On convex relaxation of graph isomorphism

Yonathan Aflalo[a,1], Alexander Bronstein[b,1,2], and Ron Kimmel[a,1]

[a]Department of Computer Science, Technion, Haifa 32000, Israel; and [b]School of Electrical Engineering, Tel Aviv University, Ramat Aviv 69978, Israel

We consider the problem of exact and inexact matching of weighted undirected graphs, in which a bijective correspondence is sought to minimize a quadratic weight disagreement. This computationally challenging problem is often relaxed as a convex quadratic program, in which the space of permutations is replaced by the space of doubly stochastic matrices. However, the applicability of such a relaxation is poorly understood. We define a broad class of friendly graphs characterized by an easily verifiable spectral property. We prove that for friendly graphs, the convex relaxation is guaranteed to find the exact isomorphism or certify its inexistence. This result is further extended to approximately isomorphic graphs, for which we develop an explicit bound on the amount of weight disagreement under which the relaxation is guaranteed to find the globally optimal approximate isomorphism. We also show that in many cases, the graph matching problem can be further harmlessly relaxed to a convex quadratic program with only n separable linear equality constraints, which is substantially more efficient than the standard relaxation involving $2n$ equality and n^2 inequality constraints. Finally, we show that our results are still valid for unfriendly graphs if additional information in the form of seeds or attributes is allowed, with the latter satisfying an easy to verify spectral characteristic.

graph isomorphism | graph matching | permutation | convex relaxation

Graphs are a natural abstraction in a variety of problems and are particularly useful for modeling structures, frequently arising in different domains of science and engineering. In many applications, graphs have to be compared or brought into correspondence. The term "graph isomorphism" or the less precise term "graph matching" (used mainly in the applied community) refers to a class of computational problems consisting of finding an optimal correspondence between the vertices of two graphs that minimizes adjacency disagreement. The uses of graph models in general and graph matching in particular are too numerous to allow a comprehensive review within the scope of this paper. In what follows, we will just list a few prominent ones, referring the reader to a (partial) review of applications with a particular emphasis on the domain of pattern recognition (1). In computer vision and pattern recognition, graph matching is used for stereo vision and 3D reconstruction (2), object detection and recognition (3, 4)—in particular, optical character recognition (5)—and image and video indexing and retrieval (6). In biometric applications, graph-based techniques have been widely used for identification tasks implemented by means of elastic graph matching. These include, among others, face recognition and pose estimation (7) and fingerprint recognition (8). In biomedical applications, graphs have been used to model vascular structures and, more recently, to represent connections between neurons (9). In data mining, graphs are used to model networks, including the Web and social networks (10).

Despite the tremendous popularity of graph models, graph matching remains a computationally intensive task. In the strict sense, it is computationally intractable as no polynomial algorithms are known for its solution, except for graphs admitting certain particular structures. The increase in the available computational power of modern computers and the remarkable development of numerous efficient graph matching heuristics have made graph matching feasible for relatively large graphs, counting

about a thousand vertices. However, novel applications such as the analysis of brain graphs—the so-called "connectomes"—and social networks require matching of graphs with millions if not billions of vertices. These scales are far beyond the reach of existing heuristics. Furthermore, a major disadvantage of graph matching heuristics is that, in general, they are not guaranteed to find the optimal matching, or at least to guarantee how far the found matching is from the optimal one.

Contributions

In this paper, we focus on the family of scalable graph matching heuristics based on continuous (in particular, convex) optimization (11). We analyze the standard convex relaxation of the graph-matching problem based on replacing the space of permutations by the space of doubly stochastic matrices, and make the following contributions.

First, we establish conditions under which the relaxation is equivalent to exact graph matching, in the sense that it is guaranteed to find the exact graph isomorphism if such exists, or certify its inexistence (Theorem 1). The class of graphs on which such an equivalence holds is characterized by an easily verifiable spectral property we call "friendliness," and is surprisingly large—practically, as large as the class of asymmetric graphs.

Second, we generalize this result to inexact graph matching, providing an explicit bound on the amount of weight disagreement under which the relaxation is guaranteed to find the globally optimal approximate isomorphism (Theorem 2).

Third, we show that equivalence of convex relaxation to exact graph matching still holds for unfriendly graphs if additional information besides the graph adjacencies is allowed to disambiguate the symmetries. Specifically, we consider such additional information in the form of a collection of knowingly corresponding functions (seeds) or vector-valued vertex attributes, and show a constructive spectral condition on the seeds/attributes under which convex relaxation of seeded/attributed graph matching is guaranteed to find one of the isomorphisms (Theorem 3).

These three contributions establish the boundaries of applicability of the convex relaxation, which have so far been poorly understood. Finally, a byproduct of our analysis is the fact that the former results are also satisfied by a simpler convex relaxation, in which the space of permutations is replaced with an affine space of matrices we call "pseudostochastic." This alternative relaxation leads to a simpler, and potentially more efficiently solvable, optimization problem.

Notation

The following notation will be used in the rest of the paper: Vectors and matrices are denoted in bold lower and uppercase, respectively, and their elements by lower and uppercase italic with appropriate subscript indices. The norm $\|\cdot\|$ will denote the standard ℓ_2 norm of a vector, and the spectral norm of a matrix (to be distinguished from the Frobenius norm, specified with the subscript $_F$). Constraints in optimization problems will be specified as "s.t." standing for "subject to". Throughout the paper, the not so rigorous term "matching" refers to the exact or inexact graph isomorphism problems rather than to the graph-theoretic notion of an independent edge set.

