Reading group meeting – February 16, 2023

Truong Son Hy

Halıcıoğlu Data Science Institute
University of California San Diego

Quote for today

The happiness of your life depends upon the quality of your thoughts.

– Marcus Aurelius

The Art Institute of Chicago, Februray 2020

Content

Paper:

- Zhou Fan, Cheng Mao, Yihong Wu, Jiaming Xu, **Spectral Graph Matching and Regularized Quadratic Relaxations: Algorithm and Theory** (ICML 2020).

  `http://proceedings.mlr.press/v119/fan20a.html`

Introduction (1)

Finding the best matching between two **weighted** graphs with adjacency matrices $A, B \in \mathbb{R}^{n \times n}$ may be formalized as the following **NP-hard** combinatorial optimization, quadratic assignment problem (QAP), problem over the set of permutation $\mathcal{S}_n$:

$$\pi_* = \max_{\pi \in \mathcal{S}_n} \sum_{i,j=1}^{n} A_{i,j} B_{\pi(i),\pi(j)}.$$

Introduction (1)

Finding the best matching between two **weighted** graphs with adjacency matrices $A, B \in \mathbb{R}^{n\times n}$ may be formalized as the following **NP-hard** combinatorial optimization, quadratic assignment problem (QAP), problem over the set of permutation $\mathcal{S}_n$:

$$\pi_* = \max_{\pi\in\mathcal{S}_n} \sum_{i,j=1}^{n} A_{i,j} B_{\pi(i),\pi(j)}.$$

Special case: Random weighted graph matching

Assumptions:

1. A, B symmetric. We aim to recover π_* from A and B.
2. Suppose that $\{(A_{ij}, B_{\pi_*(i),\pi_*(j)}) : 1 \le i < j \le n\}$ are independent pairs of positively correlated random variables, with correlation at least $1-\sigma^2$ where $\sigma \in [0,1]$.

Introduction (2)

Note

- This problem is related to other branches of mathematics: **information theory** (1-sample test), **optimal transport** (distribution matching), and **group theory** (for the general graph isomorphism).
- (Babai et al., 1982) has applied spectral methods in testing graph isomorphism.

Introduction (2)

Note

- This problem is related to other branches of mathematics: **information theory** (1-sample test), **optimal transport** (distribution matching), and **group theory** (for the general graph isomorphism).
- (Babai et al., 1982) has applied spectral methods in testing graph isomorphism.

Notable special cases of (special case) random weighted graph matching:

1. Erdos-Renyi graph model: $\{(A_{ij}, B_{\pi_*(i),\pi_*(j)})$ is a pair of standardized correlated Bernoulli random variables. This is discrete-edge model.
2. Gaussian Wigner model: $\{(A_{ij}, B_{\pi_*(i),\pi_*(j)})$ is a pair of correlated Gaussian variables (i.e. A and B are complete graphs with correlated Gaussian edge weights). This is continuous-edge model.

Spectral Methods (1)

Write the spectral decompositions of the **weighted** (complete graph) adjacency matrices A and B as

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

where the eigenvalues are ordered such that $\lambda_1 \geq ... \geq \lambda_n$ and $\mu_1 \geq ... \geq \mu_n$.

Spectral Methods (1)

Write the spectral decompositions of the **weighted** (complete graph) adjacency matrices A and B as

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

where the eigenvalues are ordered such that $\lambda_1 \geq \ldots \geq \lambda_n$ and $\mu_1 \geq \ldots \geq \mu_n$.

