HHS Public Access

Graph Matching: Relax at Your Own Risk

Vince Lyzinski, **Donniell E. Fishkind**, **Marcelo Fiori**, **Joshua T. Vogelstein**, **Carey E. Priebe**, and **Guillermo Sapiro**

V.L. is with Johns Hopkins University Human Language Technology Center of Excellence. D.E.F. and C.E.P. are with Johns Hopkins University. J.T.V. is with Johns Hopkins University and the Child Mind Institute. G.S. is with Duke University. M.F. is with Universidad de la República, Uruguay

Abstract

Graph matching—aligning a pair of graphs to minimize their edge disagreements—has received wide-spread attention from both theoretical and applied communities over the past several decades, including combinatorics, computer vision, and connectomics. Its attention can be partially attributed to its computational difficulty. Although many heuristics have previously been proposed in the literature to approximately solve graph matching, very few have any theoretical support for their performance. A common technique is to relax the discrete problem to a continuous problem, therefore enabling practitioners to bring gradient-descent-type algorithms to bear. We prove that an indefinite relaxation (when solved exactly) almost always discovers the optimal permutation, while a common convex relaxation almost always fails to discover the optimal permutation. These theoretical results suggest that initializing the indefinite algorithm with the convex optimum might yield improved practical performance. Indeed, experimental results illuminate and corroborate these theoretical findings, demonstrating that excellent results are achieved in both benchmark and real data problems by amalgamating the two approaches.

1 Introduction

Several problems related to the isomorphism and matching of graphs have been an important and enjoyable challenge for the scientific community for a long time, with applications in pattern recognition (see, for example, [1], [2]), computer vision (see, for example, [3], [4], [5]), and machine learning (see, for example, [6], [7]), to name a few. Given two graphs, the graph isomorphism problem consists of determining whether these graphs are isomorphic or not, that is, if there exists a bijection between the vertex sets of the graphs which exactly preserves the vertex adjacency. The graph isomorphism problem is very challenging from a computational complexity point of view. Indeed, its complexity is still unresolved: it is not currently classified as NP-complete or P [8]. The graph isomorphism problem is contained in the (harder) graph matching problem. The graph matching problem consists of finding the exact isomorphism between two graphs if it exists, or, in general, finding the bijection between the vertex sets that minimizes the number of adjacency disagreements. Graph matching is a very challenging and well-studied problem in the literature with applications in such diverse fields as pattern recognition, computer vision, neuroscience, etc. (see [9]).

Although polynomial-time algorithms for solving the graph matching problem are known for certain classes of graphs (e.g., trees [10], [11]; planar graphs [12]; and graphs with some spectral properties [13], [14]), there are no known polynomial-time algorithms for solving the general case. Indeed, in its most general form, the graph matching problem is equivalent to the NP-hard quadratic assignment problem.

Formally, for any two graphs on n vertices with respective $n \times n$ adjacency matrices A and B, the graph matching problem is to minimize $\|A - PBP^T\|_F$ over all $P \in \Pi$, where Π denotes the set of $n \times n$ permutation matrices, and $\|\cdot\|_F$ is the Froebenius matrix norm (other graph matching objectives have been proposed in the literature as well, this being a common one).

Note that for any permutation matrix P, $\frac{1}{2}\|A - PBP^T\|_F^2 = \frac{1}{2}\|AP - BP\|_F^2$ counts the number of adjacency disagreements induced by the vertex bijection corresponding to P.

An equivalent formulation of the graph matching problem is to minimize $-\langle AP, PB \rangle$ over all $P \in \Pi$, where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product, i.e., for all $C, D \in \mathbb{R}^{n \times n}$, $\langle C, D \rangle :=$ trace$(C^T D)$. This can be seen by expanding, for any $P \in \Pi$,

$$\|A - PBP^T\|_F^2 = \|AP - PB\|_F^2 = \|A\|_F^2 + \|B\|_F^2 - 2\langle AP, PB \rangle,$$

and noting that $\|A\|_F^2$ and $\|B\|_F^2$ are constants for the optimization problem over $P \in \Pi$.

Let $\mathscr{D}$ denote the set of $n \times n$ doubly stochastic matrices, i.e., nonnegative matrices with row and column sums each equal to 1. We define the *convex relaxed graph matching problem* to be minimizing $\|AD - DB\|_F^2$ over all $D \in \mathscr{D}$, and we define the *indefinite relaxed graph matching problem* to be minimizing $-\langle AD, DB \rangle$ over all $D \in \mathscr{D}$. Unlike the graph matching problem, which is an integer programming problem, these relaxed graph matching problems are each continuous optimization problems with a quadratic objective function subject to affine constraints. Since the quadratic objective $\|AD - DB\|_F^2$ is also convex in the variables D (it is a composition of a convex function and a linear function), there is a polynomial-time algorithm for exactly solving the convex relaxed graph matching problem (see [15]). However, $-\langle AD, DB \rangle$ is not convex (in fact, the Hessian has trace zero and is therefore *indefinite*), and nonconvex quadratic programming is (in general) NP-hard. Nonetheless the indefinite relaxation can be efficiently approximately solved with Frank-Wolfe (F-W) methodology [16], [17].

It is natural to ask how the (possibly different) solutions to these relaxed formulations relate to the solution of the original graph matching problem. Our main theoretical result, Theorem 1, proves, under mild conditions, that convex relaxed graph matching (which is tractable) almost always yields the wrong matching, and indefinite relaxed graph matching (which is intractable) almost always yields the correct matching. We then illustrate via illuminating simulations that this asymptotic result about the trade-off between tractability and correctness is amply felt even in moderately sized instances.

In light of graph matching complexity results (see for example [14], [18], [19]), it is unsurprising that the convex relaxation can fail to recover the true permutation. In our main theorem, we take this a step further and provide an answer from a probabilistic point of view, showing almost sure failure of the convex relaxation for a very rich and general family of graphs when convexly relaxing the graph matching problem. This paints a sharp contrast to the (surprising) almost sure correctness of the solution of the indefinite relaxation. We further illustrate that our theory gives rise to a new state-of-the-art matching strategy.

1.1 Correlated random Bernoulli graphs

Our theoretical results will be set in the context of correlated random (simple) Bernoulli graphs,[1] which can be used to model many real-data scenarios. Random Bernoulli graphs are the most general edge independent random graphs, and contain many important random graph families including Erdős-Rényi and the widely used stochastic block model of [21] (in the stochastic block model, Λ is a block constant matrix, with the number of diagonal blocks representing the number of communities in the network). Stochastic block models, in particular, have been extensively used to model networks with inherent community structure (see, for example, [22], [23], [24], [25]). As this model is a submodel of the random Bernoulli graph model here used, our main theorem (Theorem 1) extends to stochastic block models immediately, making it of highly practical relevance.

These graphs are defined as follows. Given $n \in \mathbb{Z}^+$, a real number $\rho \in [0, 1]$, and a symmetric, hollow matrix $\Lambda \in [0, 1]^{n \times n}$, define $\mathscr{E} := \{\{i, j\} : i \in [n], j \in [n], i \neq j\}$, where $[n] := \{1, 2, \ldots, n\}$. Two random graphs with respective $n \times n$ adjacency matrices A and B are ρ- *correlated Bernoulli*(Λ) distributed if, for all $\{i, j\} \in \mathscr{E}$, the random variables (matrix entries) $A_{i,j}$, $B_{i,j}$ are Bernoulli($\Lambda_{i,j}$) distributed, and all of these random variables are collectively independent except that, for each $\{i, j\} \in \mathscr{E}$, the Pearson product-moment correlation coefficient for $A_{i,j}$, $B_{i,j}$ is ρ. It is straightforward to show that the parameters n, ρ, and Λ completely specify the random graph pair distribution, and the distribution may be achieved by first, for all $\{i, j\} \in \mathscr{E}$, having $B_{ij} \sim$ Bernoulli($\Lambda_{i,j}$) independently drawn and then, conditioning on B, have $A_{i,j} \sim$ Bernoulli $((1 - \rho)\Lambda_{i,j} + \rho B_{i,j})$ independently drawn. While $\rho = 1$ would imply the graphs are isomorphic, this model allows for a natural vertex alignment (namely the identity function) for $\rho < 1$, i.e. when the graphs are not necessarily isomorphic.

1.2 The main result

We will consider a sequence of correlated random Bernoulli graphs for $n = 1, 2, 3, \ldots$, where Λ is a function of n. When we say that a sequence of events, $\{E_m\}_{m=1}^{\infty}$, holds *almost always* we mean that almost surely it happens that the events in the sequence occur for all but finitely many m.

Theorem 1—*Suppose A and B are adjacency matrices for ρ-correlated Bernoulli(Λ) graphs, and there is an $\alpha \in (0, 1/2)$ such that $\Lambda_{i,j} \in [\alpha, 1 - \alpha]$ for all $i \neq j$. Let $P^* \in \Pi$, and denote $A' := P^*AP^{*T}$.*

---

[1]Also known as *inhomogeneous random graphs* in [20].

a. If $(1 - \alpha)(1 - \rho) < 1/2$, *then it almost always holds that*

$$\arg \min_{D \in \mathscr{D}} - \langle A'D, DB \rangle = \arg \min_{P \in \Pi} \|A' - PBP^T\|_F = \{P^*\}.$$

b. *If the between graph correlation* $\rho < 1$*, then it almost always holds that* $P^* \notin$ arg $\min_{D \in \mathscr{D}} \|A'D - DB\|_F$.

This theorem states that: (part *a*) the unique solution of the indefinite relaxation almost always is the correct permutation matrix, while (part *b*) the correct permutation is almost always not a solution of the commonly used convex relation. Moreover, as we will show in the experiments section, the convex relaxation can lead to a doubly stochastic matrix that is not even in the Voronoi cell of the true permutation. In this case, the convex optimum is closest to an incorrect permutation, hence the correct permutation will not be recovered by projecting the doubly stochastic solution back onto Π.

In the above, ρ and α are fixed. However, the proofs follow *mutatis mutandis* if ρ and α are allowed to vary in *n*. If there exist constants $c_1, c_2 > 0$ such that $\alpha \geq c_1 \sqrt{(\log n)/n}$ and $1/2 - c_2 \sqrt{(\log n)/n} \geq (1 - \rho)(1 - \alpha)$, then Theorem 1, part *a* will hold. Note that $\alpha \geq c_1 \sqrt{(\log n)/n}$ also guarantees the corresponding graphs are almost always connected.

For the analogous result for part *b*, let us first define $\sigma(i) = \frac{1}{n-1} \sum_{k \neq i} \Lambda_{ki}(1 - \Lambda_{ki})$. If there exists an $i \in [n]$ such that $1 - \frac{3}{2\sigma(i)} \sqrt{(8 \log n)/n} > \rho$, then the results of Theorem 1, part *b* hold as proven below.

1.3 Isomorphic versus ρ-correlated graphs

There are numerous algorithms available in the literature for (approximately) solving the graph isomorphism problem (see, for example, [26], [27]), as well as for (approximately) solving the subgraph isomorphism problem (see, for example, [28]). All of the graph matching algorithms we explore herein can be used for the graph isomorphism problem as well.

We emphasize that the ρ-correlated random graph model extends our random graphs beyond isomorphic graph pairs; indeed ρ-correlated graphs G_1 and G_2 will almost surely have on the order of $[\alpha, 1 - \alpha]\rho n^2$ edgewise disagreements. As such, these graphs are a.s. *not* isomorphic. In this setting, the goal of graph matching is to align the vertices across graphs whilst simultaneously preserving the adjacency structure as best possible across graphs. However, this model does preserve a very important feature of isomorphic graphs: namely the presence of a latent alignment function (the identity function in the ρ-correlated model).

We note here that in the ρ-correlated Bernoulli(Λ) model, both G_1 and G_2 are marginally Bernoulli(Λ) random graphs, which is amenable to theoretical analysis. We note here that real data experiments across a large variety of data sets (see Section 4.3) and simulated experiments across a variety of robust random graph settings (see Section 4.4) also both

support the result of Theorem 1. Indeed, we suspect that an analogue of Theorem 1 holds over a much broader class of random graphs, and we are presently investigating this extension.

2 Proof of Theorem 1, Part a

Without loss of generality, let $P^* = I$. We will first sketch the main argument of the proof, and then we will spend the remainder of the section filling in all necessary details of the proof. The proof will proceed as follows. Almost always, $-\langle A, B \rangle < -\langle AQ, PB \rangle$ for any $P, Q \in \Pi$ such that either $P \neq I$ or $Q \neq I$. To accomplish this, we count the entrywise disagreements between AQ and PB in two steps (of course, this is the same as the number of entrywise disagreements between A and PBQ^T). We first count the entrywise disagreements between B and PBQ^T (Lemma 4), and then count the additional disagreements induced by realizing A conditioning on B. Almost always, this two step realization will result in more errors than simply realizing A directly from B without permuting the vertex labels (Lemma 5). This establishes $-\langle A, B \rangle < -\langle AQ, PB \rangle$, and Theorem 1, part a is a consequence of the Birkhoff-von Neumann theorem.

We begin with two lemmas used to prove Theorem 1. First, Lemma 2 is adapted from [29], presented here as a variation of the form found in [30, Prop. 3.2]. This lemma lets us tightly estimate the number of disagreements between B and PBQ^T, which we do in Lemma 4.

Lemma 2

For any integer $N > 0$ and constant $\alpha \in (0, \frac{1}{2})$, suppose that the random variable X is a function of at most N independent Bernoulli random variables, each with Bernoulli parameter in the interval $[\alpha, 1 - \alpha]$. Suppose that changing the value of any one of the Bernoulli random variables (and keeping all of the others fixed) changes the value of X by at most γ. Then for any t such that $0 \leq t < \sqrt{\alpha(1-\alpha)}\gamma N$, it holds that $\mathbb{P}[|X - \mathbb{E}X| > t] \leq 2 \cdot \exp\{-t^2/(\gamma^2 N)\}$.

The next result, Lemma 3, is a special case of the classical Hoeffding inequality (see, for example, [31]), which we use to tightly bound the number of additional entrywise disagreements between AQ and PB when we realize A conditioning on B.

