Tractable n-Metrics for Multiple Graphs

Sam Safavi [1] **José Bento** [1]

Abstract

Graphs are used in almost every scientific discipline to express relations among a set of objects. Algorithms that compare graphs, and output a closeness score, or a correspondence among their nodes, are thus extremely important. Despite the large amount of work done, many of the scalable algorithms to compare graphs do not produce closeness scores that satisfy the intuitive properties of metrics. This is problematic since non-metrics are known to degrade the performance of algorithms such as distance-based clustering of graphs (Bento & Ioannidis, 2018). On the other hand, the use of metrics increases the performance of several machine learning tasks (Indyk, 1999; Clarkson, 1999; Angiulli & Pizzuti, 2002; Ackermann et al., 2010). In this paper, we introduce a new family of multi-distances (a distance between more than two elements) that satisfies a generalization of the properties of metrics to multiple elements. In the context of comparing graphs, we are the first to show the existence of multi-distances that simultaneously incorporate the useful property of *alignment consistency* (Nguyen et al., 2011), and a generalized metric property. Furthermore, we show that these multi-distances can be relaxed to convex optimization problems, without losing the generalized metric property.

1. Introduction

A canonical way to check if two graphs G_1 and G_2 are similar, is to try to find a map P from the nodes of G_2 to the nodes of G_1 such that, for many pairs of nodes in G_2, their images in G_1 through P have the same connectivity relation (connected/disconnected) (Deza & Deza, 2009). For equal-sized graphs, this can be formalized as

$$d(G_1, G_2) \triangleq \min_P \{ \|\|A_1 - PA_2P^\top\|\| = \|\|A_1P - PA_2\|\| \}, \quad (1)$$

[1]Department of Computer Science, Boston College, Chestnut Hill, MA, USA. Correspondence to: José Bento <jose.bento@bc.edu>.

where A_1 and A_2 are the adjacency matrices of G_1 and G_2, P and its transpose $P^\top$ are permutation matrices, and, here, $\|\|\cdot\|\|$ is the Frobenius norm. A map P^* that minimizes (1) is called an optimal *alignment* or *match* between G_1 and G_2. If $d(G_1, G_2)$ is small (resp. large), we say G_1 and G_2 are topologically similar (resp. dissimilar). Computing d, or P^*, is hard (Klau, 2009). Determining if $d(G_1, G_2) = 0$, which is the graph isomorphism problem, is not known to be in P, or in NP-hard (Babai, 2016).

Scalable alignment algorithms, which find an approximation P to an optimal alignment P^*, or find a solution to a tractable variant of (1), e.g., (Klau, 2009; Bayati et al., 2013; Singh et al., 2008; El-Kebir et al., 2015), have mostly been developed with no concern as to whether the closeness score d obtained from the alignment P, e.g., computed via $d(G_1, G_2) = \|A_1P - PA_2\|$, results in a non-metric. An exception is the recent work in (Bento & Ioannidis, 2018). Indeed for the methods in, e.g., (Klau, 2009; Bayati et al., 2013; Singh et al., 2008; El-Kebir et al., 2015), the work of (Bento & Ioannidis, 2018) shows that one can find two graphs that are individually similar to a third one, but not similar to each other, according to d. Furthermore, (Bento & Ioannidis, 2018) shows how the lack of the metric properties can lead to a degraded performance in a clustering task to automatically classify different graphs into the categories: Barabasi Albert, Erdos-Renyi, Power Law Tree, Regular graph, and Small World. At the same time, the metric properties allow us to solve several machine learning tasks efficiently (Indyk, 1999; Clarkson, 1999; Angiulli & Pizzuti, 2002; Ackermann et al., 2010), as we now illustrate.

Diameter estimation: Given a set S with $|S|$ graphs, we can compute the maximum diameter $\Delta \triangleq \max_{G_1, G_2 \in S} d(G_1, G_2)$ by computing $\binom{|S|}{2}$ distances. However, if d is a metric, we know that there are at least $\Omega(|S|)$ pairs of graphs with $d \geqslant \Delta/2$. Indeed, if $d(G^*, G_*) = \Delta$, then, by the triangle inequality, for any $G \in S$, we cannot have both $d(G^*, G) < \Delta/2$ and $d(G_*, G) < \Delta/2$. Therefore, if we evaluate d on random pairs of graphs, we are guaranteed to find an $1/2$-approximation of Δ with only $\mathcal{O}(|S|)$ distance computations, on average.

Being able to compare two graphs is important in many fields such as biology (Kalaev et al., 2008; Zaslavskiy et al.,

2009a; Kelley et al., 2004; Weskamp et al., 2007), object recognition (Conte et al., 2004), dealing with ontologies (Hu et al., 2008; Wang et al., 2016), computer vision (Conte et al., 2004), and social networks (Zhang & S. Yu, 2015), and graph clustering (Ma et al., 2016), to name a few. In many applications, however, one needs to jointly compare multiple graphs. This is the case, for example, in aligning protein-protein interaction networks (Singh et al., 2008), recommendation systems, in the collective analysis of networks, or in the alignment of graphs obtained from brain MRI (Papo et al., 2014).

The problem of jointly comparing n graphs, $n \geqslant 3$, is harder, and has been studied far less than when $n = 2$. Examples and applications include (Pachauri et al., 2013; Douglas et al., 2018; Yan et al., 2015b; Gold & Rangarajan, 1996; Hu et al., 2016; Park & Yoon, 2016; Huang & Guibas, 2013; Solé-Ribalta & Serratosa, 2011; Williams et al., 1997; Hashemifar et al., 2016; Heimann et al., 2018; Nassar & Gleich, 2017; Feizi et al., 2016; Chen et al., 2014).

Consider the search for a function $d(G_1, ..., G_n)$ that scores how close $G_1, ..., G_n$ are. New questions arise when $n \geqslant 3$:

1. If d produces alignments between each pair of graphs in $\{G_1, \ldots, G_n\}$, should these alignments be related? What properties should they satisfy?

2. Should d satisfy similar properties to that of a metric? What properties?

3. Is it possible to find a d that is tractable? Is it possible to impose on d the properties from 1 and 2 above without losing tractability?

Multi-graph alignment scores, are important in many applications. For example, many problems require clustering using nth order interaction (Leordeanu & Sminchisescu, 2012), i.e., clustering based on the similarity of groups of n elements, not just groups of two elements, as in spectral, or hierarchical clustering. Furthermore, having a score function $d(G_1, \ldots, G_n)$ with some form of generalized metric property can have advantages, similar to what (Bento & Ioannidis, 2018) showed for metrics (cf. Section 4).

In this paper, we are the first to provide a family of similarity scores for jointly comparing multiple graphs that simultaneously (a) give intuitive joint alignments between graphs, (b) satisfy similar properties to those of metrics, and (c) can be computed using convex optimization methods.

2. Related work

Consider three graphs G_1, G_2, and G_3, and three permutation matrices $P_{1,2}$, $P_{2,3}$ and $P_{1,3}$, where the map $P_{i,j}$ is an alignment between the nodes of graphs G_i and G_j. An intuitive property that is often required for these alignments is that if $P_{1,2}$ maps (the nodes of) G_1 to G_2, and if $P_{2,3}$ maps

G_2 to G_3, then $P_{1,3}$ should map G_1 to G_3. Mathematically, $P_{1,3} = P_{1,2}P_{2,3}$. This property is often called *alignment consistency*. Papers that enforce this constraint, or variants of it, include (Huang & Guibas, 2013; Pachauri et al., 2013; Chen et al., 2014; Yan et al., 2015c;b; Zhou et al., 2015; Hu et al., 2016). Most of these papers focus on computer vision, i.e., the task of producing alignments between shapes, or reference points among different figures, although most of the ideas can be easily adapted to aligning graphs. The proposed alignment algorithms are not all equally easy to solve, some involve convex problems, others involve non-convex or integer-valued problems. None of these works care about the alignment scores satisfying metric-like properties.

There are several papers that propose procedures for generating multi-distances from pairwise distances, and prove that these multi-distances satisfy intuitive generalizations of the metric properties to $n \geqslant 3$ elements. These allow us to use the existing works on two-graph comparisons to produce distances between multiple graphs. The simplest method is to define $d(G_1, \ldots, G_n) = \sum_{i,j \in [n]} d(G_i, G_j)$. The problem with this approach is that if $d(G_i, G_j)$ also produces an alignment $P_{i,j}$, e.g., in (1), these alignments are unrelated, and hence do not satisfy consistency constrains that are usually desirable. An approach studied by (Kiss et al., 2016) is to define $d(G_1, \ldots, G_n) = \min_G \sum_{i \in [n]} d(G_i, G)$. If each $d(G_i, G)$ also produces an alignment P_i, and if we define $P_{i,j} = P_i P_j^\top$, then $\{P_{i,j}\}$ is a set of alignments that satisfy the aforementioned consistency constraint. The problem with this approach is that it tends to lead to computationally harder problems, even after several relaxations are applied (cf. Fermat distance in Section 4). A few other works that study metrics and their generalizations are (Kiss et al., 2016; Martín et al., 2011; Akleman & Chen, 1999).

The work of (Bento & Ioannidis, 2018) defines a family of metrics for comparing two graphs. Several metrics in this family are tractable, or can be reduced to solving a convex optimization problem. However, (Bento & Ioannidis, 2018) does not consider comparing $n \geqslant 3$ graphs. We refer the reader to (Khamsi, 2015) that surveys generalized metric spaces, and (Deza & Deza, 2009) that provides an extensive review of many distance functions along with their applications in different fields, and, in particular, discusses the generalizations of the concept of metrics in different areas such as topology, probability, and algebra. The authors in (Deza & Deza, 2009) also discuss several distances for comparing two graphs, most of which are not tractable.

3. Notation and preliminaries

We focus on comparing graphs of equal size. A canonical way to deal with graphs with different sizes is to add dummy nodes to make them equal-sized. Many applied papers, e.g., (Zaslavskiy et al., 2009a;b; Narayanan et al., 2011;

G_i	ith graph	$P_{i,j}$	Alig. of G_i and G_j
A_i	Adj. mat. of G_i	$\mathcal{P}$	Set of alig. mats.
n	# of graphs	d	Dist. among n graphs
m	# of nodes	Ω	Set of adj. mats.
s	Alig. score	S	Set of sets of alig. mats.
$\mathbf{P}$	Mat. of $\{P_{i,j}\}$	$\|\|\cdot\|\|$	Mat. norm
$\|\cdot\|$	Vec. norm	$\mathbf{tr}$	Trace

Table 1. Summary of main notation used.

Zaslavskiy et al., 2010; Zhou & De la Torre, 2012; Gold et al.; Yan et al., 2015a; Solé-Ribalta & Serratosa, 2010; Yan et al., 2015b), follow this approach.

Comparing equal-sized graphs, without adding dummy nodes is still important. One application in computer vision is to establish a correspondence among the nodes of n graphs, each representing a geometrical relation among m special points in n images of the same object. The user (or detection algorithm), by design, finds the same number, m, of special points in each image. See, e.g., the numerical experiments in (Hu et al., 2016; Shen et al., 2015). Other papers that only consider equal-sized graphs include: (Lyzinski et al., 2016; Pachauri et al., 2013). We also point the reader to the remark on comparing graphs of unequal size at end of Section 7.

Let $[m] = \{1, \ldots, m\}$. A graph, $G = (V \equiv [m], E)$, with node set V and edge set E, is represented by a matrix, A, whose entries are indexed by the nodes in V. We denote the set that contains all such matrices by $\Omega \subseteq \mathbb{R}^{m \times m}$. E.g., Ω can be the set of adjacency matrices, or of the matrices containing hop-distances between all pairs of nodes.

Consider a set of n graphs, $\mathcal{G} = \{G_1, G_2, \ldots, G_n\}$. Given two graphs, $G_i = (V_i, E_i)$ and $G_j = (V_j, E_j)$, from the set $\mathcal{G}$, we denote a pairwise matching matrix between G_i and G_j by $P_{i,j}$. The rows and columns of $P_{i,j}$ are indexed by the nodes in V_i and V_j, respectively. Note that we can extract a relation between E_i and E_j, from a relation between V_i and V_j. We denote the set of all pairwise matching matrices by $\mathcal{P} = \{\{P_{i,j}\}_{i,j \in [n]} : P_{i,j} \in \mathbb{R}^{m \times m}\}$. For example, $\mathcal{P}$ might be all *permutation matrices* on m elements.

Let 1:n denote the sequence $1, \ldots, n$. For $A_1, \ldots, A_n \in \Omega$, we denote the ordered sequence $(A_1, \ldots, A_n)$ by $A_{1:n}$. The notation $A_{1:n,n+1}^i$ corresponds to the sequence $A_{1:n}$, in which the ith element, A_i, is removed and replaced by A_{n+1}. If σ is a permutation, i.e., a bijection from 1:n to 1:n such that $\sigma(i) = j$, then $A_{\sigma(1:n)}$ represents a sequence, whose ith element is A_j. In this paper, we use $\|\cdot\|$ and $\|\|\cdot\|\|$ to denote vector norms and matrix norms, respectively. We now provide the following definitions that will be used in the next sections of the paper. In what follows, equality of graphs means that they are isomorphic.

Definition 1. *A map $d : \Omega^2 \mapsto \mathbb{R}$, is a metric, if and only if, for all $A, B, C \in \Omega$:* (i) $d(A, B) \geqslant 0$; (ii) $d(A, B) =$ 0, *iff $A = B$;* (iii) $d(A, B) = d(B, A)$; *and* (iv) $d(A, C) \leqslant d(A, B) + d(B, C)$.