Graph Matching

Let $\mathcal{A} = (V, \mathbf{A})$ and $\mathcal{B} = (V, \mathbf{B})$ be two undirected graphs built upon a common vertex set V, which, for convenience, is assumed to be $V = \{1, \ldots, n\}$. As $\mathcal{A}$ and $\mathcal{B}$ are fully represented by the symmetric $n \times n$ adjacency matrices $\mathbf{A}$ and $\mathbf{B}$, we will use the two notations interchangeably. We allow the adjacency matrices to contain nonbinary edge weights, and henceforth consider this case without explicitly specifying that the graphs are weighted. Let us denote by $\mathcal{P}(n) = \{\pi : V \to V\}$ the space of vertex permutations, which can be equivalently represented by $n \times n$ permutation matrices of the form $\{\mathbf{\Pi} \in \{0,1\}^{n \times n} : \mathbf{\Pi 1} = \mathbf{\Pi}^T \mathbf{1} = \mathbf{1}\}$, with $\mathbf{1}$ denoting a column vector of ones. With some abuse of notation, we will refer to both spaces as $\mathcal{P}(n)$, dropping the n whenever possible. A permutation π represents a bijective correspondence between the two graphs, mapping each vertex i in $\mathcal{A}$ to a vertex π_i in $\mathcal{B}$. Similarly, for each edge (i,j), the correspondence pulls back the adjacency weight b_{π_i, π_j}. The latter can be stated equivalently by constructing a new adjacency matrix $\mathbf{\Pi}^T \mathbf{B} \mathbf{\Pi}$, where $\mathbf{\Pi}$ is the permutation matrix representing π. To measure the adjacency disagreement under correspondence, we define on $\mathcal{P}$ a distortion function of the form $\mathrm{dis}_{\mathcal{A} \mapsto \mathcal{B}}(\mathbf{\Pi}) = \|\mathbf{A} - \mathbf{\Pi}^T \mathbf{B} \mathbf{\Pi}\|$, with $\|\cdot\|$ denoting some norm (for brevity, we will drop $\mathcal{A} \mapsto \mathcal{B}$ whenever possible). The graphs are said to be isomorphic if there exists a zero-distortion permutation. We denote the collection of all isomorphisms relating $\mathcal{A}$ and $\mathcal{B}$ by $\mathrm{Iso}(\mathcal{A} \mapsto \mathcal{B}) = \{\mathbf{\Pi} : \mathrm{dis}(\mathbf{\Pi}) = 0\}$.

In this notation, the graph-matching (GM) problem consists of finding a zero-distortion permutation; such a permutation might not be unique if the graph possesses symmetries, as we clarify in the sequel. The closely related graph isomorphism (GI) problem consists of verifying whether a zero-distortion permutation exists. This strict setting is usually referred to as "exact." Since, in practical applications, the matched graphs might be contaminated by noise, GM is frequently stated in the "inexact" flavor, consisting of finding a minimum rather than zero-distortion permutation. It is worthwhile noting that the formulation of GM based on a norm of the adjacency disagreement is extremely popular in computer vision, shape analysis (12), and neuroscience (13), where graphs are used to represent geometric structures, and the matching distortion can be interpreted as the strength of geometric deformation. While we focus exclusively on this class of problems, distortion criteria for partial and approximate isomorphism, as hinted in Fig. 1, deserve a deeper discussion beyond the scope of this paper. Several alternative formulations of inexact GM, particularly those based on edit distance (14) and maximum common subgraph (15, 16), have been extensively addressed in the literature.
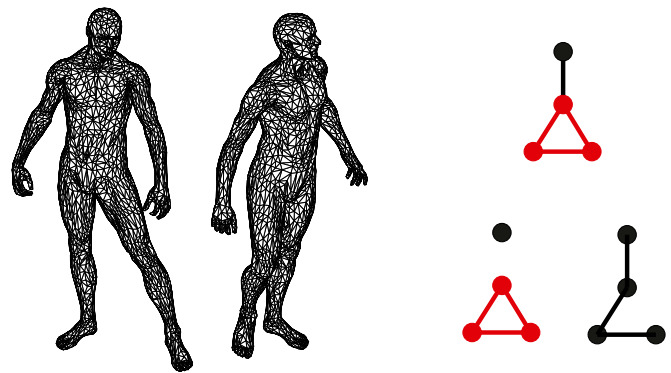

Fig. 1. (Left) Two deformable shapes are represented as weighted graphs with edge weights proportional to the pairwise geodesic distances between the corresponding vertices. Optimal isomorphism in the sense of dis = $\|\mathbf{A} - \mathbf{\Pi}^T \mathbf{B} \mathbf{\Pi}\|$ quantifies how isometric these two discrete metric spaces are and is related to the Gromov–Hausdorff distance (12). (Right) Limitation of the former distortion criterion. The bottom graphs are both 1-isomorphic to the top one, without distinguishing between exact partial isomorphism (Bottom Left; isomorphic subgraphs are marked in red) and inexact full isomorphism (Bottom Right).

Computationally, GM is at least as hard as GI, which is an NP problem (i.e., a problem whose solution can be verified in polynomial time) presently not known either to be solvable in polynomial time or to be NP complete. In fact, GI is one of the few problems which, if P ≠ NP, might reside in an intermediate "GI-complete" complexity class (17). However, the GI problem is known to be only "moderately exponential" (18); furthermore, polynomial (and even linear) time algorithms exist for checking the isomorphism of various particular types of graphs, such as planar graphs (19), graphs with bounded vertex degree (20), and trees (21). However, the constants characterizing the complexity of such algorithms are extremely large; for example, the linear time algorithm for checking the isomorphism of graphs with vertex degree bounded by 2 is over 2×10^6! Moreover, these results are largely inapplicable to inexact or weighted GM. Because of this fact, exact GM is not used in practical applications involving even moderately scaled graphs, except for very particular cases. Instead, various types of heuristics are usually used.

The common property of heuristic algorithms is that they often perform well on real problems and scale to large graphs at the expense of having no theoretical guarantee to converge to the true global minimizer of the GM problem. The wealth of literature dedicated to GM heuristics counts hundreds of studies published in the past four decades, and we will not attempt to review it within the scope of this paper. Instead, we refer the reader to ref. 1 for a comprehensive review, and focus on the popular class of continuous optimization relaxations. In these heuristics, the combinatorial GM problem is replaced by an optimization problem with continuous variables, enabling the use of efficient and scalable continuous optimization algorithms (22).

Relaxation of GM

Adopting this perspective, GM can be formulated as an optimization problem

$$\mathbf{\Pi}^* = \underset{\mathbf{\Pi} \in \mathcal{P}}{\operatorname{argmin}} \ \mathrm{dis}_{\mathcal{A} \mapsto \mathcal{B}}(\mathbf{\Pi}) = \underset{\mathbf{\Pi} \in \mathcal{P}}{\operatorname{argmin}} \|\mathbf{A} - \mathbf{\Pi}^T \mathbf{B} \mathbf{\Pi}\|. \quad [1]$$