Lemma 3

Let N_1 and N_2 be positive integers, and q_1 and q_2 be real numbers in $[0, 1]$. If $X_1 \sim$ Binomial(N_1, q_1) and $X_2 \sim$ Binomial(N_2, q_2) are independent, then for any $t \geq 0$ it holds that

$$\mathbb{P}\left[|X_1 + X_2 - \mathbb{E}(X_1 + X_2)| \geq t\right] \leq 2 \cdot \exp\left\{\frac{-2t^2}{N_1 + N_2}\right\}.$$

Setting notation for the next lemmas, let n be given. Let Π denote the set of $n \times n$ permutation matrices. Just for now, fix any $P, Q \in \Pi$ such that they are not both the identify matrix, and let τ, ω be their respective associated permutations on $[n]$; i.e. for all $i, j \in [n]$ it

holds that $\tau(i) = j$ precisely when $P_{i,j} = 1$ and, for all $i, j \in [n]$, it holds that $\omega(i) = j$ precisely when $Q_{i,j} = 1$. It will be useful to define the following sets:

$$\Delta := \{(i,j) \in [n] \times [n] : \tau(i) \neq i \text{ or } \omega(j) \neq j\},$$

$$\Delta_t := \{(i,j) \in \Delta : \tau(i) = j \text{ and } \omega(j) = i\},$$

$$\Delta_d := \{(i,j) \in \Delta : i = j \text{ or } \tau(i) = \omega(j)\},$$

$$\Delta_\tau := \{(i,j) \in [n] \times [n] : \tau(i) \neq i\},$$

$$\Delta_\omega := \{(i,j) \in [n] \times [n] : \omega(j) \neq j\}.$$

If we define m to be the maximum of $|\{i \in [n] : \tau(i) \neq i\}|$ and $|\{j \in [n] : \omega(j) \neq j\}|$, then it follows that $mn \leq |\Delta| \leq 2mn$. This is clear from noting that $\Delta_\omega, \Delta_\tau \subseteq \Delta \subseteq \Delta_\tau \cup \Delta_\omega$. Also, $|\Delta_t| \leq m$, since for $(i,j) \in \Delta_t$ it is necessary that $\tau(i) \neq i$ and $\omega(j) \neq j$. Lastly, $|\Delta_d| \leq 4m$, since

$$\Delta_d \subseteq \{(i,j) \in \Delta\} \cup \{(i,j) \in \Delta : i \neq j, \tau(i) = \omega(j)\},$$

and $|\{(i,i) \in \Delta\}| \leq 2m$, and $|\{(i,j) \in \Delta : i \neq j, \tau(i) = \omega(j)\}| \leq 2m$.

We make the following assumption in all that follows:

Assumption 1

Suppose that $\Lambda \in [0,1]^{n \times n}$ is a symmetric, hollow matrix, there is a real number $\rho \in [0,1]$, and there is a constant $\alpha \in (0, 1/2)$ such that $\Lambda_{i,j} \in [\alpha, 1-\alpha]$ for all $i \neq j$, and $(1-\alpha)(1-\rho) < 1/2$. Further, let A, B be the adjacency matrices of two random ρ-correlated Bernoulli(Λ) graphs.

Define the (random) set

$$\Theta' := \{(i,j) \in \Delta : i \neq j, \text{ and } B_{i,j} \neq B_{\tau(i),\omega(j)}\}.$$

Note that $|\Theta'|$ counts the entrywise disagreements induced *within* the off-diagonal part of B by τ and ω.

Lemma 4

Under Assumption 1, if n is sufficiently large then

$$\mathbb{P}(|\Theta'| \notin [\alpha mn/3, 2mn]) \leq 2e^{-\alpha^2 mn/128}.$$

Proof of Lemma 4

For any $(i, j) \in$, note that $(B_{i,j} - B_{\tau(i),\omega(j)})^2$ has a Bernoulli distribution; if $(i, j) \in \ _t \cup \ _d$, then the Bernoulli parameter is either 0 or is in the interval $[\alpha, 1 - \alpha]$, and if $(i, j) \in \ \backslash(\ _t \cup \ _d)$, then the Bernoulli parameter is $\Lambda_{i,j}(1 - \Lambda_{\tau(i),\omega(j)}) + (1 - \Lambda_{i,j})\Lambda_{\tau(i),\omega(j)}$, and this Bernoulli parameter is in the interval $[\alpha, 1 - \alpha]$ since it is a convex combination of values in this interval. Now, $|\Theta'| = \sum_{(i,j) \in \ ,i \neq j}(B_{i,j} - B_{\tau(i),\omega(j)})^2$, so we obtain that $\alpha(|\ | - |\ _t| - |\ _d|) \leq \mathbb{E}|\Theta'|) \leq (1 - \alpha)|\ |$, and thus

$$\alpha m(n - 5) \leq \mathbb{E}(|\Theta'|) \leq 2(1 - \alpha)mn. \quad (1)$$

Next we apply Lemma 2, since $|\Theta'|$ is a function of the at-most $N := 2mn$ Bernoulli random variables $\{B_{i,j}\}_{(i,j) \in \ :i \neq j}$, which as a set (noting that $B_{i,j} = B_{j,i}$ is counted at most once for each $\{i, j\}$) are independent, each with Bernoulli parameter in $[\alpha, 1 - \alpha]$. Furthermore, changing the value of any one of these random variable would change $|\Theta'|$ by at most $\gamma := 4$,

thus Lemma 2 can be applied and, for the choice of $t := \frac{\alpha}{2}mn$, we obtain that

$$\mathbb{P}[||\Theta'| - \mathbb{E}(|\Theta'|)| > \alpha mn/2] \leq 2e^{-\alpha^2 mn/128}. \quad (2)$$

Lemma 4 follows from (1) and (2), since

$$\mathbb{P}[||\Theta'| - \mathbb{E}(|\Theta'|)| > \alpha mn/2]$$
$$= \mathbb{P}[|\Theta'| \notin [\mathbb{E}(|\Theta'|) - \alpha mn/2, \mathbb{E}(|\Theta'|) + \alpha mn/2]]$$
$$\geq \mathbb{P}[|\Theta'| \notin [\alpha m(n - 5) - \alpha mn/2, 2(1 - \alpha)mn + \alpha mn/2]]$$
$$\geq \mathbb{P}[|\Theta'| \notin [\alpha m(n - 5) - \alpha mn/2, 2mn]],$$

and $5\alpha mn/6 \quad \alpha m(n - 5)$ when n is sufficiently large (e.g. $n \quad 30$).

With the above bound on the number of (non-diagonal) entrywise disagreements between B and PBQ^T, we next count the number of additional disagreements introduced by realizing A conditioning on B. In Lemma 5, we prove that this two step realization will almost always result in more entrywise errors than simply realizing A from B without permuting the vertex labels.

Lemma 5

Under Assumption 1, it almost always holds that, for all $P, Q \in \Pi$ such that either $P \quad I$ or Q I, $\|A - PBQ^T\|_F > \|A - B\|_F$.

Note that all (i, j) such that $i = j$ are not in Γ. Also note that $\Theta' \subseteq \Theta$ can be partitioned into the disjoint union $\Theta' = \Theta_1 \cup \Theta_2$.

Equation (4) implies

$$|\Gamma_1| + |\Gamma_2| < (|\Theta_1| + |\Theta_2|)/2 \Rightarrow |\Gamma| < |\Theta|/2 \Rightarrow \|A - B\|_F^2 < \|A - PBQ^T\|_F^2.$$

In particular,

$$\left\{ \|A - B\|_F \geq \|A - PBQ^T\|_F \right\} \Rightarrow \left\{ |\Gamma_1| + |\Gamma_2| \geq (|\Theta_1| + |\Theta_2|)/2 = |\Theta'|/2 \right\}. \quad (5)$$

Now, conditioning on B (hence, conditioning on Θ'), we have, for all $i \ne j$, that (see Section 1.1), $A_{i,j} \sim$ Bernoulli $((1 - \rho)\Lambda_{i,j} + \rho B_{i,j})$. Thus $\mathbb{1}_{A_{i,j} \ne B_{i,j}}$ has a Bernoulli distribution with parameter bounded above by $(1 - \alpha)(1 - \rho)$. Thus, $|\Gamma_1|$ is stochastically dominated by a Binomial$(|\Theta_1|, (1 - \alpha)(1 - \rho))$ random variable, and the independent random variable $|\Gamma_2|$ is stochastically dominated by a Binomial$(|\Theta_2|, (1 - \alpha)(1 - \rho))$ random variable. An application of Lemma 3 with $N_1 := |\Theta_1|$, $N_2 := |\Theta_2|$, $q_1 = q_2 := (1 - \alpha)(1 - \rho)$, and

$t := \left(\dfrac{1}{2} - (1 - \alpha)(1 - \rho) \right) |\Theta'|$, yields (recall that we are conditioning on B here)

$$\begin{aligned}
\mathbb{P}[|\Gamma_1| + |\Gamma_2| &\geq |\Theta'|/2] \\
&= \mathbb{P}\left[|\Gamma_1| + |\Gamma_2| - (1 - \alpha)(1 - \rho)|\Theta'| \geq (1/2 - (1 - \alpha)(1 - \rho)) |\Theta'| \right] \\
&\leq 2 \exp \left\{ \frac{-2(1/2 - (1 - \alpha)(1 - \rho))^2 |\Theta'|^2}{|\Theta_1| + |\Theta_2|} \right\} \\
&\leq 2 \exp \left\{ -2(1/2 - (1 - \alpha)(1 - \rho))^2 |\Theta'| \right\}.
\end{aligned} \quad (6)$$

No longer conditioning (broadly) on B, Lemma 4, equations (5) and (6), and

$(1 - \alpha)(1 - \rho) < \dfrac{1}{2}$, imply that

$$\begin{aligned}
\mathbb{P}\left[\|A - PBQ^T\|_F \leq \|A - B\|_F \right] \\
&\leq \mathbb{P}(|\Theta'| \notin [\alpha mn/3, 2mn]) \\
&\quad + \mathbb{P}[|\Gamma_1| + |\Gamma_2| \geq \tfrac{1}{2}|\Theta'| \,\big|\, |\Theta'| \in [\tfrac{\alpha}{3}mn, 2mn] \\
&\leq 4 \exp \left\{ - \min \left\{ \frac{\alpha^2}{128}, \frac{2\alpha}{3} \left(\frac{1}{2} - (1 - \alpha)(1 - \rho) \right)^2 \right\} mn \right\}.
\end{aligned} \quad (7)$$

Until this point, P and Q—and their associated permutations τ and ω—have been fixed. Now, for each $m \in [n]$, define $\mathscr{H}_m$ to be the event that $\|A - PBQ^T\|_F \le \|A - B\|_F$ for *any* $P, Q \in \Pi$ with the property that their associated permutations τ, ω are such that the maximum of $|$

$\{i \in [n] : \tau(i) \neq i\}|$ and $|\{j \in [n] : \omega(j) \neq j\}|$ is exactly m. There are at most

$$\binom{n}{m} m! \binom{n}{m} m! \leq n^{2m}$$

such permutation pairs.

By (7), for every $m \in [n]$, setting

$$c_1 = \min \{\alpha^2/128, 2\alpha(1/2 - (1-\alpha)(1-\rho))^2/3\},$$

we have $\mathbb{P}(\mathscr{H}_m) \leq n^{2m} \cdot 4 \exp\{-c_1 mn\} \leq \exp\{-c_2 n\}$, for some positive constant c_2 (the last inequality holding when n is large enough). Thus, for sufficiently large n, $\mathbb{P}(\cup_{m=1}^n \mathscr{H}_m) \leq n \cdot \exp\{-c_2 n\}$ decays exponentially in n, and is thus finitely summable over $n = 1, 2, 3, \ldots$. Lemma 5 follows from the Borel-Cantelli Lemma.

Proof of Theorem 1, part a

By Lemma 5, it almost always follows that for every $P, Q \in \Pi$ not both the identity, $\langle AQ, PB \rangle < \langle A, B \rangle$. By the Birkhoff-von Neuman Theorem, $\mathscr{D}$ is the convex hull of Π, i.e., for every $D \in \mathscr{D}$, there exists constants $\{a_{D,P}\}_{P \in \Pi}$ such that $D = \sum_{P \in \Pi} a_{D,P} P$ and $\sum_{P \in \Pi} a_{D,P} = 1$. Thus, if D is not the identity matrix, then almost always

$$\langle AD, DB \rangle = \sum_{P \in \Pi} \sum_{Q \in \Pi} a_{D,P}, a_{D,Q} \langle AQ, PB \rangle < \sum_{P \in \Pi} \sum_{Q \in \Pi} a_{D,P} a_{D,Q} \langle A, B \rangle = \langle A, B \rangle,$$

and almost always $\mathrm{argmin}_{D \in \mathscr{D}} - \langle AD, DB \rangle = \{I\}$.

3 Proof of Theorem 1, part b

The proof will proceed as follows: we will use Lemma 6 to prove that the identity is almost always not a KKT (Karush-Kuhn-Tucker) point of the relaxed graph matching problem. Since the relaxed graph matching problem is a constrained optimization problem with convex feasible region and affine constraints, this is sufficient for the proof of Theorem 1, part b.

First, we state Lemma 6, a variant of Hoeffding's inequality, which we use to prove Theorem 1, part b.

Lemma 6

Let N be a positive integer. Suppose that the random variable X is the sum of N independent random variables, each with mean 0 and each taking values in the real interval $[-1, 1]$. Then for any $t \geq 0$, it holds that

$$\mathbb{P}[|X| \geq t] \leq 2 \cdot e^{\frac{-t^2}{2N}}.$$

Again, without loss of generality, we may assume $P^* = I$. We first note that the convex relaxed graph matching problem can be written as

$$\min \|AD - DB\|_F^2, \quad (8)$$

$$\text{s.t.} D\mathbf{1}=\mathbf{1}, \quad (9)$$

$$\mathbf{1}^T D=\mathbf{1}^T, \quad (10)$$

$$D \geq 0, \quad (11)$$

where (8) is a convex function (of D) subject to affine constraints (9)–(11) (i.e., $D \in \mathscr{D}$). It follows that if I is the global (or local) optimizer of the convex relaxed graph matching problem, then I must be a KKT (Karush-Kuhn-Tucker) point (see, for example, [32, Chapter 4]).

