

Goal: Show diagonally dominant whp:

$$\underbrace{\min_i X_{ii}}_{\text{similarity for true pairs}} > \underbrace{\max_{i \neq k} X_{ik}}_{\text{similarity for fake pairs}}$$

Universality: Correlated Wigner model

Standardized weighted adjacency matrices A, B where (A_{ij}, B_{ij}) are independent 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

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

Key proof technique: Resolvent and local laws

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

Denote Wigner's semicircle density and its Stieltjes transform by

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

Key proof technique: Resolvent and local laws

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

Denote Wigner's semicircle density and its Stieltjes transform by

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

- Empirical eigenvalue density of A converges to ρ

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Tr} R_A(z) = m(z)$$

Key proof technique: Resolvent and local laws

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

Denote Wigner's semicircle density and its Stieltjes transform by

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

- Empirical eigenvalue density of A converges to ρ

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Tr} R_A(z) = m(z)$$

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

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

Key proof technique: Resolvent and local laws

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

Denote Wigner's semicircle density and its Stieltjes transform by

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

- Empirical eigenvalue density of A converges to ρ

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Tr} R_A(z) = m(z)$$

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

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

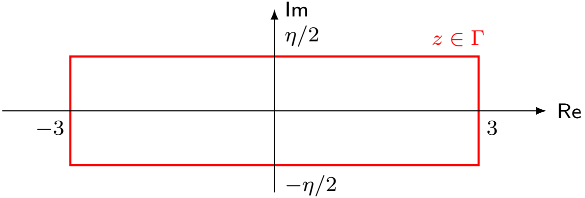
- Using similar techniques, we prove row-sum and total sum bounds:

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

Universality proof step 1: Resolvent representation

Lemma (Fan-Mao-Wu-X. '19)

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



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

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

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

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

Step 2: Leave-one-out relation

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

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

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

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

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

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

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

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

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

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

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

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

Step 3: Separating signal from noise

Diagonal entries: Apply concentration of the bilinear form

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

Off-diagonal entries:

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

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

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

Off-diagonal entries:

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

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

Step 4: Proof of diagonal dominance

- Diagonal entries:

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

- Off-diagonal entries

$$
\begin{aligned}
X_{12} &\lesssim \frac{(\log n)^{2+\varepsilon}}{n} \left\| \oint_\Gamma m(z) m(z+\mathbf{i}\eta) R_{A^{(12)}}(z) \mathbf{J} R_{B^{(12)}}(z+\mathbf{i}\eta) dz \right\|_F \\
&\lesssim \frac{(\log n)^{2+\varepsilon}}{\sqrt{\eta}}
\end{aligned}
$$

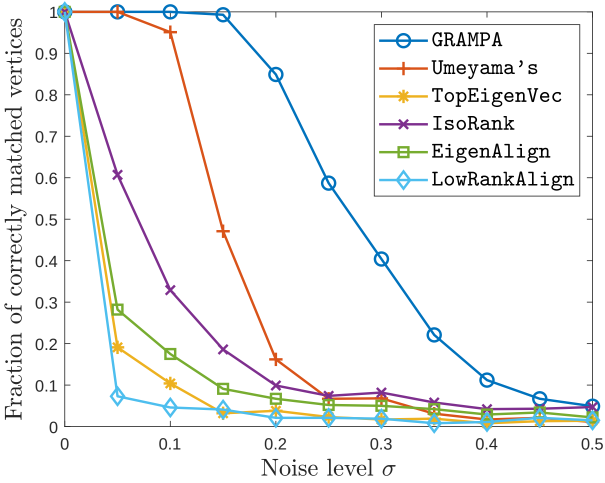
- Applying this and a union bound for every $X_{k\ell}$ shows that X is diagonally dominant when

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

Outline

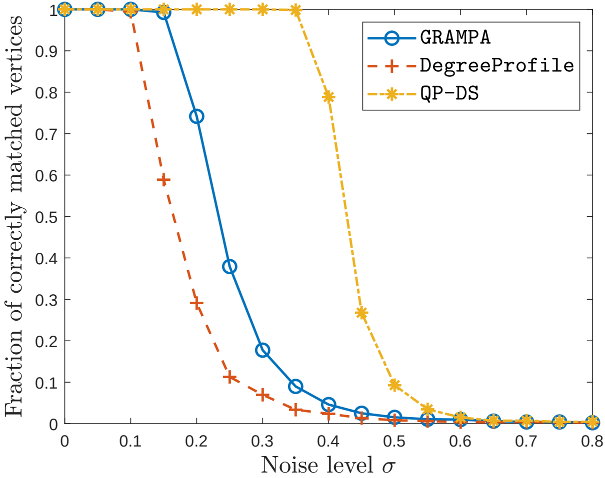
1. A new spectral algorithm
2. Analysis of our algorithm
3. Experimental results
4. Concluding remarks

Spectral algorithms on Erdős-Rényi graphs



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

Competitive methods on Erdős-Rényi graphs



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

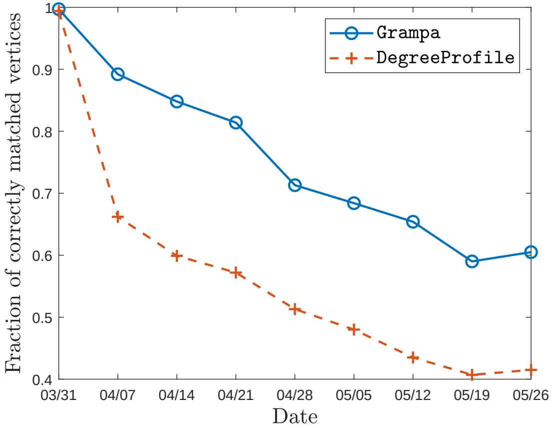
GRAMPA is $\sim$ 100–1000x faster and scalable to larger networks

Real network: Autonomous systems network

- A network of autonomous systems observed on 9 days between March 2001 and May 2001 (10K nodes, 22K-23K edges)
- Edges are added and deleted over time
- Goal: match 9 networks on 9 days to the network on day 1

Real network: Autonomous systems network

- A network of autonomous systems observed on 9 days between March 2001 and May 2001 (10K nodes, 22K-23K edges)
- Edges are added and deleted over time
- Goal: match 9 networks on 9 days to the network on day 1



Outline

1. A new spectral algorithm
2. Analysis of our algorithm
3. Experimental results
4. Concluding remarks

Concluding remarks

- Develop a new spectral graph matching algorithm

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

- Efficiently matches two graphs with average degree $\geq \mathrm{polylog}(n)$ and fraction of differred edges $\leq 1/\mathrm{polylog}(n)$
- Universality proof using resolvent representation and local laws
- Also establish a similar result for a tighter QP relaxation

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

Concluding remarks

- Develop a new spectral graph matching algorithm

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

- Efficiently matches two graphs with average degree $\geq \mathrm{polylog}(n)$ and fraction of differred edges $\leq 1/\mathrm{polylog}(n)$
- Universality proof using resolvent representation and local laws
- Also establish a similar result for a tighter QP relaxation

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

References

- Z. Fan, C. Mao, Y. Wu, J. X. *Spectral graph matching and regularized quadratic relaxations I: Algorithm and Gaussian analysis*, arxiv:1907.08880.
- Z. Fan, C. Mao, Y. Wu, J. X. *Spectral graph matching and regularized quadratic relaxations II: Erdős-Rényi graphs and universality*, arxiv:1907.08883.