
Spectral Graph Matching and Regularized Quadratic Relaxations: Algorithm and Theory

Zhou Fan¹ Cheng Mao² Yihong Wu¹ Jiaming Xu³

Abstract

Graph matching, also known as network alignment, aims at recovering the latent vertex correspondence between two unlabeled, edge-correlated weighted graphs. To tackle this task, we propose a spectral method, GRAPh Matching by Pairwise eigen-Alignments (GRAMPA), which first constructs a similarity matrix as a weighted sum of outer products between all pairs of eigenvectors of the two graphs, and then outputs a matching by a simple rounding procedure. For a universality class of correlated Wigner models, GRAMPA achieves exact recovery of the latent matching between two graphs with edge correlation $1 - 1/\text{polylog}(n)$ and average degree at least $\text{polylog}(n)$. This matches the state-of-the-art guarantees for polynomial-time algorithms established for correlated Erdős-Rényi graphs, and significantly improves over existing spectral methods. The superiority of GRAMPA is also demonstrated on a variety of synthetic and real datasets, in terms of both statistical accuracy and computational efficiency.

1. Introduction

Given a pair of graphs, the problem of *graph matching* or *network alignment* refers to finding a bijection between the vertex sets so that the edge sets are maximally aligned (Conte et al., 2004; Livi & Rizzi, 2013; Emmert-Streib et al., 2016). This is a ubiquitous problem arising in a variety of applications, including network de-anonymization (Narayanan & Shmatikov, 2008; 2009), pattern recognition (Conte et al., 2004; Schellewald & Schnörr, 2005), and

¹Department of Statistics and Data Science, Yale University, New Haven, Connecticut, USA ²School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia, USA ³The Fuqua School of Business, Duke University, Durham, North Carolina, USA. Correspondence to: Cheng Mao <cheng.mao@math.gatech.edu>.

computational biology (Singh et al., 2008; Kazemi et al., 2016). Finding the best matching between two graphs with adjacency matrices $A, B \in \mathbb{R}^{n \times n}$ may be formalized as the following combinatorial optimization problem over the set of permutations \mathcal{S}_n :

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

This is an instance of the notoriously difficult *quadratic assignment problem* (QAP) (Pardalos et al., 1994; Burkard et al., 1998), which is NP-hard to solve or to approximate within an approximation factor that even grows with n (Makarychev et al., 2010).

As the worst-case computational hardness of the QAP (1) may not be representative of typical graphs, various average-case models have been studied. For example, when the two graphs are isomorphic, the resulting *graph isomorphism* problem remains open to be solved in polynomial time for worst-case instances, but can be solved for Erdős-Rényi random graphs in linear time whenever information-theoretically possible (Bollobás, 1982; Czajka & Pandurangan, 2008). In “noisy” settings where the graphs are not exactly isomorphic, there is a recent surge of interest in computer science, information theory, and statistics for studying random graph matching (Yartseva & Grossglauser, 2013; Lyzinski et al., 2014; Kazemi et al., 2015; Cullina & Kiyavash, 2016; Shirani et al., 2017; Cullina & Kiyavash, 2017; Dai et al., 2018; Barak et al., 2018; Cullina et al., 2018; Ding et al., 2018; Feizi et al., 2019; Mossel & Xu, 2019).

1.1. Random Weighted Graph Matching

In this work, we study the following *random weighted graph matching* problem: Consider two weighted graphs with n vertices, and a latent permutation π_* on $[n] \triangleq \{1, \dots, n\}$ such that vertex i of the first graph corresponds to vertex $\pi_*(i)$ of the second. Denoting by A and B their (symmetric) weighted adjacency matrices, suppose that $\{(A_{ij}, B_{\pi_*(i),\pi_*(j)}) : 1 \leq i < j \leq n\}$ are independent pairs of positively correlated random variables, with correlation at least $1 - \sigma^2$ where $\sigma \in [0, 1]$. We aim to recover π_* from A and B .

Notable special cases of this model include the following:

- Erdős-Rényi graph model: $(A_{ij}, B_{\pi_*(i), \pi_*(j)})$ is a pair of (standardized) correlated Bernoulli random variables. Then A and B are Erdős-Rényi graphs with correlated edges. This model has been extensively studied in (Pedarsani & Grossglauser, 2011; Lyzinski et al., 2014; Korula & Lattanzi, 2014; Cullina & Kiyavash, 2016; Lubars & Srikant, 2018; Barak et al., 2018; Dai et al., 2018; Cullina et al., 2018; Ding et al., 2018).
- Gaussian Wigner model: $(A_{ij}, B_{\pi_*(i), \pi_*(j)})$ is a pair of correlated Gaussian variables. Then A and B are complete graphs with correlated Gaussian edge weights. This model was proposed in (Ding et al., 2018) as a prototype of random graph matching due to its simplicity, and certain results in the Gaussian model are expected to carry over to dense Erdős-Rényi graphs.

Spectral methods have a long history in testing graph isomorphism (Babai et al., 1982) and the graph matching problem (Umeyama, 1988). In this paper, we introduce a new spectral method for graph matching which enjoys the following exact recovery guarantee.

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

We describe the spectral method in Section 1.2 below. The method may also be interpreted as a regularized convex relaxation of the QAP program (1), and we discuss this connection in Section 1.3. The performance guarantee matches the state of the art of polynomial-time algorithms, namely, the degree profile method proposed in (Ding et al., 2018), and exponentially improves the performance of existing spectral matching algorithms, which require the edge correlation to be $1 - 1/\text{poly}(n)$ as opposed to $1 - 1/\text{polylog}(n)$.

1.2. A New Spectral Method

Write the spectral decompositions of the weighted adjacency matrices A and B as

$$A = \sum_{i=1}^n \lambda_i u_i u_i^\top \quad \text{and} \quad B = \sum_{j=1}^n \mu_j v_j v_j^\top \quad (2)$$

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

Our new spectral method is given in Algorithm 1, which we refer to as *graph matching by pairwise eigen-alignments* (GRAMPA). Therein, the linear assignment problem (5) can be solved efficiently using, e.g., the Hungarian algorithm (Kuhn, 1955). Our theoretical results in Section 2 apply equally to other potential rounding procedures. We discuss the choice of the bandwidth parameter η further in Section 3, and find in practice that the performance is not

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 $\hat{\pi} \in \mathcal{S}_n$.
- 3: Construct the similarity matrix

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

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 $\hat{\pi}$ by “rounding” \hat{X} to a permutation, e.g., by solving the *linear assignment problem* (LAP)

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

too sensitive to this choice. Moreover, note that the computational complexity of GRAMPA is $O(n^3)$ in practice, same as eigendecomposition and linear assignment.

