Metrics in the Space of High Order Networks

Weiyu Huang and Alejandro Ribeiro

***Abstract*—This paper presents methods to compare high order networks. High order networks are collections of relationship functions between elements of tuples on some nodes spaces. They can be considered as generalizations of conventional networks where only relationship functions between pairs are defined. Important properties between relationships of tuples of different lengths are established when the relationships in high order networks specifically encode dissimilarities or proximities between nodes. Two families of distances are then presented in the space of high order networks. The distances measure differences between networks and are shown to be valid metrics in the spaces of high order dissimilarity and proximity networks modulo permutation isomorphisms. Practical implications are explored by comparing the coauthorship networks of two popular signal processing researchers. The metrics succeed in identifying their respective collaboration patterns.**

I. Introduction

We consider high order networks that describe relationships between elements of a tuple and address the problem of constructing valid metric distances between them. Most often, networks are defined as structures that describe interactions between pairs of nodes [2], [3]. This is an indisputable appropriate model for networks that describe binary relationships, such as communication or influence, but not so appropriate for problems in which binary, ternary, or n-ary relationships in general, have different implications. This is, e.g., true of coauthorship networks where we count the number of joint publications by groups of scholars. Papers written by pairs of authors capture information that can be used to identify important authors and study mores of research communities. However, there is extra information to be gleaned from collaborations between triplets of authors, or even single author publications. The importance of capturing tuple proximities between groups of nodes other than pairs has been recognized and exploited in multiple domains including coverage analysis in sensor networks [4]–[6], cognitive learning and memory [7], broadcasting in wireless networks [8], image ranking [9], three-dimensional object retrieval and recognition [10], and group relationship structure in social networks [11].

The problem of defining distances between networks, or, more loosely, the problem of determining if two networks are similar or not, is important even in the case of pairwise networks. The problem is not complicated if nodes have equal labels in both networks. One can visualize the networks and study the impact of some specific edges in different networks [12] or explore the relationships between some chosen nodes [13]. A more global approach to distinguish networks could be achieved by interpreting networks as Markov Chains by proper normalizations [14], [15]. The problem, however, becomes very challenging if a common labeling doesn't exist in both networks, as we need to consider all possible mappings between nodes of each network. This complexity has motivated the use of network features as alternatives to the use of distances. Examples of features that have proved useful in particular settings are clustering coefficients [16], neighborhood topology [17], betweenness [18], motifs [19], wavelets [20], graphlet degree distribution or signatures [21], [22], and later graphlet kernels [23]. Feature analysis is valuable, but it does not allow for meaningful comparisons unless application specific features are already known to be important. A different alternative is to define actual distances [24]. Because they have to consider node correspondences, network distances are computationally intractable. Their practical value is limited to small networks and to the transformation of the problem into one of building distance approximations instead of one of searching for appropriate features.

The main problem addressed in this paper is the construction of metric distances between high order networks. Formal definitions of high order networks are presented (Section III) as a generalization of pairwise networks (Section II). Dissimilarity networks (Section III-A) and proximity networks (Section III-B) are specific high order networks where relationship functions are intended to encode dissimilarities or proximities between members of tuples. Important properties as consequences of the restrictions are established. Two families of proper metric distances are then defined in the respective space of dissimilarity and proximity networks modulo permutation isomorphisms. These distances are build as generalizations of the pairwise distances in [24] which are themselves generalizations of the Gromov-Hausdorff distance between metric spaces [25], [26]. Dissimilarity networks may be expressed in terms of proximities and the transformation of expressions preserves the metrics defined (Section III-C). Similarly proximity networks can also be conveyed by dissimilarities. We use these distances to compare the coauthorship networks of two popular signal processing researchers and show that they succeed in discriminating their collaboration patterns (Section IV). As in the case of pairwise networks these distances can be computed only when the number of nodes is small. Ongoing work is focused on the problem of finding bounds on these network distances that are computable in networks with large numbers of nodes.

II. Pairwise Networks

Conventionally, a network is defined as a pair $N_X = (X, d_X^1)$, where X is a finite set of nodes and $d_X^1 : X^2 = X \times X \to \mathbb{R}_+$ is a function that may encode similarity or

Supported by NSF CCF-1217963 and AFOSR MURI FA9550-10-1-0567.
The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania, 200 South 33rd Street, Philadelphia, PA 19104. Email: whuang, aribeiro@seas.upenn.edu. Part of the results in this paper appeared in [1].

dissimilarity between elements. For points $x, x' \in X$, values of this function are denoted as $d_X^1(x, x')$. We assume that $d_X^1(x, x') = 0$ if and only if $x = x'$ and we further restrict attention to symmetric networks where $d_X^1(x, x') = d_X^1(x', x)$ for all pairs of nodes $x, x' \in X$. The set of all such networks is denoted as $\mathcal{N}$.

When defining a distance between networks we need to take into consideration that permutations of d_X^1 amount to relabelling nodes and must not be considered as different entities. We therefore say that two networks $N_X = (X, d_X^1)$ and $N_Y = (Y, d_Y^1)$ are isomorphic whenever there exists a bijection $\phi : X \to Y$ such that for all points $x, x' \in X$,

$$d_X^1(x, x') = d_Y^1(\phi(x), \phi(x')). \tag{1}$$

Such a map is called an isometry. Since the map ϕ is bijective, (1) can only be satisfied when d_X^1 is a permutation of d_Y^1. When networks are isomorphic we write $N_X \cong N_Y$. The space of networks where isomorphic networks $N_X \cong N_Y$ are represented by the same element is termed the set of networks modulo isomorphism and denoted by $\mathcal{N} \mod \cong$. The space $\mathcal{N} \mod \cong$ can be endowed with a valid metric [24]. The definition of this distance requires introducing the prerequisite notion of correspondence [27, Def. 7.3.17].

Definition 1 *A correspondence between two sets X and Y is a subset $C \subset X \times Y$ such that $\forall\ x \in X$, there exists $y \in Y$ such that $(x, y) \in C$ and $\forall\ y \in Y$ there exists $x \in X$ such that $(x, y) \in C$. The set of all correspondences between X and Y is denoted as $\mathcal{C}(X, Y)$.*

A correspondence in the sense of Definition 1 is a map between node sets X and Y so that every element of each set has a correspondent in the other set. Correspondences include permutations as particular cases but also allow for the mapping of a single point in X to multiple correspondents in Y or, vice versa, the mapping of multiple points in X to a single correspondent in Y. Most importantly, this allows definition of correspondences between networks with different numbers of elements. We can now define the distance between two networks by selecting the correspondence that makes them most similar as we formally define next.

Definition 2 *Given two networks $N_X = (X, d_X^1)$ and $N_Y = (Y, d_Y^1)$ and a correspondence C between the node spaces X and Y define the network difference with respect to C as*

$$\Gamma_{X,Y}^1(C) := \max_{(x_1,y_1),(x_2,y_2) \in C} \left| d_X^1(x_1, x_2) - d_Y^1(y_1, y_2) \right|, \tag{2}$$

The network distance between networks N_X and N_Y is then defined as

$$d_{\mathcal{N}}^1(N_X, N_Y) := \min_{C \in \mathcal{C}(X,Y)} \Big\{ \Gamma_{X,Y}^1(C) \Big\}. \tag{3}$$

For a given correspondence $C \in \mathcal{C}(X, Y)$ the network difference $\Gamma_{X,Y}^1(C)$ selects the maximum distance difference $|d_X^1(x_1, x_2) - d_Y^1(y_1, y_2)|$ among all pairs of correspondents – we compare $d_X^1(x_1, x_2)$ with $d_Y^1(y_1, y_2)$ when the points x_1 and y_1, as well as the points x_2 and y_2, are correspondents. The distance in (3) is defined by selecting the correspondence that minimizes these maximal differences. The distance in Definition 2 is a proper metric in the space of networks modulo isomorphism. It is nonnegative, symmetric, satisfies the triangle inequality, and is null if and only if the networks are isomorphic [24]. For future reference, the notions of metric and pseudometric are formally stated next.