Theorem (Informal statement). *For the random weighted graph matching problem, if the two graphs have edge correlation at least* $1 - 1/\mathrm{polylog}(n)$ *and average degree at least* $\mathrm{polylog}(n)$, *then a spectral algorithm recovers the latent matching* π_* *exactly with high probability.*

Algorithm 1 GRAph Matching by Pairwise eigen-Alignments (GRAMPA)

1: **Input:** Weighted adjacency matrices A and B on n vertices, and a bandwidth parameter $\eta > 0$.
2: **Output:** A permutation $\widehat{\pi} \in \mathcal{S}_n$.
3: Construct the similarity matrix

$$\widehat{X} = \sum_{i,j=1}^{n} w(\lambda_i, \mu_j) \cdot u_i u_i^\top \mathbf{J} v_j v_j^\top \in \mathbb{R}^{n \times n} \quad (3)$$

where $\mathbf{J} \in \mathbb{R}^{n \times n}$ denotes the all-ones matrix and w is the Cauchy kernel of bandwidth η:

$$w(\lambda, \mu) = \frac{1}{(\lambda - \mu)^2 + \eta^2}. \quad (4)$$

4: Output the permutation estimate $\widehat{\pi}$ by "rounding" $\widehat{X}$ to a permutation, e.g., by solving the *linear assignment problem* (LAP)

$$\widehat{\pi} = \operatorname*{argmax}_{\pi \in \mathcal{S}_n} \sum_{i=1}^{n} \widehat{X}_{i,\pi(i)}. \quad (5)$$

Spectral Methods (3)

This is **not** the first spectral method. What others have done before:

- **Low-rank methods** that use a small number of eigenvectors of A and B. For example, only the leading eigenvector:

  $$\hat{X} = u_1 v_1^T$$

  (Kazemi & Grossglauser, 2016; Feizi et al., 2019).

Spectral Methods (3)

This is **not** the first spectral method. What others have done before:

- **Low-rank methods** that use a small number of eigenvectors of A and B. For example, only the leading eigenvector:

$$\hat{X} = u_1 v_1^T$$

  (Kazemi & Grossglauser, 2016; Feizi et al., 2019).

- **Full-rank methods** that use all eigenvectors of A and B:

$$\hat{X} = \sum_{i=1}^{n} s_i u_i v_i^T$$

  for some appropriately chosen signs $s_i \in \{\pm 1\}$ (Xu & King, 2001) (Finke et al., 1987) (Umeyama, 1988). Furthermore, (Umeyama, 1988) suggests

$$\hat{X} = \sum_{i=1}^{n} |u_i| |v_i|^T$$

  where $|u_i|$ denotes the entrywise absolute value of u_i.

Spectral Methods (4)

About this work **graph matching by pairwise eigen-alignments**:

- Computational complexity $O(n^3)$.

Spectral Methods (4)

About this work **graph matching by pairwise eigen-alignments**:

- Computational complexity $O(n^3)$.
- Just a new similarity matrix

$$\hat{X} = \sum_{i,j=1}^{n} w(\lambda_i, \mu_j) \cdot u_i u_i^T \boldsymbol{J} v_j v_j^T,$$

where $\boldsymbol{J}$ denotes the all-ones matrix and w is the Cauchy kernel of bandwidth η:

$$w(\lambda, \mu) = \frac{1}{(\lambda - \mu)^2 + \eta^2}.$$

Spectral Methods (4)

About this work **graph matching by pairwise eigen-alignments**:

- Computational complexity $O(n^3)$.
- Just a new similarity matrix

$$\hat{X} = \sum_{i,j=1}^{n} w(\lambda_i, \mu_j) \cdot u_i u_i^T \boldsymbol{J} v_j v_j^T,$$

where $\boldsymbol{J}$ denotes the all-ones matrix and w is the Cauchy kernel of bandwidth η:

$$w(\lambda, \mu) = \frac{1}{(\lambda - \mu)^2 + \eta^2}.$$

- **Insensitive** to the choices of signs for individual eigenvectors.

Spectral Methods (4)

About this work **graph matching by pairwise eigen-alignments**:

- Computational complexity $O(n^3)$.
- Just a new similarity matrix

$$\hat{X} = \sum_{i,j=1}^{n} w(\lambda_i, \mu_j) \cdot u_i u_i^T \boldsymbol{J} v_j v_j^T ,$$

where $\boldsymbol{J}$ denotes the all-ones matrix and w is the Cauchy kernel of bandwidth η:

$$w(\lambda, \mu) = \frac{1}{(\lambda - \mu)^2 + \eta^2}.$$

- **Insensitive** to the choices of signs for individual eigenvectors.
- Trivially, the output $\hat{\pi}(A, B)$ is **equivariant**.