Let us remark that Algorithm 1 exhibits the following two elementary but desirable properties:

- Unlike previous proposals, our spectral method is insensitive to the choices of signs for individual eigenvectors u_i and v_j in the spectral decomposition (2). More generally, it does not depend on the specific choice of eigenvectors if certain eigenvalues have multiplicity greater than one. This is because the similarity matrix (3) depends on the eigenvectors of A and B only through the projections onto their distinct eigenspaces.
- Let $\hat{\pi}(A, B)$ denote the output of Algorithm 1 with inputs A and B . For any fixed permutation π , denote by B^π the matrix with entries $B_{ij}^\pi = B_{\pi(i)\pi(j)}$, and by $\pi \circ \hat{\pi}$ the composition $(\pi \circ \hat{\pi})(i) = \pi(\hat{\pi}(i))$. Then we have the *equivariance* property $\pi \circ \hat{\pi}(A, B^\pi) = \hat{\pi}(A, B)$ and similarly for A^π . That is, the outputs given (A, B^π) and given (A, B) represent the same matching of the underlying graphs. This may be verified from (3) as a consequence of $\mathbf{J} = \mathbf{J}\Pi = \Pi\mathbf{J}$ for any permutation matrix Π .

To further motivate the construction (3), we note that Algorithm 1 follows the same general paradigm as several existing spectral methods for graph matching, which seek to recover π_* by rounding a similarity matrix \hat{X} constructed

to leverage correlations between eigenvectors of A and B . These methods include:

- *Low-rank methods* that use a small number of eigenvectors of A and B . The simplest such approach uses only the leading eigenvectors, taking as the similarity matrix

$$\widehat{X} = u_1 v_1^\top. \quad (6)$$

Then $\widehat{\pi}$ which solves (5) sorts the entries of v_1 in the order of u_1 . Other rank-1 spectral methods and low-rank generalizations have also been proposed and studied in (Kazemi & Grossglauser, 2016; Feizi et al., 2019).

- *Full-rank methods* that use all eigenvectors of A and B . A notable example is the popular method of Umeyama (Umeyama, 1988), which sets

$$\widehat{X} = \sum_{i=1}^n s_i u_i v_i^\top \quad (7)$$

for some appropriately chosen signs $s_i \in \{\pm 1\}$; see also the related approach of (Xu & King, 2001). The motivation is that (7) is the solution to the *orthogonal relaxation* of the QAP (1), where the feasible set is relaxed to the set the orthogonal matrices (Finke et al., 1987). As the correct choice of signs in (7) may be difficult to determine in practice, (Umeyama, 1988) suggests also an alternative construction

$$\widehat{X} = \sum_{i=1}^n |u_i| |v_i|^\top \quad (8)$$

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

Compared with these constructions, GRAMPA has two important features that we elaborate below:

“All pairs matter.” Departing from existing approaches, our proposal \widehat{X} in (3) uses a combination of $u_i v_j^\top$ for *all* n^2 pairs $i, j \in \{1, \dots, n\}$, rather than only $i = j$. This renders our method significantly more resilient to noise. Indeed, while all of the above methods can succeed in recovering π_* in the noiseless case, methods based only on pairs (u_i, v_i) are brittle to noise if u_i and v_i quickly decorrelate as the amount of noise increases—this may happen when λ_i is not separated from other eigenvalues by a large spectral gap. When this decorrelation occurs, u_i becomes partially correlated with v_j for neighboring indices j , and the construction (3) leverages these partial correlations in a weighted manner to provide a more robust estimate of π_* .

The eigenvector alignment is quantitatively understood for the Gaussian Wigner model $B = A + \sigma Z$: For u_i and v_j corresponding to eigenvalues in the bulk of the Wigner semicircle spectrum, and for noise levels $n^{-1+\varepsilon} \ll \sigma^2 \ll n^{-\varepsilon}$, it is known that $\mathbb{E}[\langle u_i, v_i \rangle^2] = o(1)$ and

$$\mathbb{E}[\langle u_i, v_j \rangle^2] \approx \frac{1}{n} \frac{\sigma^2}{(\lambda_i - \mu_j)^2 + C\sigma^4} \quad (9)$$

where $C \asymp 1$ depends on the Wigner semicircle density near $\lambda_i \approx \mu_j$. See (Bourgade & Yau, 2017; Benigni, 2017) for more precise statements. Thus, for this range of noise, each eigenvector u_i of A is most aligned with $O(n\sigma^2)$ eigenvectors v_j of B for which $|\lambda_i - \mu_j| \lesssim \sigma^2$, and each such alignment is of typical size $\mathbb{E}[\langle u_i, v_j \rangle^2] \asymp 1/(n\sigma^2) \ll 1$. The signal for π_* in our proposal (3) arises from a weighted average of these alignments. As a result, while existing spectral approaches are only robust up to a noise level $\sigma = \frac{1}{\text{poly}(n)}$,¹ our new spectral method is polynomially more robust and can tolerate $\sigma = O(\frac{1}{\log n})$.

Cauchy spectral weights. The performance of the spectral method depends crucially on the choice of the weight function w in (3). In fact, there are other methods of the form (3) that do not work equally well. For example, if we choose $w(\lambda, \mu) = \lambda\mu$, then (3) reduces to $\widehat{X} = A\mathbf{J}B = ab^\top$, where $a = A\mathbf{1}$ and $b = B\mathbf{1}$ are the vectors of “degrees”. Rounding such a similarity matrix is equivalent to matching by sorting the degree of the vertices, which is known to fail when $\sigma = \Omega(n^{-1})$ due to the small spacing of the order statistics (cf. Remark 1 of (Ding et al., 2018)).