Definition 2. *A map $d : \Omega^2 \mapsto \mathbb{R}$ is a pseudometric, if and only if it satisfies properties* (i), (iii) *and* (iv) *in* Definition 1, *and if $d(A, A) = 0 \;\; \forall A \in \Omega$.*

Given a pseudometric d on two graphs, we define the equivalence relation $\sim_d$ in Ω as $A \sim_d B$ if and only if $d(A, B) = 0$. Using the fact that d is a pseudometric, it is immediate to verify that the binary relation $\sim_d$ satisfies *reflexivity*, *symmetry* and *transitivity*. We denote by $\Omega' = \Omega \backslash \sim_d$ the quotient space Ω modulo $\sim_d$, and, for any $A \in \Omega$, we let $[A] \subseteq \Omega$ denote the equivalence class of A. Given $A_{1:n}$, we let $[A]_{1:n}$ denote $([A_1], \ldots, [A_n])$, an ordered set of sets.

Definition 3. *A map $s : \Omega^2 \times \mathcal{P} \mapsto \mathbb{R}$ is called a P-score, if and only if, $\mathcal{P}$ is closed under inversion, and for any $P, P' \in \mathcal{P}$, and $A, B, C \in \Omega$, s satisfies the properties:*

$$s(A, B, P) \geqslant 0, \tag{2}$$
$$s(A, A, I) = 0, \tag{3}$$
$$s(A, B, P) = s(B, A, P^{-1}), \tag{4}$$
$$s(A, B, P) + s(B, C, P') \geqslant s(A, C, PP'). \tag{5}$$

For example, if $\mathcal{P}$ is the set of permutation matrices, and $\|\|\cdot\|\|$ is an element-wise matrix p-norm, then $s(A, B, P) = \|\|AP - BP\|\|$ is a P-score.

Definition 4 ((Bento & Ioannidis, 2018)). *The SB-distance function induced by the norm $\|\|\cdot\|\| : \mathbb{R}^{m \times m} \mapsto \mathbb{R}$, the matrix $D \in \mathbb{R}^{m \times m}$, and the set $\mathcal{P} \subseteq \mathbb{R}^{m \times m}$ is the map $d_{SB} : \Omega^2 \mapsto \mathbb{R}$, such that*

$$d_{SB}(A, B) = \min_{P \in \mathcal{P}} \|\|AP - PB\|\| + tr(P^\top D).$$

The authors in (Bento & Ioannidis, 2018), prove several conditions on Ω, $\mathcal{P}$, the norm $\|\|\cdot\|\|$, and the matrix D, such that d_{SB} is a metric, or a pseudometric. For example, if $\|\|\cdot\|\|$ is an arbitrary entry-wise or operator norm, $\mathcal{P}$ is the set of $n \times n$ doubly stochastic matrices, Ω is the set of symmetric matrices, and D is a *distance matrix*, then d_{SB} is a pseudometric.

4. n-metrics for multi-graph alignment

One can generalize the notion of a (pseudo) metric to $n \geqslant 3$ elements. To this aim, we consider the following definitions.

Definition 5. *A map $d : \Omega^n \mapsto \mathbb{R}$, is an n-metric, if and only if, for all $A_1, \ldots, A_n \in \Omega$,*

$$d(A_{1:n}) \geqslant 0, \tag{6}$$
$$d(A_{1:n}) = 0, \text{ iff } A_1 = \ldots = A_n, \tag{7}$$
$$d(A_{1:n}) = d(A_{\sigma(1:n)}), \tag{8}$$
$$d(A_{1:n}) \leqslant \textstyle\sum_{i=1}^{n} d(A_{1:n,n+1}^i). \tag{9}$$

According to Definition 5, a 2-metric is a metric as per Definition 1. In the sequel, we refer to properties (6), (7), (8), and (9), as non-negativity, identity of indiscernibles, symmetry, and generalized triangle equality (GTI), respectively.

Definition 6. *A map $d : \Omega^n \mapsto \mathbb{R}$, is a pseudo n-metric, if and only if it satisfies properties (6), (8) and (9), and for any $A \in \Omega$, d satisfies the property of* self-identity

$$d(A, \cdots, A) = 0. \tag{10}$$

Revisiting diameter estimation: n-metrics have several advantages over non-n-metrics. For $n = 2$, this is shown by (Bento & Ioannidis, 2018) and references therein: metrics allow several ML algorithms to finish faster, and improve the accuracy in tasks such as clustering graphs. Some of these advantages also extend to $n > 2$. For example, it is straightforward to see that, if we generalize the diameter estimation problem in Sec. 1 to $n = 3$, we can compute a $1/3$-approximation of $\max_{G_1, G_2, G_3 \in S} d(G_1, G_2, G_3)$ in expected time $O(n^2)$, compared to $O(n^3)$ for a non-n-metric. Considering the runtime of distance-based clustering using nth *order interaction* (Purkait et al., 2017), and just like for $n = 2$, n-metrics, $n > 2$, also improve runtime, because the GTI lets us avoid dealing with all n-distances.

We now define two functions that satisfy the properties of (pseudo) n-metrics.

4.1. A first attempt: Fermat distances

Definition 7. *Given a map $d : \Omega^2 \mapsto \mathbb{R}$, the Fermat distance function induced by d, is the map $d_F : \Omega^n \mapsto \mathbb{R}$, defined by*

$$d_F(A_{1:n}) = \min_{B \in \Omega} \sum_{i=1}^n d(A_i, B). \tag{11}$$

In the context of multiple graph alignment, d is an alignment score between two graphs, and d_F aims to find a graph, represented by B, that aligns well with all the graphs, represented by $A_{1:n}$. Thus, $d_F(A_{1:n})$ can be interpreted as an alignment score computed as the sum of alignment scores between each A_i and B. If we think of $A_{1:n}$ as a cluster of graphs, we can think of B as its center.

Theorem 1. *If d is a pseudometric, then the Fermat distance function induced by d is a pseudo n-metric.*

The proof of Theorem 1 is a direct adaptation of the one in (Kiss et al., 2016), and is included in Appendix B for completeness.

For example, the Fermat distance function induced by an SB-distance function with a distance matrix $D = 0$ is

$$d_F(A_{1:n}) = \min_{B \in \Omega, \{P_i\} \in \mathcal{P}^n} \sum_{i=1}^n \interleave A_i P_i - P_i B \interleave.$$

Despite its simplicity, the above optimization problem is not easy to solve in general, even when it is a continuous smooth optimization problem. For example, if $\mathcal{P}$ is the set of doubly stochastic matrices, B is the set of real matrices with entries in $[0, 1]$, and $\interleave \cdot \interleave$ is the Frobenius norm, the problem is non-convex due to the product PB that appears in the objective function. The potential complexity of computing d_F motivates the following alternative definition.

4.2. A better approach: $\mathcal{G}$-align distances

Definition 8. *Given a map $s : \Omega^2 \times \mathcal{P} \mapsto \mathbb{R}$, the $\mathcal{G}$-align distance function induced by s, is the map $d_{\mathcal{G}} : \Omega^n \mapsto \mathbb{R}$, defined by*

$$d_{\mathcal{G}}(A_{1:n}) = \min_{P \in S} \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}), \tag{12}$$

where

$$S = \{\{P_{i,j}\}_{i,j \in [n]} : P_{i,j} \in \mathcal{P}, \forall i, j \in [n], \ P_{i,k} P_{k,j} = P_{i,j},$$
$$\forall i, j, k \in [n], P_{i,i} = I, \forall i \in [n]\}. \tag{13}$$

Remark 1. *From the definition of S, it is implied that $I \in \mathcal{P}$ and that, if $P \in S$, then $P_{i,j} P_{j,i} = P_{i,i} = I \Leftrightarrow (P_{i,j}) = (P_{j,i})^{-1} \forall i, j \in [n]$, hence $\{P_{i,j}\}$ are invertible.*

Remark 2. *In (13), we refer to the property $P_{i,j} P_{j,k} = P_{i,k}, \forall i, j, k \in [n]$, as* the alignment consistency *of $P \in S$.*

The following Lemma, provides an alternative definition for the $\mathcal{G}$-align distance function.

Lemma 1. *If s is a P-score, then*

$$d_{\mathcal{G}}(A_{1:n}) = \min_{P \in S} \sum_{i,j \in [n], \ i < j} s(A_i, A_j, P_{i,j}). \tag{14}$$

Proof.

$$\sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}) = \sum_{i \in [n]} s(A_i, A_i, P_{i,i}) +$$
$$\sum_{i,j \in [n]: \ i < j} (s(A_i, A_j, P_{i,j}) + s(A_j, A_i, P_{j,i})). \tag{15}$$

If $P \in S$, then $P_{i,i} = I$ and $P_{j,i} = (P_{i,j})^{-1}$. Thus, since s is a P-score, $s(A_i, A_i, P_{i,i}) = s(A_i, A_i, I) = 0$, by property (3), and $s(A_j, A_i, P_{j,i}) = s(A_i, A_j, P_{i,j})$, by property (4). Therefore,

$$\sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}) = 2 \sum_{i,j \in [n], \ i < j} s(A_i, A_j, P_{i,j}),$$

and the proof follows. □

Note that, if $s(A, B, P) = \interleave AP - PB \interleave$, for some element-wise matrix norm, $n = 2$, and $\mathcal{P}$ is the set of permutations on m elements, then according to Lemma 1, $d_{\mathcal{G}}(A, B) = d_{SB}(A, B)$, for $D = 0$. In general, we can define a generalized SB-distance function induced by a matrix D, a set $\mathcal{P} \subseteq \mathbb{R}^{m \times m}$ and a map $s : \Omega^2 \times \mathcal{P} \mapsto \mathbb{R}$ as

$$d_{SB}(A, B) = \min_{P \in \mathcal{P}} s(A, B, P) + \operatorname{tr}(P^\top D), \tag{16}$$

and investigate the conditions on s, $\mathcal{P}$ and D, under which (16) represents a (pseudo) metric.

The following lemma leads to an equivalent definition for the $\mathcal{G}$-align distance function, which, among other things, reduces the optimization problem in (12), to finding n different matrices rather that $n^2 - n$ matrices that need to satisfy the alignment consistency.

Lemma 2. *If $S' = \{\{P_{i,j}\}_{i,j \in [n]} : P_{i,j} \in \mathcal{P}$ and $P_{i,j} = Q_i(Q_j)^{-1}, \forall i, j \in [n]$, for some matrices $\{Q_i\} \subseteq \mathcal{P}\}$, then $S' = S$.*

Proof. We first prove that $S \subseteq S'$. Let $P \in S$. Define $Q_i = P_{i,n} \in \mathcal{P}$ for all $i \in [n]$. If $i, j \in [n-1]$, then, by definition, $P_{i,j} = P_{i,n}P_{n,j} = P_{i,n}(P_{j,n})^{-1} = Q_i(Q_j)^{-1}$. This proves that $P \in S'$.

We now prove that $S' \subseteq S$. Let $P \in S'$. For any $i, j, k \in [n]$, we have $P_{i,k}P_{k,j} = Q_i(Q_k)^{-1}Q_k(Q_j)^{-1} = Q_i(Q_j)^{-1} = P_{i,j}$. It also follows that $P_{i,j} = Q_i(Q_j)^{-1} = (Q_j(Q_i)^{-1})^{-1} = (P_{j,i})^{-1}$, and $P_{i,i} = Q_i(Q_i)^{-1} = I$. Therefore, $P \in S$. $\square$

We complete this section with the following theorem, whose detailed proof is provided in Appendix C.

Theorem 2. *If s is a P-score, then the $\mathcal{G}$-align function induced by s is a pseudo n-metric.*

In Appendix A, we discuss the special case of $\mathcal{P}$ being the set of orthogonal matrices. In this case, we can simplify both eq. (11), and eq. (12), and compute them efficiently.

5. n-metrics on quotient spaces

The theorems in Section 4 are stated for pseudometrics. However, it is easy to obtain an n-metric from a pseudo n-metric for both d_F and $d_\mathcal{G}$ using quotient spaces. In these spaces, (7) holds almost trivially (with A_i replaced by its equivalent class $[A_i]$), and the important question is whether the equivalent classes of graphs are meaningful and useful. The proofs for the theorems in this section are Appendices G and H.

Theorem 3. *Let d be a pseudometric for two graphs, d_F be the Fermat distance function for n graphs induced by d, and $\Omega' = \Omega \backslash \sim_d$. Let $d'_F : \Omega'^n \mapsto \mathbb{R}$ be such that*

$$d'_F([A]_{1:n}) = d_F(A_{1:n}). \tag{17}$$

Then, d'_F is an n-metric.

Theorem 4. *Let s be a P-score. Let $d_{\mathcal{G}_2} : \Omega^2 \mapsto \mathbb{R}$ be the $\mathcal{G}$-align distance function for two graphs induced by s, and $d_\mathcal{G} : \Omega^n \mapsto \mathbb{R}$ be the $\mathcal{G}$-align distance function for n graphs induced by s. Let $\Omega' = \Omega \backslash \sim_{d_{\mathcal{G}_2}}$, and $d'_\mathcal{G} : \Omega'^n \mapsto \mathbb{R}$ be such that*

$$d'_\mathcal{G}([A]_{1:n}) = d_\mathcal{G}(A_{1:n}). \tag{18}$$

Then, $d'_\mathcal{G}$ is an n-metric.

6. The generalized triangle inequality for $d_\mathcal{G}$: an illustrative example

While it is straightforward to show that $d_\mathcal{G}$ satisfies the properties of non-negativity, symmetry and self-identity, the proof for the generalized triangle inequality is more involved. To give the reader a flavor of the proof, we now prove that the $\mathcal{G}$-align function satisfies the generalized triangle inequality when $n = 4$.