A Fourier space algorithm for solving QAPs (1)

Another "spectral" way to look at QAPs

Risi Kondor, *A Fourier space algorithm for solving quadratic assignment problems*, SODA 2010

http://people.cs.uchicago.edu/~risi/papers/KondorSODA10.pdf

The Fourier transform of a general function $f : \mathbb{S}_n \to \mathbb{C}$ is the collection of matrices

$$\hat{f}(\lambda) = \sum_{\sigma \in \mathbb{S}_n} f(\sigma)\rho_\lambda(\sigma)$$

where λ extends over the integer partitions of n, and $\rho_\lambda : \mathbb{S}_n \to \mathbb{C}^{d_\lambda \times d_\lambda}$ is the corresponding irreducible represetation (irrep) of $\mathbb{S}_n$ (given in **Young's Orthogonal Representation** – YOR). The inverse transform:

$$f(\sigma) = \frac{1}{n!} \sum_{\lambda \vdash n} d_\lambda \mathrm{trace}[\rho_\lambda(\sigma)^{-1} \hat{f}(\lambda)].$$

A Fourier space algorithm for solving QAPs (2)

QAP:

$$\hat{\sigma} = \arg\max_{\sigma \in \mathbb{S}_n} \sum_{i,j=1}^{n} A_{\sigma(i),\sigma(j)} B_{i,j}$$

The objective function – graph correlation:

$$f(\sigma) = \sum_{i,j=1}^{n} A_{\sigma(i),\sigma(j)} B_{i,j}$$

The objective function can be expressed as:

$$f(\sigma) = \frac{1}{(n-2)!} \sum_{\pi \in \mathbb{S}_n} f_A(\sigma\pi) f_B(\pi)$$

where $f_A : \mathbb{S}_n \to \mathbb{R}$ is defined as:

$$f_A(\sigma) = A_{\sigma(n),\sigma(n-1)},$$

and similarly for f_B.

A Fourier space algorithm for solving QAPs (3)

Given a pair of graphs A and B of n vertices with graph Fourier transforms (that is **different** from GFT in graph literature) $\hat{f}_A$ and $\hat{f}_B$, the Fourier transform of their graph correltation is:

$$\hat{f}(\lambda) = \frac{1}{(n-2)!} \hat{f}_A(\lambda) \cdot (\hat{f}_B(\lambda))^T, \qquad \lambda \vdash n.$$

For any function $f : \mathbb{S}_n \to \mathbb{R}$:

$$\max_{\sigma \in \mathbb{S}_n} f(\sigma) \leq \frac{1}{n!} \sum_{\lambda \vdash n} d_\lambda ||\hat{f}(\lambda)||_1$$

where $||M||_1$ denotes the trace norm of the matrix M.

A Fourier space algorithm for solving QAPs (4)

Key result – Upper bound:

$$\max_{\tau \in \mathbb{S}_k} f_{i_n, i_{n-1}, .., i_{k+1}}(\tau) \leq \mathcal{B}(\hat{f}_{i_n, i_{n-1}, .., i_{k+1}})$$

where $f_{i_n, i_{n-1}, .., i_{k+1}} : \mathbb{S}_k \to \mathbb{R}$ is defined as

$$f_{i_n, i_{n-1}, .., i_{k+1}}(\tau) = f([i_n, n][i_{n-1}, n-1]...[i_{k+1}, k+1]\tau)$$

where $[i, j]$ denotes the contiguous cycle.

Efficient computation – Branch & Bound searching algorithm

Each upper bound

$$\mathcal{B}(\hat{f}^k_I) = \frac{1}{n!} \sum_{\lambda \vdash k} d_\lambda ||\hat{f}^k_I(\lambda)||_1$$

can be computed in $O(k^3)$ time.

Regularized Quadratic Programming (1)

Claim

The similarity matrix $\hat{X}$ corresponds to the solution to a convex relaxation of the QAP, regularized by an added ridge penalty.