The gradient of $\|AD - DB\|_F^2$ (as a function of D) is

$$\nabla(D) := 2(A^T AD + DBB^T - A^T DB - ADB^T).$$

Hence, a $\hat{D}$ satisfying (9)–(11) (i.e., $\hat{D}$ is primal feasible) is a KKT point if it satisfies

$$\nabla(\hat{D}) + \boldsymbol{\mu} + \boldsymbol{\mu}' - \boldsymbol{\nu}=0, \quad (12)$$

where $\boldsymbol{\mu}, \boldsymbol{\mu}'$, and $\boldsymbol{\nu}$ are as follows:

$$\boldsymbol{\mu} := \begin{bmatrix} \mu_1 & \mu_1 & \cdots & \mu_1 \\ \mu_2 & \mu_2 & \cdots & \mu_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_n & \cdots & \mu_n \end{bmatrix} \in \mathbb{R}^{n \times n},$$

noting that the dual variables $\mu_1, \mu_2, \ldots, \mu_n$ are not restricted. They correspond to the equality primal constraints (9) that the row-sums of a primal feasible D are all one;

$$\boldsymbol{\mu}' := \begin{bmatrix} \mu_1' & \mu_2' & \cdots & \mu_n' \\ \mu_1' & \mu_2' & \cdots & \mu_n' \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1' & \mu_2' & \cdots & \mu_n' \end{bmatrix} \in \mathbb{R}^{n \times n},$$

noting that the dual variables $\mu_1', \mu_2', \ldots, \mu_n'$ are not restricted. They correspond to the equality primal constraints (10) that the column-sums of a primal feasible D are all one;

$$\nu := \begin{bmatrix} 0 & \nu_{1,2} & \cdots & \nu_{1,n} \\ \nu_{2,1} & 0 & \cdots & \nu_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{n,1} & \nu_{n,2} & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n},$$

noting that the dual variables $\nu_{i,j}$ are restricted to be nonnegative. They correspond to the inequality primal constraints (11) that the entries of a primal feasible D be nonnegative. Complementary slackness further constrains the $\nu_{i,j}$, requiring that $\hat{D}_{i,j}\nu_{i,j} = 0$ for all i, j.

At the identity matrix I, the gradient $\nabla(I)$, denoted ∇, simplifies to $\nabla = [\nabla_{i,j}] = 2A^2 + 2B^2 - 4AB \in \mathbb{R}^{n \times n}$; and I being a KKT point is equivalent to:

$$\nabla + \boldsymbol{\mu} + \boldsymbol{\mu}' - \boldsymbol{\nu} = 0, \quad (13)$$

where $\boldsymbol{\mu}, \boldsymbol{\mu}'$, and $\boldsymbol{\nu}$ are as specified above. At the identity matrix, complimentary slackness translates to having $\nu_{1,1} = \nu_{2,2} = \cdots = \nu_{n,n} = 0$.

Now, for Equation (13) to hold, it is necessary that there exist $\mu_1, \mu_2, \mu_1', \mu_2'$ such that

$$\nabla_{1,1} + \mu_1 + \mu_1' = 0, \quad (14)$$

$$\nabla_{2,2} + \mu_2 + \mu_2' = 0, \quad (15)$$

$$\nabla_{1,2} + \mu_1 + \mu_2' \geq 0, \quad (16)$$

$$\nabla_{2,1} + \mu_2 + \mu_1' \geq 0. \quad (17)$$

Adding equations (16), (17) and subtracting equations (14), (15), we obtain

$$\nabla_{1,2} + \nabla_{2,1} \geq \nabla_{1,1} + \nabla_{2,2}. \quad (18)$$

Note that $\frac{1}{2}\nabla + \frac{1}{2}\nabla^T = 2(A - B)^T(A - B)$, hence Equation (18) is equivalent to (where $X := (A - B)^T(A - B)$)

$$2[X]_{1,2} \geq [X]_{1,1} + [X]_{2,2}. \quad (19)$$

Next, referring back to the joint distribution of A and B (see Section 1.1), we have, for all i j,

$$\mathbb{P}[A_{i,j}=0, B_{i,j}=1] = \mathbb{P}[A_{i,j}=1, B_{i,j}=0] = (1-\rho)\Lambda_{i,j}(1-\Lambda_{i,j}).$$

Now, since

$$[X]_{1,1} + [X]_{2,2} \sum_{i \neq 1}(A_{i,1} - B_{i,1})^2 + \sum_{i \neq 2}(A_{i,2} - B_{i,2})^2,$$

is the sum of $(n-1) + (n-1)$ Bernoulli random variables which are collectively independent—besides the two of them which are equal, namely $(A_{12} - B_{12})^2$ and $(A_{21} - B_{21})^2$—we have that $[X]_{1,1} + [X]_{2,2}$ is stochastically greater than or equal to a Binomial $(2n - 3, 2(1 - \rho)\alpha(1- \alpha))$ random variable. Also note that

$$[X]_{1,2} = \sum_{i \neq 1,2}(A_{i,1} - B_{i,1})(A_{i,2} - B_{i,2})$$

is the sum of $n - 2$ independent random variables (namely, the $(A_{i1} - B_{i1})(A_{i2} - B_{i2})$'s) each with mean 0 and each taking on values in $\{-1, 0, 1\}$. Applying Lemma 3 and Lemma 6, respectively, to $X_{11} + X_{22}$ and to X_{12}, with $t := (2n - 3)2(1 - \rho)\alpha(1 - \alpha)/4$, yields

$$\mathbb{P}(2[X]_{1,2} \geq [X]_{1,1} + [X]_{2,2}) \leq \mathbb{P}(2[X]_{1,2} \geq 2t) + \mathbb{P}([X]_{1,1} + [X]_{2,2} \leq 2t) \leq 2 \cdot e^{\frac{-2t^2}{2n-3}} + 2 \cdot e^{\frac{-t^2}{2(n-2)}} \leq e^{-cn},$$

for some positive constant c (the last inequality holds when n is large enough). Hence the probability that Equation (19) holds is seen to decay exponentially in n, and is finitely summable over $n = 1, 2, 3, \ldots$. Therefore, by the Borel-Cantelli Lemma we have that almost always Equation (19) does not hold. Theorem 1, part b is now shown, since Equation (19) is a necessary condition for $I \in \arg\min_{D \in \mathscr{D}} \|AD - DB\|_F^2$.

4 Experimental results

In the preceding section, we presented a theoretical result exploring the trade-off between tractability and correctness when relaxing the graph matching problem. On one hand, we have an optimistic result (Theorem 1, part a) about an indefinite relaxation of the graph matching problem. However, since the objective function is nonconvex, there is no efficient algorithm known to exactly solve this relaxation. On the other hand, Theorem 1, part b, is a pessimistic result about a commonly used efficiently solvable convex relaxation, which almost always provides an incorrect/non-permutation solution.

After solving (approximately or exactly) the relaxed problem, the solution is commonly projected to the nearest permutation matrix. We have not theoretically addressed this projection step yet. It might be that, even though the solution in $\mathcal{D}$ is not the correct permutation, it is very close to it, and the projection step fixes this. We will numerically illustrate this not being the case.

We next present simulations that corroborate and illuminate the presented theoretical results, address the projection step, and provide intuition and practical considerations for solving the graph matching problem. Our simulated graphs have $n = 150$ vertices and follow the Bernoulli model described above, where the entries of the matrix Λ are i.i.d. uniformly distributed in $[\alpha, 1 - \alpha]$ with $\alpha = 0.1$. In each simulation, we run 100 Monte Carlo replicates for each value of ρ. Note that given this α value, the threshold ρ in order to fulfill the hypothesis of the first part of Theorem 1 (namely that $(1 - \alpha)(1 - \rho) < 1/2$) is $\rho = 0.44$. As in Theorem 1, for a fixed $P* \in \Pi$, we let $A' := P*AP*^T$, so that the correct vertex alignment between A' and B is provided by the permutation matrix $P*$.

We then highlight the applicability of our theory and simulations in a series of real data examples. In the first set of experiments, we match three pairs of graphs with known latent alignment functions. We then explore the applicability of our theory in matching graphs without a pre-specified latent alignment. Specifically, we match 16 benchmark problems (those used in [17], [33]) from the QAPLIB library of [34]. See Section 4.3 for more detail. As expected by the theory, in all of our examples a smartly initialized local minimum of the indefinite relaxation achieves best performance.

We summarize the notation we employ in Table 1. To find $D*$, we employ the F-W algorithm ([16], [33]), run to convergence, to exactly solve the convex relaxation. We also use the Hungarian algorithm ([35]) to compute P_c, the projection of $D*$ to Π. To find a local minimum of $\min_{D \in \mathcal{D}} -\langle A'D, DB \rangle$, we use the FAQ algorithm of [17]. We use FAQ:$P*$, FAQ:$D*$, and FAQ:J to denote the FAQ algorithm initialized at $P*$, $D*$, and $J := \mathbf{1} \cdot \mathbf{1}^T/n$ (the barycenter of $\mathcal{D}$). We compare our results to the GLAG and PATH algorithms, implemented with off-the-shelf code provided by the algorithms' authors. We restrict our focus to these algorithms (indeed, there are a *multitude* of graph matching algorithms present in the literature) as these are the prominent relaxation algorithms; i.e., they all first relax the graph matching problem, solve the relaxation, and then project the solution onto Π.

4.1 On the convex relaxed graph matching problem

Theorem 1, part *b*, states that we cannot, in general, expect $D* = P*$. However, $D*$ is often projected onto Π, which could potentially recover $P*$. Unfortunately, this projection step suffers from the same problems as rounding steps in many integer programming solvers, namely that the distance from the best interior solution to the best feasible solution is not well understood.

In Figure 1, we plot $\left\| A'D* - D*B \right\|_F^2$ versus the correlation between the random graphs, with 100 replicates per value of ρ. Each experiment produces a pair of dots, either a red/blue pair or a green/grey pair. The energy levels corresponding to the red/green dots correspond

to $\|A'D^* - D^*B\|_F^2$, while the energies corresponding to the blue/grey dots correspond $\|A'P_c - P_cB\|_F^2$. The colors indicate whether P_c was (green/grey pair) or was not (red/blue pair) P^*. The black dots correspond to the values of $\|A'P^* - P^*B\|_F^2$.

Note that, for correlations $\rho < 1$, $D^* \neq P^*$, as expected from Theorem 1, part *b*. Also note that, even for correlations greater than $\rho = 0.44$, we note $P_c \neq P^*$ after projecting to the closest permutation matrix, even though with high probability P^* is the solution to the unrelaxed problem.

We note the large gap between the pre/post projection energy levels when the algorithm fails/succeeds in recovering P^*, the fast decay in this energy (around $\rho \approx 0.8$ in Figure 1), and the fact that the value for $\|A'P^* - P^*B\|_F^2$ can be easily predicted from the correlation value. These together suggest that $\|A'P_c - P_cB\|_F^2 - \|A'D^* - D^*B\|_F^2$ can be used *a posteriori* to assess whether or not graph matching recovered P^*. This is especially true if ρ is known or can be estimated.

How far is D^* from P^*? When the graphs are isomorphic (i.e., $\rho = 1$ in our setting), then for a large class of graphs, with certain spectral constraints, then P^* is the unique solution of the convex relaxed graph matching problem [14]. Indeed, in Figure 1, when $\rho = 1$ we see that $P^* = D^*$ as expected. On the other hand, we know from Theorem 1, part *b* that if $\rho < 1$, it is often the case that $D^* \neq P^*$. We may think that, via a continuity argument, if the correlation ρ is very close to one, then D^* will be very close to P^*, and P_c will probably recover P^*.

We empirically explore this phenomena in Figure 2. For $\rho \in [0.1, 1]$, with 100 MC replicates for each ρ, we plot the (Frobenius) distances from D^* to P_c (in blue), from D^* to P^* (in red), and from D^* to a uniformly random permutation in Π (in black). Note that all three distances are very similar for $\rho < 0.8$, implying that D^* is very close to the barycenter and far from the boundary of $\mathscr{D}$. With this in mind, it is not surprising that the projection fails to recover P^* for $\rho < 0.8$ in Figure 1, as at the barycenter, the projection onto Π is uniformly random.

For very high correlation values ($\rho > 0.9$), the distances to P_c and to P^* sharply decrease, and the distance to a random permutation sharply increases. This suggests that at these high correlation levels D^* moves away from the barycenter and towards P^*. Indeed, in Figure 1 we see for $\rho > 0.9$ that P^* is the closest permutation to D^*, and is typically recovered by the projection step.

4.2 On indefinite relaxed graph matching problem

The continuous problem one would like to solve, $\min_{D \in \mathscr{D}} -\langle A'D, DB \rangle$ (since its optimum is P^* with high probability), is indefinite. One option is to look for a local minimum of the objective function, as done in the FAQ algorithm of [17]. The FAQ algorithm uses F-W methodology ([16]) to find a local minimum of $-\langle A'D, DB \rangle$. Not surprisingly (as there are many local minima), the performance of the algorithm is heavily dependent on the initialization. Below we study the effect of initializing the algorithm at the non-informative barycenter, at D^* (a principled starting point), and at P^*. We then compare performance of

the different FAQ initializations to the PATH algorithm [33] and to the GLAG algorithm [36].

The GLAG algorithm presents an alternate formulation of the graph matching problem. The algorithm convexly relaxes the alternate formulation, solves the relaxation and projects it onto Π. As demonstrated in [36], the algorithm's main advantage is in matching weighted graphs and multimodal graphs. The PATH algorithm begins by finding D^*, and then solves a sequence of concave and convex problems in order to improve the solution. The PATH algorithm can be viewed as an alternative way of projecting D^* onto Π. Together with FAQ, these algorithms achieve the current best performance in matching a large variety of graphs (see [36], [17], [33]). However, we note that GLAG and PATH often have significantly longer running times than FAQ (even if computing D^* for FAQ:D^*); see [17], [37].

Figure 3 shows the success rate of the graph matching methodologies in recovering P^*. The vertical dashed red line at $\rho = 0.44$ corresponds to the threshold in Theorem 1 part a (above which P^* is optimal whp) for the parameters used in these experiments, and the solid lines correspond to the performance of the different methods: from left to right in gray, FAQ:P^*, FAQ:D^*, FAQ:J; in black, the success rate of P_C; the performance of GLAG and PATH are plotted in blue and red respectively.

Observe that, when initializing with P^*, the fact that FAQ succeeds in recovering P^* means that P^* is a local minimum, and the algorithm did not move from the initial point. From the theoretical results, this was expected for $\rho > 0.44$, and the experimental results show that this is also often true for smaller values of ρ. However, this only means that P^* is a local minimum, and the function could have a different global minimum. On the other hand, for very lowly correlated graphs ($\rho < 0.3$), P^* is not even a local minimum.

