A Family of Tractable Graph Distances

Jose Bento* Stratis Ioannidis†

Abstract

Important data mining problems such as nearest-neighbor search and clustering admit theoretical guarantees when restricted to objects embedded in a metric space. Graphs are ubiquitous, and clustering and classification over graphs arise in diverse areas, including, e.g., image processing and social networks. Unfortunately, popular distance scores used in these applications, that scale over large graphs, are not metrics and thus come with no guarantees. Classic graph distances such as, e.g., the chemical and the CKS distance are arguably natural and intuitive, and are indeed also metrics, but they are intractable: as such, their computation does not scale to large graphs. We define a broad family of graph distances, that includes both the chemical and the CKS distance, and prove that these are all metrics. Crucially, we show that our family includes metrics that are tractable. Moreover, we extend these distances by incorporating auxiliary node attributes, which is important in practice, while maintaining both the metric property and tractability.

1 Introduction

Graph similarity and the related problem of graph isomorphism have a long history in data mining, machine learning, and pattern recognition [21, 43, 38]. *Graph distances* naturally arise in this literature: intuitively, given two (unlabeled) graphs, their distance is a score quanitifying their structural differences. A highly desirable property for such a score is that it is a *metric*, i.e., it is non-negative, symmetric, positive-definite, and, crucially, satisfies the triangle inequality. Metrics exhibit significant computational advantages over non-metrics. For example, operations such as nearest-neighbor search [20, 19, 11], clustering [3], outlier detection [7], and diameter computation [31] admit fast algorithms precisely when performed over objects embedded in a metric space. To this end, proposing *tractable* graph metrics is of paramount importance in applying such algorithms to graphs.

Unfortunately, graph metrics of interest are often computationally expensive. A well-known example is the *chemical distance* [40]. Formally, given graphs G_A and G_B, represented by their adjacency matrices $A, B \in \{0,1\}^{n\times n}$, the chemical distance is $d_{\mathbb{P}^n}(A,B)$ is defined in terms of a mapping between the two graphs that minimizes their edge discrepancies, i.e.:

$$d_{\mathbb{P}^n}(A,B) = \min_{P\in\mathbb{P}^n} \|AP - PB\|_F, \tag{1.1}$$

where $\mathbb{P}^n$ is the set of permutation matrices of size n and $\|\cdot\|_F$, is the Frobenius norm (see Sec. 2 for definitions). The *Chartrand-Kubiki-Shultz (CKS)* [18] distance is an alternative: CKS is again given by (1.1) but, instead of edges, matrices A and B contain the pairwise shortest path distances between any two nodes.

The chemical and CKS distances have important properties. First, they are zero if and only if the graphs are isomorphic, which appeals to both intuition and practice; second, as desired, they are metrics; third, they have a natural interpretation, capturing global structural similarities between graphs. However, finding an optimal permutation P is notoriously hard; graph isomorphism, which is equivalent to deciding if there exists a permutation P s.t. $AP = PB$ (for both adjacency and path matrices), is famously a problem that is neither known to be in P nor shown to be NP-hard [8]. There is a large and expanding literature on scalable heuristics to estimate the optimal permutation P [34, 9, 42, 23]. Despite their computational advantages, unfortunately, using them to approximate $d_{\mathbb{P}^n}(A,B)$ breaks the metric property.

This significantly degrades the performance of many important tasks that rely on computing distances between graphs. For example, there is a clear separation on the approximability of clustering over metric and non-metric spaces [3]. We also demonstrate this empirically in Section 5 (c.f. Fig. 1): attempting to cluster graphs sampled from well-known families based on non-metric distances significantly increases the misclassification rate, compared to clustering using metrics.

An additonal issue that arises in practice is that nodes often have attributes not associated with adjacency. For example, in social networks, nodes may contain profiles with a user's age or gender; similarly, nodes in molecules may be labeled by atomic numbers. Such attributes are not captured by the chemical or CKS distances. However, in such cases, only *label-preserving*

*Boston College, jose.bento@bc.edu
†Northeastern University, ioannidis@ece.neu.edu

permutations P may make sense (e.g., mapping females to females, oxygens to oxygens, etc.). Incorporating attributes while preserving the metric property is thus important from a practical perspective.

Contributions. We seek generalization of the chemical and CKS distances that (a) *satisfy the metric property* and (b) are *tractable*: by this, we mean that they can be computed either by solving a convex optimization problem, or by a polynomial time algorithm. Specifically, we study generalizations of (1.1) of the form:

$$d_S(A, B) = \min_{P \in S} \|AP - PB\| \tag{1.2}$$

where $S \subset \mathbb{R}^{n \times n}$ is closed and bounded, $\|\cdot\|$ is a matrix norm, and $A, B \in \mathbb{R}^{n \times n}$ are arbitrary real matrices (representing adjacency, path distances, weights, etc.). We make the following contributions:

- We prove sufficient conditions on S and norm $\|\cdot\|$ for which (1.2) is a metric. In particular, we show that d_S is a so-called *pseudo-metric* (see Sec. 2) when:
  (i) $S = \mathbb{P}^n$ and $\|\cdot\|$ is any entry-wise or operator norm;
  (ii) $S = \mathbb{W}^n$, the set of *doubly stochastic* matrices, $\|\cdot\|$ is an arbitrary entry-wise norm, and A, B are symmetric; a modification on d_S extends this result to both operator norms as well as arbitrary matrices (capturing, e.g., directed graphs); and
  (iii) $S = \mathbb{O}^n$, the set of orthogonal matrices, and $\|\cdot\|$ is the operator or entry-wise 2-norm.

  Relaxations *(ii)* and *(iii)* are very important from a practical standpoint. For all matrix norms, computing (1.2) with $S = \mathbb{W}^n$ is tractable, as it is a convex optimization. For $S = \mathbb{O}^n$, (1.2) is non-convex but is still tractable, as it reduces to a spectral decomposition. This was known for the Frobenius norm [56]; we prove this is the case for the operator 2-norm also.
- We include node attributes in a natural way in the definition of d_S as both *soft* (i.e., penalties in the objective) or *hard* constraints in Eq. (1.2). Crucially, we do this *without affecting the metric property and tractability*. This allows us to explore label or feature preserving permutations, that incorporate both (a) exogenous node attributes, such as, e.g., user age or gender in a social network, as well as (b) endogenous, structural features of each node, such as its degree or the number of triangles that pass through it. We numerically show that adding these constraints can speed up the computation of d_S.

From an experimental standpoint, we extensively compare our tractable metrics to several existing heuristic approximations. We also demonstrate the tractability of our metrics by parallelizing their execution using the alternating method of multipliers [15], which we implement over a compute cluster using Apache Spark [62].