We consider a set of $n = 4$ graphs, $\mathcal{G} = \{G_1, G_2, G_3, G_4\}$, and a reference graph G_5, represented by matrices, $A_1, A_2, A_3, A_4 \in \Omega$ and $A_5 \in \Omega$, respectively. We will show that

$$d_\mathcal{G}(A_{1:4}) \leqslant \sum_{\ell=1}^{4} d_\mathcal{G}(A^\ell_{1:4,5}). \tag{19}$$

Let $P^* = \{P^*_{i,j}\} \in S$ be an optimal value for P in the optimization problem corresponding to the left-hand-side (l.h.s) of (19). We define $s^*_{i,j} = s(A_i, A_j, P^*_{i,j})$ for all $i, j \in [4]$. We also define $s^{\ell*}_{i,j} = s(A_i, A_j, P^{\ell*}_{i,j})$ for all $i, j \in [5]$, $\ell \in [4] \backslash \{i, j\}$, in which $P^{\ell*} = \{P^{\ell*}_{i,j}\} \in S$ is an optimal value for P in the optimization problem associated to $d_\mathcal{G}(A^\ell_{1:4,5})$ on the r.h.s of (19). Note that, according to (4), and the fact that $P^*_{i,j} = (P^*_{j,i})^{-1}$ (since $P^* \in S$), we have

$$s^*_{i,j} = s^*_{j,i}, \text{ and } s^{\ell*}_{i,j} = s^{\ell*}_{j,i}. \tag{20}$$

Moreover, according to (5), we have

$$s(A_i, A_j, P^{\ell*}_{i,k}P^{\ell'*}_{k,j}) \leqslant s^{\ell*}_{i,k} + s^{\ell'*}_{k,j}, \tag{21}$$

and, in the particular case when $\ell = \ell'$, we have

$$s^{\ell*}_{i,j} \leqslant s^{\ell*}_{i,k} + s^{\ell*}_{k,j}. \tag{22}$$

From the definition of $d_\mathcal{G}$ in Lemma 1, we have

$$\sum_{i,j \in [4], \ i<j} s^*_{i,j} \leqslant \sum_{i,j \in [4], \ i<j} s(A_i, A_j, \Gamma_{i,j}), \tag{23}$$

where $\Gamma_{i,j} = \Gamma_i \Gamma_j^{-1}$, and $\{\Gamma_i\}$ are any set of invertible matrices in $\mathcal{P}$. Note that from Lemma 2, $\{\Gamma_{i,j}\} \in S$. Consider the following choices for Γ_i's :

$$\Gamma_1 = P^{4*}_{1,5}; \ \Gamma_2 = P^{1*}_{2,5}; \ \Gamma_3 = P^{2*}_{3,5}; \ \Gamma_4 = P^{3*}_{4,5}. \tag{24}$$

We define $g^*_{i,j} = s(A_i, A_j, \Gamma_i\Gamma_j^{-1})$, in which Γ_i's are chosen according to (24). We can then rewrite (23) as

$$\sum_{i,j \in [4], \ i<j} s^*_{i,j} \leqslant \sum_{i,j \in [4], \ i<j} g^*_{i,j}. \tag{25}$$

We use Fig. 1 to bookkeep all the terms involved in proving (19). In particular, the first inequality in Fig. 1 provides a pictorial representation of (25). In this figure, each circle represents a graph in $\mathcal{G}$, and a line between G_i and G_j represents the P-score between A_i and A_j. In the diagram
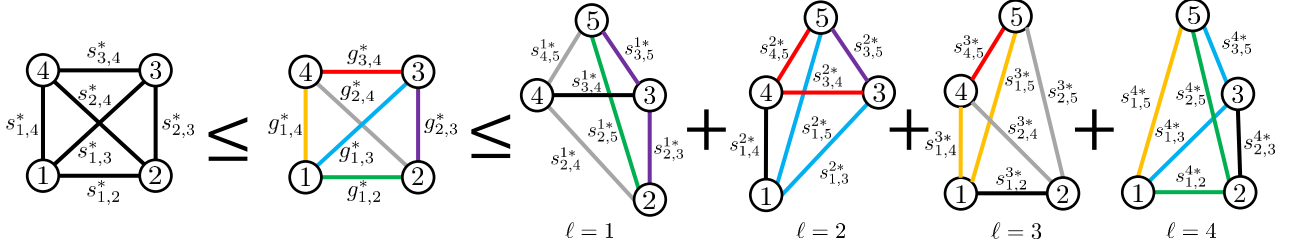
Figure 1. Generalized triangle equality of $d_{\mathcal{G}}$ for $n = 4$ graphs.

on the left, each P-score corresponds to the optimal pairwise matching between G_i and G_j associated to $d_{\mathcal{G}}(A_{1:4})$ in (19), whereas in the diagram in the middle, each P-score corresponds to the suboptimal matching between G_i and G_j, where the pairwise matching matrices are chosen according to (24). Using (21), followed by (20) we get

$$\sum_{i,j \in [4], \ i<j} g_{i,j}^* \leqslant (s_{1,5}^{4*} + s_{2,5}^{1*}) + (s_{1,5}^{4*} + s_{3,5}^{2*}) + (s_{1,5}^{4*} + s_{4,5}^{3*}) + (s_{2,5}^{1*} + s_{3,5}^{2*}) + (s_{2,5}^{1*} + s_{4,5}^{3*}) + (s_{3,5}^{2*} + s_{4,5}^{3*}).$$

The above inequality is also depicted in Fig. 1, where each diagram on the r.h.s of the second inequality represents $d_{\mathcal{G}}(A_{1:5}^\ell)$ in (19) for a different $\ell \in [4]$. Applying (22) to the r.h.s. of the above inequality, one can see that each one of the terms in parenthesis, distinguished with a different color, is upper bounded by the sum of the terms with the same color in the diagram in the r.h.s of the second inequality in Fig. 1. This completes the proof.

7. Moving towards tractability

The following lemmas are the building blocks towards a relaxation of $d_{\mathcal{G}}$ that is also easy to compute for choices of $\mathcal{P}$ other than orthonormal matrices. In this section, $\|\|\cdot\|\|_*$ denotes the nuclear norm.

Lemma 3. *Given* $\{P_{i,j}\}_{i,j \in [n]}$ *such that* $P_{i,j} \in \mathbb{R}^{m \times m}$ *for all* $i, j \in [n]$, *let* $\mathbf{P} \in \mathbb{R}^{nm \times nm}$ *have* n^2 *blocks, such that the* (i,j)*th block is* $P_{i,j}$. *Let*

$$S'' = \{\{P_{i,j}\}_{i,j \in [n]} : rank(\mathbf{P}) = m, P_{i,j} \in \mathcal{P}, \forall i, j \in [n], \quad P_{i,i} = I, \forall i \in [n]\}. \tag{26}$$

We have that $S'' = S$, *where* S *is as defined in* (13).

Proof. Let $\mathbf{P} \in \mathbb{R}^{nm \times nm}$, with blocks $\{P_{i,j}\}_{i,j \in [n]} \in S''$. Since $rank(\mathbf{P}) = m$, from the singular value decomposition of $\mathbf{P}$, we can write $\mathbf{P} = AB^\top$ where $A, B \in \mathbb{R}^{mn \times m}$. Let $A = [A_1; \ldots; A_n]$, where $A_i \in \mathbb{R}^{m \times m}$ and, similarly, let $B = [B_1; \ldots; B_n]$, where $B_i \in \mathbb{R}^{m \times m}$. It follows that $P_{i,j} = A_i B_j^\top$. Since $P_{i,i} = I$, we have $A_i B_i^\top = I$, which implies that $P_{i,j} = A_i A_j^{-1}$. By Lemma 2, this in turn implies that $\{P_{i,j}\}_{i,j \in [n]}$ satisfy the alignment consistency property. Therefore, $\{P_{i,j}\}_{i,j \in [n]} \in S$, and thus $S'' \subseteq S$.

Let $P = \{P_{i,j}\}_{i,j \in [n]} \in S$. By Lemma 2, $P_{i,j} = Q_i Q_j^{-1}$ for some invertible matrices $\{Q_i\}_{i \in [n]}$. Let $A, B \in \mathbb{R}^{mn \times m}$, with $A = [Q_1; \ldots; Q_n]$ and $B = [(Q_1^{-1})^\top, \ldots, (Q_n^{-1})^\top]$. Let $\mathbf{P}$ denote the $mn \times mn$ block matrix with $P_{i,j}$ as the (i,j)th block. We have $\mathbf{P} = AB^\top$. Thus $m \geqslant rank(\mathbf{P}) \geqslant rank(A) \geqslant rank(Q_1) = m$, which implies that $\{P_{i,j}\}_{i,j \in [n]} \in S''$, and therefore $S \subseteq S''$. $\square$

Lemma 4. [(Huang & Guibas, 2013), Proposition 1] *Let* $\mathcal{P}$ *be the set of* $m \times m$ *permutation matrices. Given* $\{P_{i,j}\}_{i,j \in [n]}$ *such that* $P_{i,j} \in \mathcal{P}$ *for all* $i, j \in [n]$, *let* $\mathbf{P} \in \mathbb{R}^{nm \times nm}$ *have* n^2 *blocks, such that the* (i,j)*th block is* $P_{i,j}$. *Let*

$$S''' = \{\{P_{i,j}\}_{i,j \in [n]} : P_{i,j} \in \mathcal{P}, \forall i, j \in [n], \mathbf{P} \succeq 0, \quad P_{i,i} = I, \forall i \in [n]\}. \tag{27}$$

We have that $S''' = S$, *where* S *is as defined in* (13).

Lemma 5. *For any* $\mathbf{P} \in \mathbb{R}^{nm \times nm}$ *with* $\mathbf{P}_{ii} = 1$ *for all* $i \in [nm]$, *we have* $\|\|\mathbf{P}\|\|_* \geqslant nm$.

Proof. Let $\mathbf{P}' = \frac{1}{2}(\mathbf{P} + \mathbf{P}^\top)$. We have $nm = tr(\mathbf{P}) = tr(\mathbf{P}') = \sum_{i \in [nm]} \lambda_i(\mathbf{P}') \leqslant \sum_{i \in [nm]} |\lambda_i(\mathbf{P}')| = \sum_{i \in [nm]} \sigma_i(\mathbf{P}') = \|\|\mathbf{P}'\|\|_* \leqslant \frac{1}{2}(\|\|\mathbf{P}\|\|_* + \|\|\mathbf{P}^\top\|\|_*) = \|\|\mathbf{P}\|\|_*$, where $\lambda_i(\cdot)$ and $\sigma_i(\cdot)$ denote the ith eigenvalue and the ith singular value of $(\cdot)$, respectively. $\square$

Lemma 6. *Let* $\mathcal{P}$ *be a subset of the orthogonal matrices. Let* $\{P_{i,j}\}_{i,j \in [n]} \in S$, *and* $\mathbf{P}$ *be the* $mn \times nm$ *block matrix with* $P_{i,j}$ *as the* (i,j)*th block. We have* $\|\|\mathbf{P}\|\|_* = mn$.

Proof. Since $\{P_{i,j}\}_{i,j \in [n]} \in S$ are alignment-consistent, we can write $P_{i,j} = P_{i,n} P_{j,n}^{-1}$ for all $i, j \in [n]$. Since $P_{j,n} \in \mathcal{P}$, it must be orthogonal. Hence, $P_{i,j} = P_{i,n} P_{j,n}^\top$, and we can write $\mathbf{P} = AA^\top$, where $A = [Q_1; \ldots; Q_n] \in \mathbb{R}^{mn \times m}$, and $Q_i = P_{i,n}$. Since $\mathbf{P}$ is positive semi-definite, its eigenvalues are equal to its singular values, which are non-negative, and thus $\|\|\mathbf{P}\|\|_* = tr(AA^\top) = tr(A^\top A) = \sum_{i \in [n]} tr(Q_i^\top Q_i) = \sum_{i \in [n]} tr(I) = mn$. $\square$

Inspired by Lemmas 3, 5, and 6, to obtain a continuous relaxation of $d_{\mathcal{G}}$, we relax the rank constraint $rank(\mathbf{P}) \leqslant m$ to $\|\|\mathbf{P}\|\|_* \leqslant mn$, use a function s that is a continuous function of P, and use a set $\mathcal{P}$ that is compact and contains

a non-empty ball around I. Alternatively, we can impose that $P_{j,i} = P_{i,j}^\top$, which was the case when $\mathcal{P}$ only contained orthonormal matrices, and relax the rank constraint to $\mathrm{tr}(\mathbf{P}) \leqslant mn$ and $\mathbf{P} \succeq 0$, i.e., $\mathbf{P}$ is a symmetric matrix with non-negative eigenvalues. Note that since we want $P_{i,i} = I$ for all $i \in [n]$, we can drop the trace constraint. The relaxation to $\mathbf{P} \succeq 0$ can also be justified by Lemma 4 and relaxing the constraint that $\mathcal{P}$ must be the set of permutations.

Definition 9. *Let $\mathcal{P} \subseteq \mathbb{R}^{m \times m}$ be compact and contain a non-empty ball around I. Let $P_{i,j} \in \mathcal{P}$ for all $i, j \in [n]$, and $\mathbf{P}$ be the $mn \times nm$ block matrix with $P_{i,j}$ as the (i,j)th block. Given a map $s : \Omega^2 \times \mathcal{P} \mapsto \mathbb{R}$, such that $s(\cdot, \cdot, P)$ is continuous for all $P \in \mathcal{P}$, the continuous $\mathcal{G}$-align distance function induced by s, is the map $d_{c\mathcal{G}} : \Omega^n \mapsto \mathbb{R}$, defined by*

$$d_{c\mathcal{G}}(A_{1:n}) = \min_{\substack{P_{i,j} \in \mathcal{P} \ \forall i,j \in [n], \\ P_{i,i} = I \ \forall i \in [n], \\ \||\mathbf{P}\||_* \leqslant mn}} \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}), \quad (28)$$

and the symmetric continuous $\mathcal{G}$-align distance function induced by s, is the map $d_{sc\mathcal{G}} : \Omega^n \mapsto \mathbb{R}$, defined by

$$d_{sc\mathcal{G}}(A_{1:n}) = \min_{\substack{P_{i,j} \in \mathcal{P} \ \forall i,j \in [n], \\ P_{i,i} = I \ \forall i \in [n], \\ \mathbf{P} \succeq 0}} \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}). \quad (29)$$

Remark 3. *Both optimization problems are continuous optimization problems, although they are potentially non-convex. However, for several natural choices of s, e.g., $s(A, B, P) = \||AP - PB\||$, and convex $\mathcal{P}$, both (28) and (29) can be computed via convex optimization.*

We finish this section, by showing that the above continuous distance functions, $d_{c\mathcal{G}}$ and $d_{sc\mathcal{G}}$, are pseudo n-metrics. In what follows, we let $\| \cdot \|$ and $\||\cdot\||_2$ denote the Euclidean norm and matrix operator norm, respectively. We will use the following definition.