The norm in the objective is typically chosen to be the standard ℓ_1 norm $\|\mathbf{X}\|_1 = \sum_{i,j} |x_{ij}|$, the Frobenius ($\ell_2$) norm $\|\mathbf{X}\|_F^2 = \sum_{i,j} x_{ij}^2$, or the minimum–maximum (ℓ_∞) norm $\|\mathbf{X}\|_\infty = \max_{i,j} |x_{ij}|$. In what follows, we will adopt the Frobenius norm, henceforth defining

$$\text{dis}_{\mathcal{A} \mapsto \mathcal{B}}(\mathbf{\Pi}) = \left\| \mathbf{A} - \mathbf{\Pi}^T \mathbf{B} \mathbf{\Pi} \right\|_F = \left\| \mathbf{\Pi A} - \mathbf{B} \mathbf{\Pi} \right\|_F,$$

where the second identity is possible due to unitarity of permutation matrices. For this particular choice, Eq. **1** can be rewritten as

$$\mathbf{\Pi}^* = \underset{\mathbf{\Pi} \in \mathcal{P}}{\text{argmin}} \left\| \mathbf{\Pi A} - \mathbf{B} \mathbf{\Pi} \right\|_F^2 = \underset{\mathbf{\Pi} \in \mathcal{P}}{\text{argmax}} \, \text{tr} \left(\mathbf{B} \mathbf{\Pi} \mathbf{A} \mathbf{\Pi}^T \right), \qquad [2]$$

known as a quadratic assignment problem (QAP).

Both Eqs. **1** and **2** are NP due to the combinatorial complexity of the constraint $\mathbf{\Pi} \in \mathcal{P}$. Relaxation techniques reduce this complexity by replacing the latter constraint with a more tractable continuous set. Since the practically used norms in Eq. **1** can yield a convex minimization objective, convex relaxation techniques consist of replacing $\mathcal{P}$ with a larger convex set, resulting in a tractable convex program. Various techniques differ mainly in the choice of the norm, the choice of the convex set (i.e., the relaxation), and the particular numerical algorithm used to solve the resulting convex program (11).

A popular choice is to relax $\mathcal{P}$ to the space of doubly stochastic matrices $\mathcal{D} = \{ \mathbf{P} \geq 0 : \mathbf{P1} = \mathbf{P}^T \mathbf{1} = \mathbf{1} \}$ constituting the convex hull of permutation matrices in $\mathbb{R}^{n \times n}$. Combined with the ℓ_1 or the ℓ_∞ norms, such a relaxation leads to a linear program (23), while the use of the Frobenius norm results in a linearly constrained quadratic program (LCQP or QP for short) (13). Both types of optimization problems are solvable using polynomial time algorithms, very efficient in practice (22).

Along with convex relaxations of the GM problem (Eq. **1**), there exist numerous techniques for relaxing its QAP formulation (Eq. **2**). Note that after the relaxation, the two problems are generally not equivalent! Unlike Eq. **1**, the objective of Eq. **2** is nonconvex, and hence, even if $\mathcal{P}$ is replaced by a convex set, the resulting optimization problem is nonconvex. One of the most celebrated relaxations of QAP is the spectral relaxation (24), in which the solution matrix is constrained to constant Frobenius norm, which transforms the relaxed problem to the maximum eigenvector problem. The latter is one of the few nonconvex optimization problems for which there exist algorithms with global convergence guarantees. Other nonconvex relaxations of the QAP have been proposed, including restricting the matrix $\mathbf{P}$ to the nonnegative simplex (25), or to the space of doubly stochastic matrices (13). All such relaxations have only local convergence guarantees.

In this paper, we consider the convex QP relaxation of GM,

$$\mathbf{P}^* = \underset{\mathbf{P} \in \mathcal{D}}{\text{argmin}} \left\| \mathbf{PA} - \mathbf{BP} \right\|_F^2. \qquad [3]$$

In the sequel, we show that the double stochasticity constraint can be further harmlessly relaxed. Since the solution $\mathbf{P}^*$ of the relaxation is, generally, not a permutation matrix, it has to be projected back onto $\mathcal{P}$ (n). The orthogonal projection $\mathbf{P}^*$ onto $\mathcal{P}$ has to maximize the standard Euclidean inner product, which can be stated as the optimization problem

$$\hat{\mathbf{\Pi}} = P_\mathcal{P} \mathbf{P}^* = \underset{\mathbf{\Pi} \in \mathcal{P}}{\text{argmax}} \, \langle \mathbf{\Pi}, \mathbf{P}^* \rangle = \underset{\mathbf{\Pi} \in \mathcal{P}}{\text{argmax}} \, \text{tr} \left(\mathbf{\Pi}^T \mathbf{P}^* \right). \qquad [4]$$

This problem is called a "linear assignment problem" (LAP) and, unlike QAP, is solvable in polynomial time using a family of techniques collectively known as the Hungarian method (26). LAP can also be formulated and solved as a linear program, in which the linear objective is minimized over the polytope $\mathcal{D}$ instead of $\mathcal{P}$. The solution of such a linear program is guaranteed to be in $\mathcal{P}$ due to a particular property of the constraints called total unimodularity.

The considered convex relaxation of GM can be thus summarized as the following two-step procedure, which we henceforth call the relaxed GM or RGM: (1) Solve QP (Eq. **3**) and (2) project the obtained solution onto the space of permutation matrices by solving the LAP (Eq. **4**). We will henceforth refer to the permutation matrix $\hat{\mathbf{\Pi}}$ obtained from step 2 above as the solution of the RGM.

Variants of the described procedure are often used in practice; due to their relatively low computational complexity, they scale to large graphs. There is a considerable practical evidence that the RGM produces a good approximation to the exact solution of the GM problem, in the sense that $\text{dis}(\hat{\mathbf{\Pi}}) \approx \text{dis}(\mathbf{\Pi}^*)$, and often $\hat{\mathbf{\Pi}} \approx \mathbf{\Pi}^*$. It is therefore astonishing that no theoretical bounds exist on $\left| \text{dis}(\hat{\mathbf{\Pi}}) - \text{dis}(\mathbf{\Pi}^*) \right|$, and practically nothing is known about $\left\| \hat{\mathbf{\Pi}} - \mathbf{\Pi}^* \right\|$! One of the main goals of this paper is to establish conditions under which RGM is equivalent to the exact GM, in the sense that the projection of the solution space of Eq. **3** onto $\mathcal{P}$ coincides with $\text{Iso}(\mathcal{A} \mapsto \mathcal{B})$. We also investigate conditions for the converse situation, when the solution space of the relaxation contains nonzero distortion permutations, making the relaxation unusable.

