n-metrics for multiple graph alignment

Sam Safavi
safavisa@bc.edu

José Bento
jose.bento@bc.edu

***Abstract*—The work of Ioannidis et al. 2018 introduces a family of distances between two graphs that provides tractable graph alignment strategies. Importantly, the alignment scores produced by this family satisfy the properties of metrics, which is very useful in several learning tasks. In this paper, we generalize this work to compare n graphs by introducing a family of distances, which is an n-metric, i.e., an extension of a metric to n elements that includes a generalization of the triangle inequality. Our new family of distances, includes the ones in the work of Ioannidis et al. 2018 as a special case, and can produce tractable alignments between multiple graphs.**

I. Introduction

Graph alignment refers to the process of evaluating the structural similarity between two graphs. It aims to find a mapping between the nodes of two graphs such that if two nodes are connected in one graph, their images through this map are connected in the other graph. Graph alignment, also referred to as graph matching, finds applications in a variety of fields including but not limited to computational biology [1]–[3], object recognition [4], ontology alignment [5], computer vision [4], and social networks [6], to name a few.

In many applications, it is desirable to jointly align multiple graphs. This is the case, for example, in aligning protein-protein interaction networks [7], recommendation systems, in the collective analysis of networks, or in the alignment of graphs obtained from brain Magnetic Resonance Imaging (MRI) [8]. The problem of collectively aligning multiple graphs at once, is refereed to as *multiple graph alignment*, which is the focus of this work. Although the literature on graph alignment, formulated for a pair of graphs, has been extensively studied, the problem of multiple graph alignment is still under-exploited. Multiple graph alignment has been studied in, e.g., [9]–[21].

Graph distances naturally arise in graph alignment literature as a measure of the similarity between two graphs. This similarity is also a measure of how well two graphs might be aligned. A highly desirable property for such a score is that it is a *metric*, i.e., it is non-negative, symmetric, and satisfies the triangle inequality, as well as, the identity of indiscernibles. Metrics provide significant computational advantages over non-metrics. For example, operations such as nearest-neighbor search, clustering, and outlier detection admit fast algorithms precisely when performed over objects embedded in a metric space [22]. In the context of multiple graph alignment, the notion of graph distance can be generalized to define a score quantifying the structural differences among a set of n graphs. This score is also a measures of how well multiple graphs might be aligned.

Related work: The notion of metric space was first introduced by Fréchet [23] and later developed by Hausdorff [24]. Ever since, devising suitable distance metrics, to quantify the similarity between a set of objects, has been studied by many researchers. For example, [25] generalizes the classical notion of distance to n elements, by replacing the triangle inequality with the *simplex inequality*. By introducing the concept of *functionally expressible multi-distance*, Ref. [26] deals with the problem of aggregating pairwise distance values in order to obtain a multi-argument distance function. Ref [26] obtains multidimensional distance functions from Euclidean, Manhattan, and Chebyshev distance functions in $\mathbb{R}^2$. Ref. [27], obtains a generalized version of the well-known distance function family L_p norm, and shows that the new functions satisfy metric properties. Ref. [28] defines mathematical metrics for sets of trajectories using convex programs. We also refer the reader to [29] that surveys generalized metric spaces, and [30] 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. Of our particular interest is the work of [22] that defines a broad family of graph distances that include computationally tractable metrics. The distances in [22] include as special cases classic graph distances, such as *chemical* and *CKS distances*, satisfy the metric property, and are tractable, i.e., can be computed either by solving a convex optimization problem, or by a polynomial time algorithm.

In this paper, we generalize [22] to compare n graphs. We generalize the notion of a metric to $n \geqslant 3$ elements, which we refer to as an n-metric. We then introduce the $\mathcal{G}$-*align* distance function that can be used for multiple graph alignment. We also consider the *Fermat* distance function, which was introduced by [25]. Most importantly, we show that both the *Fermat* distance function and the $\mathcal{G}$-*align* distance function are n-metrics.

II. Notation and Preliminaries

Let $[m] = \{1, \ldots, m\}$. A graph, $G = (V \equiv [m], E)$, with node (vertex) 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 $\Omega \subseteq \mathbb{R}^{m\times m}$. For example, Ω might be the set of adjacency matrices, or the set of matrices containing hop-distances between all pairs of nodes in a graph.

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 among the edges in E_i and E_j, from the one among the vertices in V_i and V_j. We denote the set of all pairwise matching matrices by $\mathcal{P} = \{\{P_{i,j}\}_{i,j\in[n]} : P_{i,j} \subseteq \mathbb{R}^{m\times m}\}$. For example, $\mathcal{P}$ might be the set of 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 $A^i_{1:n,n+1}$ 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.