Related Work. Graph distance (or similarity) scores find applications in varied fields such as in image processing [21], chemistry [6, 40], and social network analysis [43, 38]. Graph distances are easy to define when, contrary to our setting, the correspondence between graph nodes is known, i.e., graphs are *labeled* [46, 38, 55]. Beyond the chemical distance, classic examples of distances between unlabeled graphs are the *edit distance* [28, 51] and the *maximum common subgraph distance* [17, 16], both of which also have versions for labeled graphs. Both are metrics and are hard to compute, while existing heuristics [48, 26] are not metrics. The *reaction distance* [36] is also a metric directly related to the chemical distance [40] when edits are restricted to edge additions and deletions. Jain [32] also considers an extension of the chemical distance, limited to the Frobenius norm, that incorporates edge attributes. However, it is not immediately clear how to relax the above metrics [32, 36] to attain tractability.

A metric can also be induced by embedding graphs in a metric space and measuring the distance of these embeddings [50, 27, 49]. Several works follow such an approach, mapping graphs, e.g., to spaces determined by their spectral decomposition [63, 60, 24]. In general, in contrast to our metrics, such approaches are not as discriminative, as embeddings summarize graph structure. Continuous relaxations of graph isomorphism, both convex and non-convex [42, 4, 56], have found applications in a variety of contexts, including social networks [37], computer vision [52], shape detection [53, 30], and neuroscience [57]. None of the above works focus on metric properties of resulting relaxations, which several fail to satisfy [57, 37, 53, 30].

Metrics naturally arise in data mining tasks, including clustering [61, 29], NN search [20, 19, 11], and outlier detection [7]. Some of these tasks become tractable or admit formal guarantees precisely when performed over a metric space. For example, finding the nearest neighbor [20, 19, 11] or the diameter of a dataset [31] become polylogarithimic under metric assumptions; similarly, approximation algorithms for clustering (which is NP-hard) rely on metric assumptions, whose absence leads to a deterioration on known bounds [3]. Our search for metrics is motivated by these considerations.

2 Notation and Preliminaries

Graphs. We represent an undirected graph $G(V, E)$ with node set $V = [n] \equiv \{1, \ldots, n\}$ and edge set $E \subseteq [n] \times [n]$ by its *adjacency matrix*, i.e. $A = [a_{i,j}]_{i,j \in [n]} \in \{0, 1\}^{n \times n}$ s.t. $a_{ij} = a_{ji} = 1$ if and only if $(i, j) \in E$. In particular, A is symmetric, i.e. $A = A^\top$. We denote the set of all real, symmetric matrices by $\mathbb{S}^n$. *Directed* graphs are represented by (possibly non-

symmetric) binary matrices $A \in \{0,1\}^{n\times n}$, and *weighted* graphs by real matrices $A \in \mathbb{R}^{n\times n}$.

Matrix Norms. Given a matrix $A = [a_{ij}]_{i,j\in[n]} \in \mathbb{R}^{n\times n}$ and a $p \in \mathbb{N}_+ \cup \{\infty\}$, its *induced* or *operator p-norm* is defined in terms of the vector p-norm through $\|A\|_p = \sup_{x\in\mathbb{R}^n:\|x\|_p=1} \|Ax\|_p$, while its *entry-wise p-norm* is given by $\|A\|_p = (\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^p)^{1/p}$, for $p \in \mathbb{N}_+$, and $\|A\|_\infty = \max_{i,j} |a_{i,j}|$. We denote the entry-wise 2-norm (i.e., the *Frobenius* norm) as $\|\cdot\|_F$.

Permutation, Doubly Stochastic, and Orthogonal Matrices. We denote the set of *permutation* matrices as $\mathbb{P}^n = \{P \in \{0,1\}^{n\times n} : P\mathbf{1} = \mathbf{1}, P^\top\mathbf{1} = \mathbf{1}\}$, the set of *doubly-stochastic* matrices (i.e., the *Birkhoff polytope*) as $\mathbb{W}^n = \{W \in [0,1]^{n\times n} : W\mathbf{1} = \mathbf{1}, W^\top\mathbf{1} = \mathbf{1}\}$, and the set of *orthogonal matrices* (i.e., the *Stiefel manifold*) as $\mathbb{O}^n = \{U \in \mathbb{R}^{n\times n} : UU^\top = U^\top U = I\}$. Note that $\mathbb{P}^n = \mathbb{W}^n \cap \mathbb{O}^n$. Moreover, the Birkoff-von Neumann Theorem [12] states that $\mathbb{W}^n = \texttt{conv}(\mathbb{P}^n)$, i.e., the Birkoff polytope is the convex hull of $\mathbb{P}^n$.

Metrics. Given a set Ω, a function $d : \Omega \times \Omega \to \mathbb{R}$ is called a *metric*, and the pair (Ω, d) is called a *metric space*, if for all $x, y, z \in \Omega$:

(2.3a) $d(x,y) \geq 0$ (non-negativity)
(2.3b) $d(x,y) = 0$ iff $x = y$ (pos. definiteness)
(2.3c) $d(x,y) = d(y,x)$ (symmetry)
(2.3d) $d(x,y) \leq d(x,z) + d(z,y)$ (triangle inequality)

A function d is called a *pseudometric* if it satisfies (2.3a), (2.3c), and (2.3d), but the positive definiteness property (2.3b) is replaced by the (weaker) property:

(2.3e) $d(x,x) = 0$ for all $x \in \Omega$.

If d is a pseudometric, then $d(x,y) = 0$ defines an equivalence relation $x \sim_d y$ over Ω. A pseudometric is then a metric over $\Omega/\sim_d$, the quotient space of $\sim_d$. A d that satisfies (2.3a), (2.3b), and (2.3d) *but not* the symmetry property (2.3c) is called a *quasimetric*. If d is a quasimetric, then its *symmetric extension* $\bar{d} : \Omega\times\Omega \to \mathbb{R}$, defined as $\bar{d}(x,y) = d(x,y) + d(y,x)$, is a metric over Ω.