The Cauchy spectral weight (4) is a particular instance of the more general form $w(\lambda, \mu) = K(\frac{|\lambda - \mu|}{\eta})$, where K is a monotonically decreasing kernel function and η is a bandwidth parameter. Such a choice upweights the eigenvector pairs whose eigenvalues are close and penalizes those whose eigenvalues are separated more than η . The specific choice of the Cauchy kernel matches the form of $\mathbb{E}[\langle u_i, v_j \rangle^2]$ in (9), and is in a sense optimal as explained by a heuristic signal-to-noise calculation in Appendix H. In addition, the Cauchy kernel has its genesis as a regularization term in the associated convex relaxation, which we explain next.

1.3. Connection to Regularized Quadratic Programming

Our new spectral method is also rooted in optimization, as the similarity matrix \widehat{X} in (3) corresponds to the solution to a convex relaxation of the QAP (1), regularized by an added ridge penalty.

Denote the set of permutation matrices in $\mathbb{R}^{n \times n}$ by \mathfrak{S}_n . Then (1) may be written in matrix notation as one of the following equivalent optimization problems:

$$\max_{\Pi \in \mathfrak{S}_n} \langle A, \Pi B \Pi^\top \rangle \iff \min_{\Pi \in \mathfrak{S}_n} \|A\Pi - \Pi B\|_F^2.$$

Note that the objective $\|A\Pi - \Pi B\|_F^2$ above is a convex function in Π . Relaxing the set of permutations to its convex

¹For example, a recent result of (Ganassali et al., 2019) shows that in the Gaussian Wigner model, the rank-one method (6) based on the top eigenvector pairs can only correctly match $o(n)$ vertices once the noise level reaches $\sigma = \Omega(n^{-6/7+\varepsilon})$.

hull (the Birkhoff polytope of doubly stochastic matrices)

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

we arrive at the quadratic programming (QP) relaxation

$$\min_{X \in \mathcal{B}_n} \|AX - XB\|_F^2, \quad (10)$$

which was proposed in (Zaslavskiy et al., 2008; Aflalo et al., 2015), following an earlier LP relaxation using the ℓ_1 -objective proposed in (Almohamad & Duffuaa, 1993). Although this QP relaxation has achieved empirical success (Aflalo et al., 2015; Vogelstein et al., 2015; Lyzinski et al., 2016; Dym et al., 2017), understanding its performance theoretically is a challenging task yet to be accomplished.

Our spectral method can be viewed as the solution of a *regularized* further relaxation of the doubly stochastic QP (10). Indeed, we show in Lemma A.1 that the matrix \widehat{X} in (3) is the minimizer of

$$\min_{X \in \mathbb{R}^{n \times n}} \frac{1}{2} \|AX - XB\|_F^2 + \frac{\eta^2}{2} \|X\|_F^2 - \mathbf{1}^\top X \mathbf{1}. \quad (11)$$

Equivalently, \widehat{X} is a positive scalar multiple of the solution \widetilde{X} to the constrained program

$$\begin{aligned} \min_{X \in \mathbb{R}^{n \times n}} \quad & \|AX - XB\|_F^2 + \eta^2 \|X\|_F^2 \\ \text{s.t.} \quad & \mathbf{1}^\top X \mathbf{1} = n \end{aligned} \quad (12)$$

which further relaxes (10) and adds a ridge penalty term $\eta^2 \|X\|_F^2$. Note that \widehat{X} and \widetilde{X} are equivalent as far as the rounding step (5) is concerned. In contrast to (10), for which there is currently limited theoretical understanding, we are able to provide an exact recovery analysis for the rounded solutions to (11) and (12).

Note that the total-sum constraint in (12) is a significant relaxation of the double stochasticity (10). To make this further relaxed program work, the regularization term plays a key role. If η were zero, the similarity matrix \widehat{X} in (3) would involve the eigengap $|\lambda_i - \mu_j|$ in the denominator which can be polynomially small in n . Hence the regularization is crucial for reducing the variance of the estimate and making \widehat{X} stable, a rationale reminiscent of the ridge regularization in high-dimensional linear regression.

1.4. Diagonal Dominance of the Similarity Matrix

Equipped with this optimization point of view, we now explain the typical structure of solutions to the above quadratic programs including the spectral similarity matrix (3). It is well known that even the solution to the most stringent relaxation (10) is *not* the latent permutation matrix, which has been shown in (Lyzinski et al., 2016) by proving that the KKT conditions cannot be fulfilled with high probability.

In fact, a heuristic calculation explains why the solution to (10) is far from any permutation matrix: Let us consider the “population version” of (10), where the objective function is replaced by its expectation over the random instances A and B . Consider $\pi_* = \text{id}$ the identity permutation, and the Gaussian Wigner model $B = A + \sigma Z$, where A and Z are independent GOE matrices with $N(0, \frac{1}{n})$ off-diagonal entries and $N(0, \frac{2}{n})$ diagonal entries. Then the expectation of the objective function is

$$\begin{aligned} \mathbb{E} \|AX - XB\|_F^2 \\ = (2 + \sigma^2) \frac{n+1}{n} \|X\|_F^2 - \frac{2}{n} \text{Tr}(X)^2 - \frac{2}{n} \langle X, X^\top \rangle, \end{aligned}$$

so the population version of the quadratic program (10) is

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

whose solution is

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

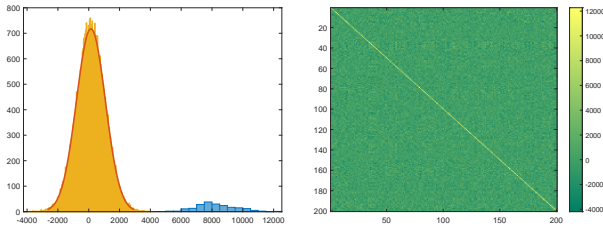
This is a convex combination of the true permutation matrix and the center of the Birkhoff polytope $\mathbf{F} = \frac{1}{n} \mathbf{J}$. Therefore, the population solution \overline{X} is in fact a very “flat” matrix, with each entry on the order of $\frac{1}{n}$, and is close to the center of the Birkhoff polytope and far from any of its vertices.

This calculation nevertheless provides us with important structural information about the solution to such a QP relaxation: \overline{X} is *diagonally dominant* for small σ , with diagonals about $2/\sigma^2$ times the off-diagonals. Although the actual solution of the relaxed program (10) or (11) is not equal to the population solution \overline{X} in expectation, it is reasonable to expect that it inherits the diagonal dominance property in the sense that $\widehat{X}_{i, \pi_*(i)} > \widehat{X}_{ij}$ for all $j \neq \pi_*(i)$, which enables rounding procedures such as (5) to succeed.