Definition 1. *A map $d : \Omega^2 \mapsto \mathbb{R}$, is a metric, if and only if, for all $A,B,C \in \Omega$:*

$$d(A,B) \geqslant 0, \tag{1}$$
$$d(A,B) = 0, \text{ iff } A = B, \tag{2}$$
$$d(A,B) = d(B,A), \tag{3}$$
$$d(A,C) \leqslant d(A,B) + d(B,C). \tag{4}$$

Definition 2. *A map $d : \Omega^2 \mapsto \mathbb{R}$, is a pseudometric, if and only if it satisfies properties* (1)*,* (3) *and* (4)*, and*

$$d(A,A) = 0 \ \ \forall A \in \Omega. \tag{5}$$

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 $A,B,C \in \Omega$, $s(\cdot,\cdot,P)$ satisfies the following properties:*

$$s(A,B,P) \geqslant 0, \tag{6}$$
$$s(A,A,I) = 0, \tag{7}$$
$$s(A,B,P) = s(B,A,P^{-1}), \tag{8}$$
$$s(A,B,P) + s(B,C,P') \geqslant s(A,C,PP'). \tag{9}$$

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 ([22])**.** *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*

$$d_{SB}(A,B) = \min_{P\in\mathcal{P}} |||AP - PB||| + tr(P^\top D). \tag{10}$$

The authors in [22], 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.

Definition 5. *A map $d : \Omega^n \mapsto \mathbb{R}$ is an $(n-1)$-hemimetric, if and only if, for all $A_1,\ldots,A_n \in \Omega$,*

$$d(A_{1:n}) \geqslant 0, \tag{11}$$
$$d(A_{1:n}) = 0, \text{ iff } A_{1:n} \text{ are not pairwise distinct}, \tag{12}$$
$$d(A_{1:n}) = d(A_{\sigma(1:n)}), \tag{13}$$
$$d(A_{1:n}) \leqslant \textstyle\sum_{i=1}^n d(A^i_{1:n,n+1}). \tag{14}$$

III. n-METRICS FOR MULTI-GRAPH ALIGNMENT

It is possible to generalize the notion of a (pseudo) metric to $n \geqslant 3$ elements. Motivated by the definition of a hemimetric, we consider the following definitions:

Definition 6. *A map $d : \Omega^n \mapsto \mathbb{R}$, is an n-metric, if and only if, for all $A_1,\ldots,A_n \in \Omega$ it satisfies properties* (11)-(14)*, with property* (12) *replaced by,*

$$d(A_{1:n}) = 0, \text{ iff } A_1 = A_2 = \ldots = A_n. \tag{15}$$

Note that according to this definition, a 2-metric is a metric as in Definition 1. In the sequel, we refer to properties (11), (13), (14), and (15) as non-negativity, symmetry, generalized triangle equality, and identity of indiscernibles, respectively.

Definition 7. *A map $d : \Omega^n \mapsto \mathbb{R}$, is a pseudo n-metric, if and only if it satisfies properties* (11)*,* (13) *and* (14)*, and for any $A \in \Omega$, d satisfies the property of* self-identity *defined as,*

$$d(A,\cdots,A) = 0. \tag{16}$$

We now define two functions that, as we show in Section VII, satisfy the properties of (pseudo) n-metrics.

A. Fermat distance function

Definition 8. *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*

$$d_F(A_{1:n}) = \min_{B\in\Omega} \sum_{i=1}^n d(A_i, B). \tag{17}$$

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 [25], and is included in Section VII for completeness.

For example, the Fermat distance function induced by an SB-distance function with one D_i per pair (A_i, B) is

$$d_F(A_{1:n}) = \min_{\substack{B\in\Omega,\\ P=\{P_i\}\in\mathcal{P}^n}} \sum_{i=1}^{n} |||A_iP_i - P_iB||| + \mathrm{tr}(P_i^\top D_i). \quad (18)$$

Despite its simplicity, (18) 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.

B. $\mathcal{G}$-align distance function

Definition 9. *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*

$$d_\mathcal{G}(A_{1:n}) = \min_{P\in S} \frac{1}{2} \sum_{i,j\in[n]} s(A_i, A_j, P_{i,j}), \quad (19)$$

where

$$S = \{\{P_{i,j}\}_{i,j\in[n]} \colon P_{i,j} \in \mathcal{P}, \forall i, j \in [n],\ P_{i,k}P_{k,j} = P_{i,j}, \forall i\ j, k \in [n], P_{i,i} = I, \forall i \in [n]\}. \quad (20)$$

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

Remark 2. *In* (20)*, 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*

$$d_\mathcal{G}(A_{1:n}) = \min_{P\in S} \sum_{i,j\in[n],\ i<j} s(A_i, A_j, P_{i,j}). \quad (21)$$

Proof.

$$\sum_{i,j\in[n]} s(A_i, A_j, P_{i,j}) = \sum_{i\in[n]} s(A_i, A_i, P_{i,i}) + \sum_{i,j\in[n]:\ i<j} (s(A_i, A_j, P_{i,j}) + s(A_j, A_i, P_{j,i})). \quad (22)$$

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 (7), and $s(A_j, A_i, P_{j,i}) = s(A_i, A_j, P_{i,j})$, by property (8). 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 $s(A, B, P) = |||AP - PB|||$, for some element-wise matrix norm, $n = 2$, and $\mathcal{P}$ is the set of permutations on m elements, then according to Lemma 1, $d_\mathcal{G}(A, B) = d_{SB}(A, B)$, 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

$$d_{SB}(A, B) = \min_{P\in\mathcal{P}} s(A, B, P) + \mathrm{tr}(P^\top D), \quad (23)$$

and investigate the conditions on s, $\mathcal{P}$ and D, under which (23) 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 (19), 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, $P \in S$. ☐

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

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

IV. n-METRICS ON QUOTIENT SPACES

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, (15) 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.