Finding the best matching between two graphs with adjs $A, B \in \mathbb{R}^{n \times n}$:

$$\max_{\pi \in \mathcal{S}_n} \sum_{i,j=1}^{n} A_{i,j} B_{\pi(i),\pi(j)},$$

that is equivalent to:

$$\max_{\Pi \in \mathcal{S}_n} \langle A, \Pi B \Pi^T \rangle \Leftrightarrow \min_{\Pi \in \mathcal{S}_n} ||A\Pi - \Pi B||_F^2.$$

Regularized Quadratic Programming (2)

Relaxing the set of permutations to its convex hull (the Birkhoff polytope of doubly stochastic matrices)

$$\mathcal{B}_n \triangleq \{X \in \mathbb{R}^{n\times n} : X\mathbf{1} = \mathbf{1}, X^T\mathbf{1} = \mathbf{1}, X_{ij} \geq 0 \forall i,j\}$$

arrives at the QAP relaxation:

$$\min_{X\in\mathcal{B}_n} ||AX - XB||_F^2.$$

This is called **doubly stochastic QP**.

Regularized Quadratic Programming (3)

The similarity matrix $\hat{X}$ is the solution of a **regularized** futher relaxation of the doubly stochastic QP:

- $\hat{X}$ is the minimizer of of

$$\min_{X \in \mathbb{R}^{n \times n}} \frac{1}{2}||AX - XB||_F^2 + \frac{\eta^2}{2}||X||_F^2 - \mathbf{1}^T X \mathbf{1}.$$

Regularized Quadratic Programming (3)

The similarity matrix $\hat{X}$ is the solution of a **regularized** futher relaxation of the doubly stochastic QP:

- $\hat{X}$ is the minimizer of of

$$\min_{X \in \mathbb{R}^{n \times n}} \frac{1}{2}||AX - XB||_F^2 + \frac{\eta^2}{2}||X||_F^2 - \mathbf{1}^T X \mathbf{1}.$$

- Equivalently, $\hat{X}$ is a positive scalar multiple of the solution $\tilde{X}$ to the constrained program

$$\min_{X \in \mathbb{R}^{n \times n}} ||AX - XB||_F^2 + \eta^2 ||X||_F^2$$

$$s.t. \mathbf{1}^T X \mathbf{1} = n$$

$\hat{X}$ and $\tilde{X}$ are equivalent as far as the rounding step by the Hungarian matching is concerned.

The similarity matrix is diagonal dominant (1)

Let consider the Gaussian Wigner model:

$$B = A + \sigma Z,$$

where A and Z are independent Gaussian Orthogonal Ensemble (GOE) matrices with $\mathcal{N}(0, \frac{1}{n})$ off-diagonal and $\mathcal{N}(0, \frac{2}{n})$ diagonal. The permutation solution is indeed $\pi_* =$ the identity matrix.

Note: I think all the Wigner models mentioned in this paper do not reflect any realistic examples.

The similarity matrix is diagonal dominant (2)

The population version of the doubly stochastic quadratic programming is:

$$\min_{X \in \mathcal{B}_n} \mathbb{E}\left\{ ||AX - XB||_F^2 \right\}$$

$$\Leftrightarrow \min_{X \in \mathcal{B}_n} (2 + \sigma^2)(n+1)||X||_F^2 - 2\text{trace}(X)^2 - 2\langle X, X^T \rangle$$

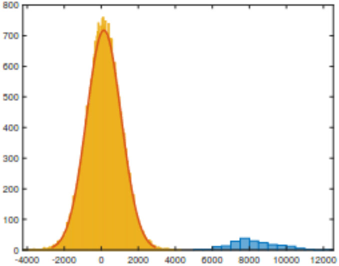
has solution

$$\bar{X} \triangleq \epsilon \boldsymbol{I} + (1 - \epsilon)\boldsymbol{F}, \quad \epsilon = \frac{2}{2 + (n+1)\sigma^2} \approx \frac{2}{n\sigma^2}$$