Graph Isomorphism, Chemical, and CKS Distance. Let $A, B \in \mathbb{R}^{n\times n}$ be the adjacency matrices of two graphs G_A and G_B. Then, G_A and G_B are *isomorphic* if and only if there exists $P \in \mathbb{P}^n$ s.t. $P^\top AP = B$ or, equivalently, $AP = PB$. The *chemical distance*, given by (1.1), extends the latter relationship to capture distances between graphs. Let $\|\cdot\|$ be a matrix norm in $\mathbb{R}^{n\times n}$. For some $\Omega \subseteq \mathbb{R}^{n\times n}$, define $d_S : \Omega \times \Omega \to \mathbb{R}_+$ as:

$$d_S(A,B) = \min_{P\in S} \|AP - PB\|, \tag{2.4}$$

where $S \subset \mathbb{R}^{n\times n}$ is a closed and bounded set, so that the infimum is indeed attained. Note that d_S is the chemical distance (1.1) when $\Omega = \mathbb{R}^{n\times n}$, $S = \mathbb{P}^n$ and $\|\cdot\| = \|\cdot\|_F$. In CKS distance [18], matrices A, B contain pairwise path distances between any two nodes; equivalently, CKS is the chemical distance of two weighted complete graphs with path distances as edge weights. Our main contribution is determining general conditions on S and $\|\cdot\|$ under which d_S is a metric over Ω, for arbitrary weighted graphs, thereby including both the chemical and CKS distances as special cases.

For concreteness, we focus on distances between graphs of equal size. Extensions to graphs of unequal size are described in [10].

3 A Family of Graph Metrics

Our first result establishes that $d_{\mathbb{P}^n}$ is a pseudometric over *all* weighted graphs when $\|\cdot\|$ is an *arbitrary* entry-wise or operator norm.

THEOREM 3.1. *If $S = \mathbb{P}^n$ and $\|\cdot\|$ is an arbitrary entry-wise or operator norm, then d_S given by* (2.4) *is a pseudometric over $\Omega = \mathbb{R}^{n\times n}$.*

Hence, $d_{\mathbb{P}^n}$ is a pseudometric under any entry-wise or operator norm over arbitrary directed, weighted graphs. Our second result states that this property extends to the *relaxed* version of the chemical distance, in which permutations are replaced by doubly stochastic matrices.

THEOREM 3.2. *If $S = \mathbb{W}^n$ and $\|\cdot\|$ is an arbitrary entry-wise norm, then d_S given by* (2.4) *is a pseudometric over $\Omega = \mathbb{S}^{n\times n}$. If $\|\cdot\|$ is an arbitrary entry-wise or operator norm, then its symmetric extension $\bar{d}_S(A,B) = d_S(A,B) + d_S(B,A)$ is a pseudometric over $\Omega = \mathbb{R}^{n\times n}$.*

Hence, if $S = \mathbb{W}^n$ and $\|\cdot\|$ is an arbitrary entry-wise norm, then (2.4) defines a pseudometric over *undirected* graphs. The symmetry property (2.3c) breaks if $\|\cdot\|$ is an operator norm or graphs are directed. In either case, d_S is a quasimetric over the quotient space $\Omega/\sim_d$, and symmetry is attained via the symmetric extension $\bar{d}_S$.

Theorem 3.2 has significant practical implications. In contrast to $d_{\mathbb{P}^n}$ and its extensions implied by Theorem 3.1, computing $d_{\mathbb{W}^n}$ under any operator or entry-wise norm is *tractable* [14]: it involves minimizing a convex function subject to linear constraints. A more limited result extends to the Stiefel manifold:

THEOREM 3.3. *If $S = \mathbb{O}^n$ and $\|\cdot\|$ is either the operator or the entry-wise (i.e., Frobenius) 2-norm, then d_S given by* (2.4) *is a pseudometric over $\Omega = \mathbb{R}^{n\times n}$.*

Though (2.4) is not a convex problem when $S = \mathbb{O}^n$, it is also tractable. Umeyama [56] shows that the optimization can be solved exactly when $\|\cdot\| = \|\cdot\|_F$ and $\Omega = \mathbb{S}^n$ (i.e., for undirected graphs) by performing

a spectral decomposition on A and B. We extend this result, showing that the same procedure also applies when $\|\cdot\|$ is the operator 2-norm (see Thm. 7 in [10]). In the general case of directed graphs, (2.4) is a classic example of a problem that can be solved through optimization on manifolds [2].

Equivalence Classes. The equivalence of matrix norms implies that all pseudometrics d_S defined through (2.4) for a given S have the same quotient space $\Omega/\sim_{d_S}$: if $d_S(A,B)=0$ for one matrix norm $\|\cdot\|$ in (2.4), it will be so for all. When $S=\mathbb{P}^n$, $\Omega/\sim_{d_{\mathbb{P}^n}}$ is the quotient space defined by graph isomorphism: any two adjacency matrices $A,B\in\mathbb{R}^{n\times n}$ satisfy $d_{\mathbb{P}^n}(A,B)=0$ if and only if their (possibly weighted) graphs are isomorphic. When $S=\mathbb{W}^n$, the quotient space $\Omega/\sim_{d_{\mathbb{W}^n}}$ has a connection to the Weisfeiler-Lehman (WL) algorithm [59] described in [10]: Ramana et al. [47] show that $d_{\mathbb{W}^n}(A,B)=0$ if and only if G_A and G_B receive identical colors by the WL algorithm. If $S=\mathbb{O}^n$ and $\Omega=\mathbb{S}^n$, i.e., graphs are undirected, then $\Omega/\sim_{d_{\mathbb{O}^n}}$ is determined by *co-spectrality*: $d_{\mathbb{O}^n}(A,B)=0$ if and only if A,B have the same spectrum. When $\Omega=\mathbb{R}^{n\times n}$, $d_{\mathbb{O}^n}(A,B)=0$ implies that A,B are co-spectral, but co-spectral matrices A,B do not necessarily satisfy $d_{\mathbb{O}^n}(A,B)=0$.

3.1 Proof of Theorems 3.1–3.3. We define several properties that play a crucial role in our proofs. We say that a set $S\subseteq\mathbb{R}^{n\times n}$ is *closed under multiplication* if $P,P'\in S$ implies that $P\cdot P'\in S$. We say that S is *closed under transposition* if $P\in S$ implies that $P^\top\in S$, and *closed under inversion* if $P\in S$ implies that $P^{-1}\in S$. Finally, given a matrix norm $\|\cdot\|$, we say that set S is *contractive* w.r.t. $\|\cdot\|$ if $\|AP\|\le\|A\|$ and $\|PA\|\le\|A\|$, for all $P\in S$ and $A\in\mathbb{R}^{n\times n}$. Put differently, S is contractive if and only if every $P\in S$ is a contraction w.r.t. $\|\cdot\|$. We rely on several lemmas, whose proofs can be found in [10]. The first three establish conditions under which (2.4) satisfies the triangle inequality (2.3d), symmetry (2.3c), and weak property (2.3e), respectively:

LEMMA 3.1. *Given a matrix norm $\|\cdot\|$, suppose that set S is (a) contractive w.r.t. $\|\cdot\|$, and (b) closed under multiplication. Then, for any $A,B,C\in\mathbb{R}^{n\times n}$, d_S given by (2.4) satisfies $d_S(A,C)\le d_S(A,B)+d_S(B,C)$.*