Definition 10. *A map $s : \Omega^2 \times \mathcal{P} \mapsto \mathbb{R}$ is called a modified P-score, if and only if, $\mathcal{P}$ is closed under transposition and multiplication, for any $P \in \mathcal{P}$, $\||P\||_2 \leqslant 1$, and for any $P, P' \in \mathcal{P}$, and $A, B, C \in \Omega$, s satisfies the properties:*

$$s(A, B, P) \geqslant 0, \quad (30)$$

$$s(A, A, I) = 0, \quad (31)$$

$$s(A, B, P) = s(B, A, P^\top), \quad (32)$$

$$s(A, B, P) + s(B, C, P') \geqslant s(A, C, PP'). \quad (33)$$

For example, if $\mathcal{P}$ is the set of doubly stochastic matrices, Ω is a subset of the symmetric matrices, and $\||\cdot\||$ is an element-wise matrix p-norm, then $s(A, B, P) = \||AP - BP\||$ is a modified P-score.

We now provide the main result of this section.

Theorem 5. *If s is a modified P-score, then the symmetric continuous $\mathcal{G}$-align distance function induced by s is a pseudo n-metric.*

Remark 4. *A theorem with slightly different assumptions can be stated and proved about the $d_{c\mathcal{G}}$. Under appropriately defined equivalent classes, we can also obtain n-metrics from (28) and (29) (cf. Section 5).*

Graphs of different sizes: We note that in this section, unlike in Sec. 4, P_{ij} does not need to be invertible. Therefore, it is possible to extend the (symmetric) continuous $\mathcal{G}$-align distance function to consider graphs of unequal sizes. We could, e.g., allow P_{ij} to be rectangular of size m_i by m_j (resp. the node sizes of graph G_i and G_j), which would still result in $\mathbf{P}$ being square. If P_{ij}'s were previously doubly stochastic matrices, now, the row sums (or column sums, but not both) would be allowed to be $\leqslant 1$. This would model unmatched nodes, and avoid non-trivial solutions for Eqs. (28) and (29), i.e., $P_{i,j} = 0$ when $i \neq j$.

8. Numerical experiments

We do two experiments comparing our tool against two state-of-the-art non-n-metrics (from computer vision) and one simpler approach. Code for these comparison can be found in `http://github.com/bentoayr/n-metrics`. This repository includes code to compute some of our n-metrics, as well as code for the other methods, which is publicly available and that can be found through links in their respective papers, and which was copied into our repository for convenience.

The two competing algorithms are *matchSync* (Pachauri et al., 2013), and *mOpt* (Yan et al., 2015b). The simpler approach, *Pairwise*, defines $d(G_1, ..., G_n) = \sum_{i > j} d(G_i, G_j)$, where each $d(G_i, G_j)$, is computed using (Cho et al., 2010). All of these algorithms output a set of permutation matrices $\{P_{i,j}\}$, where $P_{i,j}$ tells how the nodes of graph i and j are matched. Both *matchSync*, and *mOpt* try to enforce the alignment consistency property on $\{P_{i,j}\}$, while *Pairwise* computes each $P_{i,j}$ independently. For our algorithm, we use (28), with $\mathcal{P}$ being the set of doubly stochastic matrices, and $s(A, B, P) = \||AP - PB\||_{\mathrm{Fro}}$. For comparison sake, after we compute $\{P_{i,j}\}$ using our algorithm, we sometimes project each $P_{i,j}$ onto the set of permutation matrices, which amounts to solving a *maximum weight matching problem*.

8.1. Multiple graph alignment experiment

We generate one Erdös-Rényi graph with edge probability 0.5, and 7 other graphs which are a small perturbation of the original graph (we flip edges with 0.05 probability), such that we know the joint optimal alignment of these $n = 8$ graphs, i.e. $P_{i,j}^* = I$. We then randomly permute the labels of these graphs such that the new joint optimal alignment

is known but non-trivial, i.e. $P_{i,j}^* \neq I$. We then use our n-metric, and the other non-n-metrics, to find an alignment between the graphs. Finally, we compare the alignments produced by the different methods to the optimal alignment. We repeat this 30 times, on random instances.

For each set of permutations $\{P_{i,j}\}$ given by the different algorithms we compute the *alignment quality* (AQ) and the *alignment consistency* (AC).

$$AQ = 1 - \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \|P_{i,j} - P_{i,j}^*\|/2}{mn(n-1)/2},$$

$$AC = 1 - \frac{\sum_{r=1}^{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \|P_{i,j} - P_{i,r}P_{r,j}\|/2}{mn^2(n-1)/2},$$

where $\|\cdot\|$ is the Frobenius norm. We obtain the following average accuracy (over 30 tests), and standard deviations. Note that, by design, *mOpt* and *matchSync* have AC = 1.

	Ours	*mOpt*	*matchSync*	*Pairwise*
AQ	0.94±0.01	0.91±0.02	0.90±0.02	0.88±0.02
AC	0.92±0.07	1.0±0.0	1.0±0.0	0.85±0.02

In Appendix J, we include an histogram with the distribution of values for these two quantities.

8.2. Graph clustering via hypergraph cut experiment

We build two clusters of graphs, each obtained by generating (i) a Erdös-Rényi graph with edge probability 0.7 as the cluster center, and (ii) 9 other graphs that are a small perturbation of (i). Graphs in (ii) are generated just like in Section 8.1. We then try to recover the true clusters using different n-distances.

For each n-distance, we build a hypergraph with 20 nodes (1 node per graph) and 100 hyperedges. Each hyperedge is built by randomly connecting 3 nodes (out of 20), for which the distance between their graphs is below a certain threshold. This threshold is later tuned to minimize each algorithm's clustering error (define below). Ideally, most hyperedges should not include graphs in different clusters. We then use the algorithm of (Vazquez, 2009b), whose code can be found in (Vazquez, 2009a) and which is included in our repositories for convenience, to find a minimum cut of the hypergraph that divides it into two equal-sized parts. These hyper-subgraphs are our predicted clusters. The *clustering error* is the fraction of misclassified graphs times two, such that the worst possible algorithm, a random guess, gives an avg. error of 1. We repeat this 50 times. For each algorithm, we use the same threshold in all 50 repetitions.

This experiment does not require an alignment between graphs but only a distance d. For algorithms that output an alignment $\{P_{i,j}\}$, this distance is computed as $\frac{1}{2}\sum_{i,j} \|A_i P_{i,j} - P_{i,j} A_j\|_{\text{Fro}}$. For our algorithm, we calculate this distance by first projecting $\{P_{i,j}\}$ onto the permutation matrices, which we denote as *Ours*, and we also calculate this distance directly as in (28), which we denote as *Ours**.

We report the average error in the following table. The standard deviation of the mean are all 0.04 except for Ours* which is 0.05.

Ours*	*Ours*	*mOpt*	*matchSync*	*Pairwise*
0.40	0.44	0.44	0.49	0.46

In Appendix K we include an histogram with the distribution of errors for the different algorithms.

9. Future work

It is possible to define the notion of a (pseudo) (C,n)-*metric*, as a map that satisfies the following more stringent generalization of the generalized triangle inequality: $d(A_{1:n}) \leqslant C \times \sum_{i=1}^{n} d(A_{1:n,n+1}^i)$.

The authors in (Kiss et al., 2016) prove that the d_F is a (pseudo) (C,n)-metric with $\frac{1}{n-1} \leqslant C \leqslant \frac{1}{\lfloor \frac{n}{2} \rfloor}$. Any (pseudo) (C,n)-metric with $C \leqslant 1$ is also a (pseudo) n-metric. It is an open problem to determine the largest constant C, for which $d_{\mathcal{G}}$, $d_{c\mathcal{G}}$ or $d_{sc\mathcal{G}}$ are a (pseudo) (C,n)-metric, and whether $C < 1$?

We also plan to test if the claim in (Vijayan et al., 2017), which states that in several scenarios calculating and using pairwise alignments is better than calculating and using joint alignments, holds for the n-metrics we introduced.

We plan to develop fast and scalable solvers to compute our n-metrics. The objective function of our n-metrics involves a large number of sums, in turn involving variables that are coupled by the alignment consistency constraint, or its relaxed equivalent. This makes the use of decomposition-coordination methods very attractive. In particular, we plan to test solvers based on the Alternating Direction Method of Multipliers (ADMM). Although not strictly a first-order method, it is very fast and, with proper tuning, it achieves a convergence rate that is as fast as the fastest possible first-order method (França & Bento, 2016; Nesterov, 2013). Furthermore, it has been used as an heuristic to solve many non-convex, even combinatorial, problems (Bento et al., 2013; 2015; Zoran et al., 2014; Mathy et al., 2015), and can be less affected by the topology of the communication network in a cluster than, e.g. Gradient Descent (França & Bento, 2017b;a). Finally, ADMM parallelizes well on share-memory multiprocessor systems, GPUs, and computer clusters (Boyd et al., 2011; Parikh & Boyd, 2014; Hao et al., 2016).

References

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

Akleman, E. and Chen, J. Generalized distance functions. In *Shape Modeling and Applications, 1999. Proceedings. Shape Modeling International'99. International Conference on*, pp. 72–79. IEEE, 1999.

Angiulli, F. and Pizzuti, C. Fast outlier detection in high dimensional spaces. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 15–27. Springer, 2002.

Babai, L. Graph isomorphism in quasipolynomial time. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pp. 684–697. ACM, 2016.

Bajaj, C. Proving geometric algorithm non-solvability: An application of factoring polynomials. *Journal of Symbolic Computation*, 2(1):99–102, 1986.

Bayati, M., Gleich, D. F., Saberi, A., and Wang, Y. Message-passing algorithms for sparse network alignment. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(1):3, 2013.

Bento, J. and Ioannidis, S. A family of tractable graph distances. In *Proceedings of the 2018 SIAM International Conference on Data Mining*, pp. 333–341. SIAM, 2018.

Bento, J., Derbinsky, N., Alonso-Mora, J., and Yedidia, J. S. A message-passing algorithm for multi-agent trajectory planning. In *Advances in neural information processing systems*, pp. 521–529, 2013.

Bento, J., Derbinsky, N., Mathy, C., and Yedidia, J. S. Proximal operators for multi-agent path planning. In *AAAI*, pp. 3657–3663, 2015.

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

Chen, Y., Guibas, L. J., and Huang, Q.-X. Near-optimal joint object matching via convex relaxation. *arXiv preprint arXiv:1402.1473*, 2014.

Cho, M., Lee, J., and Lee, K. M. Reweighted random walks for graph matching. In *European conference on Computer vision*, pp. 492–505. Springer, 2010.

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

Cockayne, E. J. and Melzak, Z. A. Euclidean constructibility in graph-minimization problems. *Mathematics Magazine*, 42(4):206–208, 1969.

Cohen, M. B., Lee, Y. T., Miller, G., Pachocki, J., and Sidford, A. Geometric median in nearly linear time. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pp. 9–21. ACM, 2016.

Conte, D., Foggia, P., Sansone, C., and Vento, M. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.

Deza, M. M. and Deza, E. Encyclopedia of distances. In *Encyclopedia of Distances*, pp. 1–583. Springer, 2009.

Douglas, J., Zimmerman, B., Kopylov, A., Xu, J., Sussman, D., and Lyzinski, V. Metrics for evaluating network alignment. 2018.

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

Feizi, S., Quon, G., Recamonde-Mendoza, M., Medard, M., Kellis, M., and Jadbabaie, A. Spectral alignment of graphs. *arXiv preprint arXiv:1602.04181*, 2016.

França, G. and Bento, J. An explicit rate bound for over-relaxed admm. In *Information Theory (ISIT), 2016 IEEE International Symposium on*, pp. 2104–2108. IEEE, 2016.

França, G. and Bento, J. How is distributed admm affected by network topology? *arXiv preprint arXiv:1710.00889*, 2017a.

França, G. and Bento, J. Markov chain lifting and distributed ADMM. *IEEE Signal Processing Letters*, 24(3):294–298, 2017b.

Gold, S. and Rangarajan, A. A graduated assignment algorithm for graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 18(4):377–388, 1996.

Gold, S., Rangarajan, A., et al. Softmax to softassign: Neural network algorithms for combinatorial optimization.

Hao, N., Oghbaee, A., Rostami, M., Derbinsky, N., and Bento, J. Testing fine-grained parallelism for the admm on a factor-graph. In *Parallel and Distributed Processing Symposium Workshops, 2016 IEEE International*, pp. 835–844. IEEE, 2016.

Hashemifar, S., Huang, Q., and Xu, J. Joint alignment of multiple protein–protein interaction networks via convex optimization. *Journal of Computational Biology*, 23(11): 903–911, 2016.

Heimann, M., Shen, H., and Koutra, D. Node representation learning for multiple networks: The case of graph alignment. *arXiv preprint arXiv:1802.06257*, 2018.

Hoffman, A. J., Wielandt, H. W., et al. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20(1):37–39, 1953.