With this intuition in mind, let us revisit the regularized quadratic program (11) whose solution is the spectral similarity matrix (3). By a similar calculation, the solution to the population version of (11) is given by $\alpha \mathbf{I} + \beta \mathbf{J}$, with $\alpha = \frac{2n^2}{(n(\eta^2 + \sigma^2) + \sigma^2)(n(\eta^2 + \sigma^2 + 2) + \sigma^2)} \approx \frac{2}{(\eta^2 + \sigma^2)(\eta^2 + \sigma^2 + 2)}$ and $\beta = \frac{n}{n(\eta^2 + \sigma^2 + 2) + \sigma^2} \approx \frac{1}{\eta^2 + \sigma^2 + 2}$, which is diagonally dominant for small σ and η . In turn, the basis of our theoretical guarantee is to establish the diagonal dominance of the actual solution \widehat{X} ; see Figure 1 for an empirical illustration.

Although the ridge penalty $\eta^2 \|X\|_F^2$ guarantees the stability of the solution as discussed in Section 1.3, it may seem counterintuitive since it moves the solution closer to the center of the Birkhoff polytope and further away from the vertices (permutation matrices). In fact, several works in the literature (Fogel et al., 2013; Dym et al., 2017) advocate adding a negative ridge penalty, in order to make the solution closer to a permutation at the price of potentially making the optimization non-convex. This consideration,

however, is not necessary, as the ensuing rounding step can automatically map the solution to the correct permutation, even if they are far away in the Euclidean distance.



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

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

Figure 1: Diagonal dominance of the similarity matrix \widehat{X} defined by (3) or (11) for the Gaussian Wigner model $B = A + \sigma Z$ with $n = 200$, $\sigma = 0.05$ and $\eta = 0.01$.

It is worth noting that, in contrast to the prevalent analysis of convex relaxations in statistics and machine learning (where the goal is to show that the relaxed solution is close to the ground truth in a certain distance) or optimization (where the goal is to bound the gap of the objective value to the optimum), here our goal is not to show that the optimal solution *per se* constitutes a good estimator, but to show that it exhibits a diagonal dominance structure, which guarantees the success of the subsequent rounding procedure. For this reason, it is unclear from first principles that the guarantees obtained for one program, such as (12), automatically carry over to a tighter program, such as (10). Indeed, in Appendix G, a separate analysis is required when we study a tighter relaxation with row-sum constraints that achieves similar performance guarantees as (12).

2. Guarantees for Exact Recovery

In this section, we first formalize the general correlated Wigner model, and state the universal guarantees of exact recovery achieved by GRAMPA. Then the guarantees for the correlated Erdős-Rényi graph model follow as a corollary.

2.1. Correlated Wigner Model

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 $(A_{ij})_{i \leq j}$ are independent. Suppose that

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

$$\mathbb{E}[|A_{ij}|^k] \leq \frac{C^k}{nd^{(k-2)/2}} \text{ for all } i, j \text{ and } k \geq 2, \quad (14)$$

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

With the moment conditions (13) and (14) specified, we are ready to introduce the correlated Wigner model, which encompasses both the correlated Erdős-Rényi graph model proposed in (Pedarsani & Grossglauser, 2011) and the correlated Gaussian model (Ding et al., 2018) as special cases.

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 $\{(A_{ij}, B_{\pi_*(i)\pi_*(j)}) : i \leq j\}$ are independent pairs of random variables such that both $A = (A_{ij})$ and $B = (B_{ij})$ satisfy (13) and (14),*

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

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

$$\mathbb{P}\{\|A - \Pi_* B \Pi_*^\top\| \leq C\sigma\} \geq 1 - n^{-D} \quad (16)$$

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

The parameter σ measures the effective noise level in the model. In the special case of sparse Erdős-Rényi model defined below, A and B are the centered and normalized adjacency matrices of two Erdős-Rényi graphs, which differ by a fraction $2\sigma^2$ of edges approximately.

We now state the exact recovery guarantees for GRAMPA, making the informal statement in Section 1.1 precise.

Theorem 2.2 (Universal graph matching). *Fix constants $\kappa > 2$ and $a > 2\kappa$. Consider the correlated Wigner model with $n \geq d \geq (\log n)^{32+7a}$. Then there exist constants $c, c' > 0$ and $n_0 = n_0(\kappa, a)$ such that for all $n \geq n_0$, if*

$$(\log n)^{-a} \leq \eta \leq c(\log n)^{-2\kappa} \text{ and } \sigma \leq c'\eta, \quad (17)$$

then with probability at least $1 - n^{-10}$,

$$\min_k \widehat{X}_{k\pi_*(k)} > \max_{\ell \neq \pi_*(k)} \widehat{X}_{k\ell}, \quad (18)$$

and hence $\widehat{\pi}$ which solves the linear assignment problem (5) equals π_* .

The proof of this theorem is deferred to Appendix D.

2.2. Correlated (Sparse) Erdős-Rényi Graphs

An important application of the above universality result is matching two correlated sparse Erdős-Rényi graphs. Let G be an Erdős-Rényi graph with n vertices and edge probability q , denoted by $G \sim \mathcal{G}(n, q)$. Let \mathbf{A} and \mathbf{B}' be two copies of Erdős-Rényi graphs that are i.i.d. conditional on G , each of which is obtained from G by deleting every edge of G with probability $1 - s$ independently where $s \in [0, 1]$. Then we have that $\mathbf{A}, \mathbf{B}' \sim \mathcal{G}(n, p)$ marginally where $p \triangleq qs$.

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

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

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

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

Then we have the following result whose proof is deferred to Appendix F.

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 $\sigma^2 = \max\left(\frac{1-s}{1-p}, \frac{(\log n)^7}{d}\right)$.*

Then we can combine Theorem 2.2 and Lemma 2.3 to obtain the following exact recovery guarantees for the correlated Erdős-Rényi graph model.