LEMMA 3.2. *Given a matrix norm $\|\cdot\|$, suppose that $S\subset\mathbb{R}^{n\times n}$ is (a) contractive w.r.t. $\|\cdot\|$, and (b) closed under inversion. Then, for all $A,B\in\mathbb{R}^{n\times n}$, $d_S(A,B)=d_S(B,A)$.*

LEMMA 3.3. *If $I\in S$, then $d_S(A,A)=0$ for all $A\in\mathbb{R}^{n\times n}$.*

Both the set of permutation matrices $\mathbb{P}^n$ *and* the Stiefel manifold $\mathbb{O}^n$ are *groups* w.r.t. matrix multiplication: they are closed under multiplication, contain the identity I, and are closed under inversion. Hence, if they are also contractive w.r.t. a matrix norm $\|\cdot\|$, $d_{\mathbb{P}^n}$ and $d_{\mathbb{O}^n}$ defined in terms of this norm satisfy all assumptions of Lemmas 3.1–3.3. We therefore turn our attention to this property.

LEMMA 3.4. *Let $\|\cdot\|$ be any operator or entry-wise norm. Then, $S=\mathbb{P}^n$ is contractive w.r.t. $\|\cdot\|$.*

Hence, Theorem 3.1 follows as a direct corollary of Lemmas 3.1–3.4. Indeed, $d_{\mathbb{P}^n}$ is non-negative, symmetric by Lemmas 3.2 and 3.4, satifies the triangle inequality by Lemmas 3.1 and 3.4, as well as property (2.3e) by Lemma 3.3; hence $d_{\mathbb{P}^n}$ is a pseudometric over $\mathbb{R}^{n\times n}$. Our next lemma shows that the Stiefel manifold $\mathbb{O}^n$ is contractive for 2-norms:

LEMMA 3.5. *Let $\|\cdot\|$ be the operator 2-norm or the Frobenius norm. Then, $S=\mathbb{O}^n$ is contractive w.r.t. $\|\cdot\|$.*

Theorem 3.3 follows from Lemmas 3.1–3.3 and Lemma 3.5, along with the the fact that $\mathbb{O}^n$ is a group. Note that $\mathbb{O}^n$ is *not* contractive w.r.t. other norms, e.g., $\|\cdot\|_1$ or $\|\cdot\|_\infty$. Lemma 3.4 along with the Birkoff-von Neumann theorem imply that $\mathbb{W}^n$ is also contractive:

LEMMA 3.6. *Let $\|\cdot\|$ be any operator or entry-wise norm. Then, $\mathbb{W}^n$ is contractive w.r.t. $\|\cdot\|$.*

The Birkhoff polytope $\mathbb{W}^n$ is *not* a group, as it is not closed under inversion. Nevertheless, it is closed under transposition; in establishing (partial) symmetry of $d_{\mathbb{W}^n}$, we leverage the following lemma:

LEMMA 3.7. *Suppose that $\|\cdot\|$ is transpose invariant, and S is closed under transposition. Then, $d_S(A,B)=d_S(B,A)$ for all $A,B\in\mathbb{S}^n$.*

The first part of Theorem 3.2 therefore follows from Lemmas 3.1, 3.3, and 3.6, as $\mathbb{W}^n$ is closed under transposition, contains the identity I, and is closed under multiplication, while all entry-wise norms are transpose invariant. Operator norms are not transpose invariant. However, if $\|\cdot\|$ is an operator norm, or $\Omega=\mathbb{R}^{n\times n}$, then Lemma 3.6 and Lemma 3.1 imply that $d_{\mathbb{W}^n}$ satisfies non-negativity (2.3a) and the triangle inequality (2.3d), while Lemma 3.3 implies that it satisfies (2.3e). These properties are inherited by extension $\bar{d}_S$, which also satisfies symmetry (2.3c), and Theorem 3.2 follows. □

4 Incorporating Metric Embeddings

We have seen that the chemical distance $d_{\mathbb{P}^n}$ can be *relaxed* to $d_{\mathbb{W}^n}$ or $d_{\mathbb{O}^n}$, gaining tractability while still maintaining the metric property. In practice, nodes in a

graph often contain additional atributes that one might wish to leverage when computing distances. In this section, we show that such attributes can be seamlessly incorporated in d_S either as soft or hard constraints, *without violating the metric property.*

Metric Embeddings. Given a graph G_A of size n, a *metric embedding* of G_A is a mapping $\psi_A : [n] \to \tilde{\Omega}$ from the nodes of the graph to a metric space $(\tilde{\Omega}, \tilde{d})$. That is, ψ_A maps nodes of the graph to $\tilde{\Omega}$, where $\tilde{\Omega}$ is endowed with a metric $\tilde{d}$. We refer to a graph endowed with an embedding ψ_A as an *embedded graph*, and denote this by (A, ψ_A), where $A \in \mathbb{R}^{n\times n}$ is the adjacency matrix of G_A. We list two examples:

Example 1: Node Attributes. Consider an embedding of a graph to $(\mathbb{R}^k, \|\cdot\|_2)$ in which every node $v \in V$ is mapped to a k-dimensional vector describing "local" attributes. These can be *exogenous*: e.g., features extracted from a user's profile (age, binarized gender, etc.) in a social network. Alternatively, attributes may be *endogenous* or *structural*, extracted from the adjacency matrix A, e.g., the node's degree, the size of its k-hop neigborhood, its page-rank, etc.

Example 2: Node Colors. Let $\tilde{\Omega}$ be an arbitrary finite set endowed with the Kronecker delta as a metric, that is, for $s, s' \in \tilde{\Omega}$, $\tilde{d}(s, s') = 0$ if $s = s'$, while $\tilde{d}(s, s') = \infty$ if $s \neq s'$. Given a graph G_A, a mapping $\psi_A : [n] \to \tilde{\Omega}$ is then a metric embedding. The values of $\tilde{\Omega}$ are invariably called *colors* or *labels*, and a graph embedded in $\tilde{\Omega}$ is a *colored* or *labeled* graph. Colors can again be *exogenous* or *structural*: e.g., if the graph represents an organic molecule, colors can correspond to atoms, while structural colors can be, e.g., the output of the WL algorithm [59] after k iterations.

As discussed below, node attributes translate to *soft* constraints in metric (2.4), while node colors correspond to *hard* constraints. The unified view through embeddings allows us to establish metric properties for both simultaneously (c.f. Thm. 4.1 and 4.2) .