Hu, N., Thibert, B., and Guibas, L. Distributable consistent multi-graph matching. *arXiv preprint arXiv:1611.07191*, 2016.

Hu, W., Qu, Y., and Cheng, G. Matching large ontologies: A divide-and-conquer approach. *Data & Knowledge Engineering*, 67(1):140–160, 2008.

Huang, Q. and Guibas, L. Consistent shape maps via semidefinite programming. In *Computer Graphics Forum*, volume 32, pp. 177–186. Wiley Online Library, 2013.

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

Kalaev, M., Smoot, M., Ideker, T., and Sharan, R. Networkblast: comparative analysis of protein networks. *Bioinformatics*, 24(4):594–596, 2008.

Kelley, B. P., Yuan, B., Lewitter, F., Sharan, R., Stockwell, B. R., and Ideker, T. Pathblast: a tool for alignment of protein interaction networks. *Nucleic acids research*, 32 (suppl_2):W83–W88, 2004.

Khamsi, M. A. Generalized metric spaces: a survey. *Journal of Fixed Point Theory and Applications*, 17(3):455–475, 2015.

Kiss, G., Marichal, J., and Teheux, B. A generalization of the concept of distance based on the simplex inequality. *arXiv preprint arXiv:1611.07826*, 2016.

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

Leordeanu, M. and Sminchisescu, C. Efficient hypergraph clustering. In *Artificial Intelligence and Statistics*, pp. 676–684, 2012.

Lyzinski, V., Fishkind, D., Fiori, M., Vogelstein, J., Priebe, C., and Sapiro, G. Graph matching: Relax at your own risk. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (1):1–1, 2016.

Ma, G., He, L., Cao, B., Zhang, J., Philip, S. Y., and Ragin, A. B. Multi-graph clustering based on interior-node topology with applications to brain networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 476–492. Springer, 2016.

Martín, J., Mayor, G., and Valero, O. Functionally expressible multidistances. In *EUSFLAT Conf.*, pp. 41–46, 2011.

Mathy, C. J., Gonda, F., Schmidt, D., Derbinsky, N., Alemi, E. A., Bento, J., Delle Fave, F. M., and Yedidia, J. S. Sparta: Fast global planning of collision-avoiding robot trajectories. 2015.

Narayanan, A., Shi, E., and Rubinstein, B. I. Link prediction by de-anonymization: How we won the kaggle social network challenge. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pp. 1825–1834. IEEE, 2011.

Nassar, H. and Gleich, D. F. Multimodal network alignment. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pp. 615–623. SIAM, 2017.

Nesterov, Y. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, Berlin/Heidelberg, Germany, 2013.

Nguyen, A., Ben-Chen, M., Welnicka, K., Ye, Y., and Guibas, L. An optimization approach to improving collections of shape maps. In *Computer Graphics Forum*, volume 30, pp. 1481–1491. Wiley Online Library, 2011.

Pachauri, D., Kondor, R., and Singh, V. Solving the multiway matching problem by permutation synchronization. In *Advances in neural information processing systems*, pp. 1860–1868, 2013.

Papo, D., Buldú, J. M., Boccaletti, S., and Bullmore, E. T. Complex network theory and the brain, 2014.

Parikh, N. and Boyd, S. Block splitting for distributed optimization. *Mathematical Programming Computation*, 6(1):77–102, 2014.

Park, H. and Yoon, K. Encouraging second-order consistency for multiple graph matching. *Machine Vision and Applications*, 27(7):1021–1034, 2016.

Purkait, P., Chin, T., Sadri, A., and Suter, D. Clustering with hypergraphs: The case for large hyperedges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1697–1711, Sep. 2017. ISSN 0162-8828. doi: 10.1109/TPAMI.2016.2614980.

Shen, Y., Lin, W., Yan, J., Xu, M., Wu, J., and Wang, J. Person re-identification with correspondence structure learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3200–3208, 2015.

Singh, R., Xu, J., and Berger, B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences*, 105(35):12763–12768, 2008.

Solé-Ribalta, A. and Serratosa, F. Graduated assignment algorithm for finding the common labelling of a set of graphs. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pp. 180–190. Springer, 2010.

Solé-Ribalta, A. and Serratosa, F. Models and algorithms for computing the common labelling of a set of attributed graphs. *Computer Vision and Image Understanding*, 115 (7):929–945, 2011.

Vazquez, A. Hypergraph clustering. `http://www.sns.ias.edu/~vazquez/hgc.html`, 2009a. [Online; accessed 10-May-2019].

Vazquez, A. Finding hypergraph communities: a bayesian approach and variational solution. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):P07006, 2009b.

Vijayan, V., Krebs, E., Meng, L., and Milenkovic, T. Pairwise versus multiple network alignment. *arXiv preprint arXiv:1709.04564*, 2017.

Wang, X., Liu, S., Liu, J., Chen, J., Zhu, J., and Guo, B. Topicpanorama: A full picture of relevant topics. *IEEE transactions on visualization and computer graphics*, 22 (12):2508–2521, 2016.

Weskamp, N., Hullermeier, E., Kuhn, D., and Klebe, G. Multiple graph alignment for the structural analysis of protein active sites. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(2):310–320, 2007.

Williams, M. L., Wilson, R. C., and Hancock, E. R. Multiple graph matching with bayesian inference. *Pattern Recognition Letters*, 18(11-13):1275–1281, 1997.

Yan, J., Cho, M., Zha, H., Yang, X., and Chu, S. A general multi-graph matching approach via graduated consistency-regularized boosting. *arXiv preprint arXiv:1502.05840*, 2015a.

Yan, J., Wang, J., Zha, H., Yang, X., and Chu, S. Consistency-driven alternating optimization for multigraph matching: A unified approach. *IEEE Transactions on Image Processing*, 24(3):994–1009, 2015b.

Yan, J., Xu, H., Zha, H., Yang, X., Liu, H., and Chu, S. A matrix decomposition perspective to multiple graph matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 199–207, 2015c.

Zaslavskiy, M., Bach, F., and Vert, J. P. Global alignment of protein–protein interaction networks by graph matching methods. *Bioinformatics*, 25(12):i259–1267, 2009a.

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

Zaslavskiy, M., Bach, F., and Vert, J.-P. Many-to-many graph matching: a continuous relaxation approach. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 515–530. Springer, 2010.

Zhang, J. and S. Yu, P. Multiple anonymized social networks alignment. In *Data Mining (ICDM), 2015 IEEE International Conference on*, pp. 599–608. IEEE, 2015.

Zhou, F. and De la Torre, F. Factorized graph matching. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 127–134. IEEE, 2012.

Zhou, X., Zhu, M., and Daniilidis, K. Multi-image matching via fast alternating minimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4032–4040, 2015.

Zoran, D., Krishnan, D., Bento, J., and Freeman, B. Shape and illumination from shading using the generic viewpoint assumption. In *Advances in Neural Information Processing Systems*, pp. 226–234, 2014.

Supplementary material for "Tractable n-Metrics for Multiple Graphs"

A. Special case of orthogonal matrices

In this section, we discuss the special case, where the pairwise matching matrices are orthogonal. This will further illustrate why computing d_F is harder than computing d_G. We consider the following assumption.

Assumption 1. Ω *is the set of real symmetric matrices, namely,* $\Omega = \{A \in \mathbb{R}^{m \times m} : A = A^\top\}$. $\mathcal{P}$ *is the set of orthogonal matrices, namely,* $\mathcal{P} = \{P \in \mathbb{R}^{m \times m} : P^\top = P^{-1}\}$. $s(A, B, P) = \|\!|AP - PB\|\!| \ \forall A, B \in \Omega, P \in \mathcal{P}$, *where* $\|\!| \cdot \|\!|$ *is the Frobenius norm or the operator norm, which are orthogonal invariant, and* $d(A, B) = \min_{P \in \mathcal{P}} s(A, B, P)$.

We now provide the main results of this section in the following theorems, and provide the detailed proofs in Appendix D-F.

Theorem 6. *Under Assumption 1, d_F induced by d, and d_G induced by s, are pseudo n-metrics.*

Theorem 7. *Let $\Lambda_{A_i} \in \mathbb{R}^m$ be the vector of eigenvalues of A_i, ordered from largest to smallest. Then, under Assumption 1,*

$$d_F(A_{1:n}) = \min_{\Lambda_C \in \mathbb{R}^m} \sum_{i=1}^n \|\Lambda_{A_i} - \Lambda_C\|. \qquad (34)$$

Theorem 8. *Let $\Lambda_{A_i} \in \mathbb{R}^m$ be the vector of eigenvalues of A_i, ordered from largest to smallest. Then, under Assumption 1,*

$$d_G(A_{1:n}) = \frac{1}{2} \sum_{i,j \in [n]} \|\Lambda_{A_i} - \Lambda_{A_j}\|. \qquad (35)$$

Note that $d_F = d_G = 0$ if and only if $A_{1:n}$ share the same spectrum.

The function d_F is related to the geometric median of the spectra of $A_{1:n}$. In order to write (35) as an optimization problem similar to d_F in (34), it is tempting to define d_G using s^2 instead of s, and take a square root. Let us call the resulting function $\bar{d}_G$. A straightforward calculation allows us to write

$$(\bar{d}_G(A_{1:n}))^2 = \frac{1}{2} \sum_{i,j \in [n]} \|\Lambda_{A_i} - \Lambda_{A_j}\|^2$$

$$= n^2 \left(\frac{1}{n} \sum_{i \in [n]} \left\| \Lambda_{A_i} - \frac{1}{n} \sum_{j \in [n]} \Lambda_{A_j} \right\|^2 \right) \equiv n^2 \text{Var}(\Lambda_{A_{1:n}})$$

$$= n \min_{\Lambda_C \in \mathbb{R}^m} \frac{1}{2} \sum_{i \in [n]} \|\Lambda_{A_i} - \Lambda_C\|^2,$$

where we use $\text{Var}(\Lambda_{A_{1:n}})$ to denote the geometric sample variance of the vectors $\{\Lambda_{A_i}\}$. This leads to a definition very close to (34), and a connection between $\bar{d}_G$ and the geometric sample variance.

At this point it is important to note that sample variances can be computed exactly in $\mathcal{O}(n)$ steps involving only sums and products of numbers. Contrastingly, although there are fast approximation algorithms for the geometric median (Cohen et al., 2016), there are no procedures to compute it exactly in a finite number of simple algebraic operations (Bajaj, 1986; Cockayne & Melzak, 1969).

B. Proof of Theorem 1

In the following lemmas, we show that the Fermat distance function satisfies properties (6), (8), (9), and (10), and hence is a pseudo n-metric.

Lemma 7. d_F *is non-negative.*

Proof. If d is a pseudo metric, it is non-negative. Thus, (11) is the sum of non-negative functions, and hence also non-negative. $\qquad \square$

Lemma 8. d_F *satisfies the self-identity property.*

Proof. If $A_1 = A_2 = \ldots = A_n$, then $d_F(A_{1:n}) = \min_B n \times d(A_1, B)$, which is zero if we choose $B = A_1 \in \Omega$, and (10) follows. $\qquad \square$

Lemma 9. d_F *is symmetric.*

Proof. Property (8) simply follows from the commutative property of summation. $\qquad \square$

Lemma 10. d_F *satisfies the generalized triangle inequality.*

Proof. Note that the following proof is a direct adaptation of the one in (Kiss et al., 2016), and is included for the sake of completeness. We show that the Fermat distance satisfies (9), i.e.,

$$d_F(A_{1:n}) \leqslant \sum_{i=1}^n d_F(A_{1:n,n+1}^i). \qquad (36)$$

Consider $B_{1:n} \in \Omega$ such that,

$$d_F(A_{1:n,n+1}^i) = d(A_{n+1}, B_i) + \sum_{j \in [n] \setminus i} d(A_j, B_i). \qquad (37)$$

Equation (37) implies that

$$\sum_{i=1}^n d_F(A_{1:n,n+1}^i) \geqslant \sum_{i=1}^n \sum_{j \in [n] \setminus i} d(A_j, B_i) \geqslant d(A_1, B_n) + $$

$$d(A_2, B_n) + \sum_{i=2}^{n-1} (d(A_1, B_i) + d(A_{i+1}, B_i)). \qquad (38)$$

Using triangle inequality, we have $d(A_1, B_n) + d(A_2, B_n) \geqslant d(A_1, A_2)$, and, $d(A_1, B_i) + d(A_{i+1}, B_i) \geqslant d(A_1, A_{i+1})$. Thus, from (38),

$$\sum_{i=1}^{n} d_F(A_{1:n,n+1}^i) \geqslant \sum_{i=2}^{n} d(A_1, A_i) = \sum_{i=1}^{n} d(A_1, A_i) \geqslant d_F(A_{1:n}),$$

where we used $d(A_1, A_1) = 0$ in the equality. The last inequality follows from Definition 7, and completes the proof. $\qquad\square$

C. Proof of Theorem 2

In the following lemmas, we show that the $\mathcal{G}$-align distance function satisfies properties (6), (8), (9), and (10), and hence is a pseudo n-metric.