Definition 3 *Given a space $\mathcal{S}$ and an isomorphism $\cong$, a function $d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_+$ is a metric in $\mathcal{S} \mod \cong$ if for any $a, b, c \in \mathcal{S}$ the function d satisfies:*

(i) **Nonnegativity.** $d(a, b) \geq 0$.
(ii) **Symmetry.** $d(a, b) = d(b, a)$.
(iii) **Identity.** $d(a, b) = 0$ *if and only if* $a \cong b$.
(iv) **Triangle inequality.** $d(a, b) \leq d(a, c) + d(c, b)$.

The function is a pseudometric in $\mathcal{S} \mod \cong$ if for any $a, b, c \in \mathcal{S}$ the function d satisfies (i), (ii), (iv), and

(iii') **Relaxed Identity.** $d(a, b) = 0$ *if* $a \cong b$.

A metric d in $\mathcal{S} \mod \cong$ gives a proper notion of distance. Since zero distances imply elements being isomorphic, the distance between elements reflects how far they are from being isomorphic. Pseudometrics are relaxed since elements not isomorphic may still have zero distance measured by the pseudometric. The distance in Definition 2 is a metric in $\mathcal{N} \mod \cong$. Observe that since correspondences may be between networks with different number of elements, Definition 2 defines a distance $d_{\mathcal{N}}^1(N_X, N_Y)$ when the node cardinalities $|X|$ and $|Y|$ are different. In the particular case when the functions d_X^1 satisfy the triangle inequality, the set of networks $\mathcal{N}$ reduces to the set of metric spaces $\mathcal{M}$. In this case the metric in Definition 2 reduces to the Gromov-Hausdorff (GH) distance between metric spaces. The distances $d_{\mathcal{N}}^1(N_X, N_Y)$ in (3) are valid metrics even if the triangle inequalities are violated by d_X^1 or d_Y^1 [24].

In this paper we consider higher order networks where the specification of functions $d_X^k : X^{k+1} \to \mathbb{R}_+$ are meant to encode similarity or dissimilarity between node $(k+1)$-tuples. The goal of this paper is to devise generalizations of Definition 2 to high order networks and to prove that they define valid metrics in the space of high order networks modulo isomorphism; see Definitions 7, 8, 10, and 11.

III. High Order Networks

A network of order K over the node space X is defined as a collection of $K + 1$ relationship functions $\{d_X^k : X^{k+1} \to \mathbb{R}_+\}_{k=0}^K$ from the space X^{k+1} of $(k + 1)$-tuples to the nonnegative reals,

$$N_X^K = \left(X, d_X^0, d_X^1, \ldots, d_X^K\right). \tag{4}$$

For point collections $x_{0:k} := (x_0, x_1, \ldots, x_k) \in X^{k+1}$, values of this function are denoted as $d_X^k(x_{0:k})$ and are intended to represent a measure of similarity or dissimilarity for members of the group. In particular, the zeroth order function d_X^0 encodes relative weights of different nodes and the first order function d_X^1 represents the pairwise information discussed in Section II. Observe however that pairwise networks are not particular cases of networks of order 1 because a network of order K not only requires the definition of relationships between $(K+1)$-tuples but also of relationships between $(k+1)$-tuples for all integers $0 \leq k \leq K$. A network of order 0 is one in which only node weights are given, a network of order 1 is one in which weights and pairwise relationships are defined, a network of order 2 adds relationships between triplets and so on. We assume that relationship values are normalized so that $0 \leq d_X^k(x_{0:k}) \leq 1$ for all k and $x_{0:k}$.

We restrict attention to symmetric networks for which dissimilarities are invariant to permutations. Formally, if we let $x_{[0:k]} = ([x_1], [x_2], \ldots, [x_k])$ be a reordering of $x_{0:k} := (x_0, x_1, \ldots, x_k)$ symmetric relationship functions are such that $d_X^k(x_{[0:k]}) = d_X^k(x_{0:k})$ for all point collections $x_{0:k}$. A symmetric K-order network is one in which all the $K+1$ functions d_X^k in (4) are symmetric. The set of all symmetric networks of order K is denoted as $\mathcal{N}^K$. As in the case of pairwise networks we consider K-order networks N_X^K and N_Y^K to be equivalent for their k-order relationship functions if d_X^k is a permutation of d_Y^k given k as a nonnegative integer less than or equal to K. Specifically, we say that two networks N_X^K and N_Y^K are k-isomorphic if there exists a bijection $\phi: X \to Y$ such that for all $x_{0:k} \in X^{k+1}$ we have

$$d_Y^k(\phi(x_{0:k})) = d_X^k(x_{0:k}), \tag{5}$$

where we use the shorthand notation $d_Y^k(\phi(x_{0:k})) := d_Y^k(\phi(x_1), \phi(x_2), \ldots, \phi(x_k))$. The map ϕ is called a k-isometry. When networks N_X^K and N_Y^K are k-isomorphic we write $N_X^K \cong_k N_Y^K$. The space of K-order networks modulo k-isomorphism is denoted by $\mathcal{N}^K \mod \cong_k$. For each nonnegative integer $0 \leq k \leq K$, the space $\mathcal{N}^K \mod \cong_k$ of networks of order K modulo k-isomorphism can be endowed with a pseudometric. The definition of this family of pseudometrics is a generalization of Definition 2 as we formally state next.

Definition 4 *Given networks N_X^K and N_Y^K, a correspondence C between the node spaces X and Y, and an integer $0 \leq k \leq K$ define the k-order network difference with respect to C as*

$$\Gamma_{X,Y}^k(C) := \max_{(x_{0:k}, y_{0:k}) \in C} \left| d_X^k(x_{0:k}) - d_Y^k(y_{0:k}) \right| \tag{6}$$

The k-order network distance between networks N_X^K and N_Y^K is then defined as

$$d_{\mathcal{N}}^k(N_X^K, N_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \left\{ \Gamma_{X,Y}^k(C) \right\}. \tag{7}$$

The distance vector between N_X^K and N_Y^K across all orders can also be defined as

$$\begin{aligned} &\mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K) \\ &= \left(d_{\mathcal{N}}^0(N_X^K, N_Y^K), d_{\mathcal{N}}^1(N_X^K, N_Y^K), \ldots, d_{\mathcal{N}}^K(N_X^K, N_Y^K) \right)^T. \end{aligned} \tag{8}$$