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

$$d'_F([A]_{1:n}) = d_F(A_{1:n}). \quad (24)$$

Then, d'_F is an n-metric.

Proof. We first show that (24) is well defined. Let $A'_i \in [A_i]$. Since d satisfies the triangle inequality, (4), 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 (24) is well defined.

We now prove that d'_F satisfies (15). 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 (16).

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 $i, j \in [n]$. □

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 induced by s, and $d_{\mathcal{G}} : \Omega^n \mapsto \mathbb{R}$ be the $\mathcal{G}$-align distance function induced by s. Let $\Omega' = \Omega \backslash \sim_d$, and $d'_{\mathcal{G}} : \Omega'^n \mapsto \mathbb{R}$ be such that*

$$d'_{\mathcal{G}}([A]_{1:n}) = d_{\mathcal{G}}(A_{1:n}). \tag{25}$$

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

Proof. In the proof, we let S_2 denote the set S in definition (20) for the distance d on two graphs and we let S_n denote the set S in definition (20) for the distance $d_{\mathcal{G}}$ on n graphs.

We first verify that (25) 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 (9) 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 further 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 (25) is well defined.

Now we show that $d'_{\mathcal{G}}$ satisfies (15). Recall that, by Theorem 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 pseudo-metric, and hence satisfies the property of self-identity (16).

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 $i, j \in [n]$. □

V. Special case of orthogonal matrices

In this section, we discuss the special case, where the pairwise matching matrices are orthogonal. 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 orthogonal invariant. $d(A, B) = \min_{P \in \mathcal{P}} s(A, B, P)$.*

We now provide the main results of this section.

Theorem 5. *Under Assumption 1, d_F induced by d, and $d_{\mathcal{G}}$ induced by s, are pseudo n-metrics.*

Proof. 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 [22].

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,

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

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

$$\begin{aligned}&s(A, B, PP') = \|\|APP' - PP'B\|\| =\\ &\|\|APP' - PCP' + PCP' - PP'B\|\| \leqslant\\ &\|\|APP' - PCP'\|\| + \|\|PCP' - PP'B\|\| =\\ &\|\|AP - PC\|\| + \|\|CP' - P'B\|\| =\\ &s(A, C, P) + s(C, B, P').\end{aligned}$$

□

Theorem 6. *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,*

$$d_F(A_{1:n}) = \min_{\Lambda_C \in \mathbb{R}^m} \sum_{i=1}^{n} \|\Lambda_{A_i} - \Lambda_C\|. \tag{26}$$

Proof. The proof uses the following lemmas by [31] and [22].

Lemma 3. *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 4. *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*

$$\|\|A - B\|\| \geqslant \sqrt{\sum_{i \in [m]} (a_i - b_i)^2}. \tag{27}$$

Lemma 5. *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*

$$\|\|A - B\|\| \geqslant \max_{i \in [m]} |a_i - b_i|. \tag{28}$$

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*

$$\|a - b\| \leqslant \|a - Pb\|. \tag{29}$$

Proof. This follows directly from Lemma 4 and Lemma 5 by letting A and B be diagonal matrices with a and Pb in the diagonal, respectively. □

We now proceed with the proof of Theorem 6. 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 3, we have that

$$\begin{aligned} &|||A_i P_i - P_i C||| = |||(A_i - P_i C (P_i)^{-1}) P_i||| \\ &= |||A_i - P_i C (P_i)^{-1}||| \\ &= |||U_i(\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C (P_i)^{-1} U_i) U_i^{-1}||| \\ &= |||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C (P_i)^{-1} U_i||| \geqslant \|\Lambda_{A_i} - \Lambda_C\|, \end{aligned} \tag{30}$$

where the last inequality follows from Lemma 4 or Lemma 5 (depending on the norm).

It follows from (30) 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 (30), we have

$$\begin{aligned} d_F(A_{1:n}) &= \min_{P \in \mathcal{P}^n, C \in \Omega} \sum_{i=1}^n |||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_i C (P_i)^{-1} U_i||| \\ &\leqslant |||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_C)|||, \end{aligned} \tag{31}$$

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

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,*

$$d_{\mathcal{G}}(A_{1:n}) = \frac{1}{2} \sum_{i,j \in [n]} \|\Lambda_{A_i} - \Lambda_{A_j}\|. \tag{32}$$

Proof. 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 3, we have

$$\begin{aligned} &|||A_i P_{i,j} - P_{i,j} A_j||| = |||(A_i - P_{i,j} A_j (P_{i,j})^{-1}) P_{i,j}||| \\ &= |||A_i - P_{i,j} A_j (P_{i,j})^{-1}||| \\ &= |||U_i(\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i) U_i^{-1}||| \\ &= |||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i||| \geqslant \|\Lambda_{A_i} - \Lambda_{A_j}\|, \end{aligned} \tag{33}$$

where the last inequality follows from Lemma 4 or Lemma 5 (depending on the norm).