Lemma 11. $d_{\mathcal{G}}$ *is non-negative.*

Proof. Since s is a P-score, it satisfies (2), i.e., $s \geqslant 0$, which implies $d_{\mathcal{G}} \geqslant 0$, since it is a sum of P-scores. $\quad\square$

Lemma 12. $d_{\mathcal{G}}$ *satisfies the self-identity property.*

Proof. If $A_1 = A_2 = \ldots = A_n$, then, if we choose $P \in S$ such that $P_{i,j} = I$ for all $i, j \in [n]$, we have $s(A_i, A_j, P_{i,j}) = 0$ by (3), for all $i, j \in [n]$. Therefore,

$$0 \leqslant d_{\mathcal{G}}(A_{1:n}) \leqslant \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}) = 0.$$
$\qquad\square$

Lemma 13. $d_{\mathcal{G}}$ *is symmetric.*

Proof. The definition, (12), involves summing $s(A_i, A_j, P_{i,j})$ over all pairs $i, j \in [n]$, which clearly makes $d_{\mathcal{G}}$ invariant to permuting $\{A_i\}$. $\qquad\square$

Lemma 14. $d_{\mathcal{G}}$ *satisfies the generalized triangle inequality.*

Proof. We now show that $d_{\mathcal{G}}$ satisfies (9), i.e.,

$$d_{\mathcal{G}}(A_{1:n}) \leqslant \sum_{\ell=1}^{n} d_{\mathcal{G}}(A_{1:n,n+1}^i). \qquad (39)$$

Let $P^* = \{P_{i,j}^*\} \in S$ be an optimal value for P in the optimization problem corresponding to the l.h.s of (39). Henceforth, just like Section 6, we use $s_{i,j}^* = s(A_i, A_j, P_{i,j}^*)$ for all $i, j \in [n]$. Note that according to (3) and (4), we have $s_{i,i}^* = 0$, and $s_{i,j}^* = s_{j,i}^*$, respectively. From (14), we have,

$$d_{\mathcal{G}}(A_{1:n}) = \sum_{i,j \in [n],\ i<j} s(A_i, A_j, P_{i,j}^*) = \sum_{i,j \in [n],\ i<j} s_{i,j}^*. \qquad (40)$$

Let $P^{k*} = \{P_{i,j}^{k*}\} \in S$ be an optimal value for P in the optimization problem associated to $d_{\mathcal{G}}(A_{1:n,n+1}^i)$ on the

r.h.s of (39). Henceforth, just like Section 6, we use $s_{i,j}^{\ell*} = s(A_i, A_j, P_{i,j}^{\ell*})$ for all $i, j \in [n+1], \ell \in [n]\backslash\{i,j\}$. Note that $s_{i,i}^{\ell*} = 0$, and $s_{i,j}^{\ell*} = s_{j,i}^{\ell*}$. From (14), we can write,

$$\sum_{\ell=1}^{n} d_{\mathcal{G}}(A_{1:n,n+1}^i) = \sum_{\ell=1}^{n} \sum_{\substack{i,j \in [n+1],\ i<j \\ \ell \notin \{i,j\}}} s_{i,j}^{\ell*}. \qquad (41)$$

We will show that,

$$\sum_{i,j \in [n],\ i<j} s_{i,j}^* \leqslant \sum_{\ell=1}^{n} \sum_{\substack{i,j \in [n+1],\ i<j \\ \ell \notin \{i,j\}}} s_{i,j}^{\ell*}. \qquad (42)$$

From the definition of $d_{\mathcal{G}}$ in Lemma 1,

$$\sum_{i,j \in [n],\ i<j} s_{i,j}^* \leqslant \sum_{i,j \in [n],\ i<j} s(A_i, A_j, \Gamma_{i,j}), \qquad (43)$$

for any matrices $\{\Gamma_{i,j}\}_{i,j \in [n]}$ in S, where S satisfies Definition 8. Hence, from Lemma 2, we also know that

$$\sum_{i,j \in [n],\ i<j} s_{i,j}^* \leqslant \sum_{i,j \in [n],\ i<j} s(A_i, A_j, \Gamma_i \Gamma_j^{-1}), \qquad (44)$$

for any invertible matrices $\{\Gamma_i\}_{i \in [n]}$ in $\mathcal{P}$.

Consider the following choice for Γ_i :

$$\Gamma_i = P_{i,n+1}^{i-1*}, \qquad 2 \leqslant i \leqslant n, \qquad (45)$$
$$\Gamma_1 = P_{1,n+1}^{n*}. \qquad (46)$$

Remark 5. *To simplify notation, we will just use $\Gamma_i = P_{i,n+1}^{i-1*}$ for all $i \in [n]$. It is assumed that when we writing $P_{i,j}^{\ell*}$, the index in superscript satisfies $\ell = 0 \Leftrightarrow \ell = n$.*

Note that since $P^{i-1*} \in S$, then $\Gamma_i = P_{i,n+1}^{i-1*}$ is invertible and belongs to $\mathcal{P}$. Using (45) to replace Γ_i and Γ_j in (44), and the fact that $(P_{j,n+1}^{j-1*})^{-1} = P_{n+1,j}^{j-1*}$, along with property (5) of the P-score s, we have

$$\sum_{\substack{i,j \in [n] \\ i<j}} s(A_i, A_j, \Gamma_i \Gamma_j^{-1}) = \sum_{\substack{i,j \in [n] \\ i<j}} s(A_i, A_j, P_{i,n+1}^{i-1*} P_{n+1,j}^{j-1*})$$

$$\leqslant \sum_{\substack{i,j \in [n] \\ i<j}} s_{i,n+1}^{i-1*} + s_{n+1,j}^{j-1*}.$$

We now show that

$$\sum_{\substack{i,j \in [n] \\ i<j}} s_{i,n+1}^{i-1*} + s_{n+1,j}^{j-1*} \leqslant \sum_{\ell=1}^{n} \sum_{\substack{i,j \in [n+1],\ i<j \\ \ell \notin \{i,j\}}} s_{i,j}^{\ell*}, \qquad (47)$$

which will prove (42) and complete the proof of the generalized triangle inequality for $d_{\mathcal{G}}$.

To this end, let $I_1 = \{(i,j) \in [n]^2 : i < j,\ j-1 = i\}$, $I_2 = \{(i,j) \in [n]^2 : i = 1, j = n\}$, $I_3 = \{(i,j) \in [n]^2 : i < j,\ j-1 \neq i \text{ and } (i,j) \neq (1,n)\}$. We will make use of the following three inequalities, which follow directly from property (5) of the P-score s.

$$\sum_{(i,j)\in I_1} s_{i,n+1}^{i-1*} \leqslant \sum_{(i,j)\in I_1} s_{i,j}^{i-1*} + s_{j,n+1}^{i-1*}. \tag{48}$$

$$\sum_{(i,j)\in I_2} s_{n+1,j}^{j-1*} \leqslant \sum_{(i,j)\in I_2} s_{n+1,i}^{j-1*} + s_{i,j}^{j-1*}. \tag{49}$$

$$\sum_{(i,j)\in I_3} s_{i,n+1}^{i-1*} + s_{n+1,j}^{j-1*} \leqslant \sum_{(i,j)\in I_3} \left(s_{i,j}^{i-1*} + s_{j,n+1}^{i-1*} + \right.$$
$$\left. s_{n+1,i}^{j-1*} + s_{i,j}^{j-1*} \right). \tag{50}$$

Since I_1, I_2 and I_3 are pairwise disjoint, we have

$$\sum_{i,j\in[n]} (\cdot) = \sum_{(i,j)\in I_1} (\cdot) + \sum_{(i,j)\in I_2} (\cdot) + \sum_{(i,j)\in I_3} (\cdot). \tag{51}$$

Using (48)-(50), and (51) we have

$$\sum_{i,j\in[n],\ i<j} s_{i,n+1}^{i-1*} + s_{n+1,j}^{j-1*} \leqslant \sum_{(i,j)\in I_1} s_{i,j}^{i-1*} + s_{j,n+1}^{i-1*} + s_{n+1,j}^{j-1*} +$$
$$\sum_{(i,j)\in I_2} s_{i,n+1}^{i-1*} + s_{n+1,i}^{j-1*} + s_{i,j}^{j-1*} +$$
$$\sum_{(i,j)\in I_3} s_{i,j}^{i-1*} + s_{j,n+1}^{i-1*} + s_{n+1,i}^{j-1*} + s_{i,j}^{j-1*}. \tag{52}$$

To complete the proof, we show that the r.h.s of (52) is less than, or equal to

$$\sum_{\ell=1}^{n} \sum_{\substack{i,j\in[n+1],\ i<j \\ \ell\notin\{i,j\}}} s_{i,j}^{\ell*}. \tag{53}$$

To establish this, we show that each term on the r.h.s of (52) is: (i) not repeated; and (ii) is included in (53).

Definition 11. *We call two P-scores, $s_{a_1,b_1}^{c_1*}$ and $s_{a_2,b_2}^{c_2*}$, coincident, and denote it by $s_{a_1,b_1}^{c_1*} \sim s_{a_2,b_2}^{c_2*}$, if and only if $c_1 = c_2$, and $\{a_1,b_1\} = \{a_2,b_2\}$.*

Checking (i) amounts to verifying that there are no coincident terms on the r.h.s. of (52). Checking (ii) amounts to verifying that for each P-score $s_{a_1,b_1}^{c_1*}$ on the r.h.s. of (52), there exists a P-score $s_{a_2,b_2}^{c_2*}$ in (53) such that $s_{a_1,b_1}^{c_1*} \sim s_{a_2,b_2}^{c_2*}$.

Note that the r.h.s of (52) consists of three summations. To verify (i), we first compare the terms within each summation, and then compare the terms among different summations. Consider the first summation on the r.h.s of (52). We have $s_{i,j}^{i-1*} \not\sim s_{j,n+1}^{i-1*}$ because $i \in [n]$ and therefore $i \neq n+1$. We have $s_{i,j}^{i-1*} \not\sim s_{n+1,j}^{j-1*}$ because $i-1 \neq j-1$ in this case, since $i < j$. We can similarly infer that $s_{j,n+1}^{i-1*} \not\sim s_{n+1,j}^{j-1*}$.

Now consider the second summation on the r.h.s of (52). Taking the definition of I_2 and (46) into account, we can rewrite this summation as,

$$s_{1,n+1}^{n*} + s_{n+1,1}^{n-1*} + s_{1,n}^{n-1*}. \tag{54}$$

Since $n \neq n-1$, we have $s_{1,n+1}^{n*} \not\sim s_{n+1,1}^{n-1*}$, and $s_{1,n+1}^{n*} \not\sim s_{1,n}^{n-1*}$. Also, since $n \neq n+1$ we have $s_{n+1,1}^{n-1*} \not\sim s_{1,n}^{n-1*}$.

Finally, consider the third summation on the r.h.s of (52). Since $i < j$, by comparing the superscripts we immediately see that the first and second terms in the summation cannot be equal to either the third or the forth term. On the other hand, since $n + 1 \neq i \in [n]$ and $n + 1 \neq j \in [n]$, we have $s_{i,j}^{i-1*} \not\sim s_{j,n+1}^{i-1*}$, and $s_{n+1,i}^{j-1*} \not\sim s_{i,j}^{j-1*}$, respectively.

We proceed by showing that the summands are not coincident among three summations. We first make the following observations:

Observation 1: since in all summations $i, j \in [n]$, we have $i \neq n+1$, $j \neq n+1$, and therefore each term with $n + 1$ in the subscript is not coincident with any term with $\{i, j\}$ in the subscript, e.g., on the r.h.s of (52), the first terms in the first and second summations cannot be coincident.

Observation 2: since I_1, I_2 and I_3 are pairwise disjoint, any two terms from different summations with the same indices cannot be coincident, e.g., on the r.h.s of (52), the third term in the second summation cannot be coincident with the third term in third summation.

Considering the above observations, the number of pairs we need to compare reduces from $3 \times 7 + 3 \times 4 = 33$ (in (52)) pairs to only 13 pairs, whose distinction may not seem trivial. To be specific, Obs. 1, excludes 16 comparisons and Obs. 2 excludes 4 comparisons. We now rewrite the r.h.s of (52) as

$$\sum_{(i,j)\in I_1} s_{i,j}^{i-1*} + s_{j,n+1}^{i-1*} + s_{n+1,j}^{j-1*} +$$
$$s_{1,n+1}^{n*} + s_{n+1,1}^{n-1*} + s_{1,n}^{n-1*} +$$
$$\sum_{(i',j')\in I_3} s_{i',j'}^{i'-1*} + s_{j',n+1}^{i'-1*} + s_{n+1,i'}^{j'-1*} + s_{i',j'}^{j'-1*}. \tag{55}$$

In what follows, we discuss the non-trivial comparisons, and refer to the first, second and third summations in (55) as Σ_1, Σ_2, and Σ_3, respectively.

1. $s_{i,j}^{i-1*}$ in Σ_1 vs. $s_{1,n}^{n-1*}$ in Σ_2: for these two terms to be coincident we need $i = n$. We also need $\{n, j\} = \{1, n\}$, i.e., $j = 1$, which cannot be true, since in S_1 we have $i = j - 1$ according to I_1.

2. $s_{i,j}^{i-1*}$ in Σ_1 vs. $s_{i',j'}^{j'-1*}$ in Σ_3: since $(i, j) \in I_1 = \{(i,j) \in [n]^2 : i < j,\ j-1 = i\}$, we have $j = i + 1$. Thus, we can write the first term as $s_{i,i+1}^{i-1*}$. For the two terms to be coincident, their superscripts must be the

same so $i = j'$. On the other hand, for their subscripts to match, we need $j = i + 1 = i'$. The last two equalities imply that $i' = j' + 1$, which contradicts $(i', j') \in I_3$.

3. $s_{j,n+1}^{i-1*}$ in Σ_1 vs. $s_{1,n+1}^{n*}$ in Σ_2: for the superscripts to match, we need $i = 1$. We also need $j = 1$ for the equality of subscripts, which cannot be true since $i < j$.

4. $s_{j,n+1}^{i-1*}$ in Σ_1 vs. $s_{n+1,1}^{n-1*}$ in Σ_2: we need $i = n$ for the equality of superscripts, and $j = 1$ for the equality of subscripts, which cannot be true since $(i, j) \in I_1$, and therefore $i = j - 1$.

5. $s_{j,n+1}^{i-1*}$ in S_1 vs. $s_{n+1,i'}^{j'-1*}$ in S_3: we can write the first term as $s_{i+1,n+1}^{i-1}$. The equality of superscripts requires $i = j'$. The equality of subscripts requires $i' = i + 1$. Therefore, $i' = j' + 1$, which contradicts $(i', j') \in I_3$.