The difference in the performance illustrated by the gray lines indicates that the resultant graph matching solution can be improved by using D^* as an initialization to find a local minimum of the indefinite relaxed problem. We see in the figure that FAQ:D^* achieves best performance, while being computationally less intensive than PATH and GLAG, see Figure 4 for the runtime result. This amalgam of the convex and indefinite methodologies (initialize indefinite with the convex solution) is an important tool for obtaining solutions to graph matching problems, providing a computationally tractable algorithm with state-of-the-art performance.

However, for all the algorithms there is still room for improvement. In these experiments, for $\rho \in [0.44, 0.7]$ theory guarantees that with high probability the global minimum of the indefinite problem is P^*, and we cannot find it with the available methods.

When FAQ:D^* fails to recover P^*, how close is the objective function at the obtained local minima to the objective function at P^*? Figure 5 shows $-\langle A'D, DB \rangle$ for the true permutation, P^*, and for the pre-projection doubly stochastic local minimum found by FAQ:D^*. For $0.35 < \rho < 0.75$, the state-of-the-art algorithm not only fails to recover the correct bijection, but also the value of the objective function is relatively far from the optimal one. There is a transition (around $\rho \approx 0.75$) where the algorithm moves from getting a wrong local minimum to obtaining P^* (without projection!). For low values of ρ, the

objective function values are very close, suggesting that both $P*$ and the pre-projection FAQ solution are far from the true global minima. At $\rho \approx 0.3$, we see a separation between the two objective function values (agreeing with the findings in Figure 3). As $\rho > 0.44$, we expect that $P*$ is the global minima and the pre-projection FAQ solution is far from $P*$ until the phase transition at $\rho \approx 0.75$.

4.3 Real data experiments

We further demonstrate the applicability of our theory in a series of real data examples. First we match three pairs of graphs where a latent alignment is known. We further compare different graph matching approaches on a set of 16 benchmark problems (those used in [17], [33]) from the QAPLIB QAP library of [34], where no latent alignment is known a priori. Across all of our examples, an intelligently initialized local solution of the indefinite relaxation achieves best performance.

Our first example is from human connectomics. For 45 healthy patients, we have DT-MRI scans from one of two different medical centers: 21 patients scanned (twice) at the Kennedy Krieger Institute (KKI), and 24 patients scanned (once) at the Nathan Kline Institute (NKI) (all data available at http://openconnecto.me/data/public/MR/MIGRAINE_v1_0/). Each scan is identically processed via the MIGRAINE pipeline of [38] yielding a 70 vertex weighted symmetric graph. In the graphs, vertices correspond to regions in the Desikan brain atlas, which provides the latent alignment of the vertices. Edge weights count the number of neural fiber bundles connecting the regions. We first average the graphs within each medical center and then match the averaged graphs across centers.

For our second example, the graphs consist of the two-hop neighborhoods of the "Algebraic Geometry" page in the French and English Wikipedia graphs. The 1382 vertices correspond to Wikipedia pages with (undirected) edges representing hyperlinks between the pages. Page subject provides the latent alignment function, and to make the graphs of commensurate size we match the intersection graphs.

Lastly, we match the chemical and electrical connectomes of the C. elegans worm. The connectomes consist of 253 vertices, each representing a specific neuron (the same neuron in each graph). Weighted edges representing the strength of the (electrical or chemical) connection between neurons. Additionally, the electrical graph is directed while the chemical graph is not.

The results of these experiments are summarized in Table 2. In each example, the computationally inexpensive FAQ:$D*$ procedure achieves the best performance compared to the more computationally expensive GLAG and PATH procedures. This reinforces the theoretical and simulation results presented earlier, and again points to the practical utility of our amalgamated approach. While there is a canonical alignment in each example, the results point to the potential use of our proposed procedure (FAQ:$D*$) for measuring the strength of this alignment, i.e., measuring the strength of the correlation between the graphs. If the graphs are strongly aligned, as in the KKI-NKI example, the performance of FAQ:$D*$ will be close to the truth and a large portion of the latent alignment with be recovered. As the

alignment is weaker, FAQ:$D*$ will perform even better than the true alignment, and the true alignment will be poorly recovered, as we see in the C. elegans example.

What implications do our results have in graph matching problems without a natural latent alignment? To test this, we matched 16 particularly difficult examples from the QAPLIB library of [34]. We choose these particular examples, because they were previously used in [17], [33] to assess and demonstrate the effectiveness of their respective matching procedures. Results are summarized in Table 3. We see that in every example, the indefinite relaxation (suitably initialized) obtains the best possible result. Although there is no latent alignment here, if we view the best possible alignment as the "true" alignment here, then this is indeed suggested by our theory and simulations. As the FAQ procedure is computationally fast (even initializing FAQ at *both J* and $D*$ is often comparatively faster than GLAG and PATH; see [17] and [37]), these results further point to the applicability of our theory. Once again, theory suggests, and experiments confirm, that approximately solving the indefinite relaxation yields the best matching results.

4.4 Other random graph models

While the random Bernoulli graph model is the most general edge-independent random graph model, in this section we present analogous experiments for a wider variety of edge-dependent random graph models. For these models, we are unaware of a simple way to exploit pairwise edge correlation in the generation of these graphs, as was present in Section 1.1. Here, to simulate aligned non-isomorphic random graphs, we proceed as follows. We generate a graph G_1 from the appropriate underlying distribution, and then model G_2 as an errorful version of G_1; i.e., for each edge in G_1, we randomly flip the edge (i.e., bit-flip from $0 \mapsto 1$ or $1 \mapsto 0$) independently with probability $p \in [0, 1]$. We then graph match G_1 and G_2, and we plot the performance of the algorithms in recovering the latent alignment function across a range of values of p.

We first evaluate the performance of our algorithms on *power law* random graphs [39]; these graphs have a degree distribution that follows a power law, i.e., the proportion of vertices of degree d is proportional to $d^{-\beta}$ for some constant $\beta > 0$. These graphs have been used to model many real data networks, from the Internet [40], [41], to social and biological networks [42], to name a few. In general, these graphs have only a few vertices with high degree, and the great majority of the vertices have relatively low degree.

Figure 6 shows the performance comparison for the methods analyzed above: FAQ:$P*$, FAQ:$D*$, FAQ:J, P_C, PATH, and GLAG. For a range of $p \in [0, 1]$, we generated a 150 vertex power law graph with $\beta = 2$, and subsequently graph matched this graph and its errorful version. For each p, we have 100 MC replicates. As with the random Bernoulli graphs, we see from Figure 6 that the true permutation is a local minimum of the non-convex formulation for a wide range of flipping probabilities ($p \quad 0.3$), implying that in this range of p, G_1 and G_2 share significant common structure. Across all values of $p < 0.5$, FAQ:$P*$ outperforms all other algorithms considered (with FAQ:$D*$ being second best across this range). This echoes the results of Sections (4.1)–(4.3), and suggests an analogue of Theorem 1 may hold in the power law setting. We are presently investigating this.

We next evaluate the performance of our algorithms on graphs with bounded maximum degree (also called *bounded valence graphs*). These graphs have been extensively studied in the literature, and for bounded valence graphs, the graph isomorphism problem is in P[43]. For the experiments in this paper we generate a random graph from the model in [44] with maximum degree equal to 4, and vary the graph order from 50 to 350 vertices. Figure 7 shows the comparison of the different techniques and initializations for these graphs, across a range of bit-flipping parameters $p \in [0, 1]$.

It can be observed that even for isomorphic graphs ($p = 0$), all but FAQ:$P*$ fail to perfectly recover the true alignment. We did not see this phenomena in the other random graph models, and this can be explained as follows. It is a well known fact that convex relaxations fail for regular graphs [13], and also that the bounded degree model tends to generate almost regular graphs [45]. Therefore, even without flipped edges, the graph matching problem with the original graphs is very ill-conditioned for relaxation techniques. Nevertheless, the true alignment is a local minimum of the non-convex formulation for a wide range of values of p (shown by FAQ:$P*$ performing perfectly over a range of p in Figure 7). We again note that FAQ:$D*$ outperforms P_c, PATH and GLAG across all graph sizes and bit-flip parameters p. This suggests that a variant of Theorem 1 may also hold for bounded valence graphs as well, and we are presently exploring this.

We did not include experiments with any random graph models that are highly regular and symmetric (for example, mesh graphs). Symmetry and regularity have two effects on the graph matching problem. Firstly, it is well known that P_c $P*$ for non-isomorphic regular graphs (indeed, J is a solution of the convex relaxed graph matching problem). Secondly, the symmetry of these graphs means that there are potentially several isomorphisms between a graph and its vertex permuted analogue. Hence, any flipped edge could make permutations other than $P*$ into the minima of the graph matching problem.

4.5 Directed graphs

All the theory developed above is proven in the undirected graph setting (i.e., A and B are assumed symmetric). However, directed graphs are common in numerous applications. Figure 8 repeats the analysis of Figure 3 with directed graphs, all other simulation parameters being unchanged. The PATH algorithm is not shown in this new figure because it is designed for undirected graphs, and its performance for directed graphs is very poor. Recall that in Figure 3, i.e., in the undirected setting, FAQ:J performed significantly worse than P_c. In Figure 8, i.e., the directed setting, we note that the performance of FAQ:J outperforms P_c over a range of $\rho \in [0.4, 0.7]$. As in the undirected case, we again see significant performance improvement (over FAQ:J, P_c, and GLAG) when starting FAQ from $D*$ (the convex solution). Indeed, we suspect that a directed analogue of Theorem 1 holds, which would explain the performance increase achieved by the nonconvex relaxation over P_c. Here, we note that the setting for the remainder of the examples considered is the undirected graphs setting.

4.6 Seeded graphs

In some applications it is common to have some *a priori* information about partial vertex correspondences, and seeded graph matching includes these known partial matchings as constraints in the optimization (see [46], [47], [14]). However, seeds do more than just reducing the number of unknowns in the alignment of the vertices. Even a few seeds can dramatically increase performance graph matching performance, and (in the ρ-correlated Erd s-Rényi setting) a logarithmic (in *n*) number of seeds contain enough signal in their seed–to–nonseed adjacency structure to a.s. perfectly align two graphs [47]. Also, as shown in the deterministic graph setting in [14], very often $D*$ is closer to $P*$.

In Figure 9, the graphs are generated from the ρ-correlated random Bernoulli model with random Λ (entrywise uniform over [0.1, 0.9]). We run the Frank-Wolfe method (modified to incorporate the seeds) to solve the convex relaxed graph matching problem, and the method in [46], [47] to approximately solve the non-convex relaxation, starting from *J, D**, and *P**. Note that with seeds, perfect matching is achieved even below the theoretical bound on ρ provided in Theorem 1 (for ensuring *P** is the global minimizer). This provides a potential way to improve the theoretical bound on ρ in Theorem 1, and the extension of Theorem 1 for graphs with seeds is the subject of future research. With the exception of the nonconvex relaxation starting from *P**, each of the different FAQ initializations and the convex formulation all see significantly improved performance as the number of seeds increases. We also observe that the nonconvex relaxation seems to benefit much more from seeds than the convex relaxation. Indeed, when comparing the performance with no seeds, the P_c performs better than FAQ:*J*. However, with just five seeds, this behavior is inverted. Also of note, in cases when seeding returns the correct permutation, we've empirically observed that merely initializing the FAQ algorithm with the seeded start, and not enforcing the seeding constraint, also yields the correct permutation as its solution (not shown).

Figure 10 shows the running time (to obtain a solution) when starting from *D** for the nonconvex relaxation, using different numbers of seeds. For a fixed seed level, the running time is remarkably stable across ρ when FAQ does not recover the true permutation. On the other hand, when FAQ does recover the correct permutation, the algorithm runs significantly faster than when it fails to recover the truth. This suggests that, across all seed levels, the running time might, by itself, be a good indicator of whether the algorithm succeeded in recovering the underlying correspondence or not. Also note that as seeds increase, the overall speed of convergence of the algorithm decreases and, unsurprisingly, the correct permutation is obtained for lower correlation levels.

4.7 Features

Features are additional information that can be utilized to improve performance in graph matching methods, and often these features are manifested as additional vertex characteristics besides the connections with other vertices. For instance, in social networks we may have have a complete profile of a person in addition to his/her social connections.

We demonstrate the utility of using features with the nonconvex relaxation, the standard convex relaxation and the GLAG method, duely modified to include the features into the

optimization. Namely, the new objective function to minimize is $\lambda F(P) + (1 - \lambda)\text{trace}(C^T P)$, where $F(P)$ is the original cost function ($-\langle AP, PB \rangle$ in the nonconvex setting, $\|AP - PB\|_F^2$ for the convex relaxation and $\sum_{i,j} \|([AP]_{i,j}, [PB]_{i,j})\|_2$ for the GLAG method), the matrix C codes the features fitness cost, and the parameter λ balances the trade-off between pure graph matching and fit in the features domain. For each of the matching methodologies, the optimization is very similar to the original featureless version.

For the experiments, we generate ρ-correlated Bernoulli graphs as before, and in addition we generate a Gaussian random vector (zero mean, unit variance) of 5 features for each node of one graph, forming a $5 \times n$ matrix of features; we permute that matrix according to P^* to align new features vectors with the nodes of the second graph. Lastly, additive zero-mean Gaussian noise with a range of variance values is added to each feature matrix independently. If for each vertex $\upsilon \in [n]$ the resulting noisy feature for G_i, $i = 1, 2$, is $x_\upsilon^{(i)}$, then the entries of C are defined to be $C_{\upsilon,w} = \|x_\upsilon^{(1)} - x_w^{(2)}\|_2$, for $\upsilon, w \in [n]$. Lastly, we set $\lambda = 0.5$.

Figure 11 shows the behavior of the methods when using features for different levels of noise in the feature matrix. Even for highly noisy features (recalling that both feature matrices are contaminated with noise), this external information still helps in the graph matching problem. For all noise levels, all three methods improve their performance with the addition of features, and of course, the improvement is greater when the noise level decreases. Note that, as before, *FAQ* outperforms both P_c and GLAG across all noise levels. It is also worth noting that for low noise, FAQ:D^* performs comparably to FAQ:P^*, which we did not observe in the seeded (or unseeded) setting.

