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

\[
d(G_1, G_2) = \min_P \left\{ \| A_1 - PA_2 P^T \| = \| A_1 - PA_2 \| \right\}, \quad (1)
\]

where \( A_1 \) and \( A_2 \) are the adjacency matrices of \( G_1 \) and \( G_2 \), \( P \) and its transpose \( P^T \) 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 \( d(G_1, G_2) = \| A_1 P - PA_2 \| \), 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 \( \Delta = \max_{G_1, G_2 \in S} d(G_1, G_2) \) 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 \( d \geq \Delta/2 \). Indeed, if \( d(G^*, G_a) = \Delta \), then, by the triangle inequality, for any \( G \in S \), we cannot have both \( d(G^*, G) < \Delta/2 \) and \( d(G_a, G) < \Delta/2 \). Therefore, if we evaluate \( d \) on random pairs of graphs, we are guaranteed to find an 1/2-approximation of \( \Delta \) with only \( O(|S|) \) 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.,...
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 $n \geq 3$ 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 constraints that are usually desirable. An approach studied by (Kiss et al., 2016) is to define $d(G_1, \ldots, G_n) = \min_{\epsilon \in [n]} \sum_{i \in [n]} d(G_i, G)$. If each $d(G_i, G)$ also produces an alignment $P_i$, and if we define $P_{i,j} = P_iP_j^T$, 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; Martin 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 $n \geq 3$ 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.
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, if, for all \( A, B, C \), \( A \) maps to \( B \), then \( A \) maps to \( C \), and \( A \) maps to \( A \). The equivalence relation \( \sim \) in \( \Omega \) as \( A \sim B \) 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 \( \sim \) satisfies reflexivity, symmetry and transitivity. We denote by \( \Omega' = \Omega \setminus \sim \) the quotient space \( \Omega \) modulo \( \sim \), and, for any \( A \in \Omega \), we let \([A] \subseteq \Omega\) denote the equivalence class of \( A \). Given \( A_1:n \), we let \([A_1], \ldots, [A_n]\) denote an ordered set of sets.

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

\[
\begin{align*}
    s(A, B, P) & \geq 0, \\
    s(A, A, I) & = 0, \\
    s(A, B, P) & = s(B, A, P^{-1}), \\
    s(A, B, P) + s(B, C, P') & \geq s(A, C, PP').
\end{align*}
\]

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 \( \| \cdot \| : \mathbb{R}^{m \times m} \to \mathbb{R} \), the matrix \( D \in \mathbb{R}^{m \times m} \), and the set \( \mathcal{P} \subseteq \mathbb{R}^{m \times m} \) is the map \( d_{SB} : \Omega^2 \to \mathbb{R} \), such that

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

The authors in (Bento & Ioannidis, 2018), prove several conditions on \( \Omega, \mathcal{P} \), 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, \( \Omega \) 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 \( n \geq 3 \) elements. To this aim, we consider the following definitions.

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

\[
\begin{align*}
    d(A_1) & \geq 0, \\
    d(A_1) = 0, & \text{ iff } A_1 = \ldots = A_n, \\
    d(A_1) & = d(A_{\sigma(1:n)}), \\
    d(A_1) & \leq \sum_{i=1}^{n} d(A_{1:n,i+1}).
\end{align*}
\]
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 \( d : \Omega^n \to \mathbb{R} \), is a pseudo \( n \)-metric, if and only if it satisfies properties (6), (8) and (9), and for any \( A \in \Omega \), \( d(\cdot, \cdot, \cdot, A) = 0 \). (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 \( n \)th 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 \( d : \Omega^2 \to \mathbb{R} \), the Fermat distance function induced by \( d \), is the map \( d_F : \Omega^n \to \mathbb{R} \), defined by

\[
  d_F(A_1, n) = \min_{B \in \mathcal{P}} \sum_{i=1}^{n} d(A_i, B).
\]

(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

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

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 \( \|\cdot\| \) 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: \( G \)-align distances**

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

\[
  d_G(A_1, n) = \min_{P \in \mathcal{P}} \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}),
\]

(12)

where

\[
  S = \{\{P_{i,j}\}_{i,j \in [n]} : P_{i,j} \in \mathcal{P}, \forall i, j \in [n], P_{i,k} P_{k,j} = P_{i,j}, \forall i, j, k \in [n], P_{i,i} = 1, \forall i \in [n]\}.
\]

(13)

**Remark 1.** From the definition of \( S \), it is implied that \( I \in \mathcal{P} \) and that, if \( P \in S \), then \( P_{i,j} P_{j,k} = P_{i,k} = I \iff (P_{i,j})^{-1} \forall i, j \in [n], \) hence \( (P_{i,j})^{-1} \) 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 \( G \)-align distance function.

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

\[
  d_G(A_1, n) = \min_{P \in \mathcal{P}} \sum_{i,j \in [n], i < j} s(A_i, A_j, P_{i,j}).
\]

(14)

**Proof.**

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

(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 \( s(A_j, A_j, P_{j,i}) = s(A_i, A_j, P_{j,i}) \), by property (4). Therefore,

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

and the proof follows.