6. $s_{n+1,j}^{j-1*}$ in Σ_1 vs. $s_{1,n+1}^{n*}$ in Σ_2: the equality of superscripts requires $j = 1$, which is impossible since $j > i \in [n]$.

7. $s_{n+1,j}^{j-1*}$ in Σ_1 vs. $s_{n+1,1}^{n-1*}$ in Σ_2: for the equality of superscripts, we need $j = n$, in which case the subscripts will not match, since $\{n+1, n\} \neq \{n+1, 1\}$.

8. $s_{n+1,j}^{j-1*}$ in Σ_1 vs. $s_{j',n+1}^{i'-1*}$ in Σ_3: the equality of superscripts requires $i' = j$. The equality of the subscripts requires $j' = j$. The two equalities imply that $i' = j'$, which contradicts $i' < j'$.

9. $s_{n+1,j}^{j-1*}$ in Σ_1 vs. $s_{n+1,i'}^{j'-1*}$ in Σ_3: the equality of superscripts requires $j' = j$. The equality of the subscripts requires $i' = j$. The two equalities imply that $i' = j'$, which contradicts $i' < j'$.

10. $s_{1,n+1}^{n*}$ in Σ_2 vs. $s_{j',n+1}^{i'-1*}$ in Σ_3: for the equality of superscripts, we need $i' = 1$, and for the equality of subscripts, we need $j' = 1$. This contradicts $i' < j'$.

11. $s_{1,n+1}^{n*}$ in Σ_2 vs. $s_{n+1,i'}^{j'-1*}$ in the Σ_3: for equality of superscripts, we need $j' = 1$. For the equality of subscripts, we need $i' = 1$, which contradicts $i' \neq j'$.

12. $s_{n+1,1}^{n-1*}$ in Σ_2 vs. $s_{j',n+1}^{i'-1*}$ in Σ_3: for equality of superscripts, we need $i' = n$. For the equality of subscripts, we need $j' = 1$, which contradicts $i' < j'$.

13. $s_{1,n}^{n-1*}$ in Σ_2 vs. $s_{i',j'}^{i'-1*}$ in Σ_3: for the equality of superscripts, we need $i' = n$. This in turn requires $j' = 1$ for the equality of subscripts, which contradicts $i' < j'$.

What is left to show is (ii), i.e., that all terms in (55) are included in the summation in (53). To this aim, we will show that for each $s_{a,b}^c$ in (55), the indices $\{a, b, c\}$ satisfy

$$c \in [n], a, b \in [n+1] \backslash \{c\} \text{ and } a \neq b, \tag{56}$$

which is enough to prove that either $s_{a,b}^c$ or $s_{b,a}^c$ exist in (53).

We first note that the superscripts in (55) are in $[n]$, see Remark 5. Moreover, all the subscripts in (55) are either 1, $n+1$, or $i, j, i', j' \in [n]$. Thus, for any $s_{a,b}^c$ in (55), we have $a, b \in [n+1]$. Also note that, for any $s_{a,b}^c$ in (55), we have $a \neq b$, since the definition of I_1, I_2 and I_3 implies that $i < j$, $i' < j'$ and $i, j, i', j' < n+1$. Therefore, all we need to verify is that for any $s_{a,b}^c$ in (55), $a \neq c$ and $b \neq c$.

We start with the first summation, where the first term is $s_{i,j}^{i-1*}$. Clearly $i \neq i-1$ and $j \neq i-1$, from the definition of I_1. In the second term, $s_{j,n+1}^{i-1*}$, $j \neq i-1$, from the definition of I_1, and $i-1 \neq n+1$, because otherwise $i = n+2 \notin [n]$. In the third term, $s_{n+1,j}^{j-1*}$, we have $n+1 \neq j \in [n]$. Moreover, clearly $j \neq j-1$.

For any term $s_{a,b}^c$, in the second summation, we clearly see in (55) that $a \neq c$ and $b \neq c$.

We now consider the last summation in (55). In the first term, $s_{i',j'}^{i'-1*}$, clearly $i' \neq i'-1$. Moreover, $i'-1 < i' < j'$, since $(i', j') \in I_3$. In the second term, $s_{j',n+1}^{i'-1*}$, $j' \neq i'-1$, because since $i'-1 < i' < j'$. Moreover, $n+1 \neq i'-1$ because otherwise $i' = n+2 \notin [n]$. In the third term, $s_{n+1,i'}^{j'-1*}$, we have $n+1 \neq j'-1$ because otherwise $j' = n+2 \notin [n]$. On the other hand, $i' \neq j'-1$ since $(i', j') \in I_3$. In the fourth term, $s_{i',j'}^{j'-1*}$, we have $i' \neq j'-1$ since $(i', j') \in I_3$. Also, clearly $j' \neq j'-1$. $\qquad \square$

D. Proof of Theorem 6

To show that d_F is a pseudo n-metric, it suffices to show that d is a pseudometric, and evoke Theorem 1. To show that d is a pseudometric, we can evoke Theorem 3 in (Bento & Ioannidis, 2018).

To show that $d_{\mathcal{G}}$ is a pseudo n-metric, it suffices to show that s is a P-score, and evoke Theorem 2. Clearly, s is non-negative, and also $s(A, A, I) = 0$. Recall that, if P is orthogonal then, for any matrix M, we have $\|\|PM\|\| = \|\|MP\|\| = \|\|M\|\|$. Thus,

$$s(A, B, P) = \|\|AP - PB\|\| = \|\|P^{-1}(AP - PB)P^{-1}\|\|$$
$$= \|\|P^{-1}A - BP^{-1}\|\| = s(B, A, P^{-1}).$$

Finally, for any $P, P' \in \mathcal{P}$,

$$s(A, B, PP') = \|\|APP' - PP'B\|\| =$$
$$\|\|APP' - PCP' + PCP' - PP'B\|\| \leqslant$$
$$\|\|APP' - PCP'\|\| + \|\|PCP' - PP'B\|\| =$$
$$\|\|AP - PC\|\| + \|\|CP' - P'B\|\| =$$
$$s(A, C, P) + s(C, B, P').$$

E. Proof of Theorem 7

The proof uses the following lemmas by (Hoffman et al., 1953) and (Bento & Ioannidis, 2018).

Lemma 15. *For any matrix $M \in \mathbb{R}^{m \times m}$, and any orthogonal matrix $P \in \mathbb{R}^{m \times m}$, we have that $\||PM\|| = \||MP\|| = \||M\||$.*

Lemma 16. *Let $\||\cdot\||$ be the Frobenius norm. If A and B are Hermitian matrices with eigenvalues $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_m$ and $b_1 \leqslant b_2 \leqslant \ldots \leqslant b_m$ then*

$$\||A - B\|| \geqslant \sqrt{\sum_{i \in [m]} (a_i - b_i)^2}. \qquad (57)$$

Lemma 17. *Let $\||\cdot\||$ be the operator 2-norm. If A and B are Hermitian matrices with eigenvalues $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_m$ and $b_1 \leqslant b_2 \leqslant \ldots \leqslant b_m$ then*

$$\||A - B\|| \geqslant \max_{i \in [m]} |a_i - b_i|. \qquad (58)$$

We also need the following result.
Corollary 1. *If $a \in \mathbb{R}^m$, with $a_1 \leqslant a_2 \leqslant \cdots \leqslant a_m$, $b \in \mathbb{R}^m$, with $b_1 \leqslant b_2 \leqslant \cdots \leqslant b_m$, and $P \in \mathbb{R}^{m \times m}$ is a permutation matrix, then*

$$\|a - b\| \leqslant \|a - Pb\|. \qquad (59)$$

Proof. This follows directly from Lemma 16 and Lemma 17 by letting A and B be diagonal matrices with a and Pb in the diagonal, respectively. $\square$

We now proceed with the proof of Theorem 7. Let $A_i = U_i \text{diag}(\Lambda_{A_i}) U_i^{-1}$ and $C = V \text{diag}(\Lambda_C) V^{-1}$ be the eigendecomposition of the real and symmetric matrices A_i and C, respectively. The eigenvalues in the vectors Λ_{A_i} and Λ_C are ordered in increasing order, and U_i and V are orthonormal matrices. Using Lemma 15, we have that

$$\begin{aligned} \||A_i P_i - P_i C\|| &= \||(A_i - P_i C(P_i)^{-1}) P_i\|| \qquad (60) \\ &= \||A_i - P_i C(P_i)^{-1}\|| \\ &= \||U_i(\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C(P_i)^{-1} U_i) U_i^{-1}\|| \\ &= \||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C(P_i)^{-1} U_i\|| \geqslant \||\Lambda_{A_i} - \Lambda_C\||, \end{aligned}$$

where the last inequality follows from Lemma 16 or Lemma 17 (depending on the norm).

It follows from (60) that $d_F(A_{1:n}) \geqslant \min_{\Lambda_C \in \mathbb{R}^m : (\Lambda_C)_i \leqslant (\Lambda_C)_{i+1}} \sum_{i=1}^n \|\Lambda_{A_i} - \Lambda_C\| = \min_{\Lambda_C \in \mathbb{R}^m} \sum_{i=1}^n \|\Lambda_{A_i} - \Lambda_C\|$, where the last equality follows from Corollary 1.

Finally, notice that, by the equalities in (60), we have

$$\begin{aligned} d_F(A_{1:n}) &= \min_{P \in \mathcal{P}^n, C \in \Omega} \sum_{i=1}^n \||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C(P_i)^{-1} U_i\|| \\ &\leqslant \||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C)\||, \qquad (61) \end{aligned}$$

where the inequality follows from upper bounding $\min_{C \in \Omega}(\cdot)$ with the particular choice of $C = P_i^\top U_i \text{diag}(\Lambda_C) U_i^\top P_i \in \Omega$.

Since $\||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C)\||_{\text{Frobenius}} = \|\Lambda_{A_i} - \Lambda_C\|_{\text{Euclidian}}$ and $\||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C)\||_{\text{operator}} = \|\Lambda_{A_i} - \Lambda_C\|_{\infty\text{-norm}}$, the proof follows.

F. Proof of Theorem 8

Let $A_i = U_i \text{diag}(\Lambda_{A_i}) U_i^{-1}$ be the eigendecomposition of the real and symmetric matrix A_i. The eigenvalues in the vector Λ_{A_i} are ordered in increasing order, and U_i is an orthonormal matrix. Using Lemma 15, we get

$$\||A_i P_{i,j} - P_{i,j} A_j\|| = \||(A_i - P_{i,j} A_j (P_{i,j})^{-1}) P_{i,j}\|| \qquad (62)$$

$$\begin{aligned} &= \||A_i - P_{i,j} A_j (P_{i,j})^{-1}\|| \\ &= \||U_i(\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i) U_i^{-1}\|| \\ &= \||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i\|| \geqslant \||\Lambda_{A_i} - \Lambda_{A_j}\||, \end{aligned}$$

where the last inequality follows from Lemma 16 or Lemma 17 (depending on the norm).

From (62) we have $d_{\mathcal{G}}(A_{1:n}) \geqslant \frac{1}{2} \sum_{i,j \in [n]} \|\Lambda_{A_i} - \Lambda_{A_j}\|$.

At the same time, $d_{\mathcal{G}}(A_{1:n}) =$

$$\begin{aligned} \min_{P \in S} \frac{1}{2} \sum_{i,j \in [n]} \||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i\|| \\ \leqslant \||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_{A_j})\||, \qquad (63) \end{aligned}$$

where the inequality follows from upper bounding $\min_{P \in S}(\cdot)$ by choosing $P = \{P_{i,j}\}_{i,j \in [n]}$ such that $P_{i,j} = U_i U_j^{-1}$, which by Lemma 2 implies that $P \in S$.

Since $\||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_{A_j})\||_{\text{Frobenius}} = \|\Lambda_{A_i} - \Lambda_{A_j}\|_{\text{Euclidian}}$ and $\||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_{A_j})\||_{\text{operator}} = \|\Lambda_{A_i} - \Lambda_{A_j}\|_{\infty\text{-norm}}$, the proof follows.

G. Proof of Theorem 3

We first show that (17) is well defined. Let $A_i' \in [A_i]$. Since d satisfies the triangle inequality we have

$$\begin{aligned} d_F'([A']_{1:n}) &= d_F(A_{1:n}') = \min_{B \in \Omega} \sum_{i \in [n]} d(A_i', B) \\ &\leqslant \min_{B \in \Omega} \sum_{i \in [n]} d(A_i', A_i) + d(A_i, B) = \min_{B \in \Omega} \sum_{i \in [n]} d(A_i, B) \\ &= d_F(A_{1:n}) = d_F'([A]_{1:n}), \end{aligned}$$

where in the last equality we used $d(A_i', A_i) = 0$, since $A_i' \in [A_i]$. Similarly, we can show that $d_F'([A]_{1:n}) \leqslant d_F'([A']_{1:n})$. It follows that $d_F'([A]_{1:n}) = d_F'([A']_{1:n})$, and hence (17) is well defined.

We now prove that d'_F satisfies (7). Recall that, by Theorem 1, d_F is a pseudo n-metric. If $[A_1] = \cdots = [A_n]$, then

$$d'_F([A]_{1:n}) = d'_F([A_1], \ldots, [A_1]) = d_F(A_1, \ldots, A_1) = 0,$$

since, d_F is a pseudometric, and hence satisfies the property of self-identity (10).

On the other hand, if $d'_F([A]_{1:n}) = d_F(A_{1:n}) = 0$, then there exists $B \in \Omega$, such that $d(A_i, B) = 0$ for all $i \in [n]$. Since d is non-negative and symmetric, and also satisfies the triangle inequality, it follows that

$$\begin{aligned} 0 \leqslant d(A_i, A_j) &\leqslant d(A_i, B) + d(B, A_j) \\ &= d(A_i, B) + d(A_j, B) = 0. \end{aligned}$$

Hence, $[A_i] = [A_j]$ for all $i, j \in [n]$.