Both, Definition 2 and Definition 4 consider correspondences C that map the node space X onto the node space Y, compare dissimilarities, and set the network distance to the comparison that yields the smallest distance value in terms of maximum differences. The distinction between Definition 2 and (7) in Definition 4 is that in the latter $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ only considers d_X^k and d_Y^k out of $K+1$ relationship functions defined for the K-order networks. Definition 2 defines a single distance $d_{\mathcal{N}}(N_X, N_Y)$ between pairwise networks and (7) defines a family of $K+1$ pseudometrics $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ for each integer $0 \leq k \leq K$. Other than that the definition is not much different since $\Gamma_{X,Y}^k(C)$ selects the maximum k-order relationship difference $|d_X^k(x_{0:k}) - d_Y^k(y_{0:k})|$ among all tuples of correspondents – we compare $d_X^k(x_{0:k})$ with $d_Y^k(y_{0:k})$ when all the points $x_l \in x_{0:k}$ and $y_l \in y_{0:k}$ are correspondents. The distance $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ is defined by selecting the correspondence that minimizes these maximal differences. In general, the correspondence C minimizing $\Gamma_{X,Y}^k(C)$ is not necessarily identical as the correspondence C' minimizing $\Gamma_{X,Y}^l(C')$. The distance vector $\mathbf{d}_{\mathcal{N}}^K$ defined in (8) is a vector with each element measuring the dissimilarity between relationship functions of a specific order. We emphasize that, as in the case of Definition 2, $d_{\mathcal{N}}^k(N_X^K, N_Y^K)$ and $\mathbf{d}_{\mathcal{N}}^K(N_X^K, N_Y^K)$ are defined even if the numbers of nodes in X and Y are different. The function $d_{\mathcal{N}}^k : \mathcal{N}^K \times \mathcal{N}^K \to \mathbb{R}_+$ is a pseudometric in the space of K-order networks modulo k-isomorphism for any integer $0 \leq k \leq K$ as we show in the next proposition.

Proposition 1 *Given any nonnegative integer K, for any integers $0 \leq k \leq K$, the function $d_{\mathcal{N}}^k : \mathcal{N}^K \times \mathcal{N}^K \to \mathbb{R}_+$ defined in (7) is a pseudometric in the space $\mathcal{N}^K \mod \cong_k$.*

Proof: See Appendix A. ■

For each integer $0 \leq k \leq K$, the pseudometric $d_{\mathcal{N}}^k(D_X^K, D_Y^K)$ defined in Definition 4 in the space $\mathcal{N}^K \mod \cong_k$ measures dissimilarity between k-order functions d_X^k and d_Y^k. We can also ask the question how different two networks are by considering all their order functions. To that end we consider K-order networks to be equivalent if d_X^k is a permutation of d_X^k for all integers $0 \leq k \leq K$. Specifically, we say that two networks N_X^K and N_Y^K are isomorphic if there exists a bijection $\phi : X \to Y$ such that (5) follows for all $0 \leq k \leq K$ and $x_{0:k} \in X^{k+1}$. The map ϕ is called an isometry. When networks N_X^K and N_Y^K are isomorphic we write $N_X^K \cong N_Y^K$. The difference between isomorphism and k-isomorphism is that the bijection in the former preserves relationship functions over all orders whereas only preserves relationship functions for order k in the latter case. $N_X^K \cong N_Y^K$ implies $N_X^K \cong_k N_Y^K$ for all integers $0 \leq k \leq K$ however the other direction does not necessarily follows. The space of K-order networks modulo isomorphism is denoted by $\mathcal{N}^K \mod \cong$. A family of pseudometrics measuring the difference between networks over all order functions as a whole can be endowed in the space $\mathcal{N}^K \mod \cong$. The definition of this family of distances can be considered as an extension of Definition 2 and an aggregation of Definition 4 as we formally state next.

Definition 5 *Given networks N_X^K and N_Y^K, a correspondence C between the node spaces X and Y, and some p-norm $\|\cdot\|_p$*

define the network difference with respect to C as

$$\left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p := \left\|\left(\Gamma^0_{X,Y}(C), \Gamma^1_{X,Y}(C), \ldots, \Gamma^K_{X,Y}(C)\right)^T\right\|_p, \tag{9}$$

where for each integer $0 \leq k \leq K$, $\Gamma^k_{X,Y}(C)$ is the k-order network difference with respect to C defined in (6). *The network distance respect to the p-norm $\|\cdot\|_p$ between networks N^K_X and N^K_Y is then defined as*

$$d_{\mathcal{N},p}(N^K_X, N^K_Y) := \min_{C\in\mathcal{C}(X,Y)} \left\{\left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p\right\}. \tag{10}$$

The difference between Definition 2, Definition 4 and Definition 5 is that in the case of network distance $d_{\mathcal{N},p}(N^K_X, N^K_Y)$ we compare not only one relationship functions $d^k_X(x_{0:k})$ and $d^k_Y(y_{0:k})$ but also all the relationship functions of order not larger than K. The norm over the vector $\mathbf{\Gamma}^K_{X,Y}(C)$ formed by k-order network differences with respect to C for $0 \leq k \leq K$ is assigned as the difference between N^K_X and N^K_Y measured by the correspondence C. The distance $d_{\mathcal{N},p}(N^K_X, N^K_Y)$ is then defined as the minimum of these differences achieved by some correspondence. Similar as in the cases of Definition 2 and Definition 4, $d_{\mathcal{N},p}(N^K_X, N^K_Y)$ is defined even if the numbers of nodes in X and Y are different. The function $d_{\mathcal{N},p} : \mathcal{N}^K \times \mathcal{N}^K \to \mathbb{R}_+$ is a pseudometric in the space of K-order networks modulo isomorphism as we show in the following proposition.

Proposition 2 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{\mathcal{N},p} : \mathcal{N}^K \times \mathcal{N}^K \to \mathbb{R}_+$ defined in (10) is a pseudometric in the space $\mathcal{N}^K \mod \cong$.*

Proof: See Appendix B. ∎

Observe that in (10) we are only allowed to pick one correspondence minimizing $\|\mathbf{\Gamma}^K_{X,Y}(C)\|_p$ whereas in (7) for each k we are able to pick one correspondence minimizing the order specific $\Gamma^k_{X,Y}(C)$. This establishes a relationship between $d_{\mathcal{N},p}$ and $\|\mathbf{d}^K_{\mathcal{N}}\|_p$ and bridges a connection between Definition 4 and Definition 5 as next.

Proposition 3 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{\mathcal{N},p}$ defined in* (10) *is no smaller than $\|\mathbf{d}^K_{\mathcal{N}}\|_p$ where $\mathbf{d}^K_{\mathcal{N}}$ is defined in* (8). *I.e., for any K-order networks N^K_X, N^K_Y, we have the following relationship*

$$\begin{aligned} d_{\mathcal{N},p}(N^K_X, N^K_Y) &\geq \left\|\mathbf{d}^K_{\mathcal{N}}(N^K_X, N^K_Y)\right\|_p \\ &= \left\|\left(d^0_{\mathcal{N}}(N^K_X, N^K_Y), \ldots, d^K_{\mathcal{N}}(N^K_X, N^K_Y)\right)^T\right\|_p. \end{aligned} \tag{11}$$

Proof : From (7), for any K-order networks N^K_X, N^K_Y, a particular correspondence C between the node spaces X and Y, and an integer $0 \leq k \leq K$, it holds true that

$$\Gamma^k_{X,Y}(C) \geq d^k_{\mathcal{N}}(N^K_X, N^K_Y). \tag{12}$$

This implies the vector $\mathbf{d}^K_{\mathcal{N}}(N^K_X, N^K_Y)$ is elementwise no greater than $\mathbf{\Gamma}^K_{X,Y}(C)$. The property of p-norm guarantees that

$$\left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p \geq \left\|\mathbf{d}^K_{\mathcal{N}}(N^K_X, N^K_Y)\right\|_p. \tag{13}$$

Since (13) applies for any correspondence C, the minimum of $\left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p$ achieved by some correspondence in the set of correspondence $\mathcal{C}(X,Y)$ is still no smaller than $\left\|\mathbf{d}^K_{\mathcal{N}}(N^K_X, N^K_Y)\right\|_p$. This completes the proof. ∎

In general, the k-order function d^k_X of a given network N^K_X does not impose any constraint on the l-order function d^l_X of the same network. In practical situations, however, it is common to observe that adding nodes to tuple results in either increasing or decreasing relationships between the extended tuple. This motivates the definition and consideration of dissimilarity networks and proximity networks that we undertake in the next two sections.