Note that, if \( s(A, B, P) = \|AP - PB\| \), for some element-wise matrix norm, \( n = 2 \), and \( \mathcal{P} \) is the set of permutations on \( m \) elements, then according to Lemma 1, \( d_G(A, B) = d_{SB}(A, B) \), 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 \( s : \Omega^2 \times \mathcal{P} \to \mathbb{R} \) as

\[
  d_{SB}(A, B) = \min_{P \in \mathcal{P}} s(A, B, P) + \text{tr}(P^T D),
\]

(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 \( G \)-align distance function, which, among other things, reduces the optimization problem in (12), to finding \( n \) different matrices rather than \( n^2 - n \) matrices that need to satisfy the alignment consistency.

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

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 \( G \)-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 \( d_G \) 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 \( \Omega' = \Omega \setminus \sim_d \). Let \( d_F' : \Omega^n \rightarrow \mathbb{R} \) be such that

\[
d_F([A]_{1:n}) = d_F(A_{1:n}).
\]

Then, \( d_F' \) is an \( n \)-metric.

**Theorem 4.** Let \( s \) be a \( P \)-score. Let \( d_G : \Omega^2 \rightarrow \mathbb{R} \) be the \( G \)-align distance function for two graphs induced by \( s \), and \( d_G : \Omega' \rightarrow \mathbb{R} \) be the \( G \)-align distance function for \( n \) graphs induced by \( s \). Let \( \Omega' = \Omega \setminus \sim_{d_G} \), and \( d_G' : \Omega'^n \rightarrow \mathbb{R} \) be such that

\[
d_G([A]_{1:n}) = d_G(A_{1:n}).
\]

Then, \( d_G' \) is an \( n \)-metric.

### 6. The generalized triangle inequality for \( d_G \): an illustrative example

While it is straightforward to show that \( d_G \) 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 \( G \)-align function satisfies the generalized triangle inequality when \( n = 4 \).

We consider a set of \( n = 4 \) graphs, \( G = \{ G_1, G_2, G_3, G_4 \} \), 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

\[
d_G(A_{1:4}) \leq \sum_{\ell=1}^{4} d_G(A_{1:4,5}).
\]

Let \( P^* \subseteq \{ P_{i,j}^* \} \in \mathcal{S} \) be an optimal \( P \) for 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 \( s_{\ell,i}^* = s(A_\ell, A_i, P_{\ell,i}^*) \) for all \( i, j \in [5], \ell \in [4] \setminus \{i, j\} \), in which \( P_{\ell,i}^* = \{ P_{i,j}^* \} \in \mathcal{S} \) is an optimal value for \( P \) in the optimization problem associated to \( d_G(A_{1:4,5}) \) on the r.h.s of (19). Note that, according to (4), and the fact that \( P_{i,j}^* = (P_{i,j}^*)^{-1} \) (since \( P^* \in \mathcal{S} \)), we have

\[
s_{i,j}^* = s_{j,i}^*, \quad \text{and} \quad s_{\ell,i}^* = s_{i,\ell}^*,
\]

Moreover, according to (5), we have

\[
s_{\ell,i}^* \leq s_{i,k}^* + s_{k,j}^*.
\]

From the definition of \( d_G \) in Lemma 1, we have

\[
\sum_{i,j \in [4], i < j} s_{i,j}^* \leq \sum_{i,j \in [4], i < j} s(A_i, A_j, \Gamma_{i,j}),
\]

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, \( \{ \Gamma_{i,j} \} \in \mathcal{E} \). Consider the following choices for \( \Gamma \)’s:

\[
\Gamma_1 = P_{1,4,5}^{-1}; \quad \Gamma_2 = P_{2,3,4,5}^{-1}; \quad \Gamma_3 = P_{2,4,5}^{-1}; \quad \Gamma_4 = P_{4,3,5}^{-1}.
\]

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

\[
\sum_{i,j \in [4], i < j} g_{i,j}^* \leq \sum_{i,j \in [4], i < j} g_{i,j}^*.
\]

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
The above inequality is also depicted in Fig. 1, where each property. Therefore, "r "A "r P of the relaxation of d_G(A_{1:5}) 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
\[ \sum_{i,j \in [4], i < j} g_{i,j}^* \leq (s_{1,5}^4 + s_{2,5}^4) + \frac{1}{2}(s_{1,5}^4 + s_{1,5}^2) + \frac{1}{2}(s_{2,5}^4 + s_{2,5}^2) + \frac{1}{2}(s_{1,5}^4 + s_{1,5}^2) + \frac{1}{2}(s_{2,5}^4 + s_{2,5}^2) + \frac{1}{2}(s_{1,5}^4 + s_{1,5}^2) + \frac{1}{2}(s_{2,5}^4 + s_{2,5}^2). \]

The above inequality is also depicted in Fig. 1, where each diagram on the r.h.s of the second inequality represents d_G(A_{1:5}) 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 \( d_G \) that is also easy to compute for choices of \( P \) 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 \( P \in \mathbb{R}^{nm \times nm} \) have \( n^2 \) blocks, such that the \( (i,j) \)th block is \( P_{i,j} \). Let
\[ S'' = \{\{P_{i,j}\}_{i,j \in [n]} : \text{rank}(P) = m, P_{i,j} \in P, \forall i, j \in [n], P_{i,i} = I, \forall i \in [n]\}. \]
We have that \( S'' = S \), where \( S \) is as defined in (13).