Embedding Distance. Consider two embedded graphs (A, ψ_A), (B, ψ_B) of size n that are embedded *in the same metric space* $(\tilde{\Omega}, \tilde{d})$. For $u \in [n]$ a node in the first graph, and $v \in [n]$ a node in the second graph, the embedded distance between the two nodes is given by $\tilde{d}(\psi_A(u), \psi_B(v))$. Let $D_{\psi_A,\psi_B} = [\tilde{d}(\psi_A(u), \psi_B(v))]_{u\in V, v\in V} \in \mathbb{R}_+^{n\times n}$ be the corresponding matrix of embedded distances. After mapping nodes to the same metric space, it is natural to seek $P \in \mathbb{P}^n$ that preserve the *embedding distance*. This amounts to finding a $P \in \mathbb{P}^n$ that minimizes:

$$\mathsf{tr}\left(P^\top D_{\psi_A,\psi_B}\right) = \textstyle\sum_{u,v\in[n]} P_{u,v}\tilde{d}(\psi_A(u), \psi_B(v)). \tag{4.5}$$

Note that, in the case of colored graphs and the Kronecker delta distance, minimizing (4.5) finds a $P \in \mathbb{P}^n$ that maps nodes in A nodes in B of equal color. It is not hard to verify[1] that $\min_{P\in\mathbb{P}^n} \mathsf{tr}\left(P^\top D_{\psi_A,\psi_B}\right)$ induces a metric between graphs embedded in $(\tilde{\Omega}, \tilde{d})$. Despite the combinatorial nature of $\mathbb{P}^n$, (4.5) is a maximum weighted matching problem, which can be solved through, e.g., the Hungarian algorithm [39] in polynomial time in n. We note that this metric is not as expressive as (2.4): depending on the definition of the embeddings ψ_A, ψ_B, attributes may only capture "local" similarities between nodes, as opposed to the "global" view of a mapping attained by (2.4).

A Unified, Tractable Metric. Motivated by the above considerations, we focus on unifying the "global" metric (2.4) with the "local" metrics induced by arbitrary graph embeddings. Proofs for the two theorems below are provided in the supplement. Given a metric space $(\tilde{\Omega}, \tilde{d})$, let $\Psi_{\tilde{\Omega}}^n = \{\psi : [n] \to \tilde{\Omega}\}$ be the set of all mappings from $[n]$ to $\tilde{\Omega}$. Then, given two embedded graphs $(A, \psi_A), (B, \psi_B) \in \mathbb{R}^{n\times n} \times \Psi_{\tilde{\Omega}}^n$, we define:

$$\begin{aligned} d_S\left((A, \psi_A), (B, \psi_B)\right) = \min_{P\in S} \big[\|AP - PB\| + \dots \\ + \mathsf{tr}(P^\top D_{\psi_A,\psi_B})\big] \end{aligned} \tag{4.6}$$

for some compact set $S \subset \mathbb{R}^{n\times n}$ and matrix norm $\|\cdot\|$. Our next result states that incorporating this linear term does not affect the pseudometric property of d_S.

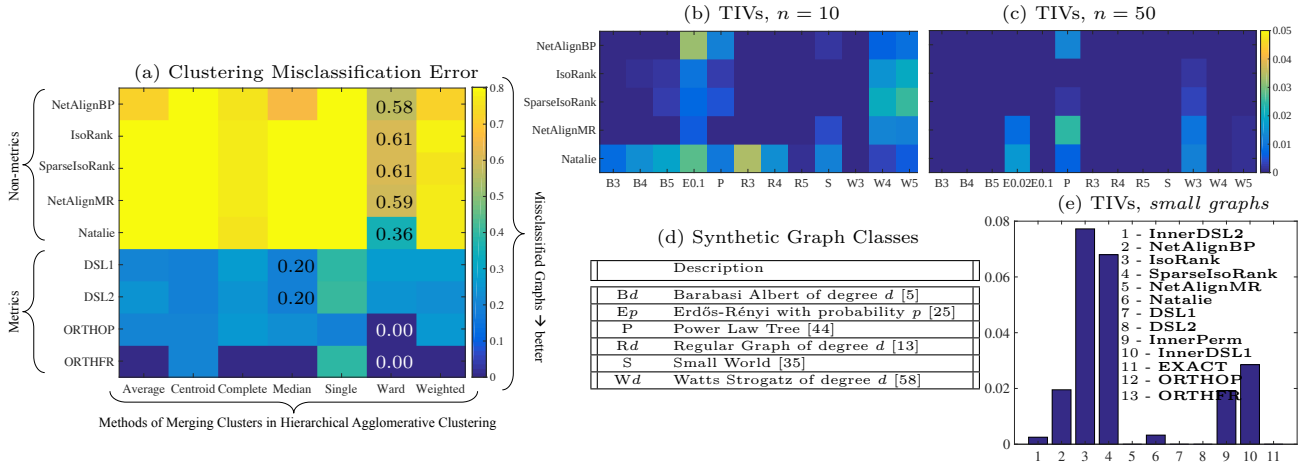
THEOREM 4.1. *If $S = \mathbb{P}^n$ and $\|\cdot\|$ is an arbitrary entry-wise or operator norm, then d_S given by* (4.6) *is a pseudometric over the set of embedded graphs $\Omega = \mathbb{R}^{n\times n} \times \Psi_{\tilde{\Omega}}^n$.*

We stress here that this result is non-obvious: is not true that adding *any* linear term to d_S leads to a quantity that satisfies the triangle inequality. It is precisely because D_{ψ_A,ψ_B} contains pairwise distances that Theorem 4.1 holds. We can similarly extend Theorem 3.2:

THEOREM 4.2. *If $S = \mathbb{W}^n$ and $\|\cdot\|$ is an arbitrary entry-wise norm, then d_S given by* (4.6) *is a pseudometric over $\Omega = \mathbb{S}^n \times \Psi_{\tilde{\Omega}}^n$, the set of symmetric graphs embedded in $(\tilde{\Omega}, \tilde{d})$. Moreover, if $\|\cdot\|$ is an arbitrary entry-wise or operator norm, then the symmetric extension $\bar{d}_S$ of* (4.6) *is a pseudometric over $\Omega = \mathbb{R}^{n\times n} \times \Psi_{\tilde{\Omega}}^n$.*

Adding the linear term (4.5) in d_S has significant practical advantages. Beyond expressing exogenous attributes, a linear term involving colors, combined with a Kronecker distance, translates into *hard* constraints: any permutation attaning a finite objective value *must* map nodes in one graph to nodes of the same color.

[1] This follows from Thm. 4.1 for $A = B = 0$, i.e., for distances between embedded graphs with no edges.