From (33) 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,

$$\begin{aligned} d_{\mathcal{G}}(A_{1:n}) &= \min_{P \in S} \frac{1}{2} \sum_{i,j \in [n]} |||\text{diag}(\Lambda_{A_i}) - U_i^{-1} P_{i,j} A_j (P_{i,j})^{-1} U_i||| \\ &\leqslant |||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_{A_j})|||, \end{aligned} \tag{34}$$

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 $|||\text{diag}(\Lambda_{A_i}) - \text{diag}(\Lambda_{A_j})|||_{\text{Frobenius}} = \|\Lambda_{A_i} - \Lambda_{A_j}\|_{\text{Eucledian}}$ 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. □

Note that $d_F = d_{\mathcal{G}} = 0$ 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 (32) as an optimization problem similar to d_F in (26), it is tempting to define $d_{\mathcal{G}}$ using s^2 instead of s, and take a square root. Let us call the resulting function $\bar{d}_{\mathcal{G}}$. A straightforward calculation allows us to write

$$(\bar{d}_{\mathcal{G}}(A_{1:n}))^2 = \frac{1}{2} \sum_{i,j\in[n]} \|\Lambda_{A_i} - \Lambda_{A_j}\|^2 \tag{35}$$

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

$$= n \min_{\Lambda_C \in \mathbb{R}^m} \frac{1}{2} \sum_{i\in[n]} \|\Lambda_{A_i} - \Lambda_C\|^2, \tag{36}$$

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 (26), and a nice connection between $\bar{d}_{\mathcal{G}}$ and the geometric sample variance.

Using *Jensen's inequality*, the following connection between d_F, $d_{\mathcal{G}}$ and the variance of the spectra of $A_{1:n}$ can be verified.

$$d_F \leqslant \sqrt{2} \bar{d}_{\mathcal{G}} = \sqrt{2} n \sqrt{\text{Var}(\Lambda_{A_{1:n}})}, \tag{37}$$

$$d_{\mathcal{G}} \leqslant \sqrt{n/2}\, \bar{d}_{\mathcal{G}} = \frac{n^{3/2}}{\sqrt{2}} \sqrt{\text{Var}(\Lambda_{A_{1:n}})}. \tag{38}$$

Another interesting difference can be noted if we let $A_{1:n}$ be fix, and define the set function $\tilde{d}_F : [n] \supseteq I = \{i_1, \ldots, i_k\} \mapsto \mathbb{R}$ as $\tilde{d}_F(I) = d_F(A_{i_1}, \ldots, A_{i_k})$, and, similarly, define the set function $\tilde{d}_{\mathcal{G}} : [n] \supseteq I = \{i_1, \ldots, i_k\} \mapsto \mathbb{R}$ as $\tilde{d}_{\mathcal{G}}(I) = d_{\mathcal{G}}(A_{i_1}, \ldots, A_{i_k})$. It is straightforward to verify that $\tilde{d}_{\mathcal{G}}$ is a *super-modular* function, while $\tilde{d}_F$ is neither super-modular nor *sub-modular*.

VI. 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, in this section, we prove that the $\mathcal{G}$-align function satisfies the generalized triangle inequality when $n = 4$.

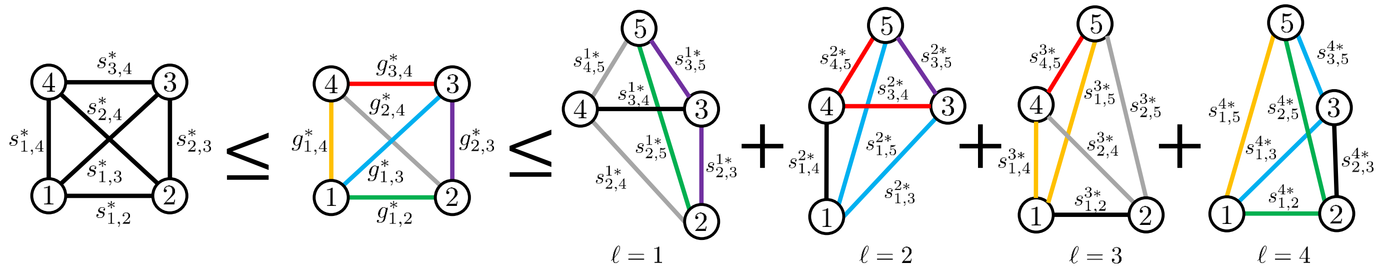

Fig. 1. Generalized triangle equality of $d_{\mathcal{G}}$ for $n = 4$ graphs.

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

$$d_{\mathcal{G}}(A_{1:4}) \leqslant \sum_{\ell=1}^{4} d_{\mathcal{G}}(A^{\ell}_{1:4,5}). \tag{39}$$

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 (39). 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 (39). Note that, according to (8), and the fact that $P^*_{i,j} = (P^*_{j,i})^{-1}$ (since $P^* \in S$), we have

$$s^*_{i,j} = s^*_{j,i}, \text{ and } s^{\ell *}_{i,j} = s^{\ell *}_{j,i}. \tag{40}$$

Moreover, according to (9), we have

$$s(A_i, A_j, P^{\ell *}_{i,k} P^{\ell' *}_{k,j}) \leqslant s^{\ell *}_{i,k} + s^{\ell' *}_{k,j}, \tag{41}$$

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

$$s^{\ell *}_{i,j} \leqslant s^{\ell *}_{i,k} + s^{\ell *}_{k,j}. \tag{42}$$

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

$$\sum_{i,j \in [4],\ i<j} s^*_{i,j} \leqslant \sum_{i,j \in [4],\ i<j} s(A_i, A_j, \Gamma_{i,j}), \tag{43}$$

where $\Gamma_{i,j} = \Gamma_i \Gamma_j^{-1}$, and $\{\Gamma_i\}$ are any set of invertible matrices. Note that from Lemma 2 we know that $\{\Gamma_{i,j}\} \in S$.