H. Proof of Theorem 4

In the proof, we let S_2 denote the set S in definition (13) for the distance d on two graphs and we let S_n denote the set S in definition (13) for the distance $d_{\mathcal{G}}$ on n graphs.

We first verify that (18) is well defined. Let $A'_i \in [A_i]$. Let $\{I, P_i^*, (P_i^*)^{-1}\} \in S_2$ be such that

$$\begin{aligned} d_{\mathcal{G}_2}(A_i, A'_i) \equiv \frac{1}{2}(s(A_i, A_i, I) &+ s(A'_i, A'_i, I) + \\ s(A'_i, A_i, P_i^*) &+ s(A_i, A'_i, (P_i^*)^{-1})) = 0. \end{aligned}$$

Since s is a P-score, $s(A'_i, A_i, P_i^*) = 0$. For any $\tilde{P} = \{\tilde{P}_{i,j}\}_{i,j \in [n]} \in S$ we have $\{P_i^* \tilde{P}_{i,j}(P_j^*)^{-1}\}_{i,j \in [n]} \in S$. Thus,

$$\begin{aligned} d'_{\mathcal{G}}([A']_{1:n}) = d_{\mathcal{G}}(A'_{1:n}) &= \min_{P \in S} \frac{1}{2} \sum_{i,j \in [n]} s(A'_i, A'_j, P_{i,j}) \\ &\leqslant \frac{1}{2} \sum_{i,j \in [n]} s(A'_i, A'_j, P_i^* \tilde{P}_{i,j}(P_j^*)^{-1}). \end{aligned}$$

By property (5) and the fact that $s(A'_i, A_i, P_i^*) = s(A_i, A'_i, (P_i^*)^{-1}) = 0$ for all $i \in [n]$, we can write

$$\begin{aligned} \frac{1}{2} \sum_{i,j \in [n]} s(A'_i, A'_j, P_i^* \tilde{P}_{i,j}(P_j^*)^{-1}) &\leqslant \frac{1}{2} \sum_{i,j \in [n]} \Big(s(A'_i, A_i, P_i^*) \\ + s(A_i, A_j, \tilde{P}_{i,j}) + s(A_j, A'_j, (P_j^*)^{-1}) \Big) &= s(A_i, A_j, \tilde{P}_{i,j}). \end{aligned}$$

Taking the minimum of the r.h.s. of the above expression over $\tilde{P}$ we get $d'_{\mathcal{G}}([A']_{1:n}) \leqslant d_{\mathcal{G}}(A_{1:n}) = d'_{\mathcal{G}}([A]_{1:n})$. Similarly, we can prove $d'_{\mathcal{G}}([A]_{1:n}) \leqslant d'_{\mathcal{G}}([A']_{1:n})$. It follows that $d'_{\mathcal{G}}([A]_{1:n}) = d'_{\mathcal{G}}([A']_{1:n})$, and hence (18) is well defined.

Now we show that $d'_{\mathcal{G}}$ satisfies (7). Recall that, by Thm. 2, $d_{\mathcal{G}}$ is a pseudo n-metric. If $[A_1] = \cdots = [A_n]$, then

$$d'_{\mathcal{G}}([A]_{1:n}) = d'_{\mathcal{G}}([A_1], \ldots, [A_1]) = d_{\mathcal{G}}(A_1, \ldots, A_1) = 0,$$

since, $d_{\mathcal{G}}$ is a pseudometric, and hence satisfies the property of self-identity (10).

On the other hand, if $d'_{\mathcal{G}}([A]_{1:n}) = d_{\mathcal{G}}(A_{1:n}) = 0$, then, for any $i, j \in [n]$, we have that $s(A_i, A_j, P_{i,j}) = 0$ for some $P_{i,j}$, and hence $d(A_i, A_j) = 0$. This implies that $[A_i] = [A_j]$ for all $i, j \in [n]$.

I. Proof of Theorem 5

The following lemma will be used later.

Lemma 18. *Let* $\Gamma_i \in \mathbb{R}^{m \times m}$, $\|\|\Gamma_i\|\|_2 \leqslant 1$ *for all* $i \in [n]$. *Let* $\mathbf{P} \in \mathbb{R}^{nm \times nm}$ *have* n^2 *blocks such that the* (i,j)*th block is* $\Gamma_i \Gamma_j^\top$ *if* $i \neq j$, *and* I *otherwise. We have that* $\mathbf{P} \succeq 0$, *and that* $\|\|\mathbf{P}\|\|_* \leqslant mn$.

Proof. Let us first prove that $\mathbf{P} \succeq 0$. Let $\mathbf{v} \in \mathbb{R}^{nm}$ have n blocks, the ith block being $v_i \in \mathbb{R}^m$. Since $\|\Gamma_i \Gamma_i^\top\|_2 \leqslant \|\Gamma_i\|_2 \|\Gamma_i^\top\|_2 \leqslant 1$, we have that $\|\Gamma_i^\top v_i\|_2^2 = \|v_i^\top \Gamma_i \Gamma_i^\top v_i\|_2 \leqslant \|v_i\|_2^2$ for all $i \in [n]$. Therefore, we have $\mathbf{v}^\top \mathbf{P} \mathbf{v} = \|\sum_{i \in [n]} \Gamma_i^\top v_i\|_2^2 + \sum_{i \in [n]} \|v_i\|_2^2 - \sum_{i \in [n]} \|\Gamma_i^\top v_i\|_2^2 \geqslant 0$, for any $\mathbf{v}$, which implies that $\mathbf{P} \succeq 0$. We now prove that $\|\|\mathbf{P}\|\|_* \leqslant mn$. Let σ_r and λ_r be the rth singular value and rth eigenvalue of $\mathbf{P}$ respectively. Since $\mathbf{P}$ is real-symmetric and positive semi-definite, we have that $\|\|\mathbf{P}\|\|_* = \sum_r \sigma_r = \sum_r |\lambda_r| = \sum_r \lambda_r = \text{tr}(\mathbf{P}) = mn$. $\square$

Proof of Theorem 5.

(Non-negativity): Since s is a modified P-score, it satisfies (30), i.e., $s \geqslant 0$, which implies $d_{sc\mathcal{G}} \geqslant 0$, since the objective function on the r.h.s of (29) is a sum of modified P-scores.

(Self-identity): If $A_1 = A_2 = \ldots = A_n$, then, if we choose $P_{i,j} = I$ for all $i, j \in [n]$, we have $s(A_i, A_j, P_{i,j}) = 0$ by (31), for all $i, j \in [n]$. Note that from the definition of $d_{sc\mathcal{G}}$, we are assuming that $I \in \mathcal{P}$. Furthermore, $\mathbf{P}$ defined using these $P_{i,j}$'s satisfies $\|\|\mathbf{P}\|\|_* \leqslant mn$. Therefore, this choice of $P_{i,j}$'s satisfies the constraints in the minimization problem in the definition of $d_{sc\mathcal{G}}(A_{1:n})$. Therefore, $d_{sc\mathcal{G}}(A_{1:n})$ is upper-bounded by 0, which along with its non-negativity leads to $d_{sc\mathcal{G}}(A_{1:n}) = 0$.

(Symmetry): The optimization problem in (29), involves summing $s(A_i, A_j, P_{i,j})$ over all pairs $i, j \in [n]$. Thus, permuting the matrices $\{A_i\}$ is the same as solving (29) with $P_{i,j}$ replaced by $P_{\sigma(i), \sigma(j)}$ for some permutation σ. Thus, all that we need to show is that $\mathbf{P} \succeq 0$ if and only if $\mathbf{P}' \succeq 0$, where $\mathbf{P}'$ is just like $\mathbf{P}$ but with its blocks' indexes permuted. To see this, note that the eigenvalues of a matrix M do not change if M is then permuted under some permutation matrix T.

(Generalized triangle inequality): We will follow exactly the same argument as in the proof of the generalized triangle inequality for Theorem 2, which is provided in Appendix C.

The only modification is in equation (44), and in a couple of steps afterwards.

Equation (44) should be replaced with

$$\sum_{i \neq j} s(A_i, A_j, P_{i,j}^*) \leqslant \sum_{i \neq j} s(A_i, A_j, \Gamma_i \Gamma_j^\top), \qquad (64)$$

where $\{\Gamma_i\}_{i \in [n]}$ are matrices in $\mathcal{P}$. This inequality holds because $P_{i,j}$ defined by $P_{i,j} = \Gamma_i \Gamma_j^\top \, \forall i \neq j$, and $P_{i,i} = I \forall i$, satisfies the constraints in (28), and hence the r.h.s. of (64) upper bounds the optimal objective value for (28). Indeed, since $\Gamma_i \in \mathcal{P}$, and since, by assumption, $\mathcal{P}$ is closed under multiplication and transposition, it follows that $\Gamma_i \Gamma_j^\top \in \mathcal{P}$. Furthermore, if we define $\mathbf{P}$ to have as the (i,j)th block, $i \neq j$, $\Gamma_i \Gamma_j^\top$, and have as the (i,i)th block the identity I, then, by Lemma 18, we know that $\mathbf{P} \geq 0$.

Starting from (64), we use (33) and (32) from the modified P-score properties and obtain

$$\sum_{i \neq j} s(A_i, A_j, \Gamma_i \Gamma_j^\top) \leqslant \sum_{i \neq j} s(A_i, A_{n+1}, \Gamma_i) +$$
$$\sum_{i \neq j} s(A_{n+1}, A_j, \Gamma_j^\top) = s(A_i, A_{n+1}, \Gamma_i) \qquad (65)$$
$$+ \sum_{i \neq j} s(A_j, A_{n+1}, \Gamma_j). \qquad (66)$$

The rest of the proof follows by choosing Γ_i has in (45) and (46), and noting that the new definition of $s_{i,j}^*$ and $s_{i,j}^{\ell *}$ satisfies the same properties as in the proof of Theorem 2. In particular, we have that $s_{i,j}^* = s_{j,i}^*$ and $s_{i,j}^{\ell *} = s_{j,i}^{\ell *}$, because $\mathbf{P}$ in (29) is symmetric, and because we are assuming that (32) holds.

J. Distribution of AQ and AC for the alignment experiment

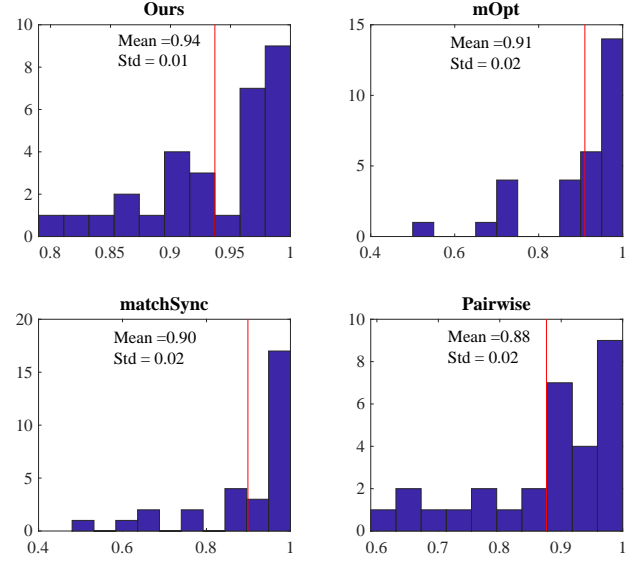## K. Distribution of clustering errors for the clustering experiment

Figure 2. Distribution of alignment quality (AQ) for the 30 tests in Section 8.1.
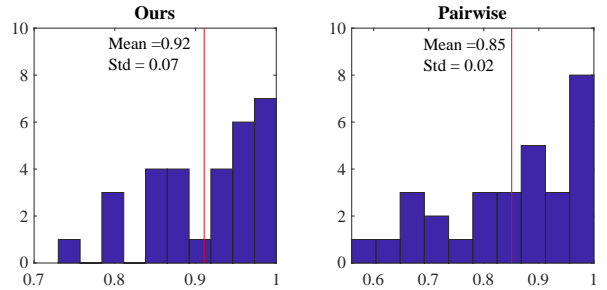
Figure 3. Distribution of alignment consistency (AC) for the 30 tests in Section 8.1. Note that, by construction, *mOpt* and *matchSync* always have AC = 1.
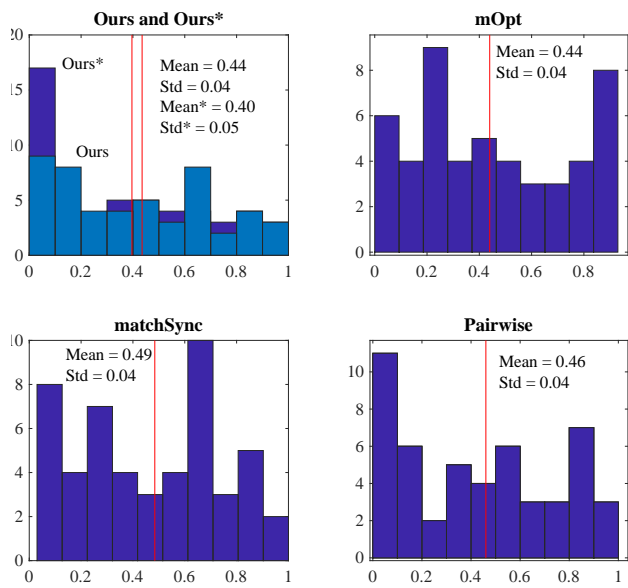
Figure 4. Distribution of errors for clustering for the 50 tests in Section 8.2. Recall that the error is the fraction of misclassified graphs times the number of clusters, which is 2 in our case. A random guess gives an average clustering error of 1.