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

Lemma 4. ([Huang & Guibas, 2013], Proposition 1) Let \( P \) be the set of \( m \times n \) permutation matrices. Given \( \{P_{i,j}\}_{i,j \in [n]} \) such that \( P_{i,j} \in P \) for all \( i, j \in [n] \), let \( P \in \mathbb{R}^{nm \times nm} \) have \( n^2 \) blocks, such that the \( (i,j) \)th block is \( P_{i,j} \). Let
\[ S'' = \{\{P_{i,j}\}_{i,j \in [n]} : P_{i,j} \in P, \forall i, j \in [n], P_{i,i} = I, \forall i \in [n]\}. \]
We have that \( S'' = S \), where \( S \) is as defined in (13).

Lemma 5. For any \( P \in \mathbb{R}^{nm \times nm} \), we have \( \|P\|_* \geq nm \).

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

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

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

Inspired by Lemmas 3, 5, and 6, to obtain a continuous relaxation of \( d_G \), we relax the rank constraint \( \text{rank}(P) \leq m \) to \( \|P\|_* \leq mn \), use a function \( s \) that is a continuous function of \( P \), and use a set \( \mathbb{P} \) that is compact and contains
We finish this section by showing that the above continuous $P$-score, which was the case when $P$ only contained orthonormal matrices, and relax the rank constraint to $\text{tr}(P) \leq mn$ and $P \succeq 0$, i.e., $P$ 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 $P \succeq 0$ can also be justified by Lemma 4 and relaxing the constraint that $P$ must be the set of permutations.

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

$$d_{cG}(A_{1:n}) = \min_{P_{i,j} = I \ \forall \ i \in [n], P_{i,j} \in P_{i,j}} \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}), \quad (28)$$

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

$$d_{scG}(A_{1:n}) = \min_{P_{i,j} = I \ \forall \ i \in [n], P_{i,j} \in P_{i,j}} \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}), \quad (29)$$

**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 $P$, both (28) and (29) can be computed via convex optimization.

We finish this section, by showing that the above continuous distance functions, $d_{cG}$ and $d_{scG}$, 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 $s : \Omega^2 \times P \rightarrow \mathbb{R}$ is called a modified $P$-score, if and only if $P$ is closed under transposition and multiplication, for any $P \in P$, $\|P\|_2 \leq 1$, and for any $P, P' \in P$, and $A, B, C \in \Omega$, $s$ satisfies the properties:

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

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

$$s(A, B, P') = s(B, A, P^T), \quad (32)$$

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

For example, if $P$ is the set of doubly stochastic matrices, $\Omega$ 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 $G$-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 $d_{cG}$. 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, $P_{i,j}$ does not need to be invertible. Therefore, it is possible to extend the (symmetric) continuous $G$-align distance function to consider graphs of unequal sizes. We could, e.g., allow $P_{i,j}$ 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 $P$ being square. If $P_{i,j}$’s were previously doubly stochastic matrices, now, the row sums (or column sums, but not both) would be allowed to be $\leq 1$. 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 $d(G_1, \ldots, G_n) = \sum_{i \neq j} d(G_i, G_j)$, 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 $P$ being the set of doubly stochastic matrices, and $s(A, B, P) = \|AP - BP\|_{Fro}$ 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 Erdős-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).

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

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

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 $AC = 1$.

<table>
<thead>
<tr>
<th></th>
<th>Ours</th>
<th>$mOpt$</th>
<th>matchSync</th>
<th>Pairwise</th>
</tr>
</thead>
<tbody>
<tr>
<td>AQ</td>
<td>0.94±0.01</td>
<td>0.91±0.02</td>
<td>0.90±0.02</td>
<td>0.88±0.02</td>
</tr>
<tr>
<td>AC</td>
<td>0.92±0.07</td>
<td>1.0±0.0</td>
<td>1.0±0.0</td>
<td>0.85±0.02</td>
</tr>
</tbody>
</table>

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 Erdős-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

\[
\frac{1}{2} \sum_{i,j} \|A_i P_{i,j} - P_{i,j} A_j\|_{F0}.
\]

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.

<table>
<thead>
<tr>
<th></th>
<th>Ours</th>
<th>$mOpt$</th>
<th>matchSync</th>
<th>Pairwise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.40</td>
<td>0.44</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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:

\[d(A_{1:n}) \leq C \times \sum_{i=1}^{n-1} d(A_{1:n,i+1}).\]

The authors in (Kiss et al., 2016) prove that the $d_{F}$ is a (pseudo) $(C, n)$-metric with $C = 1$. Any (pseudo) $(C, n)$-metric with $C = 1$ is also a (pseudo) $n$-metric. It is an open problem to determine the largest constant $C$, for which $d_{Q}$. $d_{G}$ or $d_{C}$ 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 (Fraňa & 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 (Fraňa & 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


Tractable \( n \)-Metrics for Multiple Graphs


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.** $\Omega$ is the set of real symmetric matrices, namely, $\Omega = \{ A \in \mathbb{R}^{m \times m} : A = A^T \}$, $\mathcal{P}$ is the set of orthogonal matrices, namely, $\mathcal{P} = \{ P \in \mathbb{R}^{m \times m} : P^T = P \}$.