Corollary 2.4 (Erdős-Rényi graph matching). *Suppose that $np(1-p) \geq (\log n)^{61}$ and $\frac{1-s}{1-p} \leq (\log n)^{-8-\delta}$ for a constant $\delta \in (0, 0.1)$. Then there exists $n_0 = n_0(\delta)$ such that for all $n \geq n_0$, if $\eta = (\log n)^{-4-\delta/3}$, then with probability at least $1 - n^{-10}$, the solution $\hat{\pi}$ to the linear assignment problem (5) coincides with π_* .*

Proof. Choose a and κ so that $4.1 > a > 4 + \delta/3 > 2\kappa > 4$. Then all conditions of Theorem 2.2 are satisfied for large n , so Corollary 2.4 follows. \square

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 $\sigma \leq (\log n)^{-4-\delta}$. In fact, GRAMPA recovers the latent permutation Π_* under a milder condition $\sigma \leq c(\log n)^{-1}$ 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.*

2.3. Proof Outline for Theorem 2.2

We now discuss the overall strategy for proving Theorem 2.2, which relies on techniques from random matrix theory. For a real symmetric matrix A with spectral decomposition (2), its resolvent is defined as

$$R_A(z) \triangleq (A - z\mathbf{I})^{-1} = \sum_i \frac{1}{\lambda_i - z} u_i u_i^\top \quad \text{for } z \in \mathbb{C} \setminus \mathbb{R}.$$

Moreover, let $m(z) \triangleq \int \frac{1}{x-z} \rho(x) dx = \frac{-z + \sqrt{z^2 - 4}}{2}$ denote the Stieltjes transform of the Wigner semicircle density $\rho(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \mathbf{1}_{\{|x| \leq 2\}}$.

The first step of the proof is to apply the Cauchy integral formula to obtain the following integral representation of the similarity matrix (3) via resolvents:

$$X \triangleq \eta \hat{X} = \frac{1}{2\pi} \text{Re} \oint_{\Gamma} R_A(z) \mathbf{J} R_B(z + i\eta) dz, \quad (20)$$

where Γ is the rectangular contour with vertices $\pm 3 \pm i\eta/2$.

Results from random matrix theory known as *local laws* (Erdős et al., 2013a;b) provide entrywise bound on the resolvents. To establish the diagonal dominance of the matrix X , directly applying the local law to the integrand in (20) turns out to be too crude. Instead, we employ a leave-one-out technique using the Schur complement to obtain

$$\begin{aligned} X_{ij} &\approx \frac{1}{2\pi} (a^{(ij)})^\top M^{(ij)} b^{(ji)}, \text{ where } M^{(ij)} \triangleq \\ &\text{Re} \oint_{\Gamma} m(z) m(z + i\eta) R_{A^{(ij)}}(z) \mathbf{J} R_{B^{(ij)}}(z + i\eta) dz. \end{aligned}$$

Here $a^{(ij)}$ denotes the i -th column of A with the i -th and j -th entries removed, $b^{(ji)}$ denotes the j -th column of B with the i -th and j -th entries removed, and $A^{(ij)}$ denotes the matrix A with the i -th and j -th columns and rows removed.

Crucially, the matrix $M^{(ij)}$ is independent of the vectors $(a^{(ij)}, b^{(ji)})$, which allows us to apply a Hanson-Wright-like inequality for concentration of bilinear forms to obtain that with high probability,

- $X_{ii} \approx \frac{1-\sigma^2}{2\pi n} \text{Re} \text{Tr} M^{(ii)} \approx \frac{1-\sigma^2}{\eta} + O\left(\frac{\sigma}{\eta^2}\right)$, and
- $|X_{ij}| \lesssim \frac{(\log n)^\kappa}{n} \|M^{(ij)}\|_F \lesssim \frac{(\log n)^\kappa}{\sqrt{\eta}}$ for $i \neq j$,

where $\kappa > 2$ is any fixed constant. The above quantitative control on $\text{Tr} M^{(ii)}$ and $\|M^{(ij)}\|_F$ is achieved by analyzing the resolvents involved, in particular, applying their local laws.

Comparing the diagonal entries X_{ii} to off-diagonal entries X_{ij} where $i \neq j$, by a union bound, we conclude that with high probability the similarity matrix X (and thus \hat{X}) is diagonally dominant under the assumptions of Theorem 2.2, which ensures the correctness of Algorithm 1.

3. Numerical Experiments

This section is devoted to comparing our spectral method to various existing algorithms for graph matching, using both synthetic examples and real datasets.

Similar to the last step of Algorithm 1, many algorithms we compare to also involve rounding a similarity matrix to

produce a matching in the final step. In the experiments, for the sake of comparison we use the linear assignment (5) for rounding, unless otherwise specified.

3.1. Comparison to Existing Graph Matching Methods

We now compare the performance of GRAMPA to several existing methods in the literature. For GRAMPA, Theorem 2.2 suggests that the bandwidth parameter η needs to be chosen within a certain range. In practice, one may compute estimates $\hat{\pi}^\eta$ for different values of η and select the one with the minimum objective value $\|A - B\hat{\pi}^\eta\|_F^2$. We find in simulations that the performance of GRAMPA is in fact not very sensitive to the choice of η , unless η is extremely close to zero or larger than one. For simplicity and consistency, we fix $\eta = 0.2$ in the following synthetic experiments.

First, we compare GRAMPA to other spectral methods. Besides the rank-1 method of rounding the outer product of top eigenvectors (6) (denoted by TopEigenVec), we consider the IsoRank algorithm of (Singh et al., 2008), the EigenAlign and LowRankAlign² algorithms of (Feizi et al., 2019), and Umeyama’s method (Umeyama, 1988) which rounds the similarity matrix (8). In Figure 2(a), we apply these algorithms to match Erdős-Rényi graphs with 100 vertices³ and edge density 0.5. For each spectral method, we plot the fraction of correctly matched pairs of vertices of the two graphs versus the noise level σ , averaged over 10 independent repetitions. While all estimators recover the exact matching in the noiseless case, it is clear that GRAMPA is more robust to noise than all previous spectral methods by a wide margin.