Description	
Bd	Barabasi Albert of degree d [5]
Ep	Erdős-Rényi with probability p [25]
P	Power Law Tree [44]
Rd	Regular Graph of degree d [13]
S	Small World [35]
Wd	Watts Strogatz of degree d [58]

Figure 1: A clustering experiment using metrics and non-metrics (y-axis) for different clustering parameters (x-axis) is shown in *(a), left*. We sample graphs with $n = 50$ nodes from the six classes, shown in the adjacent table in *(d), bottom-center*. We compute distances between them using nine different algorithms from Table 1. Only the distances in our family (DSL1, DSL2, ORTHOP, and ORTHFR) are metrics. The resulting graphs are clustered using hierarchical agglomerative clustering [29] using *Average*, *Centroid*, *Complete*, *Median*, *Single*, *Ward*, *Weighted* as a means of merging clusters. Colors represent the fraction of misclassified graphs, with the minimal misclassification rate per distance labeled explicitly. Metrics outperform other distance scores across all clustering methods. The error rate of a random guess is ≈ 0.8. Subfigures *(b)* and *(c), top center and right,* shows that non-metric distances produce triangle inequality violations (TIVs) which contribute to poor clustering results; the figure shows the fraction of TIVs within different 10-node and 50 node graph families under these algorithms. Finally, subfigure *(e), bottom right*, shows the fraction of triangle inequality violations for different algorithms on the *small graphs* dataset of all 7-node graphs.

Theorem 4.2 therefore implies that such constraints can thus be added to the optimization problem, while maintaining the metric property. In practice, as the number of variables in optimization problem (2.4) is n^2, incorporating such hard constraints can significantly reduce the problem's computation time; we illustrate this in the next section. Note that adding (4.5) to $d_{\mathbb{O}^n}$ does *not* preserve the metric propery.

5 Experiments

Graphs. We use *synthetic graphs* from six classes summarized in the table in Fig. 1(d). In addition, we use a dataset of *small graphs*, comprising all 853 connected graphs of 7 nodes [45]. Finally, we use a *collaboration graph* with 5242 nodes and 14496 edges representing author collaborations [41].

Algorithms. We compare our metrics to several competitors outlined in Table 1 (see also [10] for additional details). All receive only two unlabeled undirected simple graphs A and B and output a matching a matrix $\hat{P}$ either in $\mathbb{W}^n$ or in $\mathbb{P}^n$ estimating P^*. If $\hat{P} \in \mathbb{P}^n$, we compute $\|A\hat{P} - \hat{P}B\|_1$. If $\hat{P} \in \mathbb{W}^n$, then we compute both $\|A\hat{P} - \hat{P}B\|_1$ and $\|A\hat{P} - \hat{P}B\|_F$; all norms are entry-wise. We also implement our two relaxations $d_{\mathbb{W}}$ and $d_{\mathbb{O}^n}$, for two different matrix norm combinations.

Clustering Graphs. The difference between our metrics and non-metrics is striking when clustering graphs. This is illustrated by the clustering experiment shown in Fig. 1(a). Graphs of size $n = 50$ from the 6 classes in Fig. 1(d) are clustered together through hierarchical agglomerative clustering. We compute distances between them using nine different algorithms; only the distances in our family (DSL1, DSL2, ORTHOP, and ORTHFR) are metrics. The quality of clusters induced by our metrics are far superior than clusters induced by non-metrics; in fact, **ORTHOP** and **ORTHFR** can lead to no misclassifications. This experiment strongly suggests our produced metrics correctly capture the topology of the metric space between these larger graphs.

(Non-metric) Distance Score Algorithms	
NetAlignBP	Network Alignment using Belief Propagation [9, 33]
IsoRank	Neighborhood Topology Isomorphism using Page Rank [54, 33]
SparseIsoRank	Neighborhood Topology Sparse Isomorphism using Page Rank [9, 33]
InnerPerm	Inner Product Matching with Permutations [42]
InnerDSL1	Inner Product Matching with Matrices in $\mathbb{W}^n$ and entry-wise 1-norm [42]
InnerDSL2	Inner Product Matching with Matrices in $\mathbb{W}^n$ and Frobenius norm [42]
NetAlignMR	Iterative Matching Relaxation [34, 33]
Natalie (V2.0)	Improved Iterative Matching Relaxation [23, 22]
Metrics from our Family (2.4)	
EXACT	Chemical Distance via brute force search over GPU
DSL1	Doubly Stochastic Chemical Distance $d_{\mathbb{W}^n}$ with entry-wise 1-norm
DSL2	Doubly Stochastic Chemical Distance $d_{\mathbb{W}^n}$ with Frobenius norm
ORTHOP	Orthogonal Relaxation of Chemical Distance $d_{\mathbb{O}^n}$ with operator 2-norm
ORTHFR	Orthogonal Relaxation of Chemical Distance $d_{\mathbb{O}^n}$ with Frobenius norm

Table 1: Competitor Distance Scores & Our Metrics

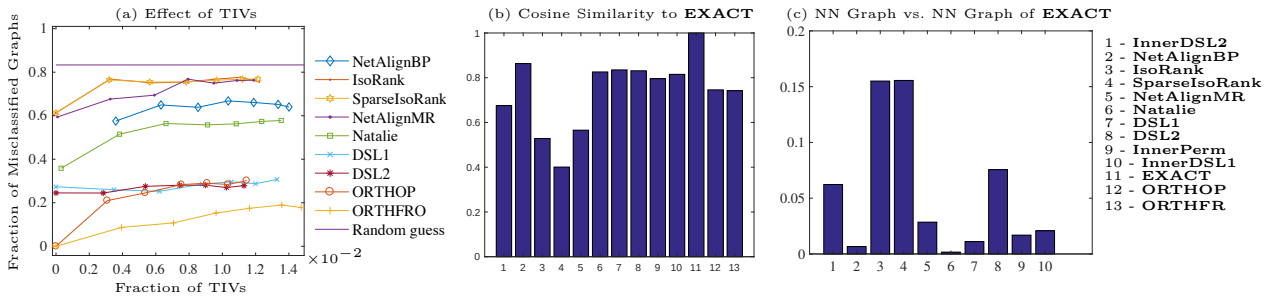


Figure 2: *(a)* Effect of introducing TIVs on the performance of different algorithms on the clustering experiment of Figure 1(a) when using the Ward method. *(b)* Cosine similarity between the Laplacian of distances produced by each algorithm and the one by **EXACT**. *(c)* Distance between nearest neighbor (NN) graphs induced by different algorithms and NN graph induced by **EXACT**.