Even for modestly errorful features, including these features improves downstream matching performance versus the setting without features. This points to the utility of high fidelity features in the matching task. Indeed, given that the state-of-the-art graph matching algorithms may not achieve the optimal matching for even modestly correlated graphs, the use of external information like seeds and features can be critical.

5 Conclusions

In this work we presented theoretical results showing the surprising fact that the indefinite relaxation (if solved exactly) obtains the optimal solution to the graph matching problem with high probability, under mild conditions. Conversely, we also present the novel result that the popular convex relaxation of graph matching almost always fails to find the correct (and optimal) permutation. In spite of the apparently negative statements presented here, these results have an immediate practical implication: the utility of intelligently initializing the indefinite matching algorithm to obtain a good approximate solution of the indefinite problem.

The experimental results further emphasize the trade-off between tractability and correctness in relaxing the graph matching problem, with real data experiments and simulations in non edge-independent random graph models suggesting that our theory could be extended to more general random graph settings. Indeed, all of our experiments corroborate that best

results are obtained via approximately solving the intractable indefinite problem. Additionally, both theory and examples point to the utility of combining the convex and indefinite approaches, using the convex to initialize the indefinite.

Acknowledgments

References

1. Caelli T, Kosinov S. An eigenspace projection clustering method for inexact graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004; 26(4):515–519. [PubMed: 15382655]

2. Berg, AC.; Berg, TL.; Malik, J. Shape matching and object recognition using low distortion correspondences; 2005 IEEE Conference on Computer Vision and Pattern Recognition; 2005. p. 26-33.

3. Xiao B, Hancock ER, Wilson RC. A generative model for graph matching and embedding. Computer Vision and Image Understanding. 2009; 113(7):777–789.

4. Cho, M.; Lee, KM. Progressive graph matching: Making a move of graphs via probabilistic voting; 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012. p. 398-405.

5. Zhou, F.; De la Torre, F. Factorized graph matching; 2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012. p. 127-134.

6. Huet B, Cross ADJ, Hancock ER. Graph matching for shape retrieval. Advances in Neural Information Processing Systems. 1999:896–902.

7. Cour, T.; Srinivasan, P.; Shi, J. Advances in Neural Information Processing Systems. Vol. 19. MIT; 1998. Balanced graph matching; p. 313-320.2007

8. Garey, M.; Johnson, D. Computers and Intractability: A Guide to the Theory of NP-completeness. Freeman, WH., editor. 1979.

9. Conte D, Foggia P, Sansone C, Vento M. Thirty years of graph matching in pattern recognition. International Journal of Pattern Recognition and Artificial Intelligence. 2004; 18(03):265–298.

10. Torsello A, Hidovic-Rowe D, Pelillo M. Polynomial-time metrics for attributed trees. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2005; 27(7):1087–1099.

11. Ullman JD, Aho AV, Hopcroft JE. The design and analysis of computer algorithms. Addison-Wesley, Reading. 1974; 4:1–2.

12. Hopcroft, JE.; Wong, J-K. Linear time algorithm for isomorphism of planar graphs (preliminary report). Proceedings of the sixth annual ACM symposium on Theory of computing; ACM; 1974. p. 172-184.

13. Fiori M, Sapiro G. On spectral properties for graph matching and graph isomorphism problems. arXiv preprint arXiv:1409.6806. 2014

14. Aflalo Y, Bronstein A, Kimmel R. Graph matching: Relax or not? arXiv:1401.7623. 2014

15. Goldfarb D, Liu S. An $o(n^3 l)$ primal interior point algorithm for convex quadratic programming. Mathematical Programming. 1990; 49(1–3):325–340.

16. Frank M, Wolfe P. An algorithm for quadratic programming. Naval Research Logistics Quarterly. 1956; 3(1–2):95–110.

17. Vogelstein J, Conroy J, Lyzinski V, Podrazik L, Kratzer S, Harley E, Fishkind D, Vogelstein R, Priebe C. Fast approximate quadratic programming for graph matching. arXiv:1112.5507. 2012

18. O'Donnell R, Wright J, Wu C, Zhou Y. Hardness of robust graph isomorphism, Lasserre gaps, and asymmetry of random graphs. arXiv:1401.2436. 2014

19. Atserias A, Maneva E. Sherali–Adams relaxations and indistinguishability in counting logics. SIAM Journal on Computing. 2013; 42(1):112–137.

20. Bollobás, B. Random Graphs. Springer; 1998.

21. Holland PW, Laskey KB, Leinhardt S. Stochastic block-models: First steps. Social Networks. 1983; 5(2):109–137.

22. Snijders TA, Nowicki K. Estimation and prediction for stochastic blockmodels for graphs with latent block structure. Journal of classification. 1997; 14(1):75–100.

23. Nowicki K, Snijders TAB. Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association. 2001; 96(455):1077–1087.

24. Newman ME, Girvan M. Finding and evaluating community structure in networks. Physical review E. 2004; 69(2):026113.

25. Airoldi EM, Blei DM, Fienberg SE, Xing EP. Mixed membership stochastic blockmodels. Advances in Neural Information Processing Systems. 2009:33–40.

26. Cordella LP, Foggia P, Sansone C, Vento M. A (sub) graph isomorphism algorithm for matching large graphs. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2004; 26(10): 1367–1372.

27. Fankhauser S, Riesen K, Bunke H, Dickinson P. Suboptimal graph isomorphism using bipartite matching. International Journal of Pattern Recognition and Artificial Intelligence. 2012; 26(06)

28. Ullmann JR. An algorithm for subgraph isomorphism. Journal of the ACM (JACM). 1976; 23(1): 31–42.

29. Alon N, Kim J, Spencer J. Nearly perfect matchings in regular simple hypergraphs. Israel Journal of Mathematics. 1997; 100:171–187.

30. Kim JH, Sudakov B, Vu VH. On the asymmetry of random regular graphs and random graphs. Random Structures and Algorithms. 2002; 21:216–224.

31. Chung F, Lu L. Concentration inequalities and Martingale inequalities: A survey. Internet Mathematics. 2006; 3:79–127.

32. Bazaraa, M.; Mokhtar, S.; Sherali, H.; Shetty, CM. Nonlinear Programming: Theory and Algorithms. John Wiley & Sons; 2013.

33. Zaslavskiy M, Bach F, Vert J. A path following algorithm for the graph matching problem. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2009; 31(12):2227–2242.

34. Burkard RE, Karisch SE, Rendl F. Qaplib–a quadratic assignment problem library. Journal of Global Optimization. 1997; 10(4):391–403.

35. Kuhn HW. The Hungarian method for the assignment problem. Naval Research Logistic Quarterly. 1955; 2:83–97.

36. Fiori M, Sprechmann P, Vogelstein J, Musé P, Sapiro G. Robust multimodal graph matching: Sparse coding meets graph matching. Advances in Neural Information Processing Systems. 2013; 26:127–135.

37. Lyzinski V, Sussman DL, Fishkind DE, Pao H, Chen L, Vogelstein JT, Park Y, Priebe CE. Spectral clustering for divide-and-conquer graph matching. stat. 2014; 1050:22.

38. Gray WR, Bogovic JA, Vogelstein JT, Landman BA, Prince JL, Vogelstein RJ. Magnetic resonance connectome automated pipeline: an overview. Pulse, IEEE. 2012; 3(2):42–48.

39. Barabási A-L, Albert R. Emergence of scaling in random networks. Science. 1999; 286(5439): 509–512. [PubMed: 10521342]

40. Albert R, Jeong H, Barabási A-L. Internet: Diameter of the world-wide web. Nature. 1999; 401(6749):130–131.

41. Faloutsos, M.; Faloutsos, P.; Faloutsos, C. ACM SIGCOMM Computer Communication Review. Vol. 29. ACM; 1999. On power-law relationships of the internet topology; p. 251-262.

42. Girvan M, Newman ME. Community structure in social and biological networks. Proceedings of the National Academy of Sciences. 2002; 99(12):7821–7826.

43. Luks EM. Isomorphism of graphs of bounded valence can be tested in polynomial time. Journal of Computer and System Sciences. 1982; 25(1):42–65.

44. Balinska K, Quintas L. Algorithms for the random f-graph process. Communications in Mathematical and in Computer Chemistry/MATCH. 2001; (44):319–333.

45. Koponen V. Random graphs with bounded maximum degree: asymptotic structure and a logical limit law. arXiv preprint arXiv:1204.2446. 2012

46. Fishkind DE, Adali S, Priebe CE. Seeded graph matching. arXiv:1209.0367. 2012

47. Lyzinski V, Fishkind DE, Priebe CE. Seeded graph matching for correlated Erdos-Renyi graphs. Journal of Machine Learning Research. 2014; 15:3513–3540. [Online]. Available: http://jmlr.org/papers/v15/lyzinski14a.html.

Biographies

Vince Lyzinski received the B.Sc. degree in mathematics from University of Notre Dame in 2006, the M.Sc. degree in mathematics from John Hopkins University (JHU) in 2007, the M.Sc.E. degree in applied mathematics and statistics from JHU in 2010, and the Ph.D. degree in applied mathematics and statistics from JHU in 2013. From 2013–2014 he was a postdoctoral fellow in the Applied Mathematics and Statistics (AMS) Department at JHU. Since 2014, he has been a Senior Research Scientist at the JHU Human Language Technology Center of Excellence and an Assistant Research Professor in the AMS Department at JHU. His research interests include graph matching, statistical inference on random graphs, pattern recognition, dimensionality reduction, stochastic processes, and high-dimensional data analysis.

Donniell Fishkind received his PhD from the Department of Applied Mathematics and Statistics (AMS) at Johns Hopkins University (JHU) in 1998. After two years on the faculty of the Department of Mathematics and Statistics at University of Southern Maine, he returned to the Department of AMS at JHU, where he has been for the past 15 years, currently with rank Associate Research Professor and with the position of Associate Director of Undergraduate Studies. He is interested in the nexus of linear algebra, graph theory, and optimization. Other projects he has worked on include topology correction for digital images, probabilistic path planning, and season game scheduling for some of the leagues in Minor League Baseball. He has taught over 4000 undergraduate and graduate students at JHU, and was awarded university and/or departmental teaching awards in each of the past 6 years, for a total of 8 awards in this time span.

Marcelo Fiori received the Electrical Engineering (2008), and the M.Sc. degree (2011) from the Universidad de la Repblica, Uruguay (UdelaR), and he is currently a Ph.D. student. He

holds an Assistantship with the Institute of Mathematics at UdelaR. His main research interest are machine learning, graph matching problems, and sparse linear models, with special focus in signal processing.

Joshua T. Vogelstein received a B.S. degree from the Department of Biomedical Engineering (BME) at Washington University in St. Louis, MO in 2002, a M.S. degree from the Department of Applied Mathematics and Statistics (AMS) at Johns Hopkins University (JHU) in Baltimore, MD in 2009, and a Ph.D. degree from the Department of Neuroscience at JHU in 2009. He was a Postdoctoral Fellow in AMS at JHU from 2009 until 2011, at which time he was appointed an Assistant Research Scientist, and became a member of the Institute for Data Intensive Science and Engineering. He spent 2 years at Information Initiative at Duke, before coming home to his current appointment as Assistant Professor in BME at JHU, and core faculty in both the Institute for Computational Medicine and the Center for Imaging Science. His research interests primarily include computational statistics, focusing on ultrahigh-dimensional and non-Euclidean neuroscience data, especially connectomics. His research has been featured in a number of prominent scientific and engineering journals and conferences including Annals of Applied Statistics, IEEE PAMI, NIPS, SIAM Journal of Matrix Analysis and Applications, Science Translational Medicine, Nature Methods, and Science.


Carey E. Priebe received the B.S. degree in mathematics from Purdue University, West Lafayette, IN, in 1984, the M.S. degree in computer science from San Diego State University, San Diego, CA, in 1988, and the Ph.D. degree in information technology (computational statistics) from George Mason University, Fairfax, VA, in 1993. From 1985 to 1994, he worked as a mathematician and scientist in the U.S. Navy research and development laboratory system. Since 1994, he has been a Professor in the Department of Applied Mathematics and Statistics, Whiting School of Engineering, The Johns Hopkins University (JHU), Baltimore, MD. At JHU, he holds joint appointments in the Department of Computer Science, the Department of Electrical and Computer Engineering, the Center for Imaging Science, the Human Language Technology Center of Excellence, and the Whitaker Biomedical Engineering Institute. His research interests include computational statistics, kernel and mixture estimates, statistical pattern recognition, statistical image analysis, dimensionality reduction, model selection, and statistical inference for high-dimensional and graph data. Dr. Priebe is a Lifetime Member of the Institute of Mathematical Statistics, an Elected Member of the International Statistical Institute, and a Fellow of the American Statistical Association.