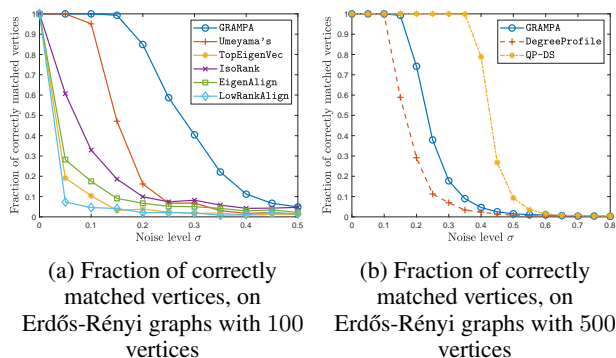


Figure 2: Comparison of GRAMPA to existing methods for matching Erdős-Rényi graphs with expected edge density 0.5. Each experiment is averaged over 10 repetitions.

²We implement the rank-2 version of LowRankAlign here because a higher rank does not appear to improve its performance in the experiments.

³This experiment is not run on larger graphs because IsoRank and EigenAlign involve taking Kronecker products of graphs and are thus not as scalable as the other methods.

Next, we consider more competitive graph matching algorithms outside the spectral class. Since our method admits an interpretation through the regularized QP (11) or (12), it is of interest to compare its performance to (the algorithm that rounds the solution to) the full QP (10) with full doubly stochastic constraints, denoted by QP-DS. Another recently proposed method for graph matching is Degree Profile (Ding et al., 2018), for which theoretical guarantees comparable to our results have been established for the Gaussian Wigner and Erdős-Rényi models.

Figure 2(b) plots the fraction of correctly matched vertex pairs by the three algorithms, on Erdős-Rényi graphs with 500 vertices and edge density 0.5, averaged over 10 independent repetitions. GRAMPA outperforms DegreeProfile, while QP-DS is clearly the most robust, albeit at a much higher computational cost. Since off-the-shelf QP solvers are extremely slow on instances with n larger than several hundred, we resort to an alternating direction method of multipliers (ADMM) procedure used in (Ding et al., 2018). Still, solving (10) is more than 350 times slower than computing the similarity matrix (3) for the instances in Figure 2(b). Moreover, DegreeProfile is about 15 times slower. We argue that GRAMPA achieves a desirable balance between speed and robustness when implemented on large networks.

3.2. Networks of Autonomous Systems

As suggested by both theory and the preceding experiments, GRAMPA and DegreeProfile have been the two fast and robust methods. We corroborate the improvement of GRAMPA over DegreeProfile using quantitative benchmarks on a time-evolving real-world network of $n = 10000$ vertices. Here, for simplicity, we apply both methods to the *unnormalized* adjacency matrices, and set $\eta = 1$ for GRAMPA. We find that the results are not very sensitive to this choice of η . Although QP-DS yields better performance in Figure 2(b), it is extremely slow to run on such a large network, so we omit it from the comparison here.

We use a subset of the Autonomous Systems dataset from the University of Oregon Route Views Project (University of Oregon Route Views Project), available as part of the Stanford Network Analysis Project (Leskovec & Krevl, 2014; Leskovec et al., 2005). The data consists of instances of a network of autonomous systems observed on nine days between March 31, 2001 and May 26, 2001. Edges and (a small fraction of) vertices of the network were added and deleted over time. In particular, the number of vertices of the network on the nine days ranges from 10,670 to 11,174 and the number of edges from 22,002 to 23,409. The labels of the vertices are known.

To test the graph matching methods, we consider 10,000 vertices of the network that are present on all nine days. The resulting nine graphs can be seen as noisy versions of

each other, with correlation decaying over time. We apply GRAMPA and DegreeProfile to match each graph to that on the first day of March 31, with vertices randomly permuted.

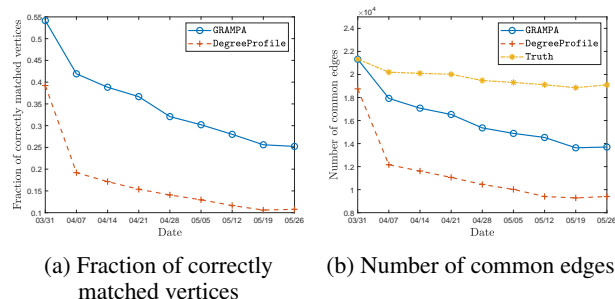


Figure 3: Comparison of GRAMPA with DegreeProfile for matching networks of autonomous systems on nine days to that on the first day

In Figure 3(a), we plot the fraction of correctly matched pairs of vertices against the chronologically ordered dates. GRAMPA correctly matches many more pairs of vertices than DegreeProfile for all nine days. As expected, the performance of both methods degrades over time as the network becomes less correlated with the original one.

Note that even the matching of the graph on the first day to itself is not exact for both methods. This is due to the fact that there are over 3,000 degree-one vertices in this graph, and some of them are attached to the same high-degree vertices. Therefore the exact matching is non-identifiable. Thus for a given matching $\hat{\pi}$, an arguably more relevant figure of merit is the number of common edges, i.e. the (rescaled) objective value $\langle A, B^{\hat{\pi}} \rangle / 2$. We plot this in Figure 3(b) together with the value for the ground-truth matching. The values of GRAMPA and of the ground-truth matching are the same on the first day, indicating that GRAMPA successfully finds an automorphism of this graph, while DegreeProfile fails. Furthermore, GRAMPA consistently recovers a matching with more common edges over the nine days.

3.3. Deformable Shape Matching

Matching 3D deformable shapes is a central problem in computer vision with a wide variety of applications, and has been extensively studied for decades. In this section, we evaluate the performance of GRAMPA on the SHREC’16 dataset (Löhner et al., 2016), which provides a realistic and challenging experiment setup for evaluating shape matching methods on noisy data. The test data in SHREC’16 contains 25 kid shapes (15 for training and 10 for testing), undergoing non-rigid deformations (e.g. change in pose) and large topological changes. At the lower resolution, each shape is represented by a triangulated mesh graph consisting of $\sim 10K$ vertices with 3D coordinates and $\sim 20K$ triangular

faces.