Exact Matching of Asymmetric Graphs

We start with the case of exact (i.e., distortionless) matching of asymmetric graphs. An undirected graph $\mathcal{A}$ with the adjacency matrix $\mathbf{A}$ is said to possess a symmetry $\mathbf{\Pi} \in \mathcal{P}$ if $\text{dis}_{\mathcal{A} \mapsto \mathcal{A}}(\mathbf{\Pi}) = 0$. This notation emphasizes that symmetries are self-isomorphisms. Symmetries form a group with the matrix multiplication operation (or with the function composition, if permutations are interpreted as bijective functions), which we refer to as the "symmetry group" (also known as automorphism group) of $\mathcal{A}$ and denote by $\text{Sym} \, \mathcal{A}$. The graph is called asymmetric if its symmetry group is trivial, $\text{Sym} \, \mathcal{A} = \{\mathbf{I}\}$.

It is straightforward to show that two isomorphic graphs $\mathcal{A}$ and $\mathcal{B}$ have identical (isomorphic) symmetry groups, and if $\mathbf{\Pi} \in \mathcal{P}$ is an isomorphism, then $\mathbf{\Pi} \text{Sym}^T \mathcal{A} = \{ \mathbf{\Pi} \mathbf{\Pi}'^T : \mathbf{\Pi}' \in \text{Sym} \, \mathcal{A} \}$ (or, equivalently, $\text{Sym} \mathcal{B} \mathbf{\Pi}$) are also isomorphisms. The converse is also true: If the two graphs are related by a collection of isomorphisms $\text{Iso}(\mathcal{A} \mapsto \mathcal{B}) = \{ \mathbf{\Pi}_1, \ldots, \mathbf{\Pi}_k \in \mathcal{P} \}$, then they are symmetric with $\text{Sym} \, \mathcal{A}$ generated by $\text{Iso}(\mathcal{B} \mapsto \mathcal{A}) \circ \text{Iso}(\mathcal{A} \mapsto \mathcal{B}) = \{ \mathbf{\Pi}_i^T \mathbf{\Pi}_j \}$, and $\text{Sym} \, \mathcal{B}$ by $\text{Iso}(\mathcal{A} \mapsto \mathcal{B}) \circ \text{Iso}(\mathcal{B} \mapsto \mathcal{A}) = \{ \mathbf{\Pi}_i \mathbf{\Pi}_j^T \}$. Consequently, if $\mathcal{A}$ is asymmetric and $\mathcal{B}$ is isomorphic to it, they are related by a unique isomorphism which is the global minimizer of Eq. **1**. In what follows, we denote this unique isomorphism by $\mathbf{\Pi}^*$.

We emphasize that the symmetry or asymmetry of a graph has nothing to do with the fact that the adjacency matrix $\mathbf{A}$ is symmetric. The latter property stems from the fact that the graph is undirected, and because of it, $\mathbf{A}$ admits unitary diagonalization of the form $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$, with an orthonormal $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ containing the eigenvectors in its columns, and a diagonal $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_n\}$ containing the corresponding eigenvalues.

The uniqueness of the isomorphism relating isomorphic asymmetric graphs is crucial for the results we present next. However, the existence or the absence of symmetry is not an easy property to verify. To overcome this difficulty, instead of considering the class of asymmetric graphs, we consider another class of graphs characterized by the following spectral property:

Definition. *A graph $\mathcal{A}$ is called friendly if its adjacency matrix $\mathbf{A}$ has simple spectrum (i.e., all of the λ_i are distinct), and all its eigenvectors satisfy $\mathbf{u}_i^T \mathbf{1} \neq 0$.*

We note the following important consequence of friendliness:

Lemma 1. *A friendly graph is asymmetric.*

Proof. Let $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$ denote the adjacency matrix of the graph, and let us assume by contradiction that there exists $\mathbf{\Pi} \neq \mathbf{I}$ such that $\mathbf{\Pi A} = \mathbf{A} \mathbf{\Pi}$. Then, for every eigenvector $\mathbf{u}_i$ of $\mathbf{A}$, we have $\mathbf{A} \mathbf{\Pi} \mathbf{u}_i = \mathbf{\Pi} \mathbf{A} \mathbf{u}_i = \lambda_i \mathbf{\Pi} \mathbf{u}_i$, that is, $\mathbf{\Pi} \mathbf{u}_i$ is also an eigenvector of $\mathbf{A}$. Since, due to friendliness, $\mathbf{A}$ has a simple spectrum, the only two possibilities are $\mathbf{\Pi} \mathbf{u}_i = \pm \mathbf{u}_i$. Since we assumed $\mathbf{\Pi} \neq \mathbf{I}$, there must

be at least one eigenvector $\mathbf{u}_i$ for which $\mathbf{\Pi}\mathbf{u}_i = -\mathbf{u}_i$. Then, $\mathbf{1}^T\mathbf{\Pi}\mathbf{u}_i = -\mathbf{1}^T\mathbf{u}_i$. On the other hand, since $\mathbf{\Pi}$ is a permutation matrix, $\mathbf{1}^T\mathbf{\Pi}\mathbf{u}_i = \mathbf{1}^T\mathbf{u}_i$. Hence, $\mathbf{u}_i^T\mathbf{1} = 0$ in contradiction to friendliness of $\mathbf{A}$.

The converse is not true, as there might exist an asymmetric graph with an unfriendly adjacency matrix. For example, any regular unweighted graph has a constant eigenvector and is thus highly unfriendly; on the other hand, there exist asymmetric regular graphs such as the Frucht graph with $n = 12$. However, unfriendliness still seems to be a singular property, and intuition suggests that unfriendly graphs should have measure zero among random asymmetric weighted graphs, and the class of friendly graphs should be almost as big as that of asymmetric graphs. We do not pursue the rigorous proof of this claim, since it might delicately depend on what is meant by "random." We only emphasize that, in contrast to asymmetry, friendliness is an easily verifiable property.