Guillermo Sapiro was born in Montevideo, Uruguay, on April 3, 1966. He received his B.Sc. (summa cum laude), M.Sc., and Ph.D. from the Department of Electrical Engineering at the Technion, Israel Institute of Technology, in 1989, 1991, and 1993 respectively. After post-doctoral research at MIT, Dr. Sapiro became Member of Technical Staff at the research facilities of HP Labs in Palo Alto, California. He was with the Department of Electrical and Computer Engineering at the University of Minnesota, where he held the position of Distinguished McKnight University Professor and Vincentine Hermes-Luh Chair in Electrical and Computer Engineering. Currently he is the Edmund T. Pratt, Jr. School Professor with Duke University. G. Sapiro works on theory and applications in computer vision, computer graphics, medical imaging, image analysis, and machine learning. He has authored and co-authored over 300 papers in these areas and has written a book published by Cambridge University Press, January 2001. G. Sapiro was awarded the Gutwirth Scholarship for Special Excellence in Graduate Studies in 1991, the Ollendorff Fellowship for Excellence in Vision and Image Understanding Work in 1992, the Rothschild Fellowship for Post-Doctoral Studies in 1993, the Office of Naval Research Young Investigator Award in 1998, the Presidential Early Career Awards for Scientist and Engineers (PECASE) in 1998, the National Science Foundation Career Award in 1999, and the National Security Science and Engineering Faculty Fellowship in 2010. He received the test of time award at ICCV 2011. G. Sapiro is a Fellow of IEEE and SIAM. G. Sapiro was the founding Editor-in-Chief of the SIAM Journal on Imaging Sciences.
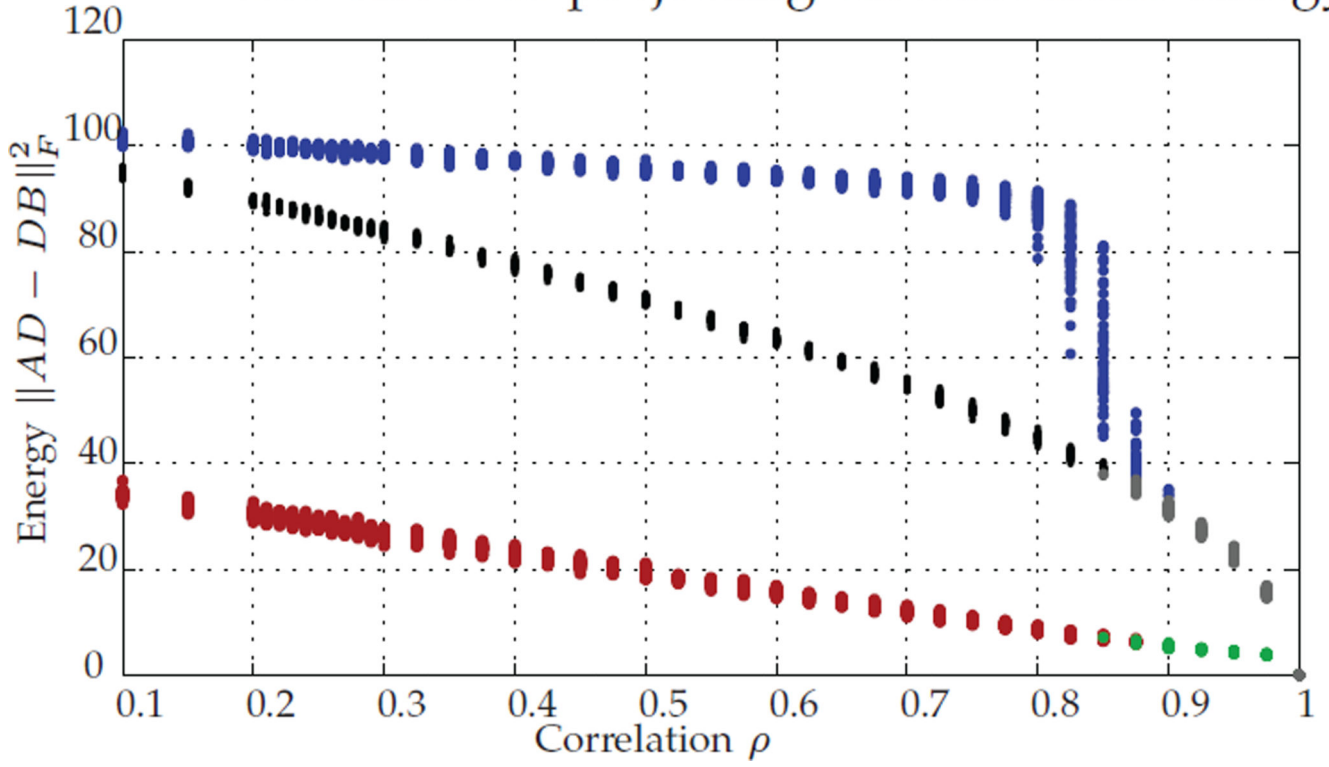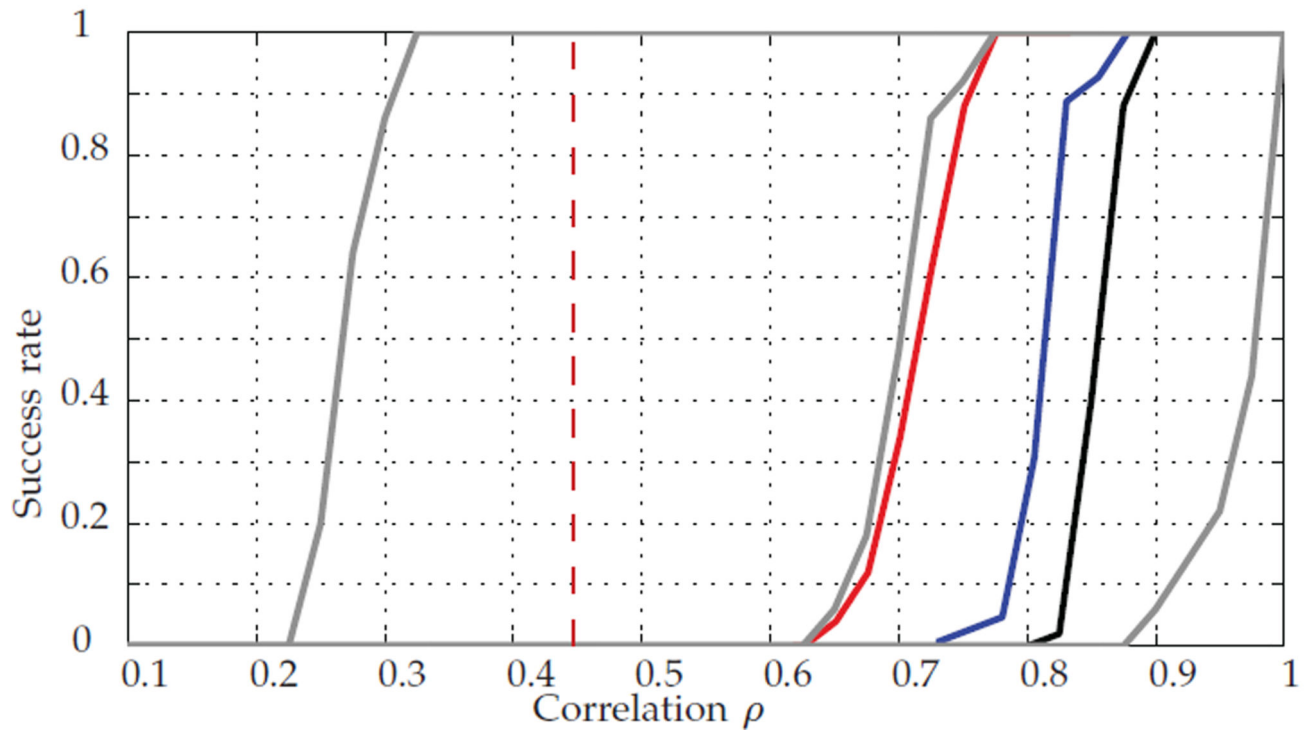
Fig. 1.

For ρ ∈ [0.1, 1], we plot $\|A'D^* - D^*B\|_F^2$ (red/green) and $\|A'P_c - P_cB\|_F^2$ (blue/gray). Red/blue dots correspond to simulations where $P_c \neq P^*$, and grey/green dots to $P_c = P^*$. Black dots correspond to $\|A'P^* - P^*B\|_F^2$. For each ρ, we ran 100 MC replicates.
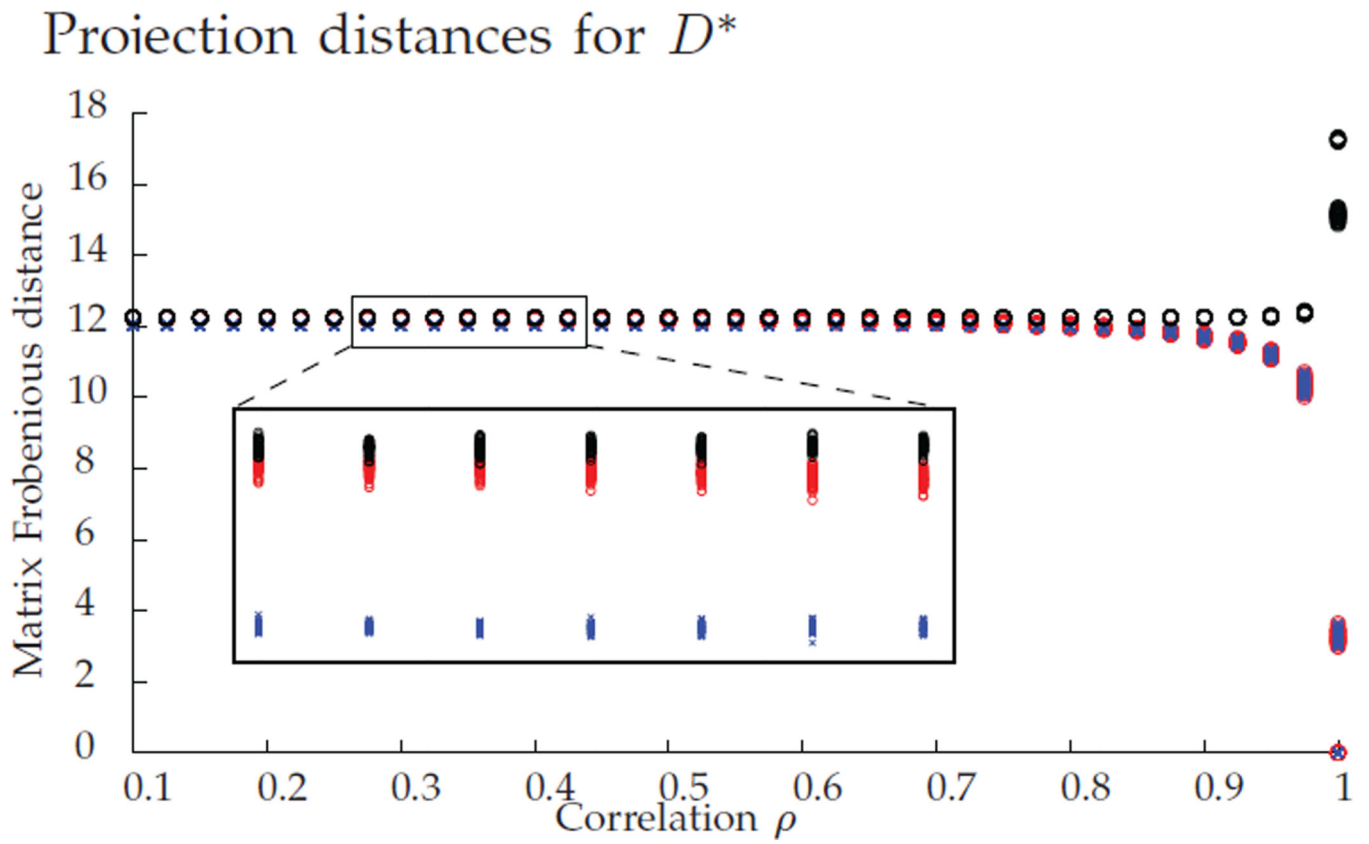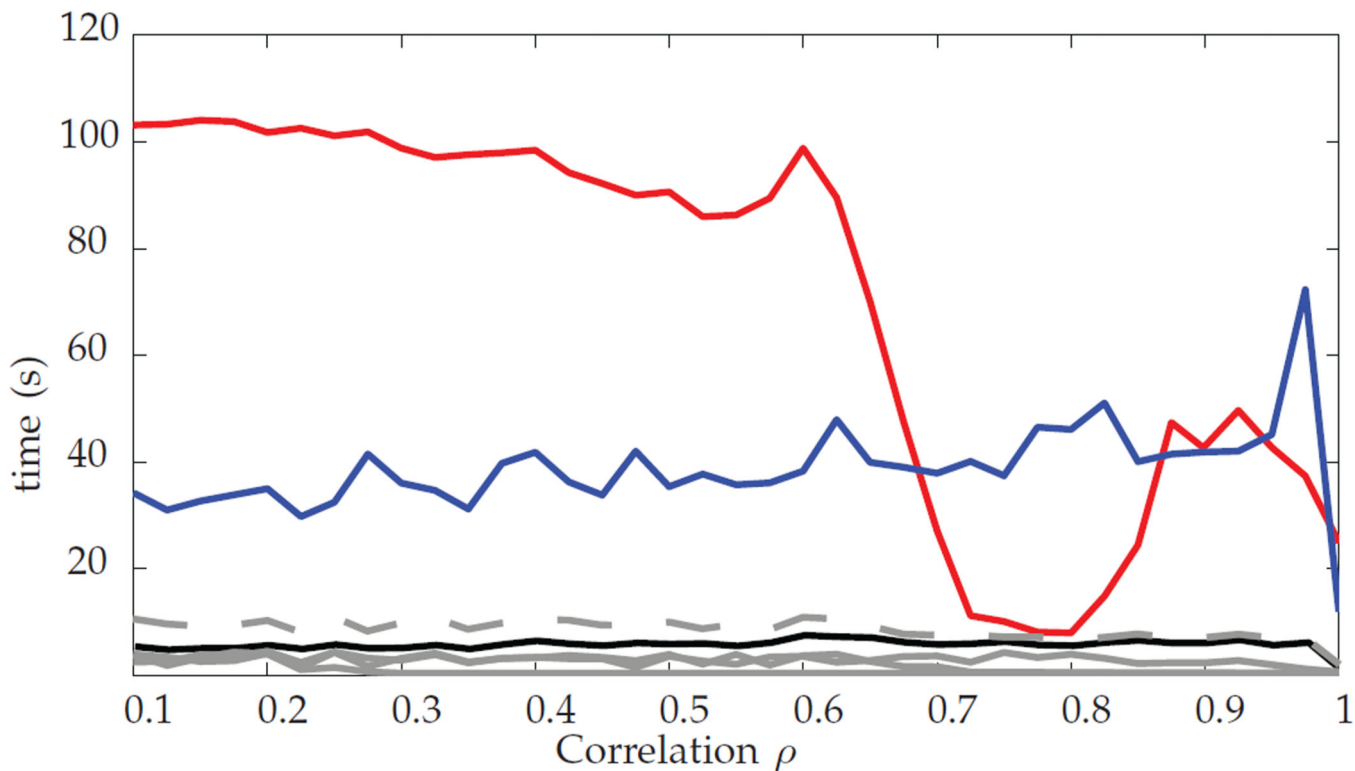
Fig. 2.
Distance from D^* to P_c (in blue), to P^* (in red), and to a random permutation (in black). For each value of ρ, we ran 100 MC replicates.

Fig. 3.

Success rate in recovering P^*. In gray, FAQ starting at, from left to right, P^*, D^*, and J; in black, P_c; in red, PATH; in blue, GLAG. For each ρ, we ran 100 MC replicates.

Running times across methods

Fig. 4.