A. Dissimilarity Networks

In dissimilarity networks the function $d^k_X(x_{0:k})$ encodes a level of dissimilarity between elements of the $x_{0:k}$ tuple. In this scenario it is reasonable to assume that adding elements to a tuple makes the group more dissimilar. This restriction along with a standard identity property makes up the formal definition that we introduce next.

Definition 6 *We say that the K-order network $D^K_X = \left(X, d^0_X, d^1_X, \ldots, d^K_X\right)$ is a dissimilarity network if the following two properties holds:*

Identity. *For any $0 \leq k \leq K$, $d^k_X(x_{0:k}) = 0$ if and only if all nodes in $x_{0:k}$ are identical; i.e., if and only if $x_i = x_j$ for all $x_i, x_j \in x_{0:k}$.*

Order increasing. *For any order $1 \leq k \leq K$ and tuples $x_{0:k} := (x_0, x_1, \ldots, x_k) \in X^{k+1}$ and $x_{0:k-1} := (x_0, x_1, \ldots, x_{k-1}) \in X^k$ it holds that*

$$d^k_X(x_{0:k}) \geq d^{k-1}_X(x_{0:k-1}). \tag{14}$$

The set of all dissimilarity networks of order K is denoted as $\mathcal{D}^k$.

In pairwise networks we required $d^k_X(x, x') = 0$ if and only if $x = x'$. The identity property in Definition 6 can be considered as a generalization. In pairwise dissimilarity networks dissimilarity 0 stands for most similarity and is reserved to represent the dissimilarity of a node to itself. In high order dissimilarity networks the highest similarity $d^k_X(x_{0:k}) = 0$ is also reserved to represent the closeness of a node to itself. Further note that since we restrict attention to symmetric networks a relationship akin to (14) holds if we remove an arbitrary element of the tuple $x_{0:k}$, not necessarily the last. Thus, the order increasing property implies that removing an element from a tuple can't make the group more dissimilar than it was. Equivalently, adding a node to a tuple makes the group as a whole more dissimilar or, at least, does not change the group's dissimilarity.

To see that the order increasing property in Definition 6 is reasonable consider a network describing the temporal dynamics of the formation of a research community – see Figure 1. The k-order dissimilarity function in this network marks the normalized time instant at which members of a given $(k+1)$-tuple write their first joint paper. In particular,

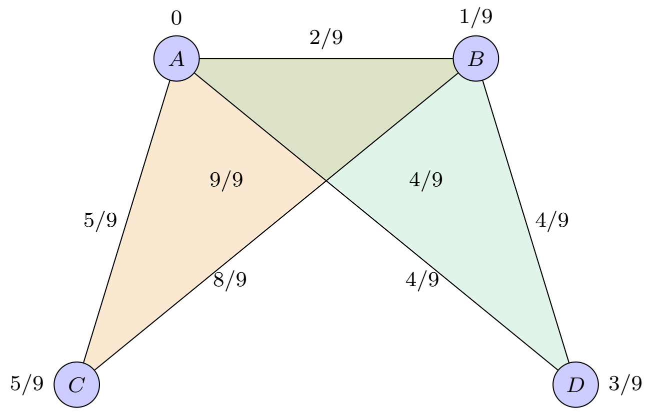


Fig. 1. Temporal dynamics for the formation of a research community. The k-order dissimilarity function in this 2-order network [cf. (4)] marks the normalized time instant at which members of a given $(k+1)$-tuple write their first joint paper. E.g., A writes her first paper at time 0, and coauthors with B, D, and C at times 1/9, 4/9, and 5/9. She also writes jointly with B and D at time 4/9 and with B and C at time 1. This is a dissimilarity network [cf. Definition 6] because the order increasing property follows from the fact that a paper can't be coauthored by three people without being at the same time coauthored by each of the three possible pairs of authors authors can't become coauthors until after they write their first paper.

the zeroth order dissimilarities d_X^0 are the normalized time instants when authors publish their first academic paper. In Figure 1 authors A, B, C, and D publish their first papers at times 0, 1/9, 3/9, and 5/9. The first order dissimilarities d_X^1 denote the normalized times at which nodes become coauthors. Since authors can't become coauthors until after they write their first paper it is certain that $d_X^1(x_0, x_1) \geq d_X^0(x_0)$ and $d_X^1(x_0, x_1) \geq d_X^0(x_1)$ for all x_0 and x_1. In Figure 1 A and B become coauthors at time $2/9$, which occurs after they publish their respective first papers at times 0 and $1/9$. Authors A and D as well as B and D become coauthors at time $4/9$, A and C become coauthors at time $5/9$ and B and C become coauthors at time $8/9$. Authors C and D never write a paper together. Observe that the first order dissimilarity between A and C is the same as the zeroth order dissimilarity of C which means that, most likely, the first paper that C writes is a joint work with A.

Second order dissimilarities d_X^2 for triplets $x_{0:2} = (x_0, x_1, x_2)$ denote the normalized time at which a paper is coauthored by the three members of the triplet. Since a paper can't be coauthored by three people without being at the same time coauthored by each of the three possible pairs of authors we must have that $d_X^2(x_0, x_1, x_2) \geq d_X^1(x_0, x_1)$, $d_X^2(x_0, x_1, x_2) \geq d_X^1(x_0, x_2)$, and $d_X^2(x_0, x_1, x_2) \geq d_X^1(x_1, x_2)$ for all x_0, x_1, and x_2. In Figure 1, authors A, B, and D publish a joint paper at time $4/9$, which is not smaller than the pairwise coauthorship times between each two of the individual authors. Notice that the second order distance between A, B, and D is the same as the first order distances between A, D and B, D, which is most likely due to the fact that the first papers coauthored by A, D and B, D is actually a joint paper by A, B, and D. Authors A, B, and C publish a joint paper at time 1, which is a time that comes after the individual paired publications that occur at times $4/9$, $5/9$, and $8/9$. Since distances up to order 2 are defined, the network in Figure 1 is a dissimilarity network of order 2.

Order increasing property is also reflected in the identity property. Observe that if the members of a nodes k-tuple $x_{l,l,\ldots,l}$ are all identical, identity property requires the dissimilarity in the group $d_X^k(x_{l,\ldots,l}) = 0$. On the other hand, adding any node $x_{l'} \neq x_l$ to the tuple makes not all the members of the tuple being identical and the identity property forces $d_X^{k+1}(x_{l,\ldots,l,l'}) > 0 = d_X^k(x_{l,\ldots,l})$. This also suggests that adding a node to a tuple makes the group as a whole more dissimilar or, at least, does not change the group's dissimilarity.

Restricting $d_{\mathcal{N}}^k$ defined in Definition 4 to dissimilarity networks gives a family of k-order dissimilarity network distance $d_{\mathcal{D}}^k$ which we define as follows.