For each of the 90 pairs from the 10 test shapes, we apply our GRAMPA algorithm with $\eta = 1$ to the unweighted adjacency matrices of the mesh graphs, without using the training data. As the graph sizes are slightly different across different shapes, we generalize GRAMPA to match a m -vertex graph A to a n -vertex graph B by letting \mathbf{J} in (3) to be the $m \times n$ all-one matrix. Moreover, to speed up rounding, we solve the LAP in (5) approximately via the greedy max-weight matching. Finally, to boost the matching accuracy, we apply an iterative clean-up procedure: $\Pi_{t+1} = \operatorname{argmax}_{\Pi \in \mathcal{S}_n} \langle \Pi, A\Pi_t B \rangle$ for $t = 0, \dots, 100$, where the initial permutation matrix Π_0 is given by GRAMPA; each LAP is again approximated by the greedy max-weighted matching. See Fig. 4(a) for a visualization of the final correspondence between two kid shapes, where the matched vertices are colored with the same color.

To evaluate the matching quality, we follow the Princeton benchmark protocol (Kim et al., 2011). Assume that a matching algorithm matches vertex i in shape \mathcal{M} to vertex j in shape \mathcal{N} , while the ground-truth correspondence is (i, j^*) . Then the *normalized geodesic error* of this correspondence at vertex i is defined as $\varepsilon(i) = \frac{d_{\mathcal{N}}(j, j^*)}{\sqrt{\operatorname{area}(\mathcal{N})}}$, where $d_{\mathcal{N}}$ denotes the geodesic distance on \mathcal{N} and $\operatorname{area}(\mathcal{N})$ is the total surface area of \mathcal{N} . Finally, we plot the cumulative distribution curve of $\{\varepsilon(i)\}_{i=1}^m$ in Fig. 4(b), where $\operatorname{curve}(x)$ is the fraction of vertices i such that $\varepsilon(i) \leq x$. In particular, $\operatorname{curve}(0)$ is the fraction of correctly matched vertices in shape \mathcal{M} .

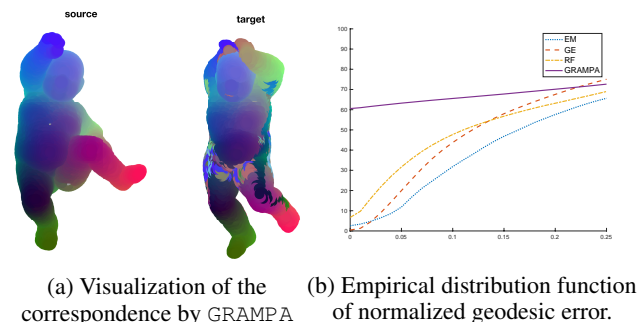


Figure 4: Comparison of GRAMPA to existing methods on SHREC’16 dataset.

We compare with three existing methods tested in the original SHREC’16 paper (Löhner et al., 2016), namely, Isometric embedding (EM) (Sahillioğlu & Yemez, 2012), Green’s function embedding alignment (GE) (Lipman et al., 2010), and random forests (RF) (Rodolà et al., 2014). From Fig. 4(b), we see that GRAMPA significantly improves over existing methods, especially in terms of the fraction of correctly matched vertices. Unlike these algorithms, GRAMPA does not use the 3D coordinates of vertices, and does not require a training set as opposed to the learning-based tech-

niques such as RF.

Acknowledgments

Y. Wu and J. Xu are grateful to Zorah Lähler for helpful suggestions on the experiment in Section 3.3. Y. Wu was supported in part by the NSF Grants CCF-1527105, CCF-1900507, an NSF CAREER award CCF-1651588, and an Alfred Sloan fellowship. J. Xu was supported by the NSF Grants IIS-1932630, CCF-1850743, and CCF-1856424. Z. Fan was supported in part by NSF Grant DMS-1916198.

References

- Aflalo, Y., Bronstein, A., and Kimmel, R. On convex relaxation of graph isomorphism. *Proceedings of the National Academy of Sciences*, 112(10):2942–2947, 2015.
- Almohamad, H. and Duffuaa, S. O. A linear programming approach for the weighted graph matching problem. *IEEE Transactions on pattern analysis and machine intelligence*, 15(5):522–525, 1993.
- Babai, L., Grigoryev, D. Y., and Mount, D. M. Isomorphism of graphs with bounded eigenvalue multiplicity. In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pp. 310–324. ACM, 1982.
- Barak, B., Chou, C.-N., Lei, Z., Schramm, T., and Sheng, Y. (Nearly) efficient algorithms for the graph matching problem on correlated random graphs. *arXiv preprint arXiv:1805.02349*, 2018.
- Benigni, L. Eigenvectors distribution and quantum unique ergodicity for deformed Wigner matrices. *arXiv preprint arXiv:1711.07103*, 2017.
- Bollobás, B. Distinguishing vertices of random graphs. *North-Holland Mathematics Studies*, 62:33–49, 1982.
- Bourgade, P. and Yau, H.-T. The eigenvector moment flow and local quantum unique ergodicity. *Communications in Mathematical Physics*, 350(1):231–278, 2017.
- Burkard, R. E., Cela, E., Pardalos, P. M., and Pitsoulis, L. S. The quadratic assignment problem. In *Handbook of combinatorial optimization*, pp. 1713–1809. Springer, 1998.
- Conte, D., Foggia, P., Sansone, C., and Vento, M. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.
- Cullina, D. and Kiyavash, N. Improved achievability and converse bounds for Erdős-Rényi graph matching. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, pp. 63–72. ACM, 2016.
- Cullina, D. and Kiyavash, N. Exact alignment recovery for correlated Erdős-Rényi graphs. *arXiv preprint arXiv:1711.06783*, 2017.
- Cullina, D., Kiyavash, N., Mittal, P., and Poor, H. V. Partial recovery of Erdős-Rényi graph alignment via k -core alignment. *arXiv preprint arXiv:1809.03553*, Nov. 2018.
- Czajka, T. and Pandurangan, G. Improved random graph isomorphism. *Journal of Discrete Algorithms*, 6(1):85–92, 2008.
- Dai, O. E., Cullina, D., Kiyavash, N., and Grossglauser, M. On the performance of a canonical labeling for matching correlated Erdős-Rényi graphs. *arXiv preprint arXiv:1804.09758*, 2018.
- Ding, J., Ma, Z., Wu, Y., and Xu, J. Efficient random graph matching via degree profiles. *arXiv preprint arXiv:1811.07821*, Nov 2018.
- Dym, N., Maron, H., and Lipman, Y. DS++: a flexible, scalable and provably tight relaxation for matching problems. *ACM Transactions on Graphics (TOG)*, 36(6):184, 2017.
- Emmert-Streib, F., Dehmer, M., and Shi, Y. Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, 346:180–197, 2016.
- Erdős, L., Knowles, A., Yau, H.-T., and Yin, J. The local semicircle law for a general class of random matrices. *Electron. J. Probab.*, 18(59):1–58, 2013a.
- Erdős, L., Knowles, A., Yau, H.-T., and Yin, J. Spectral statistics of Erdős-Rényi graphs I: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013b.
- Feizi, S., Quon, G., Mendoza, M., Medard, M., Kellis, M., and Jadbabaie, A. Spectral alignment of graphs. *IEEE Transactions on Network Science and Engineering*, 2019.
- Finke, G., Burkard, R. E., and Rendl, F. Quadratic assignment problems. In *North-Holland Mathematics Studies*, volume 132, pp. 61–82. Elsevier, 1987.
- Fogel, F., Jenatton, R., Bach, F., and d’Aspremont, A. Convex relaxations for permutation problems. In *Advances in Neural Information Processing Systems*, pp. 1016–1024, 2013.
- Ganassali, L., Lelarge, M., and Massoulié, L. Spectral alignment of correlated gaussian random matrices. *arXiv preprint arXiv:1912.00231*, 2019.
- Kazemi, E. and Grossglauser, M. On the structure and efficient computation of isorank node similarities. *arXiv preprint arXiv:1602.00668*, 2016.