Average run times. In solid gray, FAQ:D^* (note that this does not include the time to find D^*) and FAQ:J; finding P_c (first finding D^*) in black; PATH in red; and GLAG in blue. In dashed gray, the total time for FAQ:D^* (including the time to find D^*). Note that the runtime of PATH drops precipitously at $\rho = 0.6$, which corresponds to the performance increase in Figure 3
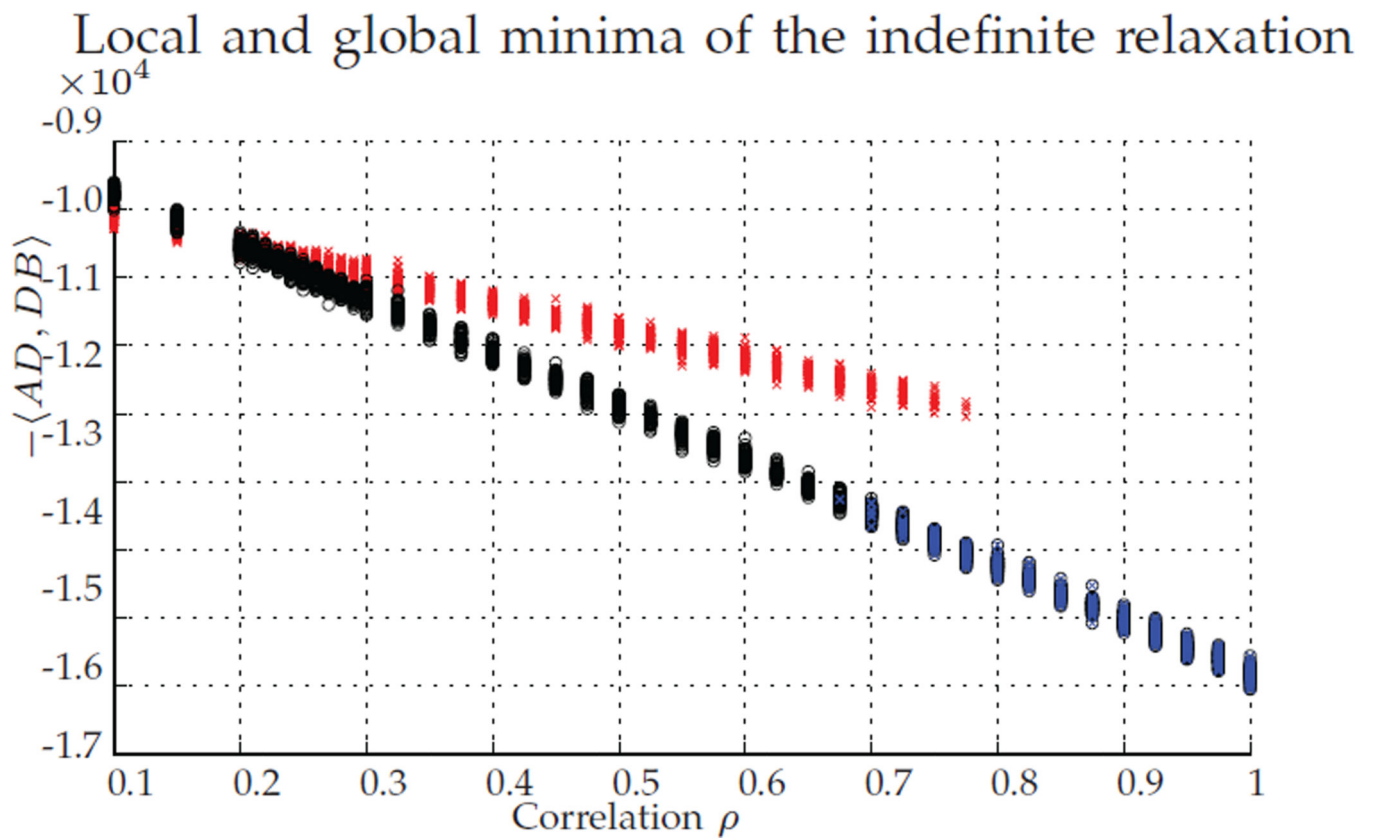
Fig. 5.
Value of $-\langle A'D, DB \rangle$ for $D = P*$ (black) and for the output of FAQ:$D*$ (red/blue indicating failure/success in recovering the true permutation). For each ρ, we ran 100 MC replicates.
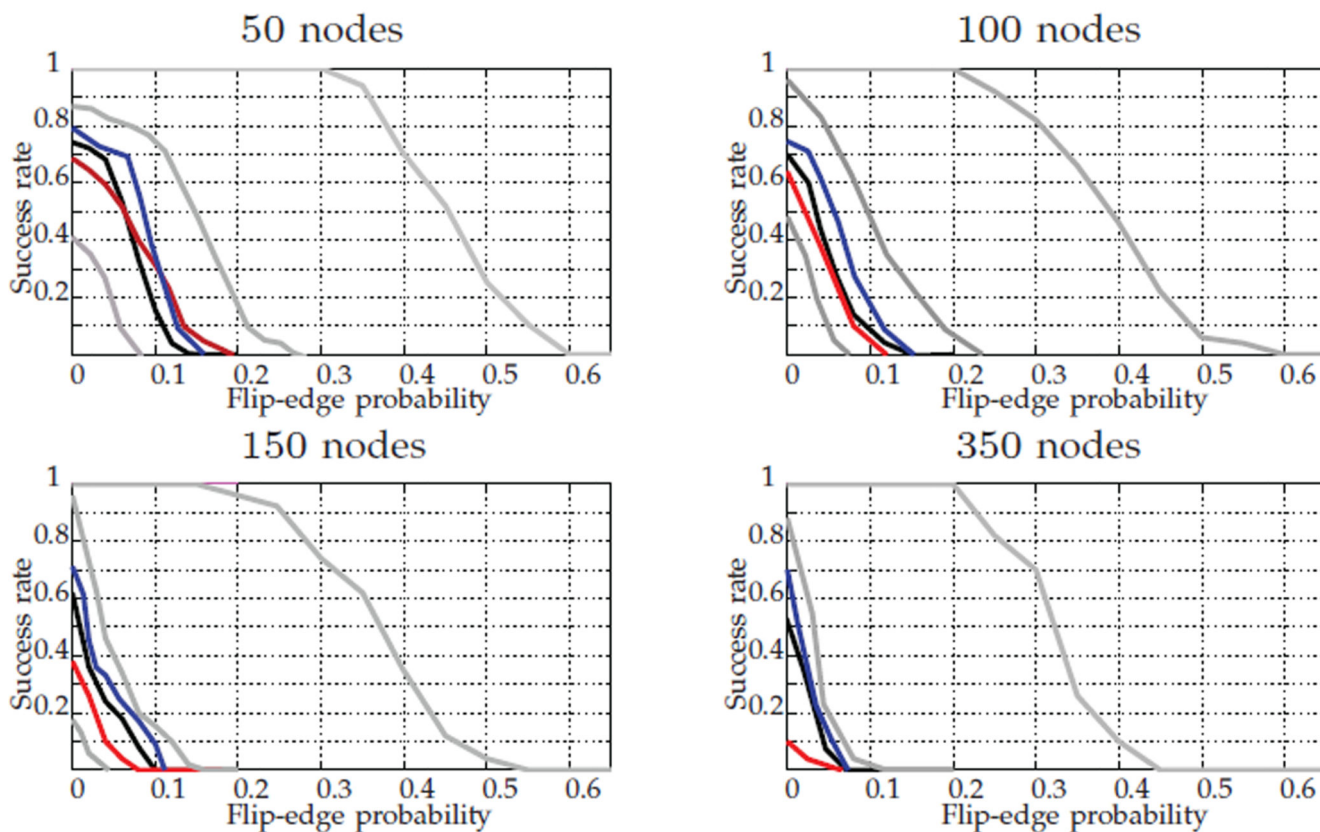
Matching performance for power law graphs
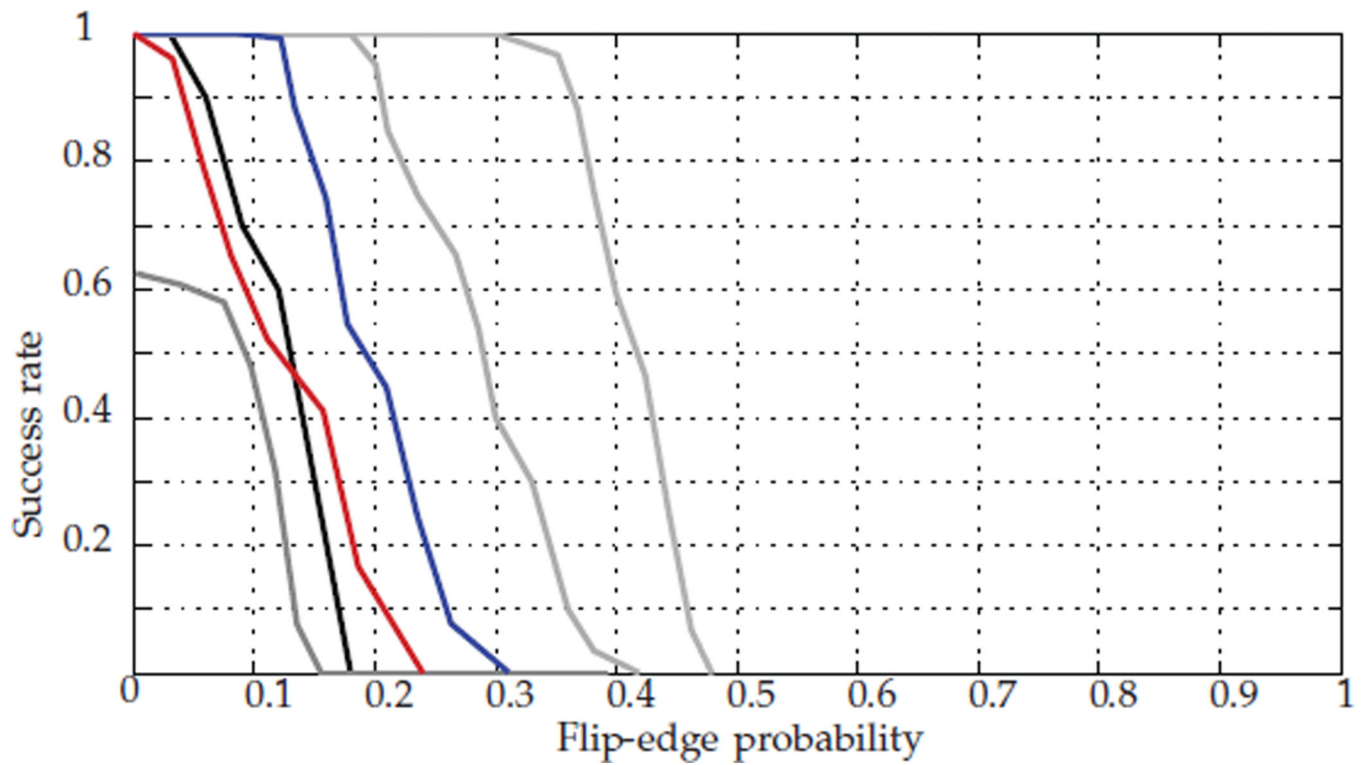
Fig. 6.
Success rate in recovering $P*$ for 150 vertex power law graphs with $\beta = 2$ for: In gray, from right to left, FAQ:$P*$, FAQ:$D*$, and FAQ:J; in black, P_c; in red, PATH; in blue, GLAG. For each value of the bit-flip parameter p, we ran 100 MC replicates.
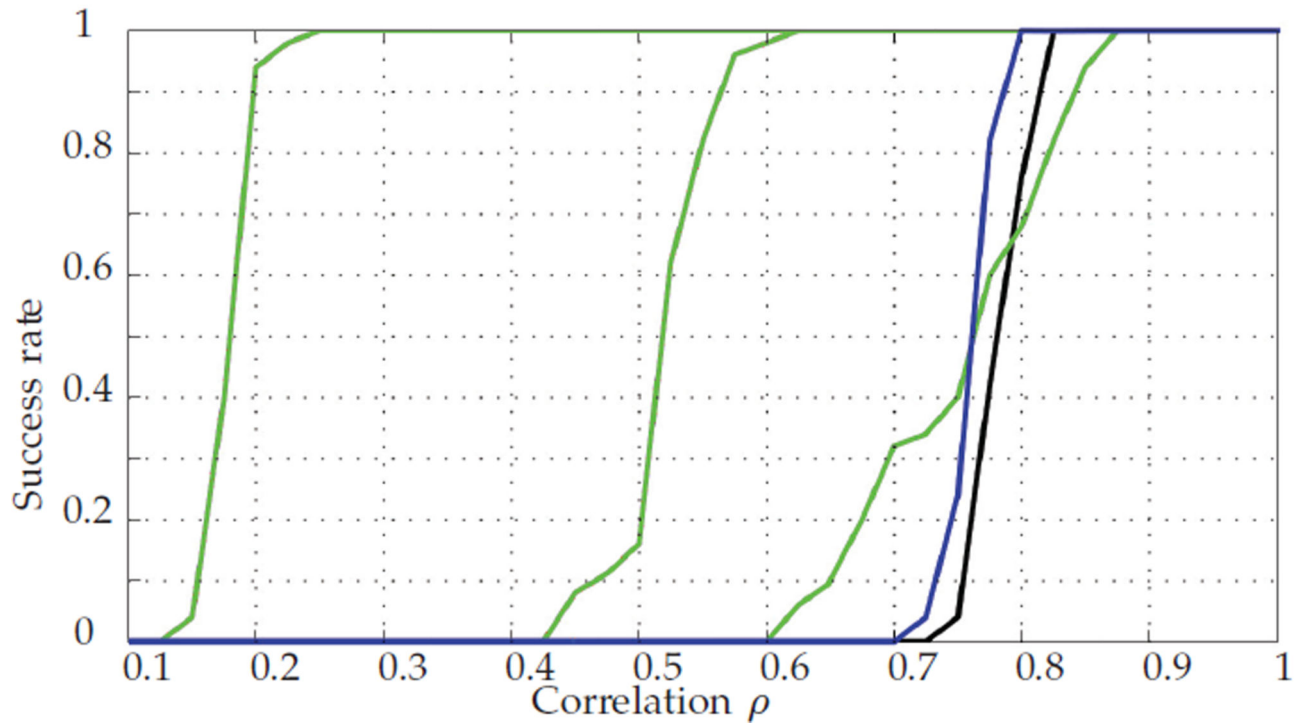
Success rates for bounded degree graphs

Fig. 7.
Success rate in recovering P^* for bounded degree graphs (max degree 4). In gray, from right to left, FAQ:P^*, FAQ:D^*, and FAQ:J; in black, P_c; in red, PATH; in blue, GLAG. For each probability we ran 100 MC replicates.