Consider the following choices for Γ_i's :

$$\Gamma_1 = P^{4*}_{1,5};\ \Gamma_2 = P^{1*}_{2,5};\ \Gamma_3 = P^{2*}_{3,5};\ \Gamma_4 = P^{3*}_{4,5}. \tag{44}$$

We define $g^*_{i,j} = s(A_i, A_j, \Gamma_i \Gamma_j^{-1})$, in which Γ_i's are chosen according to (44). We can then rewrite (43) as

$$\sum_{i,j \in [4],\ i<j} s^*_{i,j} \leqslant \sum_{i,j \in [4],\ i<j} g^*_{i,j}. \tag{45}$$

We use Fig 1 to bookkeep all the terms involved in proving (39). In particular, the first inequality in Fig. 1 provides a pictorial representation of (45). 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 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 (39), 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 (44).

Using (41), followed by (40) we get

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

The above inequality is also depicted in Fig. 1, where each diagram on the r.h.s of the second inequality represents $d_{\mathcal{G}}(A^{\ell}_{1:5})$ in (39) for a different $\ell \in [4]$. Applying (42) 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 r.h.s of This completes the proof.

VII. General Proofs

In this section, we provide detailed proofs for Theorems 1 and 2 from Section III.

A. Proof of Theorem 1

In the following lemmas, we show that the Fermat distance function satisfies properties (11), (13), (14), and (16), and hence is a pseudo n-metric.

Lemma 6. *d_F is non-negative.*

Proof. If d is a pseudo metric, it is non-negative. Thus, (17) is the sum of non-negative functions, and hence also non-negative. □

Lemma 7. *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 (16) follows. □

Lemma 8. *d_F is symmetric.*

Proof. Property (13) simply follows from the commutative property of summation. □

Lemma 9. *d_F satisfies the generalized triangle inequality.*

Proof. Note that the following proof is a direct adaptation of the one in [25], and is included for the sake of completeness. We show that the Fermat distance satisfies (14), i.e.,

$$d_F(A_{1:n}) \leqslant \sum_{i=1}^{n} d_F(A^i_{1:n,n+1}). \tag{46}$$

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

$$d_F(A^i_{1:n,n+1}) = d(A_{n+1}, B_i) + \sum_{j \in [n] \setminus i} d(A_j, B_i). \quad (47)$$

Equation (47) implies that

$$\sum_{i=1}^{n} d_F(A^i_{1:n,n+1}) \geqslant \sum_{i=1}^{n} \sum_{j \in [n] \setminus i} d(A_j, B_i) \geqslant d(A_1, B_n) + d(A_2, B_n) + \sum_{i=2}^{n-1} (d(A_1, B_i) + d(A_{i+1}, B_i)). \quad (48)$$

Using (4), 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 (48),

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

where we used $d(A_1, A_1) = 0$ in the equality. The last inequality follows from Definition 8, and completes the proof. □

B. Proof of Theorem 2

In the following lemmas, we show that the $\mathcal{G}$-align distance function satisfies properties (11), (13), (14), and (16), and hence is a pseudo n-metric.

Lemma 10. *$d_{\mathcal{G}}$ is non-negative.*

Proof. Since s is a P-score, it satisfies (6), i.e., $s \geqslant 0$, which implies $d_{\mathcal{G}} \geqslant 0$, since it is a sum of P-scores. □

Lemma 11. *$d_{\mathcal{G}}$ 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 (7), for all $i, j \in [n]$. Therefore,

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

□

Lemma 12. *$d_{\mathcal{G}}$ is symmetric.*

Proof. The definition, (19), 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 $\{A_i\}$. □

Lemma 13. *$d_{\mathcal{G}}$ satisfies the generalized triangle inequality.*

Proof. We now show that $d_{\mathcal{G}}$ satisfies (14), i.e.,

$$d_{\mathcal{G}}(A_{1:n}) \leqslant \sum_{\ell=1}^{n} d_{\mathcal{G}}(A^i_{1:n,n+1}). \quad (49)$$

Let $P^* = \{P^*_{i,j}\} \in S$ be an optimal value for P in the optimization problem corresponding to the l.h.s of (49). Henceforth, just like Section VI, we use $s^*_{i,j} = s(A_i, A_j, P^*_{i,j})$ for all $i, j \in [n]$. Note that according to (7) and (8), we have $s^*_{i,i} = 0$, and $s^*_{i,j} = s^*_{j,i}$, respectively. From (21), we have,

$$d_{\mathcal{G}}(A_{1:n}) = \sum_{i,j \in [n],\ i<j} s(A_i, A_j, P^*_{i,j}) = \sum_{i,j \in [n],\ i<j} s^*_{i,j}. \quad (50)$$