- Kazemi, E., Hassani, S. H., and Grossglauser, M. Growing a graph matching from a handful of seeds. *Proceedings of the VLDB Endowment*, 8(10):1010–1021, 2015.
- Kazemi, E., Hassani, H., Grossglauser, M., and Modarres, H. P. Proper: global protein interaction network alignment through percolation matching. *BMC bioinformatics*, 17(1):527, 2016.
- Kim, V. G., Lipman, Y., and Funkhouser, T. Blended intrinsic maps. *ACM Transactions on Graphics (TOG)*, 30(4):1–12, 2011.
- Korula, N. and Lattanzi, S. An efficient reconciliation algorithm for social networks. *Proceedings of the VLDB Endowment*, 7(5):377–388, 2014.
- Kuhn, H. W. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- Lähner, Z., Rodolà, E., Bronstein, M., Cremers, D., Burghard, O., Cosmo, L., Dieckmann, A., Klein, R., and Sahillioglu, Y. SHREC’16: Matching of deformable shapes with topological noise. *Proc. 3DOR*, 2(10.2312), 2016.
- Leskovec, J. and Krevl, A. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.
- Leskovec, J., Kleinberg, J., and Faloutsos, C. Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pp. 177–187. ACM, 2005.
- Lipman, Y., Rustamov, R. M., and Funkhouser, T. A. Biharmonic distance. *ACM Transactions on Graphics (TOG)*, 29(3):1–11, 2010.
- Livi, L. and Rizzi, A. The graph matching problem. *Pattern Analysis and Applications*, 16(3):253–283, 2013.
- Lubars, J. and Srikant, R. Correcting the output of approximate graph matching algorithms. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pp. 1745–1753. IEEE, 2018.
- Lyzinski, V., Fishkind, D. E., and Priebe, C. E. Seeded graph matching for correlated Erdős-Rényi graphs. *Journal of Machine Learning Research*, 15(1):3513–3540, 2014.
- Lyzinski, V., Fishkind, D., Fiori, M., Vogelstein, J., Priebe, C., and Sapiro, G. Graph matching: Relax at your own risk. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 38(1):60–73, 2016.
- Makarychev, K., Manokaran, R., and Sviridenko, M. Maximum quadratic assignment problem: Reduction from maximum label cover and LP-based approximation algorithm. *Automata, Languages and Programming*, pp. 594–604, 2010.
- Mossel, E. and Xu, J. Seeded graph matching via large neighborhood statistics. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1005–1014. SIAM, 2019.
- Narayanan, A. and Shmatikov, V. Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125. IEEE, 2008.
- Narayanan, A. and Shmatikov, V. De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on*, pp. 173–187. IEEE, 2009.
- Pardalos, P. M., Rendl, F., and Wolkowicz, H. The quadratic assignment problem: A survey and recent developments. In *In Proceedings of the DIMACS Workshop on Quadratic Assignment Problems, volume 16 of DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pp. 1–42. American Mathematical Society, 1994.
- Pedarsani, P. and Grossglauser, M. On the privacy of anonymized networks. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1235–1243, 2011.
- Rodolà, E., Rota Bulò, S., Windheuser, T., Vestner, M., and Cremers, D. Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4177–4184, 2014.
- Sahillioglu, Y. and Yemez, Y. Minimum-distortion isometric shape correspondence using em algorithm. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2203–2215, 2012.
- Schellewald, C. and Schnörr, C. Probabilistic subgraph matching based on convex relaxation. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 171–186. Springer, 2005.
- Shirani, F., Garg, S., and Erkip, E. Seeded graph matching: Efficient algorithms and theoretical guarantees. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pp. 253–257. IEEE, 2017.
- Singh, R., Xu, J., and Berger, B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences*, 105(35):12763–12768, 2008.

- Umeyama, S. An eigendecomposition approach to weighted graph matching problems. *IEEE transactions on pattern analysis and machine intelligence*, 10(5):695–703, 1988.
- University of Oregon Route Views Project. Autonomous Systems Peering Networks. *Online data and reports*, <http://www.routeviews.org/>.
- Vogelstein, J. T., Conroy, J. M., Lyzinski, V., Podrazik, L. J., Kratzer, S. G., Harley, E. T., Fishkind, D. E., Vogelstein, R. J., and Priebe, C. E. Fast approximate quadratic programming for graph matching. *PLOS one*, 10(4): e0121002, 2015.
- Xu, L. and King, I. A PCA approach for fast retrieval of structural patterns in attributed graphs. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(5):812–817, 2001.
- Yartseva, L. and Grossglauser, M. On the performance of percolation graph matching. In *Proceedings of the first ACM conference on Online social networks*, pp. 119–130. ACM, 2013.
- Zaslavskiy, M., Bach, F., and Vert, J.-P. A path following algorithm for the graph matching problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2227–2242, 2008.