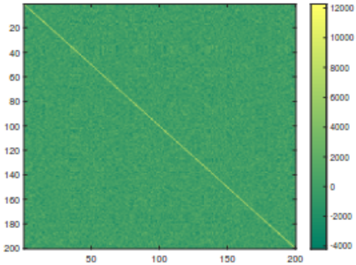
that is a convex combination of the true permutation matrix (identity) and the center of the Birkhoff polytope $\boldsymbol{F} = \frac{1}{n}\boldsymbol{J}$, a very **flat** matrix. The authors claim it is reasonable to expect that $\hat{X}$ inherits the diagonal dominance property from the population solution $\bar{X}$:

$$\hat{X}_{i,\pi_*(i)} > \hat{X}_{i,j}, \quad j \neq \pi_*(i).$$

The similarity matrix is diagonal dominant (3)



(a) Histogram of diagonal (blue) and off-diagonal (yellow with a normal fit) entries of $\widehat{X}$.



(b) Heat map of $\widehat{X}$.

Note: If the similarity matrix is diagonal dominant, the task of the Hungarian matching (rounding step) is trivial.

Correlated Wigner Model (1)

To model a general random weighted graph, we consider the following Wigner model: Let $A = (A_{ij})$ be a symmetric random matrix in $\mathbb{R}^{n \times n}$, where the entries $(A_{ij})_{i \le j}$ are independent. Suppose that

$$\mathbb{E}\left[A_{ij}\right] = 0,\ \mathbb{E}\left[A_{ij}^2\right] = 1/n \text{ for } i \neq j, \text{ and} \tag{13}$$

$$\mathbb{E}\left[\,|A_{ij}|^k\,\right] \le \frac{C^k}{nd^{(k-2)/2}} \text{ for all } i, j \text{ and } k \ge 2, \tag{14}$$

where $d \equiv d(n)$ is an n-dependent sparsity parameter and C is a positive constant.

Note: All the element-wise moments are bounded.

Correlated Wigner Model (2)

Definition 2.1 (Correlated Wigner model). *Let n be a positive integer, $\sigma \in [0,1]$ an (n-dependent) noise parameter, π_* a latent permutation on $[n]$, and $\Pi_* \in \{0,1\}^{n\times n}$ the corresponding permutation matrix such that $(\Pi_*)_{i\pi_*(i)} = 1$. Suppose that $\{(A_{ij}, B_{\pi_*(i)\pi_*(j)}) : i \le j\}$ are independent pairs of random variables such that both $A = (A_{ij})$ and $B = (B_{ij})$ satisfy (13) and (14),*

$$\mathbb{E}\left[A_{ij}B_{\pi_*(i)\pi_*(j)}\right] \ge \frac{1-\sigma^2}{n} \quad \text{for all } i \ne j, \qquad (15)$$

and for a constant $C > 0$, any $D > 0$, and all $n \ge n_0(D)$,

$$\mathbb{P}\left\{\left\|A - \Pi_* B \Pi_*^\top\right\| \le C\sigma\right\} \ge 1 - n^{-D} \qquad (16)$$

where $\|\cdot\|$ denotes the spectral norm.

Note: It is likely (high probability) that the optimal objective is small.

Correlated (sparse) Erdos-Renyi graphs (1)

Equivalently, we may first sample an Erdős-Rényi graph $\mathbf{A} \sim \mathsf{G}(n, p)$, and then define $\mathbf{B}'$ by

$$\mathbf{B}'_{ij} \sim \begin{cases} \mathrm{Bern}(s) & \text{if } \mathbf{A}_{ij} = 1 \\ \mathrm{Bern}\left(\frac{p(1-s)}{1-p}\right) & \text{if } \mathbf{A}_{ij} = 0. \end{cases}$$

Suppose that we observe a pair of graphs $\mathbf{A}$ and $\mathbf{B} = \Pi_*^\top \mathbf{B}' \Pi_*$, where Π_* is the latent permutation matrix. We then wish to recover Π_* or, equivalently, the corresponding permutation π_*.