(a) Coloring Constraints

k	$\|P\|_0$	$\|AP-PA\|_0$	τ
1	3,747,960	100.569	133s
2	239,048	3,004	104s
3	182,474	2,036	136s
4	182,016	2,030	169s
5	182,006	2,030	200s

(b) Convergence of ADMM

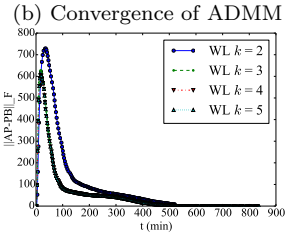


Figure 3: *(a)* Effect of coloring/hard constraints on the numbers of variables ($\|P\|_0$) and terms of objective ($\|AP - PA\|_0$) using k iterations of the WL coloring algorithm. The last column shows the execution time of WL on a 40 CPU machine using Apache Spark [62]. *(b)* Convergence of ADMM algorithm [15] computing **DSL2** on two copies of the collaboration graph as a function of time, implemented using Apache Spark [62] on a 40 CPU machine.

Triangle Inequality Violations (TIV). Given graphs A, B and C and a distance d, a TIV occurs when $d(A, C) > d(A, B) + d(B, C)$. Being metrics, none of our distances induce TIVs; this is not the case for the remaining algorithms in Table 1. Fig. 1(b) and (c) show the TIV fraction across the synthetic graphs of Fig. 1(d), while Fig. 1(e) shows the fraction of TIVs found on the 853 small graphs ($n = 7$). **NetAlignMR** also produces no TIVs on the small graphs, but it does induce TIVs in synthetic graphs. We observe that it is easier to find TIVs when graphs are close: in synthetic graphs, TIVs abound for $n = 10$. No algorithm performs well across all categories of graphs.

Effect of TIVs on Clustering. Next, to investigate the effect of TIVs on clustering, we artificially introduced triangle inequality violations into the pairs of distances between graphs. We then re-evaluated clustering performance for hierarchical agglomerative clustering using the *Ward* method, which performed best in Fig. 1(a). Fig. 2(a) shows the fraction of misclassified graphs as the fraction of TIVs introduced increases. To incur as small a perturmbation on distances as possible, we introduce TIVs as follows: For every three graphs, A, B, C, with probability p, we set $d(A, C) = d(A, B) + d(B, C)$. Although this does not introduce a TIV w.r.t. A,B, and C, this distortion does introduce TIVs w.r.t. other triplets involving A and C. We repeat this 20 times for each algorithm and each value of p, and compute the average fraction of TIVs, shown in the x-axis, and the average fraction of misclassified graphs, shown in the y-axis. As little as 1% TIVs significantly deteriorate clustering performance. We also see that, even after introducing TIVs, clustering based on metrics outperforms clustering based on non-metrics.

Comparison to Chemical Distance. We compare how different distance scores relate to the chemical distance **EXACT** through two experiments on the small graphs (computation on larger graphs is prohibitive). In Figure 2(b), we compare the distances between small graphs with 7 nodes produced by the different algorithms and **EXACT** using the DISTATIS method of [1]. Let $D \in \mathbb{R}_+^{835\times 835}$ be the matrix of distances between graphs under an algorithm. DISTATIS computes the normalized Laplacian of this matrix, given by $L = -UDU/\|UDU\|_2$ where $U = I - \frac{\mathbf{1}\mathbf{1}^\top}{n}$. The DISTATIS score is the cosine similarity of such Laplacians (vectorized). We see that our metrics produce distances attaining high similarity with **EXACT**, though **NetAlignBP** has the highest similarity. We measure proximity to **EXACT** with an additional test. Given D, we compute the nearest neighbor (NN) meta-graph by connecting a graph in D to every graph at distance less than its average distance to other graps. This results in a (labeled) meta-graph, which we can compare to the NN meta-graph induced by other algorithms, measuring the fraction of distinct edges. Fig. 2(c) shows that our algorithms perform quite well, though **Natalie** yields the smallest distance to **EXACT**.

Incorporating Constraints. Computation costs can be reduced through metric embeddings, as in (4.6). To show this, we produce a copy of the 5242 node collaboration graph with permuted node labels. We then run the WL algorithm [59] to produce structural colors, which induce coloring constraints on $P \in \mathbb{W}^n$. The support of P (i.e., the number of variables in the optimization (2.4)), the support of $AP - PA$

(i.e., the number of non-zero summation terms in the objective of (2.4)), as well as the execution time τ of the WL algorithm, are summarized in Fig. 3(b). The original unconstrained problem involves $5242^2 \approx 27.4$M variables. However, after using WL and induced costraints, the effective dimension of the optimization problem (2.4) reduces considerably. This, in turn, speeds up convergence time, shown in Fig. 3(b): including the time to compute constraints, a solution is found 110 times faster after the introduction of the constraints.

6 Conclusion

Our work suggests that incorporating soft and hard constraints has a great potential to further improve the efficiency of our metrics. In future work, we intend to investigate and characterize the resulting equivalence classes under different soft and hard constraints and to quantify these gains in efficiency, especially in parallel implementations like ADMM. Determining the necessity of the conditions used in proving that d_S is a metric is also an open problem.

Acknowledgements

The authors gratefully acknowledge the support of the National Science Foundation (grants IIS-1741197,IIS-1741129) and of the National Institutes of Health (grant 1U01AI124302).

References

[1] H Abdi, A J O'Toole, D Valentin, and B Edelman. DISTATIS: The analysis of multiple distance matrices. In *CVPR Workshops*, 2005.

[2] P-A Absil, R Mahony, and R Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[3] M R Ackermann, J Blömer, and C Sohler. Clustering for metric and nonmetric distance measures. *ACM Transactions on Algorithms (TALG)*, 6(4):59, 2010.

[4] Y Aflalo, A Bronstein, and R Kimmel. On convex relaxation of graph isomorphism. *PNAS*, 112(10):2942–2947, 2015.

[5] R Albert and A-L Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.

[6] F H Allen. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B: Structural Science*, 58(3):380–388, 2002.

[7] F Angiulli and C Pizzuti. Fast outlier detection in high dimensional spaces. In *PKDD*, 2002.

[8] L Babai. Graph isomorphism in quasipolynomial time [extended abstract]. In *STOC*, 2016.

[9] M Bayati, M Gerritsen, D F Gleich, A Saberi, and Y Wang. Algorithms for large, sparse network alignment problems. In *ICDM*, 2009.

[10] J Bento and S Ioannidis. A family of tractable graph distances, 2018. `https://arxiv.org/abs/1801.04301`.

[11] A Beygelzimer, S Kakade, and J Langford. Cover trees for nearest neighbor. In *ICML*, 2006.

[12] G Birkhoff. Three observations on linear algebra. *Univ. Nac. Tucumán. Revista A*, 5:147–151, 1946.

[13] B Bollobás. Random graphs. In *Modern Graph Theory*, pages 215–252. Springer, 1998.

[14] S Boyd and L Vandenberghe. *Convex Optimization*. Cambridge university press, 2004.