$s(A, B, P) = \| AP - PB \| \forall A, B \in \Omega, P \in \mathcal{P}$, 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,

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

**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,

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

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

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

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

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

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 $d_G$ and the geometric sample variance.

At this point it is important to note that sample variances can be computed exactly in $O(n)$ 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.

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

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

**Lemma 9.** $d_F$ is symmetric.

**Proof.** Property (8) simply follows from the commutative property of summation.

**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.,

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

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

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

Equation (37) implies that

$$\sum_{i=1}^n d_F(A_{1:n,n+1}) \geq \sum_{i=1}^n \sum_{j \in [n]\setminus i} d(A_j, B_i) \geq d(A_1, B_n) + d(A_2, B_n) + \sum_{i=2}^{n-1} (d(A_1, B_i) + d(A_{i+1}, B_i)). \quad (38)$$
Using triangle inequality, we have \( d(A_1, B_n) + d(A_2, B_n) \geq d(A_1, A_2) \), and, \( d(A_1, B_i) + d(A_{i+1}, B_i) \geq d(A_1, A_{i+1}) \). Thus, from (38),
\[
\sum_{i=1}^{n} d_F(A^i_{1:n,n+1}) \geq \sum_{i=2}^{n} d(A_1, A_i) = \sum_{i=1}^{n} d(A_1, A_i) \geq d_F(A_{1:n}),
\]
where we used \( d(A_1, A_1) = 0 \) in the equality. The last inequality follows from Definition 7, and completes the proof.

\[\square\]

\section*{C. Proof of Theorem 2}

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

\begin{lemma}
\( d_G \) is non-negative.
\end{lemma}

\begin{proof}
Since \( s \) is a \( P \)-score, it satisfies (2), i.e., \( s \geq 0 \), which implies \( d_G \geq 0 \), since it is a sum of \( P \)-scores.
\end{proof}

\begin{lemma}
\( d_G \) satisfies the self-identity property.
\end{lemma}

\begin{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,
\[
0 \leq d_G(A_{1:n}) \leq \frac{1}{2} \sum_{i,j \in [n]} s(A_i, A_j, P_{i,j}) = 0.
\]
\end{proof}

\begin{lemma}
\( d_G \) is symmetric.
\end{lemma}

\begin{proof}
The definition, (12), involves summing \( s(A_i, A_j, P_{i,j}) \) over all pairs \( i, j \in [n] \), which clearly makes \( d_G \) invariant to permuting \( \{A_i\} \).
\end{proof}

\begin{lemma}
\( d_G \) satisfies the generalized triangle inequality.
\end{lemma}

\begin{proof}
We now show that \( d_G \) satisfies (9), i.e.,
\[
d_G(A_{1:n}) \leq \sum_{i=1}^{n} d_G(A^i_{1:n,n+1}). \tag{39}
\]
Let \( P^* = \{P^*_i\} \in S \) 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 \( s^*_{i,j} = s(A_i, A_j, P^*_i) \) 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,
\[
d_G(A_{1:n}) = \sum_{i,j \in [n], i < j} s(A_i, A_j, P^*_i) = \sum_{i,j \in [n], i < j} s^*_{i,j}.
\]
Let \( P^{k*} = \{P^{k*}_{i,j}\} \in S \) be an optimal value for \( P \) in the optimization problem associated to \( d_G(A^k_{1:n,n+1}) \) on the r.h.s of (39). Henceforth, just like Section 6, we use \( s^*_{i,j} = s(A_i, A_j, P^{k*}_{i,j}) \) for all \( i, j \in [n+1], \ell \in [n]\{i, j\} \). Note that \( s^*_{i,i} = 0 \), and \( s^*_{i,j} = s^*_{j,i} \). From (14), we can write,
\[
\sum_{i,j \in [n], i < j} s^*_{i,j} = \sum_{\ell=1}^{n} \sum_{i,j \in [n+1], i < j} s^*_{i,j}. \tag{41}
\]
We will show that,
\[
\sum_{i,j \in [n], i < j} s^*_{i,j} \leq \sum_{\ell=1}^{n} \sum_{i,j \in [n+1], i < j} s^*_{i,j}. \tag{42}
\]
From the definition of \( d_G \) in Lemma 1,
\[
\sum_{i,j \in [n], i < j} s^*_{i,j} \leq \sum_{i,j \in [n], i < j} s(A_i, A_j, \Gamma_{i,j}), \tag{43}
\]
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,
\[
\sum_{i,j \in [n], i < j} s^*_{i,j} \leq \sum_{i,j \in [n], i < j} s(A_i, A_j, \Gamma_{i,j}^{-1}), \tag{44}
\]
for any invertible matrices \( \{\Gamma_{i,j}\}_{i,j \in [n]} \) in \( P \).