Using the notion of friendliness, we state our first result:

Theorem 1. *Let $\mathcal{A}$ and $\mathcal{B}$ be friendly isomorphic graphs. Then, GM and RGM are equivalent.*

Proof. We consider the relaxation (Eq. **3**) of GM, denoting by $\mathbf{\Pi}^*$ the global minimizer of the latter. The minimizer is unique due to asymmetry. For any doubly stochastic matrix $\mathbf{P}$,

$$\mathbf{PA} - \mathbf{BP} = (\mathbf{P}\mathbf{\Pi}^{*T}\mathbf{B} - \mathbf{BP}\mathbf{\Pi}^{*T})\mathbf{\Pi}^* = \mathbf{QB} - \mathbf{BQ},$$

where $\mathbf{Q} = \mathbf{P}\mathbf{\Pi}^{*T}$. We therefore reformulate Eq. **3** in terms of $\mathbf{Q}$ as the minimization of $f(\mathbf{Q}) = (1/2)\|\mathbf{QB} - \mathbf{BQ}\|_F^2$ with respect to $\mathbf{Q} \in \mathcal{D}$. Since the objective $f(\mathbf{Q})$ is convex in $\mathbf{Q}$, and so is the set of double stochastic matrices $\mathcal{D}$, the problem has a global minimum at $\mathbf{Q} = \mathbf{I}$. It remains to prove that the minimum is unique. Since $\mathbf{B}$ is symmetric, simple calculus yields the gradient of $f(\mathbf{Q})$, $\nabla_\mathbf{Q} f = \mathbf{QB}^2 + \mathbf{B}^2\mathbf{Q} - 2\mathbf{BQB}$. By omitting the nonnegativity and unit column sum constraints, we further relax the constraint $\mathbf{Q} \in \mathcal{D}$ to $\mathbf{Q1} = \mathbf{1}$, referring to such matrices as pseudostochastic. The Lagrangian of f with the pseudostochasticity constraint on $\mathbf{Q}$ is given by $L(\mathbf{Q}, \boldsymbol{\alpha}) = f(\mathbf{Q}) + \boldsymbol{\alpha}^T(\mathbf{Q1} - \mathbf{1}) = f(\mathbf{Q}) + \mathrm{tr}((\mathbf{Q1} - \mathbf{1})\boldsymbol{\alpha}^T)$, with $\boldsymbol{\alpha}$ denoting the vector of Lagrange multipliers. Problem Eq. **3** reaches a minimum when

$$\nabla_\mathbf{Q} L(\mathbf{Q}, \boldsymbol{\alpha}) = \mathbf{QB}^2 + \mathbf{B}^2\mathbf{Q} - 2\mathbf{BQB} + \boldsymbol{\alpha}\mathbf{1}^T = 0.$$

Substituting the unitary eigendecomposition $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, the latter equation can be rewritten as

$$\mathbf{F}\mathbf{\Lambda}^2 + \mathbf{\Lambda}^2\mathbf{F} - 2\mathbf{\Lambda}\mathbf{F}\mathbf{\Lambda} + \boldsymbol{\gamma}\mathbf{v}^T = 0, \qquad [5]$$

where $\boldsymbol{\gamma} = \mathbf{U}^T\boldsymbol{\alpha}$, $\mathbf{v} = \mathbf{U}^T\mathbf{1}$, and $\mathbf{F} = \mathbf{U}^T\mathbf{QU}$. It is easy to see that Eq. **5** can be expressed coordinate-wise as

$$F_{ij}(\lambda_i - \lambda_j)^2 + v_j\gamma_i = 0. \qquad [6]$$

For every $i = j$, we have $v_i\gamma_i = 0$; since the friendliness assumption implies $v_i \neq 0$ for all i, we have $\boldsymbol{\gamma} = 0$. This yields

$$F_{ij}(\lambda_i - \lambda_j)^2 = 0 \text{ for every } i \neq j. \qquad [7]$$

Since friendliness also implies $\lambda_i \neq \lambda_j$, $\mathbf{F}$ must be diagonal. Because $\mathbf{Q}$ is pseudostochastic, it has to satisfy $\mathbf{1} = \mathbf{Q1} = \mathbf{UFU}^T\mathbf{1}$ or, equivalently, $\mathbf{v} = \mathbf{Fv}$. However, since $\mathbf{F}$ is diagonal and, due to friendliness, $\mathbf{v}$ has no zero elements, the above property is satisfied only if $\mathbf{F} = \mathbf{I}$. This implies that $\mathbf{Q} = \mathbf{I}$ or, equivalently, $\mathbf{P} = \mathbf{\Pi}^*$. Hence, $\mathbf{\Pi}^*$ is the unique minimizer of Eq. **3**. Since the solution is already in $\mathcal{P}$, the projection (Eq. **4**) leaves it unchanged.

Note that in the proof, we only used the pseudostochasticity constraint $\mathbf{P1} = \mathbf{1}$. This leads to an important consequence: Instead of relaxing $\mathcal{P}$ to the space $\mathcal{D}$ of doubly stochastic matrices, a coarser relaxation to pseudostochastic matrices is equivalent in

the discussed case. Practically, this means that we can solve a simpler quadratic program

$$\mathbf{P}^* = \underset{\mathbf{P}}{\mathrm{argmin}} \|\mathbf{PA} - \mathbf{BP}\|_F^2 \text{ s.t. } \mathbf{P1} = \mathbf{1}, \qquad [8]$$

with n^2 variables and only n equality constraints, instead of $2n$ equality constraints and n^2 inequality constraints in Eq. **3**. In what follows, we focus on this simpler convex relaxation instead of Eq. **3** in the RGM.

While checking the friendliness condition in Theorem 1 is straightforward, checking whether the perfect isomorphism condition is satisfied is not (in fact, it is a graph isomorphism problem!). However, in practice, one can simply solve relaxation Eq. **8** for the two friendly graphs, project the result onto $\mathcal{P}$, and check whether $\mathrm{dis}(\hat{\mathbf{\Pi}}) = 0$. If the answer is positive, $\hat{\mathbf{\Pi}}$ is guaranteed to be the unique isomorphism; otherwise, the theorem guarantees that the graphs are not isomorphic.

Inexact Matching of Asymmetric Graphs

The case of perfectly isomorphic graphs, to which Theorem 1 is applicable, is often an unachievable mathematical idealization. Many practical applications of GM assume some amount of noise, and seek a least distortion correspondence rather than a perfect isomorphism. To formalize this notion, we say that two graphs $\mathcal{A}$ and $\mathcal{B}$ are ρ-isomorphic if there exists $\mathbf{\Pi}^* \in \mathcal{P}$ with $\mathrm{dis}(\mathbf{\Pi}^*) \leq \rho$.

Similarly, we say that a graph $\mathcal{A}$ possesses a ρ-symmetry $\mathbf{\Pi} \in \mathcal{P}$ if $\mathrm{dis}_{\mathcal{A} \mapsto \mathcal{A}}(\mathbf{\Pi}) \leq \rho$. Note that unlike their exact counterparts, ρ-symmetries do not form a group, as the composition of two ρ-symmetries might have $\mathrm{dis} > \rho$. A graph with a trivial ρ-symmetry set is called ρ-asymmetric. The lack of symmetry of such a graph is strong enough to guarantee that a bounded perturbation of the adjacency weights does not produce new, previously inexistent symmetries.