Definition 7 *Given dissimilarity networks D_X^K and D_Y^K and an integer $0 \leq k \leq K$, the k-order dissimilarity network distance between dissimilarity networks D_X^K and D_Y^K is defined as*

$$d_{\mathcal{D}}^k(D_X^K, D_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \left\{ \Gamma_{X,Y}^k(C) \right\}, \tag{15}$$

where $\Gamma_{X,Y}^k(C)$ is the k-order network difference with respect to C defined in (6).

The distance vector between D_X^K and D_Y^K across all orders is then defined as

$$\begin{aligned} &\mathbf{d}_{\mathcal{D}}^K(D_X^K, D_Y^K) \\ &= \left(d_{\mathcal{D}}^0(D_X^K, D_Y^K), d_{\mathcal{D}}^1(D_X^K, D_Y^K), \ldots, d_{\mathcal{D}}^K(D_X^K, D_Y^K)\right)^T. \end{aligned} \tag{16}$$

Inherited from $d_{\mathcal{N}}^k$, the function $d_{\mathcal{D}}^k : \mathcal{D}^K \times \mathcal{D}^K \to \mathbb{R}_+$ is a pseudometric in the space of K-order dissimilarity networks modulo k-isomorphism. Moreover, for each nonnegative integer $1 \leq k \leq K$, restricting our attention on dissimilarity networks makes $d_{\mathcal{D}}^k$ a well-defined metric, not only pseudometric, in the space $\mathcal{D}^K \mod \cong_k$ of dissimilarity networks of order K modulo k-isomorphism. We show this in the following theorem.

Theorem 1 *Given any nonnegative integer K, for any positive integers $1 \leq k \leq K$, the function $d_{\mathcal{D}}^k : \mathcal{D}^K \times \mathcal{D}^K \to \mathbb{R}_+$ defined in (15) is a metric in the space $\mathcal{D}^K \mod \cong_k$. The function $d_{\mathcal{D}}^0 : \mathcal{D}^K \times \mathcal{D}^K \to \mathbb{R}_+$ defined in (15) is a pseudometric in the space $\mathcal{D}^K \mod \cong_0$.*

Proof: See Appendix C. ■

The caveat for $d_{\mathcal{D}}^0$ is because we may have two dissimilarity networks D_X^K and D_Y^K owning different number of nodes and the zeroth other dissimilarities d_X^0 and d_Y^0 being identical for any nodes in the two dissimilarity networks. In such scenarios, $d_{\mathcal{D}}^0(D_X^K, D_Y^K) = 0$ however two dissimilarity networks are not 0-isomorphic.

Restricting $d_{\mathcal{N},p}$ defined in Definition 5 to dissimilarity networks yields a family of dissimilarity network distances $d_{\mathcal{D},p}$ as we formally state the next.

Definition 8 *Given networks D_X^K and D_Y^K and some p-norm $\|\cdot\|_p$, the dissimilarity network distance respect to the p-norm between dissimilarity networks D_X^K and D_Y^K is defined as*

$$d_{\mathcal{D},p}(D_X^K, D_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \left\{ \left\| \mathbf{\Gamma}_{X,Y}^K(C) \right\|_p \right\}, \tag{17}$$

where the born $\|\mathbf{\Gamma}_{X,Y}^K(C)\|_p$ of network differences with respect to C is defined in (9).

By restricting our attention on dissimilarity networks instead of general high order networks, $d_{\mathcal{D},p}$ becomes a well-defined metric, not only pseudometric, in the space $\mathcal{D}^K \mod \cong$ of dissimilarity networks of order K modulo isomorphism. We state this in the following theorem.

Theorem 2 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{\mathcal{D},p} : \mathcal{D}^K \times \mathcal{D}^K \to \mathbb{R}_+$ defined in (17) is a metric in the space $\mathcal{D}^K \mod \cong$.*

Proof: See Appendix C. ■

The Definition 7 and Definition 8 may also be connected in a similar way as Proposition 3.

Proposition 4 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{\mathcal{D},p}$ defined in* (17) *is no smaller than $\|\mathbf{d}_{\mathcal{D}}^K\|_p$ where $\mathbf{d}_{\mathcal{D}}^K$ is defined in* (16). *I.e., for any K-order dissimilarity networks D_X^K, D_Y^K, we have the following relationship*

$$d_{\mathcal{D},p}(D_X^K, D_Y^K) \geq \left\| \mathbf{d}_{\mathcal{D}}^K(D_X^K, D_Y^K) \right\|_p. \tag{18}$$

B. Proximity Networks

In proximity networks the relationship functions $d_X^k(x_{0:k})$ denote similarity or proximity between elements of a tuple. Thus, large values of the proximity function $d_X^k(x_{0:k})$ represent strong relationship whereas small values denote weak relationships – the exact opposite is true of dissimilarity networks. In this framework it is reasoned to assume that adding elements to a tuple forces the group to be less similar. This constraint along with an identity property makes up the formal definition we introduce as follows.

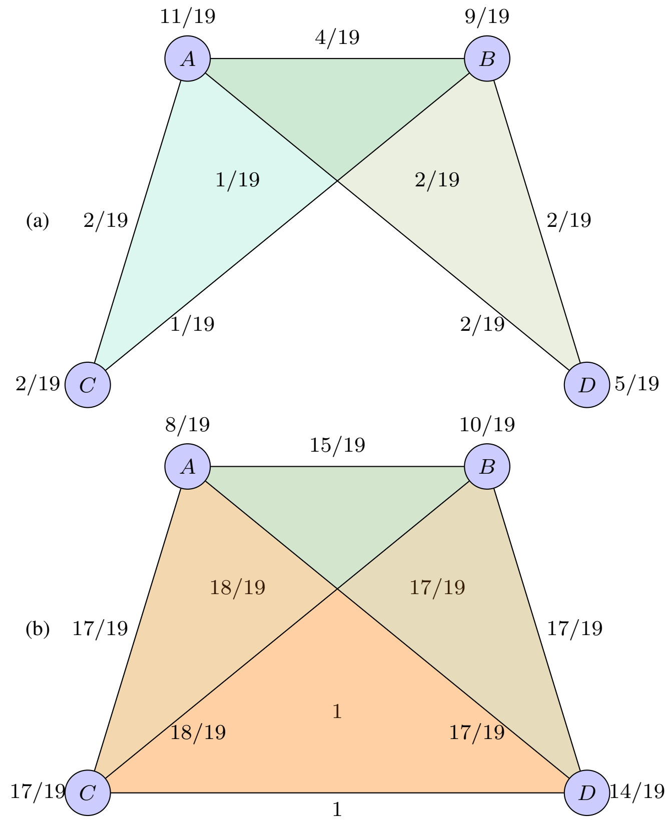


Fig. 2. (a): Collaborations between authors in a research community. The k-order proximity function in this 2-order network [cf. (4)] marks the number of publications between members of a given $(k+1)$-tuples normalized by the total number of papers. E.g., author A publishes 11 papers and there are 19 papers in total which implies $d_X^0(A) = 11/19$. Author A co-publishes 4, 1, and 2 papers with B, D, and C implying $d_X^1(A,B) = 4/19, d_X^1(A,D) = 1/19$, and $d_X^1(A,C) = 2/19$. Authors A, B and D co-publish 2 papers suggesting $d_X^2(A,B,D) = 2/19$. This is a proximity network [cf. Definition 9] because the order decreasing property follows from the fact that a paper collaborated by three authors is also a collaboration for each pair of authors and a paper written by a pair of authors is also a paper for each of the individuals. (b): Relationships between authors expressed in terms of dissimilarities constructed from the proximity network in (a). The k-order relationship function in this 2-order network denotes the level of dissimilarities between members of a given $(k+1)$-tuples. This is a dissimilarity network because the order increasing property follows. The constructed dissimilarity network has same order and identical node sets as the proximity network.