Consider the following choice for \( \Gamma_i : \)
\[
\Gamma_i = P^{i-1*}_{i,n+1}, \quad 2 \leq i \leq n, \tag{45}
\]
\[
\Gamma_1 = P^{n*}_{1,n+1}. \tag{46}
\]
\begin{remark}
To simplify notation, we will just use \( \Gamma_i = P^{i-1*}_{i,n+1} \) for all \( i \in [n] \). It is assumed that when we writing \( P^{i-1*}_{i,j} \), the index in superscript satisfies \( \ell = 0 \Leftrightarrow \ell = n \).
\end{remark}

Note that since \( P^{i-1*} \in S \), then \( \Gamma_i = P^{i-1*} \) is invertible and belongs to \( P \). Using (45) to replace \( \Gamma_i \) and \( \Gamma_j \) in (44), and the fact that \( (P^{j-1*}_{j,n+1})^{-1} = P^{j-1*}_{j,n+1} \), along with property (5) of the \( P \)-score \( s \), we have
\[
\sum_{i,j \in [n], i < j} s(A_i, A_j, \Gamma_{i,j}^{-1}) = \sum_{i,j \in [n], i < j} s(A_i, A_j, P^{i-1*}_{i,n+1} P^{j-1*}_{j,n+1})
\[
\leq \sum_{i,j \in [n], i < j} s^{i-1*}_{i,n+1} + s^{j-1*}_{n+1,j}.
\]
We now show that
\[
\sum_{i,j \in [n], i < j} s^{i-1*}_{i,n+1} + s^{j-1*}_{n+1,j} \leq \sum_{\ell=1}^{n} \sum_{i,j \in [n+1], i < j} s^*_{i,j}. \tag{47}
\]
which will prove (42) and complete the proof of the generalized triangle inequality for \( d_G \).
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 $s$.

\[
\sum_{(i,j) \in I_1} s_{i,n+1}^{i-1} \leq \sum_{(i,j) \in I_1} s_{i,j}^{i-1*} + s_{j,n+1-i}^{i-1*}.
\]  

(48)

\[
\sum_{(i,j) \in I_2} s_{n+1,i,j}^{j-1} \leq \sum_{(i,j) \in I_2} s_{n+1,j}^{j-1*} + s_{i,j}^{j-1*}.
\]  

(49)

\[
\sum_{(i,j) \in I_3} s_{i,n+1}^{i-1} + s_{j,n+1}^{j-1} \leq \sum_{(i,j) \in I_3} \left( s_{i,j}^{i-1*} + s_{j,n+1-i}^{j-1*} + s_{n+1,i}^{j-1} + s_{i,j}^{i-1} \right).
\]  

(50)

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

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

(51)

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

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

(52)

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

\[
\sum_{\ell=1}^{n} \sum_{i,j \in [n+1], i < j \ell (i,j)} s_{i,j}^{\ell*}.
\]  

(53)

To establish this, we show that each term on the r.h.s of (52) is: (i) not repeated; and (ii) 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 $s_{i,j}^{i-1*} \neq s_{n+1,i}^{i-1*}$ because $i \in [n]$ and therefore $i \neq n + 1$. We have $s_{i,j}^{j-1*} \neq s_{n+1,j}^{j-1*}$ because $i - 1 \neq j - 1$ in this case, since $i < j$. We can similarly infer that $s_{i,n+1}^{i-1*} \neq s_{n+1,i}^{i-1*}$.

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,

\[
s_{i,n+1}^{i-1*} + s_{n+1,i}^{i-1*} + s_{n+1,n}^{i-1*}.
\]  

(54)

Since $n \neq n - 1$, we have $s_{i,n+1}^{n-1*} \neq s_{n+1,i}^{n-1*}$, and $s_{n+1,n}^{n-1*} \neq s_{i,n}^{n-1*}$. Also, since $n \neq n + 1$ we have $s_{n+1,n}^{n-1*} \neq s_{i,n}^{n-1*}$.

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 $s_{i,j}^{i-1*} \neq s_{j,n+1}^{j-1*}$, and $s_{j,n+1}^{j-1*} \neq s_{i,j}^{j-1*}$, 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 term 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

\[
\sum_{(i,j) \in I_1} s_{i,j}^{i-1*} + s_{i,n+1}^{i-1*} + s_{n+1,j}^{i-1*} + s_{i,n+1,i}^{i-1*} + \sum_{(i,j) \in I_2} s_{i,j}^{j-1*} + s_{j,n+1}^{j-1*} + s_{i,j}^{j-1*} + s_{j,n+1}^{j-1*} + s_{i,j}^{j-1*} + s_{j,n+1}^{j-1*} + s_{i,j}^{j-1*} + s_{j,n+1}^{j-1*} + s_{i,j}^{j-1*}.
\]  

(55)

In what follows, we discuss the non-trivial comparisons, and refer to the first, second and third summations in (55) as $\Sigma_1, \Sigma_2$, and $\Sigma_3$, respectively.