To generalize our result to the case of nearly isomorphic graphs, we define the strength of a graph's friendliness:

Definition. *A graph $\mathcal{A}$ is (ϵ, δ)-friendly if its adjacency matrix $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ has the spectral gap $\sigma(\mathbf{A}) = \min_{i \neq j}|\lambda_i - \lambda_j| > \delta$, and $\epsilon < |\mathbf{u}_i^T\mathbf{1}| < (1/\epsilon)$ for $i = 1, \ldots, n$.*

Also note that our former definition of friendliness corresponds to $(\epsilon, \delta) = (0, 0)$. We refer to the case $\epsilon, \delta > 0$ as "strong" friendliness.

For the broad family of strongly friendly graphs, we first show that the result of Theorem 1 is stable in the sense that a bounded perturbation in the adjacency matrix results in a bounded perturbation of the solution:

Lemma 2. *Let $\mathcal{A}$ and $\mathcal{B}$ be (ϵ, δ)-friendly isomorphic graphs with spectral radius $\sigma = \max_i|\lambda_i|$, related by the unique isomorphism $\mathbf{\Pi}^*$. Let $\tilde{\mathcal{B}}$ be a perturbed version of $\mathcal{B}$ with $\tilde{\mathbf{B}} = \mathbf{B} + \rho\mathbf{R}$, where $\mathbf{R}$ is symmetric with $\|\mathbf{R}\|_F \leq 1$, and $\rho \leq \min\{\sqrt{2}\sigma, [(\delta^2\epsilon^4)/(12\sigma n^{1.5})]\}$. Then, the solution $\mathbf{P}_\rho^*$ of the perturbed problem, Eq. **8**, is unique and satisfies $\|\mathbf{P}_\rho^* - \mathbf{\Pi}^*\|_F < (1/2)$.*

The proof closely follows the proof of Theorem 1, and relies on a result from perturbation analysis of linear systems. Full proof as well as the mentioned result (summarized as Lemma S1) are presented in *Supporting Information*.

Applying the former result to matching a graph with itself ($\mathcal{A} = \mathcal{B}$), the following generalization of Lemma 1 can be straightforwardly obtained:

Corollary 1. *An (ϵ, δ)-friendly graph is ρ-asymmetric, with ρ satisfying the conditions of Lemma 2.*

In fact, this property guarantees that the perturbation creates no symmetries and, thus, the perturbed version of system Eq. **7** remains full rank.

The stability of the relaxation in Lemma 2 leads directly to our second result:

Theorem 2. *Let $\mathcal{A}$ be an (ϵ,δ)-friendly graph with the adjacency matrix normalized such that $\sigma = 1$, and let $\mathcal{B}$ be ρ-isomorphic to $\mathcal{A}$. Then, if $\rho < [(\delta^2\epsilon^4)/(12n^{1.5})]$, RGM and GM are equivalent.*

Proof. Let $\boldsymbol{\Pi}^*$ be a ρ-isomorphism relating $\mathcal{B}$ and $\mathcal{A}$, and let us denote $\mathbf{B}_0 = \boldsymbol{\Pi}^*\mathbf{A}\boldsymbol{\Pi}^{*\mathrm{T}}$ and $\mathbf{R} = (1/\rho)(\mathbf{B}-\mathbf{B}_0)$. Then, $\mathcal{B}_0$ is perfectly isomorphic to $\mathcal{A}$, and $\mathcal{B}$ is a perturbed version of $\mathcal{B}_0$ with $\mathbf{B} = \mathbf{B}_0 + \rho\mathbf{R}$ and $\|\mathbf{R}\|_{\mathrm{F}} = (1/\rho)\|\mathbf{B}-\boldsymbol{\Pi}^*\mathbf{A}\boldsymbol{\Pi}^{*\mathrm{T}}\|_{\mathrm{F}} \leq 1$. By Corollary 2, $\mathcal{B}$ is ρ-asymmetric and, hence, $\mathcal{B}_0$ is asymmetric. Denoting by $\mathbf{P}^*$ the solution of Eq. **8** applied to $\mathcal{A}$ and $\mathcal{B}$, we invoke Lemma 2, which guarantees uniqueness of $\mathbf{P}^*$ and $\|\mathbf{P}^*-\boldsymbol{\Pi}^*\|_{\mathrm{F}} < 1/2$. By standard norm inequalities, this implies $\left|\mathbf{P}^*_{ij}-\boldsymbol{\Pi}^*_{ij}\right| < 1/2$ element-wise for every i,j. Therefore, the projection of $\mathbf{P}^*$ onto $\mathcal{P}$ coincides with $\boldsymbol{\Pi}^*$.

As in the case of perfectly isomorphic graphs, checking the strong friendliness condition in Theorem 1 is straightforward, while checking the ρ-isomorphism of $\mathcal{A}$ and $\mathcal{B}$ is not. However, as in the previous case, one can again solve relaxation Eq. **8**, project the solution onto $\mathcal{P}$, and verify whether $\mathrm{dis}(\hat{\boldsymbol{\Pi}}) < \rho$. In case of a positive answer, $\hat{\boldsymbol{\Pi}}$ is guaranteed to be the unique global minimizer of the GM problem; otherwise, the graphs are guaranteed not to be ρ-isomorphic. An empirical evaluation of the bound in Theorem 2 is presented in Fig. 2.