Let $P^{k*} = \{P^{k*}_{i,j}\} \in S$ be an optimal value for P in the optimization problem associated to $d_{\mathcal{G}}(A^i_{1:n,n+1})$ on the r.h.s of (49). Henceforth, just like Section VI, we use $s^{\ell*}_{i,j} = s(A_i, A_j, P^{\ell*}_{i,j})$ for all $i, j \in [n+1], \ell \in [n] \setminus \{i, j\}$. Note that $s^{\ell*}_{i,i} = 0$, and $s^{\ell*}_{i,j} = s^{\ell*}_{j,i}$. From (21), we can write,

$$\sum_{\ell=1}^{n} d_{\mathcal{G}}(A^i_{1:n,n+1}) = \sum_{\ell=1}^{n} \sum_{\substack{i,j \in [n+1],\ i<j \\ \ell \notin \{i,j\}}} s^{\ell*}_{i,j}. \quad (51)$$

We will show that,

$$\sum_{i,j \in [n],\ i<j} s^*_{i,j} \leqslant \sum_{\ell=1}^{n} \sum_{\substack{i,j \in [n+1],\ i<j \\ \ell \notin \{i,j\}}} s^{\ell*}_{i,j}. \quad (52)$$

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

$$\sum_{i,j \in [n],\ i<j} s^*_{i,j} \leqslant \sum_{i,j \in [n],\ i<j} s(A_i, A_j, \Gamma_{i,j}), \quad (53)$$

for any matrices $\{\Gamma_{i,j}\}_{i,j \in [n]}$ in S, where S satisfies Definition 9. Hence, from Lemma 2, we also know that

$$\sum_{i,j \in [n],\ i<j} s^*_{i,j} \leqslant \sum_{i,j \in [n],\ i<j} s(A_i, A_j, \Gamma_i \Gamma_j^{-1}), \quad (54)$$

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

Consider the following choice for Γ_i :

$$\Gamma_i = P^{i-1*}_{i,n+1}, \qquad 2 \leqslant i \leqslant n, \quad (55)$$
$$\Gamma_1 = P^{n*}_{1,n+1}. \quad (56)$$

Remark 3. *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 $P^{\ell*}_{i,j}$, the index in superscript satisfies $\ell = 0 \Leftrightarrow \ell = n$.*

Note that since $P^{i-1*} \in S$, then $\Gamma_i = P^{i-1*}_{i,n+1}$ is invertible and belongs to $\mathcal{P}$. Using (55) to replace Γ_i and Γ_j in (54), and the fact that $(P^{j-1*}_{j,n+1})^{-1} = P^{j-1*}_{n+1,j}$, along with property (9) of the P-score s, we have

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

We now show that

$$\sum_{\substack{i,j \in [n] \\ i<j}} s^{i-1*}_{i,n+1} + s^{j-1*}_{n+1,j} \leqslant \sum_{\ell=1}^{n} \sum_{\substack{i,j \in [n+1],\ i<j \\ \ell \notin \{i,j\}}} s^{\ell*}_{i,j}, \quad (57)$$

which will prove (52) 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 (9) of the P-score s, and the alignment consistency of $P \in S$.

$$\sum_{(i,j)\in I_1} s^{i-1*}_{i,n+1} \leqslant \sum_{(i,j)\in I_1} s^{i-1*}_{i,j} + s^{i-1*}_{j,n+1}. \tag{58}$$

$$\sum_{(i,j)\in I_2} s^{j-1*}_{n+1,j} \leqslant \sum_{(i,j)\in I_2} s^{j-1*}_{n+1,i} + s^{j-1*}_{i,j}. \tag{59}$$

$$\sum_{(i,j)\in I_3} s^{i-1*}_{i,n+1} + s^{j-1*}_{n+1,j} \leqslant \sum_{(i,j)\in I_3} \Big(s^{i-1*}_{i,j} + s^{i-1*}_{j,n+1} + s^{j-1*}_{n+1,i} + s^{j-1*}_{i,j} \Big). \tag{60}$$

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

$$\sum_{i,j\in[n]} (\cdot) = \sum_{(i,j)\in I_1} (\cdot) + \sum_{(i,j)\in I_2} (\cdot) + \sum_{(i,j)\in I_3} (\cdot). \tag{61}$$

Using (58)-(60), and (61) we have

$$\begin{aligned}\sum_{i,j\in[n],\ i<j} s^{i-1*}_{i,n+1} + s^{j-1*}_{n+1,j} \leqslant &\sum_{(i,j)\in I_1} s^{i-1*}_{i,j} + s^{i-1*}_{j,n+1} + s^{j-1*}_{n+1,j} + \\ &\sum_{(i,j)\in I_2} s^{i-1*}_{i,n+1} + s^{j-1*}_{n+1,i} + s^{j-1*}_{i,j} + \\ &\sum_{(i,j)\in I_3} s^{i-1*}_{i,j} + s^{i-1*}_{j,n+1} + s^{j-1*}_{n+1,i} + s^{j-1*}_{i,j}.\end{aligned} \tag{62}$$

To complete the proof, we show that the r.h.s of (62) is less than, or equal to

$$\sum_{\ell=1}^{n} \sum_{\substack{i,j\in[n+1],\ i<j \\ \ell\notin\{i,j\}}} s^{\ell*}_{i,j}. \tag{63}$$

To establish this, we show that each term on the r.h.s of (62) is: (i) not repeated; and (ii) is included in (63).