Definition 9 *We say that the K-order network $P_X^K = \left(X, d_X^0, d_X^1, \ldots, d_X^K\right)$ is a proximity network if the following two properties holds:*

Identity. *For any $0 \leq k \leq K$, $d_X^k(x_{0:k}) = 1$ if and only if all nodes in $x_{0:k}$ are identical; i.e., if and only if $x_i = x_j$ for all $x_i, x_j \in x_{0:k}$.*

Order decreasing. *For any order $1 \leq k \leq K$ and tuples $x_{0:k} \in X^{k+1}$ and $x_{0:k-1} \in X^k$ it holds that*

$$d_X^k(x_{0:k}) \leq d_X^{k-1}(x_{0:k-1}). \tag{19}$$

The set of all proximity networks of order K is denoted as $\mathcal{P}^K$.

In dissimilarity networks we required $d_X^k(x_{0:k}) = 0$ if and only if all nodes in the tuple are identical. The identity property in Definition 9 can also be considered as a generalization. In dissimilarity networks dissimilarity 0 stands for most similarity and is reserved to represent the dissimilarity of a node to itself. In high order proximity networks the highest proximity $d_X^k(x_{0:k}) = 1$ is reserved to represent the closeness of a node to itself. We emphasize that, as in the case of Definition 6, we restrict to symmetric networks a relationship as in (19) holds if we remove an arbitrary node from the tuple $x_{0:k}$, not necessarily the last. Thus, the order decreasing property implies that removing an element from a tuple can't make the set less similar than it was. Alternatively speaking, adding a node to a tuple makes the set as a whole less similar or, at least, does not change the set's proximity.

To see that the order decreasing property in Definition 9 is reasoned consider a network illustrating the collaborations between authors in a research community – See Figure 2 (a). The k-order proximity function in this network labels the number of publications between members of a given $(k+1)$-tuples. In specific, the zeroth order proximities d_X^0 are the numbers of papers published by authors normalized by the total number of papers. In Figure 2 (a) authors A, B, C, D publish $11, 9, 2, 5$ papers in total and there are 19 papers in total which implies $d_X^0(A) = 11/19$, $d_X^0(B) = 9/19$, $d_X^0(C) = 2/19$, $d_X^0(D) = 5/19$. The first order proximities d_X^1 represent the number of papers co-published by nodes. Since collaboration for a pair of authors is also a paper for each of the individuals it is certain that $d_X^1(x_0, x_1) \leq d_X^0(x_0)$ and $d_X^1(x_0, x_1) \leq d_X^0(x_1)$ for all x_0 and x_1. In Figure 9 (a) A and B collaborate 4 papers, which is less than the 11 and 9 papers written by each of the individuals. Authors A and C as well as A and D as well as B and D collaborate 2 papers in total, B and C collaborate 1 paper. Authors C and D never write a paper together. Notice that the first order proximity between A and C is the same as the zeroth order proximity of C, which is mostly likely due to the fact that C spends most of her time working with A.

Second order proximities d_X^2 for triplets $x_{0:2}$ indicates the number of papers coauthored by the three members of the triplet. Since a paper with three authors is also a collaboration for the three possible pairs of authors we must have that $d_X^2(x_0, x_1, x_2) \leq d_X^1(x_0, x_1)$, $d_X^2(x_0, x_1, x_2) \leq d_X^1(x_0, x_2)$, and $d_X^2(x_0, x_1, x_2) \leq d_X^1(x_1, x_2)$ for all x_0, x_1, and x_2. In Figure 2 (a), authors A, B, and D cowrite 2 paper, which is no more than the number of pairwise collaborations between each pair of the authors. Observe that the second order proximity between A, B, and D is the same as the first order proximities between A, D and B, D which means that, most likely, the 2 papers coauthored by A, D and B, D are actually joint papers by A, B, and D. Authors A, B, C publish 1 joint paper, a number smaller than the individual paired publications of 4, 2, and 1. Since proximities up to order 2 are defined, the network in Figure 2 (a) is a proximity network of order 2.

Restricting $d_{\mathcal{N}}^k$ defined in Definition 4 to proximity networks gives a family of k-order proximity network distance $d_{\mathcal{P}}^k$ as we formally state next.

Definition 10 *Given proximity networks P_X^K and P_Y^K and an integer $0 \leq k \leq K$, the k-order proximity network distance between proximity networks P_X^K and P_Y^K is defined as*

$$d_{\mathcal{P}}^k(D_X^K, D_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \left\{ \Gamma_{X,Y}^k(C) \right\}, \tag{20}$$

where $\Gamma_{X,Y}^k(C)$ is the k-order network difference with respect to C defined in (6). *The distance vector between P_X^K and P_Y^K across all orders is then defined as*

$$\begin{aligned} &\mathbf{d}_{\mathcal{P}}^K(P_X^K, P_Y^K) \\ &= \left(d_{\mathcal{P}}^0(P_X^K, P_Y^K), d_{\mathcal{P}}^1(P_X^K, P_Y^K), \ldots, d_{\mathcal{P}}^K(P_X^K, P_Y^K) \right)^T. \end{aligned} \tag{21}$$

Similar as $d_{\mathcal{D}}^k$ defined in Definition 7, for each nonnegative integer $1 \leq k \leq K$, the function $d_{\mathcal{P}}^k : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ is a proper metric, not only pseudometric, in the space $\mathcal{P}^K \mod \cong_k$ of proximity networks of order K modulo k-isomorphism. We show this in the following theorem.

Theorem 3 *Given any nonnegative integer K, for any positive integers $1 \leq k \leq K$, the function $d_{\mathcal{P}}^k : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ defined in* (20) *is a metric in the space $\mathcal{P}^K \mod \cong_k$. The function $d_{\mathcal{P}}^0 : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ defined in (20) is a pseudometric in the space $\mathcal{P}^K \mod \cong_0$.*

Proof: See Appendix D. ■

Similar as in Theorem 1, the caveat for $d_{\mathcal{P}}^0$ is because two proximity networks P_X^K and P_Y^K may possess different number of nodes while the zeroth other dissimilarities d_X^0 and d_Y^0 are identical for any nodes in the two proximity networks. In such scenarios, $d_{\mathcal{P}}^0(P_X^K, P_Y^K) = 0$ however two proximity networks are not 0-isomorphic.

Restricting $d_{\mathcal{N},p}$ defined in Definition 5 to proximity networks gives us a family of proper proximity network metrics $d_{\mathcal{P},p}$ as we formally state next.

Definition 11 *Given networks P_X^K and P_Y^K and some p-norm $\|\cdot\|_p$, the proximity network distance respect to the p-norm between proximity networks P_X^K and P_Y^K is defined as*

$$d_{\mathcal{P},p}(P_X^K, P_Y^K) := \min_{C \in \mathcal{C}(X,Y)} \left\{ \left\| \mathbf{\Gamma}_{X,Y}^K(C) \right\|_p \right\}, \tag{22}$$

where the norm $\|\mathbf{\Gamma}_{X,Y}^K(C)\|_p$ of network differences with respect to C is defined in (9).

Theorem 4 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{\mathcal{P},p} : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ defined in (22) is a metric in the space $\mathcal{P}^K \mod \cong$.*

Proof: See Appendix D. ■

A relationship between the Definition 10 and Definition 11 can be established in a similar way as Proposition 3.