1. $s_{i,j}^{i-1*}$ in $\Sigma_1$ vs. $s_{i,n+1}^{i-1*}$ in $\Sigma_2$: for these two terms to be coincident we need $i = n$. We also need $(i, j) = \{1, n\}$, 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 $\Sigma_1$ vs. $s_{j,n+1}^{j-1*}$ in $\Sigma_3$: 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 $s_{i,j}^{i-1*}$. 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. \( s^{i-1,n}_{j,n+1} \) in \( \Sigma_1 \) vs. \( s^{n+1,n}_{i+1,j} \) in \( \Sigma_2 \): 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. \( s^{i-1,n}_{j,n+1} \) in \( \Sigma_1 \) vs. \( s^{n-1,n}_{i+1,j} \) in \( \Sigma_2 \): 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 \( i' = j + 1 \), which contradicts \((i', j') \in I_3\).

5. \( s^{i-1,n}_{j,n+1} \) in \( S_1 \) vs. \( s^{n+1,n}_{i+1,j} \) in \( S_3 \): we can write the first term as \( s^{i-1,n+1}_{j} \). 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. \( s^{n-1,n}_{i+1,j} \) in \( \Sigma_1 \) vs. \( s^{n+1,n}_{i+1,j} \) in \( \Sigma_2 \): the equality of superscripts requires \( j = 1 \), which is impossible since \( j > i \in [n] \).

7. \( s^{n-1,n}_{i+1,j} \) in \( \Sigma_1 \) vs. \( s^{n-1,n+1}_{j} \) in \( \Sigma_2 \): 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. \( s^{j-1,n}_{i+1,j} \) in \( \Sigma_1 \) vs. \( s^{i-1,n+1}_{j} \) in \( \Sigma_3 \): 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. \( s^{j-1,n}_{i+1,j} \) in \( \Sigma_1 \) vs. \( s^{j-1,n+1}_{i} \) in \( \Sigma_3 \): 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. \( s^{n+1,n}_{i+1,j} \) in \( \Sigma_2 \) vs. \( s^{j-1,n+1}_{j} \) in \( \Sigma_3 \): for the equality of superscripts, we need \( i' = 1 \), and for the equality of subscripts, we need \( j' = 1 \). This contradicts \( i' < j' \).

11. \( s^{n+1,n}_{i+1,j} \) in \( \Sigma_2 \) vs. \( s^{j-1,n+1}_{i} \) in \( \Sigma_3 \): for equality of superscripts, we need \( j' = 1 \). For the equality of subscripts, we need \( i' = 1 \), which contradicts \( i' \neq j' \).

12. \( s^{n-1,n}_{j,n+1} \) in \( \Sigma_2 \) vs. \( s^{j-1,n+1}_{j} \) in \( \Sigma_3 \): for equality of superscripts, we need \( i' = j \). For the equality of subscripts, we need \( j' = 1 \), which contradicts \( i' < j' \).

13. \( s^{n-1,n}_{j,n+1} \) in \( \Sigma_2 \) vs. \( s^{j-1,n+1}_{j} \) in \( \Sigma_3 \): 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

\[
c \in [n], \quad a, b \in [n + 1] \setminus \{c\} \quad \text{and} \quad a \neq b,
\]

which is enough to prove that either \( s^{a,b}_{c} \) or \( s^{c}_{a,b} \) exist in (55).

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 \( i < j, i', j' < j' \) 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 \( s^{1-1,n}_{1,j} \). Clearly \( i' \neq i - 1 \) and \( j' \neq i - 1 \), from the definition of \( I_1 \). In the second term, \( s^{j-1,n+1}_{j,1} \), \( j \neq i - 1 \), from the definition of \( I_1 \), and \( i - 1 \neq n + 1 \), because otherwise \( i = n + 2 \notin [n] \). In the third term, \( s^{j-1,n}_{j,n+1} \), 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, \( s^{j-1,n}_{j,n+1} \), clearly \( i' \neq i - 1 \). Moreover, \( i' < i' < j' \), since \((i', j') \in I_3 \). In the second term, \( s^{j-1,n}_{j,n+1} \), \( j' \neq i' - 1 \), because since \( i' - 1 < i' < j' \). Moreover, \( n + 1 \neq i' - 1 \) because otherwise \( i' = n + 2 \notin [n] \). In the third term, \( s^{n+1,n}_{i+1,j} \), 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, \( s^{j-1,n}_{j,n+1} \), we have \( i' \neq j' - 1 \) since \((i', j') \in I_3 \). Also, clearly \( j' \neq j' - 1 \).

\[ \square \]

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 \( d_F \) 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 \( P, P' \in \mathcal{P} \),

\[
s(A, B, PP') = ||APP' - PP'B|| =
\]

\( APP' - PCP' + PCP' - PP'B \leq \)

\( PP' - 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 \( a_1 \leq a_2 \leq \cdots \leq a_m \) and \( b_1 \leq b_2 \leq \cdots \leq b_m \) then
\[
\| A - B \| \geq \sqrt{\sum_{i \in [m]} (a_i - b_i)^2}.
\]  
(57)