Definition 10. *We call two P-scores, $s^{c_1*}_{a_1,b_1}$ and $s^{c_2*}_{a_2,b_2}$,* coincident*, and denote it by $s^{c_1*}_{a_1,b_1} \sim s^{c_2*}_{a_2,b_2}$, 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 (62). Checking (ii) amounts to verifying that for each P-score $s^{c_1*}_{a_1,b_1}$ on the r.h.s. of (62), there exists a P-score $s^{c_2*}_{a_2,b_2}$ in (63) such that $s^{c_1*}_{a_1,b_1} \sim s^{c_2*}_{a_2,b_2}$.

Note that the r.h.s of (62) 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 (62). We have $s^{i-1*}_{i,j} \nsim s^{i-1*}_{j,n+1}$ because $i \in [n]$ and therefore $i \neq n+1$. We have $s^{i-1*}_{i,j} \nsim s^{j-1*}_{n+1,j}$ because $i-1 \neq j-1$ in this case, since $i < j$. We can similarly infer that $s^{i-1*}_{j,n+1} \nsim s^{j-1*}_{n+1,j}$.

Now consider the second summation on the r.h.s of (62). Taking the definition of I_2 and (56) into account, we can rewrite this summation as,

$$s^{n*}_{1,n+1} + s^{n-1*}_{n+1,1} + s^{n-1*}_{1,n}. \tag{64}$$

Since $n \neq n-1$, we have $s^{n*}_{1,n+1} \nsim s^{n-1*}_{n+1,1}$, and $s^{n*}_{1,n+1} \nsim s^{n-1*}_{1,n}$. Also, since $n \neq n+1$ we have $s^{n-1*}_{n+1,1} \nsim s^{n-1*}_{1,n}$.

Finally, consider the third summation on the r.h.s of (62). 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 $s^{i-1*}_{i,j} \nsim s^{i-1*}_{j,n+1}$, and $s^{j-1*}_{n+1,i} \nsim s^{j-1*}_{i,j}$, 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 (62), 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 (62), 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 (62)) 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 (62) as

$$\begin{aligned}&\sum_{(i,j)\in I_1} s^{i-1*}_{i,j} + s^{i-1*}_{j,n+1} + s^{j-1*}_{n+1,j} + \\ &\quad s^{n*}_{1,n+1} + s^{n-1*}_{n+1,1} + s^{n-1*}_{1,n} + \\ &\sum_{(i',j')\in I_3} s^{i'-1*}_{i',j'} + s^{i'-1*}_{j',n+1} + s^{j'-1*}_{n+1,i'} + s^{j'-1*}_{i',j'}.\end{aligned} \tag{65}$$

In what follows, we discuss the non-trivial comparisons, and refer to the first, second and third summations in (65) as Σ_1, Σ_2, and Σ_3, respectively.

1) $s^{i-1*}_{i,j}$ in Σ_1 vs. $s^{n-1*}_{1,n}$ 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) $s^{i-1*}_{i,j}$ in Σ_1 vs. $s^{j'-1*}_{i',j'}$ 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 $s^{i-1*}_{i,i+1}$. 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) $s^{i-1*}_{j,n+1}$ in Σ_1 vs. $s^{n*}_{1,n+1}$ 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) $s^{i-1*}_{j,n+1}$ in Σ_1 vs. $s^{n-1*}_{n+1,1}$ 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) $s^{i-1*}_{j,n+1}$ in S_1 vs. $s^{j'-1*}_{n+1,i'}$ in S_3: we can write the first term as $s^{i-1}_{i+1,n+1}$. 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) $s^{j-1*}_{n+1,j}$ in Σ_1 vs. $s^{n*}_{1,n+1}$ in Σ_2: the equality of superscripts requires $j = 1$, which is impossible since $j > i \in [n]$.

7) $s^{j-1*}_{n+1,j}$ in Σ_1 vs. $s^{n-1*}_{n+1,1}$ 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) $s^{j-1*}_{n+1,j}$ in Σ_1 vs. $s^{i'-1*}_{j',n+1}$ 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) $s^{j-1*}_{n+1,j}$ in Σ_1 vs. $s^{j'-1*}_{n+1,i'}$ 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) $s^{n*}_{1,n+1}$ in Σ_2 vs. $s^{i'-1*}_{j',n+1}$ 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) $s^{n*}_{1,n+1}$ in Σ_2 vs. $s^{j'-1*}_{n+1,i'}$ 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) $s^{n-1*}_{n+1,1}$ in Σ_2 vs. $s^{i'-1*}_{j',n+1}$ 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) $s^{n-1*}_{1,n}$ in Σ_2 vs. $s^{i'-1*}_{i',j'}$ 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 (65) are included in the summation in (63). To this aim, we will show that for each $s^c_{a,b}$ in (65), the indices $\{a, b, c\}$ satisfy

$$c \in [n], a, b \in [n+1] \backslash \{c\} \text{ and } a \neq b, \tag{66}$$

which is enough to prove that either $s^c_{a,b}$ or $s^c_{b,a}$ exist in (63).

We first note that the superscripts in (65) are in $[n]$, see Remark 3. Moreover, all the subscripts in (65) are either 1, $n+1$, or $i, j, i', j' \in [n]$. Thus, for any $s^c_{a,b}$ in (65), we have $a, b \in [n+1]$. Also note that, for any $s^c_{a,b}$ in (65), 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 $s^c_{a,b}$ in (65), $a \neq c$ and $b \neq c$.