Proposition 5 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K the function $d_{\mathcal{P},p}$ defined in* (22) *is no smaller than $\|\mathbf{d}_{\mathcal{P}}^K\|_p$ where $\mathbf{d}_{\mathcal{P}}^K$ is defined in* (21). *I.e., for any K-order proximity networks P_X^K, P_Y^K, we have the following relationship*

$$d_{\mathcal{P},p}(P_X^K, P_Y^K) \geq \left\| \mathbf{d}_{\mathcal{P}}^K(P_X^K, P_Y^K) \right\|_p. \tag{23}$$

C. Connections between Dissimilarity Networks and Proximity Networks

In a proximity network the relationship functions encode a level of similarity between elements of the $x_{0:k}$ tuple. A corresponding dissimilarity network can be constructed with identical nodes as the proximity network, preserving the relationship functions between elements of the $x_{0:k}$ tuple however expressing the relationships as dissimilarities. Similarly, given a dissimilarity network, a corresponding proximity network can be constructed with identical nodes as the dissimilarity network and conveying the same relationship functions between elements in a tuple in terms of proximities. These two constructions bridges a connection between proximity networks and dissimilarity networks we formally state next.

Definition 12 *The dissimilarity network $D_{X'}^K$ constructed from a given K-order proximity network P_X^K is defined as*

$$\begin{aligned} D_{X'}^K = \big\{ &(X', d_{X'}^1, \dots, d_{X'}^K) \mid X' = X, \\ &d_{X'}^k(x_{0:k}) = 1 - d_X^k(x_{0:k}), \\ &\textit{for all integers } 0 \le k \le K, x_{0:k} \in X \big\}. \end{aligned} \quad (24)$$

The proximity network $P_{X'}^K$ constructed from a given K-order dissimilarity network D_X^K is defined as

$$\begin{aligned} P_{X'}^K = \big\{ &(X', d_{X'}^1, \dots, d_{X'}^K) \mid X' = X, \\ &d_{X'}^k(x_{0:k}) = 1 - d_X^k(x_{0:k}), \\ &\textit{for all integers } 0 \le k \le K, x_{0:k} \in X \big\}. \end{aligned} \quad (25)$$

For a given proximity network P_X^K, the constructed dissimilarity network $D_{X'}^K$ has same order and identical node sets as P_X^K. For any nodes $(k+1)$-tuples $x_{0:k} \in X$, their dissimilarity $d_{X'}^k(x_{0:k})$ is defined as 1 minus their proximity $d_X^k(x_{0:k})$. Order increasing property naturally follow for $D_{X'}^K$. Most importantly, if all nodes are identical in the sequence $x_{0:k}$, $d_X^k(x_{0:k}) = 1$, from which $d_{X'}^k(x_{0:k}) = 0$ follows. An illustration for the construction is presented in Figure 2 (b), where we construct the corresponding dissimilarity network for the coauthorship network considered in Figure 2 (a). Similarly, for a given dissimilarity network D_X^K, the constructed dissimilarity network $P_{X'}^K$ has same order and identical node sets as D_X^K. The identity and order decreasing properties holds true for $P_{X'}^K$. Both the k-order network distances and the network distance by considering relationship functions at all orders as a whole are preserved by the constructions. We show these in the following propositions.

Proposition 6 *Given two proximity networks P_X^K and P_Y^K, for any integer $0 \le k \le K$, the constructed dissimilarity networks $D_{X'}^K$ and $D_{Y'}^K$ built from* (24) *satisfy*

$$d_{\mathcal{D}}^k(D_{X'}^K, D_{Y'}^K) = d_{\mathcal{P}}^k(P_X^K, P_Y^K). \quad (26)$$

Similarly, given two dissimilarity networks D_X^K and D_Y^K, for any integer $0 \le k \le K$, the constructed proximity networks $P_{X'}^K$ and $P_{Y'}^K$ built from (25) *satisfy*

$$d_{\mathcal{P}}^k(P_{X'}^K, P_{Y'}^K) = d_{\mathcal{D}}^k(D_X^K, D_Y^K). \quad (27)$$

Proposition 7 *Given two proximity networks P_X^K and P_Y^K and some p-norm $\|\cdot\|_p$, the constructed dissimilarity networks $D_{X'}^K$ and $D_{Y'}^K$ built from* (24) *satisfy*

$$d_{\mathcal{D},p}(D_{X'}^K, D_{Y'}^K) = d_{\mathcal{P},p}(P_X^K, P_Y^K). \quad (28)$$

Similarly, given two dissimilarity networks D_X^K and D_Y^K and some p-norm $\|\cdot\|_p$, the constructed proximity networks $P_{X'}^K$ and $P_{Y'}^K$ built from (25) *satisfy*

$$d_{\mathcal{P},p}(P_{X'}^K, P_{Y'}^K) = d_{\mathcal{D},p}(D_X^K, D_Y^K). \quad (29)$$

Proof: See Appendix E. ■

Now the development of metrics in high order networks is complete. For general symmetric networks of order K, the pseudometric $d_{\mathcal{N}}^k$ in the space $\mathcal{N}^K \mod \cong_k$ measures differences between k-order relationship functions between networks for any integers $0 \le k \le K$. The pseudometric $d_{\mathcal{N},p}$ in the space $\mathcal{N}^K \mod \cong$ measures the differences between networks by considering all order functions given some p-norm. These two pseudometrics are related such that $\|\mathbf{d}_{\mathcal{N}}^K\|_p \le d_{\mathcal{N},p}$. Restricting our attention to dissimilarity networks makes $d_{\mathcal{D}}^k$ a metric in the space $\mathcal{D}^K \mod \cong_k$ for each integer $1 \le k \le K$ and $d_{\mathcal{D},p}$ a metric in the space $\mathcal{D}^K \mod \cong$. Similarly, considering only proximity networks produces $d_{\mathcal{P}}^k$ a metric in the space $\mathcal{P}^K \mod \cong_k$ for each integer $1 \le k \le K$ and $d_{\mathcal{P},p}$ a metric in the space $\mathcal{P}^K \mod \cong$. Dissimilarity networks and proximity networks may be transformed from one to the other while preserving both the k-order network distances and the network distance. The metrics defined and their relationships established in Section III are depicted in Figure 3.

IV. Comparison of Coauthorship Networks

We apply the metrics defined in Section III to compare second order coauthorship networks where relationship functions denote the number of publications of single authors, pairs of authors, and triplets of authors. These coauthorship networks are proximity networks because they satisfy the order decreasing property in Definition 9. Since both, Definition 10 and Definition 11, require searching over all possible correspondences between the node spaces, we can compute exact distances for networks with a small number of nodes only. Thus, we consider publications in the IEEE Transactions on Signal Processing (TSP) in the last decade but restrict attention to the collaboration networks of Prof. Georgios B. Giannakis (GG) of the University of Minnesota and Prof. Martin Vetterli (MV) of the École Polytechnique Fédérale de Lausanne. For each of the lead coauthors, GG and MV, we construct networks for the 2004-2008 and 2009-2013 quinquennia. These networks are referred to in the following as GG0408, GG0913, MV0408, and MV0913. For GG we also define networks for each of the biennia 2004-2005, 2006-2007, 2008-2009, 2010-2011, and 2012-2013. We denote these networks as GG0405 GG0607, GG0809, GG1011, and GG1213. Lists of publications are queried from the Engineering Village database [28].