We first normalize the adjacency matrices $\mathbf{A}$ and $\mathbf{B}$ so that they satisfy the moment conditions (13) and (14). Define the centered, rescaled versions of $\mathbf{A}$ and $\mathbf{B}$ by

$$A \triangleq (np(1-p))^{-1/2}(\mathbf{A} - \mathbb{E}[\mathbf{A}])$$
$$\text{and} \quad B \triangleq (np(1-p))^{-1/2}(\mathbf{B} - \mathbb{E}[\mathbf{B}]). \qquad (19)$$

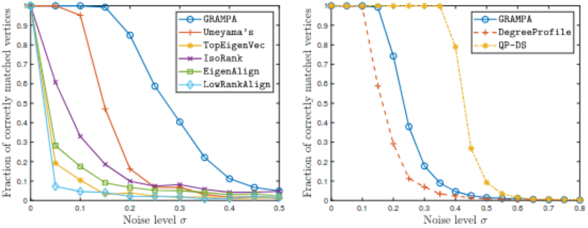
Correlated (sparse) Erdos-Renyi graphs (2)

Lemma 2.3. *For all large n, the matrices $A = (A_{ij})$ and $B = (B_{ij})$ satisfy conditions* (13), (14), (15), *and (16) with $d = np(1-p)$ and $\sigma^2 = \max\left(\frac{1-s}{1-p}, \frac{(\log n)^7}{d}\right)$.*

Note: This lemma literally says the two models correlated Wigner and correlated Erdos-Renyi are more-or-less the same. Fundamentally, the authors only provide the theoretical analysis of recovery for the case of Wigner model that is in Remark 2.5.

Remark 2.5. *From Theorem 2.2, we can obtain similar exact recovery guarantees for the correlated Gaussian Wigner model $B = \sqrt{1-\sigma^2}\Pi_*^\top A \Pi_* + \sigma Z$, where A and Z are independent GOE matrices and $\sigma \le (\log n)^{-4-\delta}$. In fact, GRAMPA recovers the latent permutation Π_* under a milder condition $\sigma \le c(\log n)^{-1}$ for a small constant $c > 0$. However, this refined result requires a dedicated analysis different from the proof of Theorem 2.2, so we defer it to a companion work.*

Experiments (1)

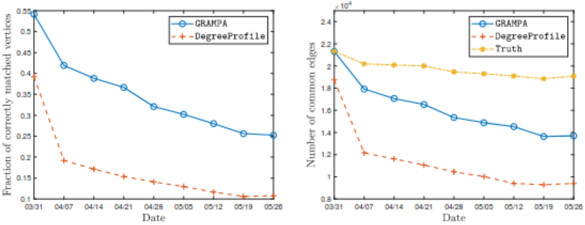


(a) Fraction of correctly matched vertices, on Erdős-Rényi graphs with 100 vertices

(b) Fraction of correctly matched vertices, on Erdős-Rényi graphs with 500 vertices

Figure 2: Comparison of GRAMPA to existing methods for matching Erdős-Rényi graphs with expected edge density 0.5. Each experiment is averaged over 10 repetitions.

Experiments (2)

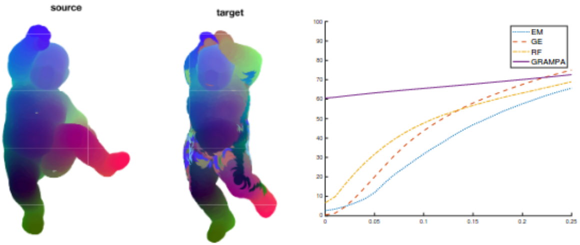


(a) Fraction of correctly matched vertices

(b) Number of common edges

Figure 3: Comparison of GRAMPA with DegreeProfile for matching networks of autonomous systems on nine days to that on the first day

Experiments (3)



(a) Visualization of the correspondence by GRAMPA

(b) Empirical distribution function of normalized geodesic error.

Figure 4: Comparison of GRAMPA to existing methods on SHREC'16 dataset.

Summary: I think the special cases, such as Wigner models and Erdos-Renyi, have been well-studied already.