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

For any term $s^c_{a,b}$, in the second summation, we clearly see in (65) that $a \neq c$ and $b \neq c$.

We now consider the last summation in (65). In the first term, $s^{i'-1*}_{i',j'}$, clearly $i' \neq i'-1$. Moreover, $i'-1 < i' < j'$, since $(i', j') \in I_3$. In the second term, $s^{i'-1*}_{j',n+1}$, $j' \neq i'-1$, because since $i'-1 < i' < j'$. Moreover, $n+1 \neq i'-1$ because otherwise $i' = n+2 \notin [n]$. In the third term, $s^{j'-1*}_{n+1,i'}$, 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, $s^{j'-1*}_{i',j'}$, we have $i' \neq j'-1$ since $(i', j') \in I_3$. Also, clearly $j' \neq j'-1$. $\square$

VIII. 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.

$$d(A_{1:n}) \leqslant C \times \sum_{i=1}^n d(A^i_{1:n,n+1}). \tag{67}$$

The authors in [25] prove that the d_F is a (pseudo) (C, n)-metric with $\frac{1}{n-1} \leqslant C \leqslant \frac{1}{\lfloor \frac{n}{2} \rfloor}$. Clearly, 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 $d_{\mathcal{G}}$ is a (pseudo) (C, n)-metric, and whether $C < 1$?

Acknowledgement

This work was partially funded by NIH/1U01AI124302 and NSF/IIS-1741129.

References

[1] M. Kalaev, M. Smoot, T. Ideker, and R. Sharan. Networkblast: comparative analysis of protein networks. *Bioinformatics*, 24(4):594–596, 2008.
[2] M. Zaslavskiy, F. Bach, and J. P. Vert. Global alignment of protein–protein interaction networks by graph matching methods. *Bioinformatics*, 25(12):i259–1267, 2009.
[3] B. P. Kelley, B. Yuan, F. Lewitter, R. Sharan, B. R. Stockwell, and T. Ideker. Pathblast: a tool for alignment of protein interaction networks. *Nucleic acids research*, 32(suppl_2):W83–W88, 2004.
[4] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *International journal of pattern recognition and artificial intelligence*, 18(03):265–298, 2004.
[5] W. Hu, Y. Qu, and G. Cheng. Matching large ontologies: A divide-and-conquer approach. *Data & Knowledge Engineering*, 67(1):140–160, 2008.
[6] J. Zhang and P. S. Yu. Multiple anonymized social networks alignment. In *Data Mining (ICDM), 2015 IEEE International Conference on*, pages 599–608. IEEE, 2015.
[7] R. Singh, J. Xu, and B. Berger. 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.
[8] D. Papo, J. M. Buldú, S. Boccaletti, and E. T. Bullmore. Complex network theory and the brain, 2014.
[9] D. Pachauri, R. Kondor, and V. Singh. Solving the multi-way matching problem by permutation synchronization. In *Advances in neural information processing systems*, pages 1860–1868, 2013.
[10] J. Douglas, B. Zimmerman, A. Kopylov, J. Xu, D. Sussman, and V. Lyzinski. Metrics for evaluating network alignment. 2018.
[11] J. Yan, J. Wang, H. Zha, X. Yang, and S. Chu. Consistency-driven alternating optimization for multigraph matching: A unified approach. *IEEE Transactions on Image Processing*, 24(3):994–1009, 2015.
[12] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 18(4):377–388, 1996.
[13] N. Hu, B. Thibert, and L. Guibas. Distributable consistent multi-graph matching. *arXiv preprint arXiv:1611.07191*, 2016.
[14] H. Park and K. Yoon. Encouraging second-order consistency for multiple graph matching. *Machine Vision and Applications*, 27(7):1021–1034, 2016.
[15] Q. Huang and L. Guibas. Consistent shape maps via semidefinite programming. In *Computer Graphics Forum*, volume 32, pages 177–186. Wiley Online Library, 2013.
[16] A. Solé-Ribalta and F. Serratosa. Models and algorithms for computing the common labelling of a set of attributed graphs. *Computer Vision and Image Understanding*, 115(7):929–945, 2011.
[17] M. L. Williams, R. C. Wilson, and E. R. Hancock. Multiple graph matching with bayesian inference. *Pattern Recognition Letters*, 18(11-13):1275–1281, 1997.
[18] S. Hashemifar, Q. Huang, and J. Xu. Joint alignment of multiple protein–protein interaction networks via convex optimization. *Journal of Computational Biology*, 23(11):903–911, 2016.

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

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

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

[22] J. Bento and S. Ioannidis. A Family of Tractable Graph Distances. *ArXiv e-prints*, January 2018.

[23] M. M. Fréchet. Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22(1):1–72, 1906.

[24] F. Hausdorff. *Mengenlehre*. Walter de Gruyter Berlin, 1927.

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

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

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

[28] J. Bento. A metric for sets of trajectories that is practical and mathematically consistent. *arXiv preprint arXiv:1601.03094*, 2016.

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

[30] M. M. Deza and E. Deza. Encyclopedia of distances. In *Encyclopedia of Distances*, pages 1–583. Springer, 2009.

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