Matching of Symmetric Graphs

The assumption of friendliness plays a crucial role in the results we have developed so far: It guarantees uniqueness of solution of the relaxation. These results cannot be directly extended to symmetric graphs, for which the solution space of the relaxation should contain all isomorphisms and their affine combinations. Unfortunately, besides the true isomorphisms, this affine subspace may also contain pseudostochastic matrices that are not permutations, some of which fall into Voronoi cells of permutations that are not isomorphisms. As a result, using convex relaxation for matching symmetric graphs may lead to a wrong solution, depending on the particular optimization algorithm and its initialization. Empirical evidence of this phenomenon is presented in Fig. S1; see ref. 27 for a formal proof of failure of convex relaxation on a particular class of random graphs. However, in what follows, we show that by providing additional information in the form of corresponding seeds or vertex attributes disambiguating the symmetry, equivalence of the relaxation to the exact GM problem still holds.
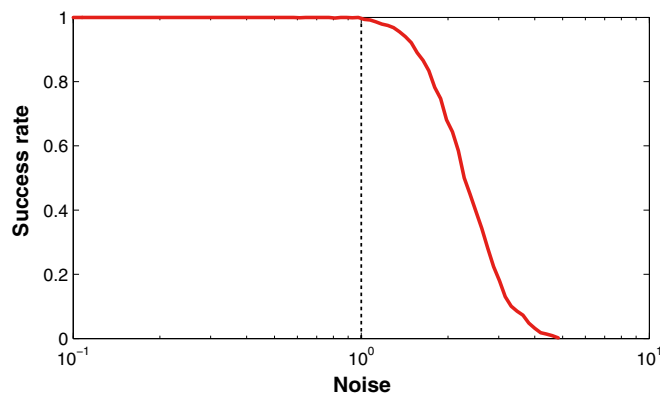
Fig. 2. Empirical evaluation of the bound in Theorem 2 on 10^3 random strongly friendly graphs. For each graph, different amount of noise was added, and the ratio of successful runs of convex relaxation (Eq. **8**) recorded on the vertical axis (a run was deemed successful if the ground truth isomorphism is recovered). The noise strength on the horizontal axis is normalized for each graph in such a way that the value of the bound is always 1. Observe that all runs with noise within the bound converged successfully, while those with stronger noise failed with probability increasing as the amount of noise grows.

Let $\mathbf{C}$ and $\mathbf{D}$ be $n\times q$ matrices, whose columns are real-valued functions on the vertices of the graphs $\mathcal{A}$ and $\mathcal{B}$, respectively. For example, an indicator function of the kth vertex in the graph is the kth vector of the standard Euclidean basis in $\mathbb{R}^n$. The matrices $\mathbf{C}$ and $\mathbf{D}$ can be alternatively interpreted as q-dimensional vector-valued vertex attributes, with the kth row of $\mathbf{C}$ representing the attribute of vertex k in $\mathcal{A}$. We say that the matrices $\mathbf{C}$ and $\mathbf{D}$ are covariant under a permutation $\boldsymbol{\Pi}$ relating between the graphs if $\boldsymbol{\Pi}\mathbf{C} = \mathbf{D}$. With this additional information, we consider the following extension of Eq. **8**:

$$\mathbf{P}^* = \underset{\mathbf{P}}{\arg\min} \|\mathbf{PA}-\mathbf{BP}\|_{\mathrm{F}}^2 + \mu\|\mathbf{PC}-\mathbf{D}\|_{\mathrm{F}}^2 \ \text{s.t.} \ \mathbf{P1} = \mathbf{1}. \quad [9]$$

This problem can be thought of as a convex relaxation of seeded GM, in which the seeds are provided through a penalty, whose strength is controlled by the parameter μ, rather than through a hard constraint; alternatively, it can be interpreted as a relaxation of attributed GM, in which a permutation is sought to minimize the aggregate of edge adjacency and vertex attribute disagreement. In light of this duality, we henceforth refer to $\mathbf{C}$ and $\mathbf{D}$ as seeds.

As before, to avoid verifying whether a graph is symmetric or not, we consider the easily verifiable friendliness property. We assume that a general adjacency matrix of a graph has d nonsimple eigenspaces with multiplicities summing up to $m + d$. To simplify notation, we will say that an eigenvalue λ_i has multiplicity m_i+1, referring to the multiplicity of the eigenspace to which λ_i belongs. Since the eigenvectors spanning an (m_i+1)-dimensional eigenspace are defined up to a rotation within it, eigenvectors shall be selected such that none of them is orthogonal to the constant vector $\mathbf{1}$, unless the entire eigenspace is orthogonal to it. We call the latter eigenspaces "hostile," and denote by k the total dimension of hostile eigenspaces. A graph is friendly if and only if both m and k are 0, and is (m,k)-unfriendly otherwise.

It is easy to observe that with $\mu = 0$, each (m_i+1)-dimensional nonsimple eigenspace of the adjacency matrix decreases the rank of Eq. **5** by m_i; if the eigenspace is hostile, the rank is further decreased by 1. This is precisely the reason for Theorem 1 not being applicable to unfriendly graphs. The introduction of the seeds disagreement term to the objective contributes to Eq. **5** a term of the form $\mu\mathbf{FG}$, where $\mathbf{G} = \mathbf{U}^{\mathrm{T}}\mathbf{DD}^{\mathrm{T}}\mathbf{U}$ is the Gram matrix of the seeds $\mathbf{D}$ represented in the eigenbasis of the adjacency matrix. Since $\mathbf{G}$ is always positive semidefinite, the rank of Eq. **5** typically increases and becomes full under the following conditions:

Theorem 3. *Let $\mathcal{A}$ and $\mathcal{B}$ be isomorphic unfriendly graphs with adjacencies $\mathbf{A}$ and $\mathbf{B}$ and seeds $\mathbf{C}$ and $\mathbf{D}$, respectively. Let $\mathbf{C}$ and $\mathbf{D}$ be covariant under a particular isomorphism $\boldsymbol{\Pi}^* \in \mathrm{Iso}(\mathcal{A} \mapsto \mathcal{B})$, and let $\mathbf{D}$ further satisfy for every nonsimple (m_i+1)-dimensional eigenspace of $\mathbf{B}$ corresponding to $\lambda_i = \cdots = \lambda_{i+m_i}$, $(\mathbf{1}^{\mathrm{T}}\mathbf{u}_i)\mathbf{DD}^{\mathrm{T}}\mathbf{u}_j \neq \mathbf{1}(\mathbf{u}_i^{\mathrm{T}}\mathbf{DD}^{\mathrm{T}}\mathbf{u}_j)$ for every $j = i+1,\ldots,i+m_i$ if the eigenspace is not hostile, or $\mathbf{DD}^{\mathrm{T}}\mathbf{u}_j \neq \mathbf{0}$ for every $j = i,\ldots,i+m_i$ otherwise. Then, $\boldsymbol{\Pi}^*$ is the unique minimizer of Eq. **9** for any $\mu > 0$.*

For a proof, see *Supporting Information*. Conditions of Theorem 3 are both easy to verify and constructive in the sense that given the spectral decomposition of the adjacency matrix of one of the graphs, the theorem specifies how to construct a set of seeds such that if a set of corresponding seeds in the other graph is further given and is covariant under a preferred isomorphism, the convex relaxation (Eq. **9**) is guaranteed to find the latter isomorphism. In particular, $\mathbf{D}$ must have at least $m + k$ linearly independent columns. In practice, we observed that it is sufficient to generate random point seeds ensuring that the matrix $\mathbf{D}$ is not invariant under any nontrivial symmetry of the graph, namely $\boldsymbol{\Pi}\mathbf{D} \neq \mathbf{D}$ for every $\boldsymbol{\Pi} \in \mathrm{Sym}\mathcal{B}\backslash\{\mathbf{I}\}$. An empirical corroboration of this result is presented in Fig. S1.