**Lemma 17.** Let \( \| \cdot \| \) be the operator 2-norm. If \( A \) and \( B \) are Hermitian matrices with eigenvalues \( a_1 \leq a_2 \leq \cdots \leq a_m \) and \( b_1 \leq b_2 \leq \cdots \leq b_m \), and \( P \in \mathbb{R}^{m \times m} \) is a permutation matrix, then
\[
\| a - b \| \leq \| a - Pb \|. 
\]  
(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 \( \Lambda_{A_i} \) and \( \Lambda_C \) are ordered in increasing order, and \( U_i \) and \( V \) are orthonormal matrices. Using Lemma 15, we have that
\[
\| A_i P_i - P_i C \| = \| (A_i - P_i C (P_i)^{-1}) P_i \| 
= \| A_i - P_i C (P_i)^{-1} \| 
= \| U_i (\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C (P_i)^{-1} U_i) U_i^{-1} \| 
= \| \text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C (P_i)^{-1} U_i \| \geq \| \Lambda_{A_i} - \Lambda_C \|,
\]  
where the last inequality follows from Lemma 16 or Lemma 17 (depending on the norm).

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

Finally, notice that, by the equalities in (60), we have
\[
d_F(A_{1:n}) = \min_{P \in \mathbb{R}^{m \times m}, C \in \mathbb{R}^{m \times m}} \sum_{i=1}^{n} \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C) \| \leq \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C) \|,
\]  
(61)

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

Since \( \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C) \|_{\text{Frobenius}} = \| \Lambda_{A_i} - \Lambda_C \|_{\text{Euclidean}} \) and \( \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C) \|_{\text{operator}} = \| \Lambda_{A_i} - \Lambda_C \|_{\infty}\), 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 \( \Lambda_{A_i} \) are ordered in increasing order, and \( U_i \) is an orthonormal matrix. Using Lemma 15, we get
\[
\| A_i P_{i,j} - P_{i,j} A_j \| = \| (A_i - P_{i,j} A_j (P_{i,j})^{-1}) P_{i,j} \| 
= \| A_i - P_{i,j} A_j (P_{i,j})^{-1} \| 
= \| U_i (\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i) U_i^{-1} \| 
= \| \text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i \| \geq \| \Lambda_{A_i} - \Lambda_A \|,
\]  
(62)

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

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

At the same time, \( d_G(A_{1:n}) = \min_{P \in S} \frac{1}{2} \sum_{i,j \in [n]} \| \text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i \| \leq \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_A) \|, \)  
(63)

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 \( \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_A) \|_{\text{Frobenius}} = \| \Lambda_{A_i} - \Lambda_A \|_{\text{Euclidean}} \) and \( \| \text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_A) \|_{\text{operator}} = \| \Lambda_{A_i} - \Lambda_A \|_{\infty} \), 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
\[
d_F([A']_{1:n}) = d_F(A_{1:n}) = \min_{B \in \Omega} \sum_{i=1}^{n} d(A_i', B) \leq \min_{B \in \Omega} \sum_{i=1}^{n} d(A_i, B) = \min_{B \in \Omega} \sum_{i=1}^{n} d(A_i, B) = d_F([A]_{1:n}) = d_F([A']_{1:n}),
\]
where in the last equality we used \( d(A_i', A_i) = 0 \), since \( A_i' \in [A_i] \). Similarly, we can show that \( d'_F([A]_{1:n}) \leq d'_F([A']_{1:n}) \). It follows that \( d_F([A]_{1:n}) = d_F([A']_{1:n}) \), and hence (17) is well defined.
We now prove that $d'_G$ satisfies (7). Recall that, by Theorem 1, $d_F$ is a pseudo $n$-metric. If $[A_1] = \cdots = [A_n]$, then

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

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

On the other hand, if $d'_F([A_1]_{1:n}) = d_F(A_1, A_1) = 0$, then there exists $B \in \Omega$, such that $d(A_1, B) = 0$ for all $i \in [n]$. Since $d$ is non-negative and symmetric, and also satisfies the triangle inequality, it follows that

$$0 \leq d(A_i, A_j) \leq d(A_i, B) + d(B, A_j) = d(A_i, B) + d(A_j, B) = 0.$$

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 $d_2$ on two graphs and we let $S_n$ denote the set $S$ in definition (13) for the distance $d_2$ 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

$$d'_{P}(A_i, A'_i) = \frac{1}{2}(s(A_i, A_i, I) + s(A'_i, A'_i, I) + s(A'_i, A_i, (P^*_i)^{-1}) + s(A_i, A'_i, P^*_i) + s(A_i, A'_i, (P^*_i)^{-1})) = 0.$$

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 $\{P^*_i \tilde{P}_{i,j} (P^*_i)^{-1}\}_{i,j \in [n]} \in S_2$. Thus,

$$d'_G([A'_1]_{1:n}) = d'_{G}(A'_1, A'_1) = \min_{P \in S} \frac{1}{2} \sum_{i,j \in [n]} s(A'_i, A'_j, P_{i,j}).$$