Fig. 8.
Success rate for directed graphs. We plot P_c (black), the GLAG method (blue), and the nonconvex relaxation starting from different points in green, from right to left: FAQ:J, FAQ:D^*, FAQ:P^*.

The effect of seeding on matching accuracy
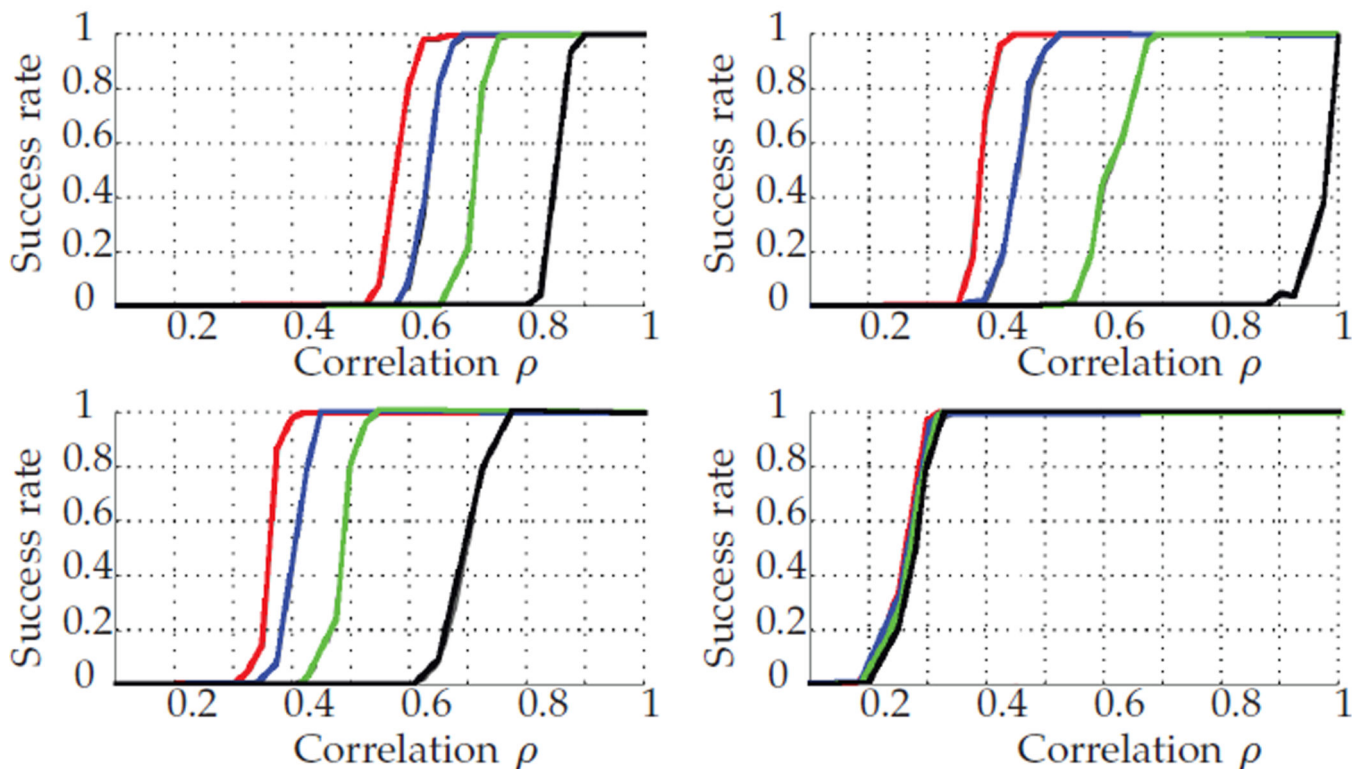
Fig. 9.
Success rate of different methods using seeds. We plot P_c (top left), FAQ:J (top right), FAQ:D* (bottom left), and FAQ:P* (bottom right). For each method, the number of seeds increases from right to left: 0 (black), 5 (green), 10 (blue) and 15 (red) seeds. Note that more seeds increases the success rate across the board.
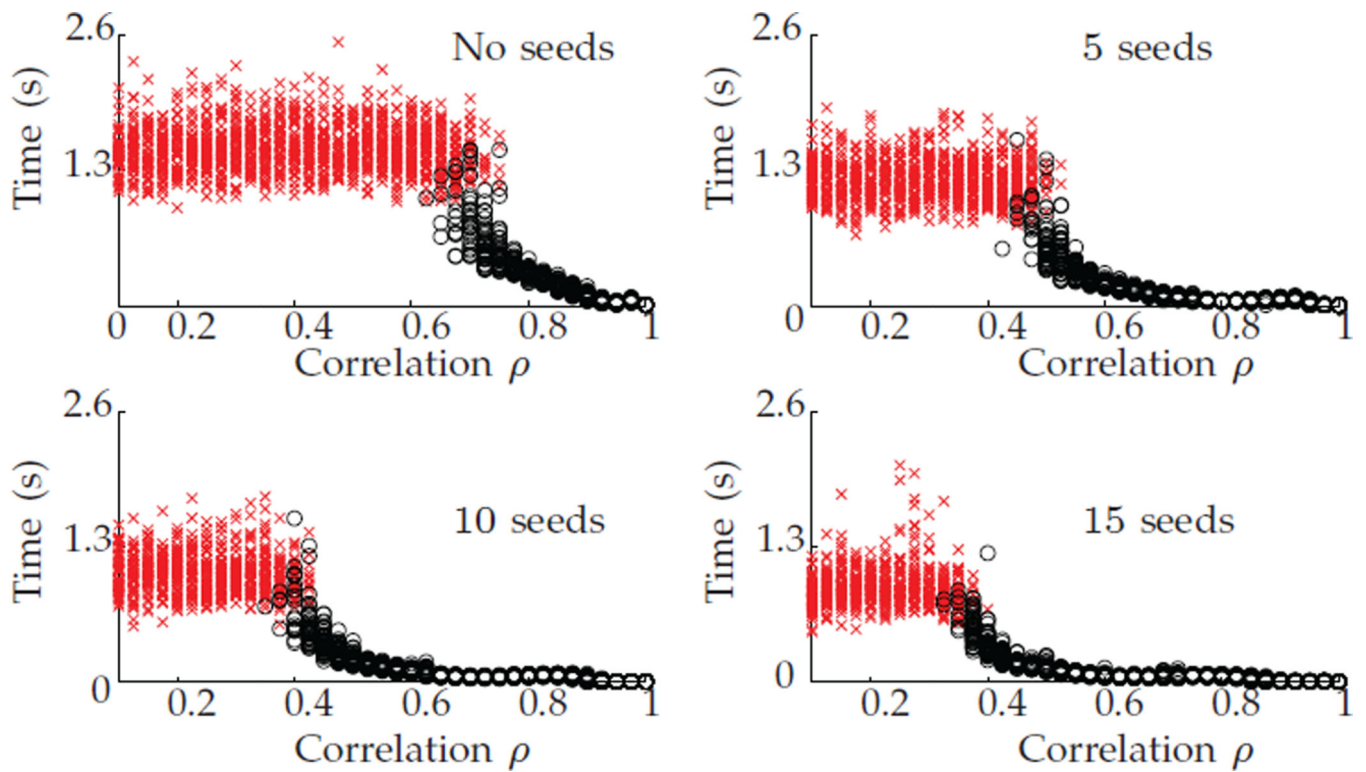
Fig. 10.
Running time for the nonconvex relaxation when starting from D^*, for different number of seeds. A red "x" indicates the algorithm failed to recover P^*, and a black "o" indicates it succeeded. In each, the algorithm was run to termination at discovery of a local min.

The effect of features on matching accuracy
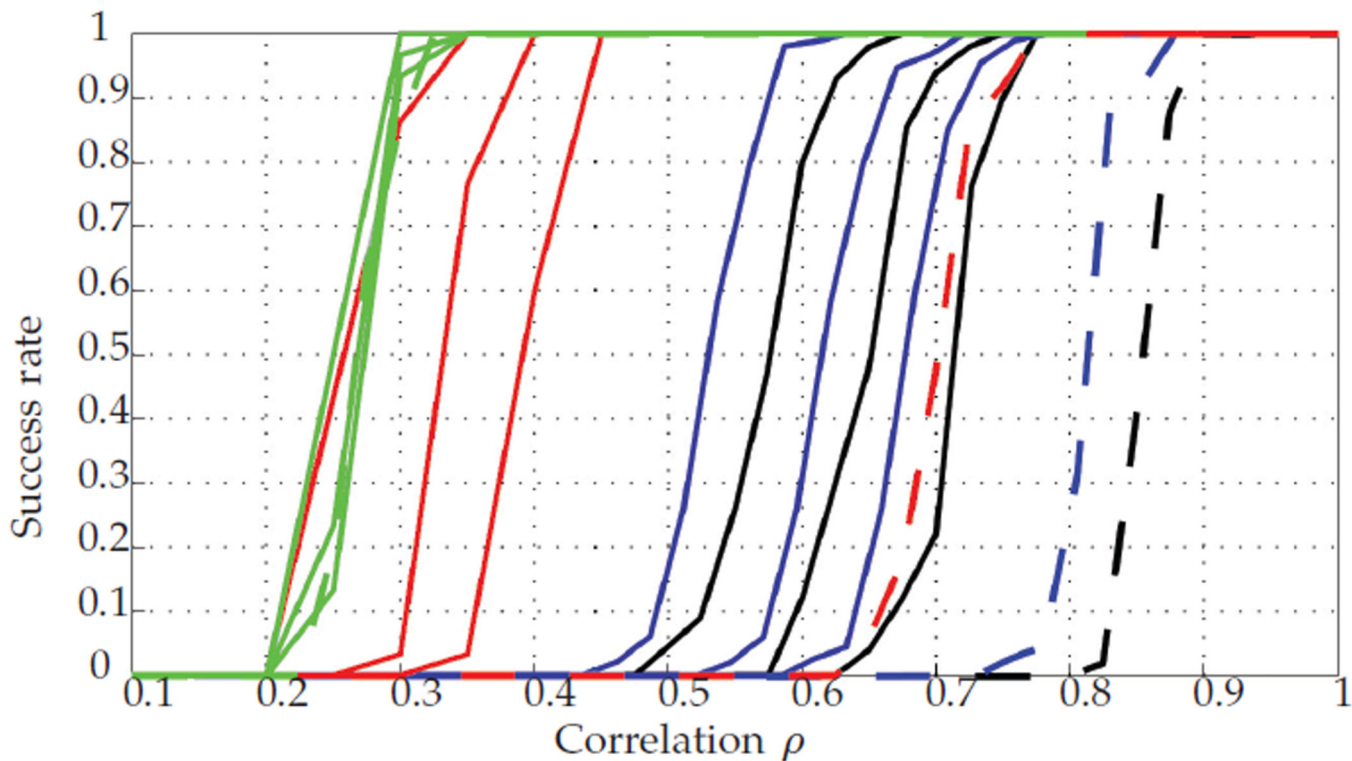
Fig. 11.
Success rate of different methods using features: P_c (in black), GLAG (in blue), FAQ:$D*$ (in red), and FAQ:$P*$ (in green). For each method, the noise level (variance of the Gaussian random noise) increases from left to right: 0.3, 0.5, and 0.7. In dashed lines, we show the success of the same methods without features.

TABLE 1

Notation

Notation	Algorithm used	Ref.
$D^* \in \operatorname{argmin}_{D \in \mathscr{D}} \lVert A'D - DB \rVert_F^2$	F-W algorithm run to convergence	[16], [33]
P_c = projecting D^* to Π	Hungarian algorithm	[35]
FAQ:$P*$	FAQ init. at $P*$	[17]
FAQ:$D*$	FAQ init. at $D*$	[17]
FAQ:J	FAQ init. at J	[17]

TABLE 2

$\|A'P - PB\|_F$ for the P given by each algorithm together with the number of vertices correctly matched ($n_{corr.}$) in real data experiments

Algorithm		KKI-NKI	Wiki.	C. elegans
Truth	$\|A'\,P - PB\|_F$	82892.87	189.35	155.00
	$n_{corr.}$	70	1381	253
Convex relax.	$\|A'\,P - PB\|_F$	104941.16	225.27	153.38
	$n_{corr.}$	41	97	2
GLAG	$\|A'\,P - PB\|_F$	104721.97	219.98	145.53
	$n_{corr.}$	36	181	4
PATH	$\|A'\,P - PB\|_F$	165626.63	252.55	158.60
	$n_{corr.}$	1	1	1
FAQ:J	$\|A'\,P - PB\|_F$	93895.21	205.28	127.55
	$n_{corr.}$	38	30	1
FAQ:D*	$\|A'\,P - PB\|_F$	**83642.64**	**192.11**	**127.50**
	$n_{corr.}$	**63**	**477**	**5**

TABLE 3

$\left\| A'P - PB \right\|_F^2$ for the different tested algorithms on 16 benchmark examples of the QAPLIB library.

QAP	OPT	Convex rel.	GLAG	PATH	Non-Convex. Initialization:	
					Barycenter	**Convex sol**
chr12c	11156	21142	61430	18048	**13088**	13610
chr15a	9896	41208	78296	19086	29018	**16776**
chr15c	9504	47164	82452	16206	**11936**	18182
chr20b	2298	9912	13728	5560	**2764**	3712
chr22b	6194	10898	21970	8500	8774	**7332**
esc16b	292	314	320	300	314	**292**
rou12	235528	283422	353998	256320	254336	**254302**
rou15	354210	413384	521882	391270	371458	**368606**
rou20	725522	843842	1019622	778284	759838	**754122**
tai10a	135028	175986	218604	152534	157954	**149560**
tai15a	388214	459480	544304	419224	**397376**	397926
tai17a	491812	606834	708754	530978	520754	**516492**
tai20a	703482	810816	1015832	753712	**736140**	756834
tai30a	1818146	2089724	2329604	1903872	1908814	**1858494**
tai35a	2422002	2859448	3083180	2555110	2531558	**2524586**
tai40a	3139370	3727402	4001224	3281830	**3237014**	3299304