Discussion and Conclusion

In this paper, we considered convex relaxation of the NP GM problem. We proposed an easy-to-verify friendliness property, and proved that for friendly graphs, convex relaxation is equivalent to the computationally intractable exact matching; the result extends to inexact matching of strongly friendly graphs. In such cases, convex relaxation is guaranteed to find the exact (or approximate) isomorphism or guarantee its inexistence. We also showed that convex relaxation is applicable to exact matching of unfriendly graphs (in particular, those possessing nontrivial symmetries), provided that additional information is supplied in the form of seeds or vertex attributes. We showed constructive spectral characteristics that such seeds/attributes have to satisfy in order for the convex relaxation of seeded GM to be guaranteed to find one of the isomorphisms. The analysis we presented is inspired in part by ref. 28, where matching surfaces is treated as matching metric spaces performed in their spectral domain. However, despite this superficial resemblance, our proofs here are based on the spectral properties of the adjacency matrices, in contrast to those of the graph Laplacian frequently studied in spectral graph theory.

A surprising observation is that none of our results is influenced by the nonnegativity constraints $\mathbf{P} \geq 0$. While the space of doubly stochastic matrices is the smallest convex set containing the space $\mathcal{P}$ of $n \times n$ permutations, and is therefore the most natural convex relaxation of the latter, our findings question the utility of the nonnegativity constraints in GM problems, and suggest relaxing $\mathcal{P}$ as a bigger affine space. Also, for the class of friendly graphs on which we were able to prove global convergence of convex relaxations, the column-wise equality constraints $\mathbf{P}^\mathrm{T}\mathbf{1} = \mathbf{1}$ have no utility and can be removed. The question of whether these constraints are at all needed, and whether they can help extend the applicability of convex relaxation, requires further investigation. From the practical perspective, the removal of the nonnegativity constraints renders the quadratic problem much easier to solve, essentially, by solving a linear system.

APPLIED MATHEMATICS

1. Conte D, Foggia P, Sansone C, Vento M (2004) Thirty years of graph matching in pattern recognition. *Int J Pattern Recognit Artif Intell* 18:265–298.
2. Christmas WJ, Kittler J, Petrou M (1995) Structural matching in computer vision using probabilistic relaxation. *IEEE Trans Pattern Anal Mach Intell* 17:749–764.
3. Lades M, et al. (1993) Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans Comput* 42:300–311.
4. Pelillo M, Siddiqi K, Zucker SW (1999) Matching hierarchical structures using association graphs. *IEEE Trans Pattern Anal Mach Intell* 21:1105–1120.
5. Rocha J, Pavlidis T (1994) A shape analysis model with applications to a character recognition system. *IEEE Trans Pattern Anal Mach Intell* 16:393–404.
6. Berretti S, Del Bimbo A, Vicario E (2001) Efficient matching and indexing of graph models in content-based retrieval. *IEEE Trans Pattern Anal Mach Intell* 23:1089–1105.
7. Wiskott L, Fellous JM, Kuiger N, Von Der Malsburg C (1997) Face recognition by elastic bunch graph matching. *IEEE Trans Pattern Anal Mach Intell* 19:775–779.
8. Isenor D, Zaky SG (1986) Fingerprint identification using graph matching. *Pattern Recognit* 19:113–122.
9. Sporns O, Tononi G, Kötter R (2005) The human connectome: A structural description of the human brain. *PLOS Comput Biol* 1(4):e42.
10. Cook DJ, Holder LB (2006) *Mining Graph Data* (Wiley, New York).
11. Schellewald C, Roth S, Schnörr C (2001) Evaluation of convex optimization techniques for the weighted graph-matching problem in computer vision. *Pattern Recognit* 2191:361–368.
12. Bronstein AM, Bronstein MM, Kimmel R (2006) Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proc Natl Acad Sci USA* 103(5):1168–1172.
13. Vogelstein JT, et al. (2011) Large (brain) graph matching via fast approximate quadratic programming. arXiv:1112.5507.
14. Gao X, Xiao B, Tao D, Li X (2010) A survey of graph edit distance. *Pattern Anal Appl* 13:113–129.
15. Pelillo M (2004) Metrics for attributed graphs based on the maximal similarity common subgraph. *Int J Pattern Recognit Artif Intell* 18:299–313.
16. Pelillo M (1998) A unifying framework for relational structure matching. *Proc Int Conf Pattern Recognit* 2:1316–1319.
17. Fortin S (1996) *The Graph Isomorphism Problem: Technical Report* (Univ Alberta, Edmonton, AB, Canada).
18. Babai L (1981) Moderately exponential bound for graph isomorphism. *Fundamentals of Computation Theory* (Springer, New York), pp 34–50.
19. Hopcroft JE, Wong JK (1974) Linear time algorithm for isomorphism of planar graphs. Proceedings of the Sixth Annual ACM Symposium on the Theory of Computing (Assoc Comput Machinery, New York), pp 172–184.
20. Luks EM (1982) Isomorphism of graphs of bounded valence can be tested in polynomial time. *J Comput Syst Sci* 25:42–65.
21. Aho AV, Hopcroft JE (1974) *The Design and Analysis of Computer Algorithms* (Addison-Wesley, Reading, MA).
22. Bertsekas DP (1999) *Nonlinear Programming* (Athena Scientific, Nashu, NH).
23. Almohamad H, Duffuaa SO (1993) A linear programming approach for the weighted graph matching problem. *IEEE Trans Pattern Anal Mach Intell* 15:522–525.
24. Leordeanu M, Hebert M (2005) A spectral technique for correspondence problems using pairwise constraints. *Proc Int Conf Comput Vision* 2:1482–1489.
25. Rota Bulò S, Pelillo M, Bomze IM (2011) Graph-based quadratic optimization: A fast evolutionary approach. *Comput Vis Image Understanding* 115:984–995.
26. Kuhn HW (1955) The Hungarian method for the assignment problem. *Nav Res Logistics Q* 2:83–97.
27. Lyzinski V, et al. (2014) Graph matching: Relax at your own risk. arXiv:1405.3133.
28. Aflalo Y, Dubrovina A, Kimmel R (2013) Spectral generalized multi-dimensional scaling. arXiv:1311.2187.