[15] S Boyd, N Parikh, E Chu, B Peleato, and J Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[16] H Bunke. On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, 18(8):689–694, 1997.

[17] H Bunke and K Shearer. A graph distance metric based on the maximal common subgraph. *Pattern Recognition Letters*, 19(3):255–259, 1998.

[18] G Chartrand, G Kubicki, and M Schultz. Graph similarity and distance in graphs. *Aequationes Mathematicae*, 55(1-2): 129–145, 1998.

[19] K L Clarkson. Nearest neighbor queries in metric spaces. *Discrete & Computational Geometry*, 22(1):63–93, 1999.

[20] K L Clarkson. Nearest-neighbor searching and metric space dimensions. *Nearest-Neighbor Methods for Learning and Vision: Theory and Practice*, pages 15–59, 2006.

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

[22] M El-Kebir, J Heringa, and G Klau. Natalie, a tool for pairwise global network alignment. `http://www.mi.fu-berlin.de/w/LiSA/Natalie`.

[23] M El-Kebir, J Heringa, and G W Klau. Natalie 2.0: Sparse global network alignment as a special case of quadratic assignment. *Algorithms*, 8(4):1035–1051, 2015.

[24] H Elghawalby and E R Hancock. Measuring graph similarity using spectral geometry. In *ICIAR*, 2008.

[25] P Erdös and A Rényi. On random graphs, i. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

[26] S Fankhauser, K Riesen, and H Bunke. Speeding up graph edit distance computation through fast bipartite matching. In *GBR*, 2011.

[27] M Ferrer, E Valveny, F Serratosa, K Riesen, and H Bunke. Generalized median graph computation by means of graph embedding in vector spaces. *Pattern Recognition*, 43(4):1642–1655, 2010.

[28] M R Garey and D S Johnson. *Computers and Intractability*, volume 29. WH Freeman New York, 2002.

[29] J A Hartigan. *Clustering algorithms*. Wiley New York, 1975.

[30] L He, C Y Han, and W G Wee. Object recognition and recovery by skeleton graph matching. In *ICME*, 2006.

[31] P Indyk. Sublinear time algorithms for metric space problems. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 428–434. ACM, 1999.

[32] B J Jain. On the geometry of graph spaces. *Discrete Applied Mathematics*, 214:126–144, 2016.

[33] A Khan, D Gleich, M Halappanavar, and A Pothen. Multicore codes for network alignment. `https://www.cs.purdue.edu/homes/dgleich/codes/netalignmc/`.

[34] G W Klau. A new graph-based method for pairwise global network alignment. *BMC bioinformatics*, 10(1):S59, 2009.

[35] J Kleinberg. The small-world phenomenon: An algorithmic perspective. In *STOC*, 2000.

[36] J Koca, M Kratochvil, V Kvasnicka, L Matyska, and J Pospichal. *Synthon model of organic chemistry and synthesis design*, volume 51. Springer Science & Business Media, 2012.

[37] D Koutra, H Tong, and D Lubensky. Big-align: Fast bipartite graph alignment. In *ICDM*, 2013.

[38] D Koutra, J T Vogelstein, and C Faloutsos. Deltacon: A principled massive-graph similarity function. In *SDM*, 2013.

[39] H W Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.

[40] V Kvasnička, J Pospíchal, and V Baláž. Reaction and chemical distances and reaction graphs. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*, 79(1):65–79, 1991.

[41] J Leskovec, J Kleinberg, and C Faloutsos. Stanford large network dataset collection. `http://snap.stanford.edu/data/ca-GrQc.html`.

[42] V Lyzinski, D E Fishkind, M Fiori, J T Vogelstein, C E Priebe, and Guillermo Sapiro. Graph matching: Relax at your own risk. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):60–73, 2016.

[43] O Macindoe and W Richards. Graph comparison using fine structure analysis. In *SocialCom*, 2010.

[44] H M Mahmoud, R T Smythe, and J Szymański. On the structure of random plane-oriented recursive trees and their branches. *Random Structures & Algorithms*, 4(2):151–176, 1993.

[45] B McKay. List of 7 node connected graphs. `http://users.cecs.anu.edu.au/~bdm/data/graphs.html`.

[46] P Papadimitriou, A Dasdan, and H Garcia-Molina. Web graph similarity for anomaly detection. *Journal of Internet Services and Applications*, 1(1):19–30, 2010.

[47] M V Ramana, E R Scheinerman, and D Ullman. Fractional isomorphism of graphs. *Discrete Mathematics*, 132(1-3):247–265, 1994.

[48] K Riesen and H Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision Computing*, 27(7):950–959, 2009.

[49] K Riesen and H Bunke. *Graph classification and clustering based on vector space embedding*, volume 77. World Scientific, 2010.

[50] K Riesen, M Neuhaus, and H Bunke. Graph embedding in vector spaces by means of prototype selection. In *GBR*, 2007.

[51] A Sanfeliu and KS Fu. A distance measure between attributed relational graphs for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, (3):353–362, 1983.

[52] C Schellewald, S Roth, and C Schnörr. Evaluation of convex optimization techniques for the weighted graph-matching problem in computer vision. In *Pattern Recognition*, 2001.

[53] T B Sebastian, P N Klein, and B B Kimia. Recognition of shapes by editing their shock graphs. *IEEE Transactions on pattern analysis and machine intelligence*, 26(5):550–571, 2004.

[54] R Singh, J Xu, and B Berger. Pairwise global alignment of protein interaction networks by matching neighborhood topology. In *RECOMB*, 2007.

[55] S Soundarajan, T Eliassi-Rad, and B Gallagher. A guide to selecting a network similarity method. In *SDM*, 2014.

[56] S Umeyama. An eigendecomposition approach to weighted graph matching problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5):695–703, 1988.

[57] J T Vogelstein, J M Conroy, L J Podrazik, S G Kratzer, E T Harley, D E Fishkind, R J Vogelstein, and C E Priebe. Large (brain) graph matching via fast approximate quadratic programming. *arXiv preprint arXiv:1112.5507*, 2011.

[58] D J Watts and S H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[59] B Weisfeiler and A A Lehman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsia*, 2(9):12–16, 1968.

[60] R C Wilson and P Zhu. A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9): 2833–2841, 2008.

[61] E P Xing, A Y Ng, M I Jordan, and S Russell. Distance metric learning with application to clustering with side-information. In *NIPS*, volume 15, page 12, 2002.

[62] M Zaharia, M Chowdhury, M J Franklin, S Shenker, and I Stoica. Spark: Cluster computing with working sets. *HotCloud*, 10(10-10):95, 2010.

[63] P Zhu and R C Wilson. A study of graph spectra for comparing graphs. In *BMVC*, 2005.