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

$$\frac{1}{2} \sum_{i,j \in [n]} s(A'_i, A'_j, P^*_i \tilde{P}_{i,j} (P^*_i)^{-1}) \leq \frac{1}{2} \sum_{i,j \in [n]} s(A'_i, A_i, P^*_i) + s(A_i, A'_j, \tilde{P}_{i,j}) + s(A_i, A'_j, (P^*_i)^{-1}) = s(A_i, A_j, \tilde{P}_{i,j}).$$

Taking the minimum of the r.h.s. of the above expression over $\tilde{P}$ we get $d'_G([A'_1]_{1:n}) \leq d'_G(A_i, A_i) = d'_G([A_i]_{1:n})$. Similarly, we can prove $d'_G([A]_{1:n}) \leq d'_G([A_i]_{1:n})$. It follows that $d'_G([A]_{1:n}) = d'_G([A_i]_{1:n})$, and hence (18) is well defined.

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

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

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

On the other hand, if $d'_G([A]_{1:n}) = d_G(A_1, A_1) = 0$, 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 $\Gamma_i \in \mathbb{R}^{nm \times nm}$, $\|\Gamma_i\|_2 \leq 1$ for all $i \in [n]$. Let $P \in \mathbb{R}^{nm \times nm}$ have $n^2$ blocks such that the $((i, j))$th block is $\Gamma_i \Gamma_j$ if $i \neq j$, and $I$ otherwise. We have that $P \succeq 0$, and that $\|P\|_* \leq mn$.

**Proof.** Let us first prove that $P \succeq 0$. Let $v \in \mathbb{R}^{nm}$ have $n$ blocks, the $i$th block being $v_i \in \mathbb{R}^m$. Since $\|\Gamma_i \Gamma_j\|_2 \leq \|\Gamma_i\|_2 \|\Gamma_j\|_2 \leq 1$, we have that $\|v_i \Gamma_i \Gamma_j v_j\|_2 \leq \|v_i\|_2 \|\Gamma_j\|_2$ for all $i \in [n]$. Therefore, we have $v^T P v = \|\sum_{i \in [n]} \Gamma_i^T v_i\|_2^2 + \sum_{i \in [n]} \|\sum_{i \in [n]} \Gamma_i^T v_i\|_2^2 \leq 0$, for any $v$, which implies that $P \succeq 0$.

We now prove that $\|P\|_* \leq mn$. Let $\sigma_r$ and $\lambda_r$ be the $r$th singular value and $r$th eigenvalue of $P$ respectively. Since $P$ is real-symmetric and positive semi-definite, we have that $\|P\|_* = \sigma_r = \sum_{i \in [n]} \lambda_i = \sum_{i \in [n]} \lambda_i = \text{tr}(P) = mn$. □

**Proof of Theorem 5.**

(Non-negativity): Since $s$ is a modified $P$-score, it satisfies (30), i.e., $s \succeq 0$, which implies $d_{scG} \succeq 0$, 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 $d_{scG}$, we are assuming that $I \in P$. Furthermore, $P$ defined using these $P_{i,j}$'s satisfies $\|P\|_* \leq mn$. Therefore, this choice of $P_{i,j}$'s satisfies the constraints in the minimization problem in the definition of $d_{scG}(A_{1:n})$. Therefore, $d_{scG}(A_{1:n})$ is upper-bounded by 0, which along with its non-negativity leads to $d_{scG}(A_{1:n}) = 0$.

(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 $\sigma$. Thus, all that we need to show is that $P \succeq 0$ if and only if $P' \succeq 0$, where $P'$ is just like $P$ 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

\[
\sum_{i \neq j} s(A_i, A_j, P_{i,j}^*) \leq \sum_{i \neq j} s(A_i, A_j, \Gamma_i \Gamma_j^T),
\]

(64)

where \( \{ \Gamma_i \} \subseteq [n] \) are matrices in \( \mathcal{P} \). This inequality holds because \( P_{i,j} \) defined by \( P_{i,j} = \Gamma_i \Gamma_j^T \) \( \forall i \neq j \), 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 \( \Gamma_i \Gamma_j^T \in \mathcal{P} \). Furthermore, if we define \( P \) to have as the \( (i,j) \)th block, \( \Gamma_i \Gamma_j^T \), and have as the \( (i,i) \)th block the identity \( I \), then, by Lemma 18, we know that \( P \geq 0 \).

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

\[
\sum_{i \neq j} s(A_i, A_j, \Gamma_i \Gamma_j^T) \leq \sum_{i \neq j} s(A_i, A_{n+1}, \Gamma_i) + \sum_{i \neq j} s(A_{n+1}, A_j, \Gamma_j^T) = s(A_i, A_{n+1}, \Gamma_i) + \sum_{i \neq j} s(A_j, A_{n+1}, \Gamma_j).
\]

(65)

(66)

The rest of the proof follows by choosing \( \Gamma_i \) 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 \( P \) 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.](image)

![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.](image)
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.