For each of these lead authors we consider all of their TSP publications in the period of interest and construct proximity networks where the node space X is formed by the lead author and the respective set of coauthors. Zeroth order proximities are defined as the total number of publications of each member of the network, first order proximities as the number of papers coauthored by given pairs, and second order proximities as the number of papers coauthored by specific triplets. To make networks with different numbers of papers comparable we normalize all distances by the total number of papers in the network. With this construction we have that the zeroth order proximity of GG or MV are 1 in all of their respective networks. There are papers with more than three coauthors but we don't record proximities of order larger than 2.

The quenquennial networks GG0408, GG0913, MV0408, and MV0913 are shown in Figure 4 and the biennial networks GG0607, GG0809, GG1011, and GG1213 in Figure 5. The

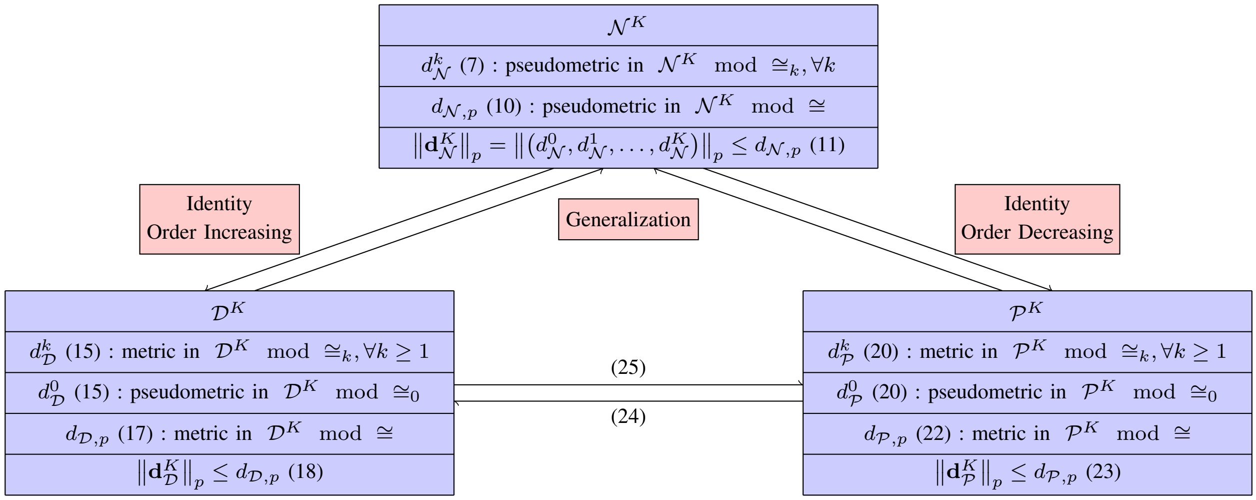


Fig. 3. Relationships between the spaces of high order networks, dissimilarity networks, and proximity networks. A family of pseudometrics can be defined to measure dissimilarities between a specific order functions between high order networks. Another family of pseudometrics can be defined to quantify distinctions between high order networks across all order functions. These two families of pseudometrics are related and become metrics in the corresponding spaces when we restrict attentions to dissimilarity networks or proximity networks.

size of the nodes is proportional to the zeroth order distances, and the width of the links to the first order distances. Second order proximities are represented by shading the triangle enclosed by the coauthor triplet and the color intensity is proportional to the second order proximities. There are clear differences in the collaboration patterns. We show here that proximity network distances succeed in identifying these patterns and distinguish between the coauthorship networks of GG and MV.

A. *Quinquennial networks*

Heat map representations of the k-order proximity network distances $d^k_{\mathcal{P}}$ for $k \in \{0, 1, 2\}$ and the proximity network distance with respect to the 1-norm, $d_{\mathcal{P},1}$, for the networks GG0408, GG0913, MV0408, and MV0913 are shown in Figure 6 (top). Two dimensional Euclidean embeddings of the same distances are also shown in Figure 6 (bottom). The two GG networks (diamonds) separate clearly from the two MV networks (circles) either by considering the individual k-order distances $d^k_{\mathcal{P}}$ or the aggregate distance $d_{\mathcal{P},1}$. The two MV networks do not group as clearly. Overall they are closer to each other than to the GG networks, but the difference is small. An unsupervised classification run across all four distances would assign all four networks correctly.

The k-order network distance $d^k_{\mathcal{P}}$ is defined by searching for a correspondence such that the maximum k-order proximity difference $|d^k_X(x_{0:k}) - d^k_Y(y_{0:k})|$ among all tuple of correspondents is minimized [cf. (6) and (7)]. For the optimal correspondence $C^\star = \operatorname{argmin}_{C \in \mathcal{C}(X,Y)} \Gamma^k_{X,Y}(C)$, define the pair of correspondent tuples that achieve the maximum k-order difference as

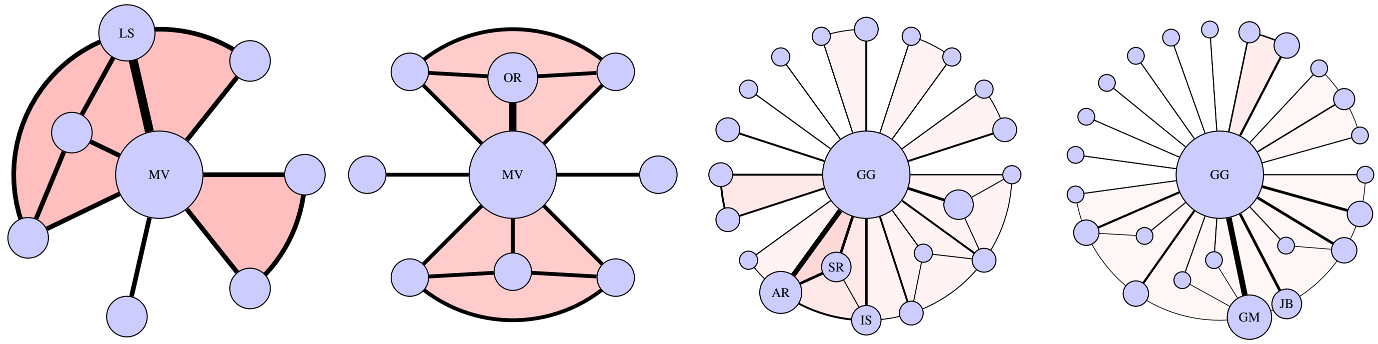
$$(x^\star_{0:k}, y^\star_{0:k}) = \operatorname*{argmax}_{(x_{0:k}, y_{0:k}) \in C^\star} \left| d^k_X(x_{0:k}) - d^k_Y(y_{0:k}) \right|. \quad (30)$$

The tuple pair $(x^\star_{0:k}, y^\star_{0:k})$ is the bottleneck that prevents making the networks closer to each other. Examining these bottleneck pairs for each k-order distance reveals what are the differences between proximity networks to which $d^k_{\mathcal{P}}$ is most sensitive about. In general, k-order bottleneck pairs tend to be pairs of tuples with high proximity values in their respective networks. Minimizing correspondences $C^\star$ map tuples with high proximity as closely as possible. Therefore, network distances are typically determined by large proximity values in one of the networks that can't be matched closely to proximity values in the other network.

In the coauthorship networks of networks of Figure 4 the bottleneck pair for 0-order distances $d^0_{\mathcal{P}}$, is formed by nodes with high zero order proximities and $d^0_{\mathcal{P}}$ reflects the difference between their zero order proximities. Since the networks are normalized so that the lead nodes have size 1, $d^0_{\mathcal{P}}$ is determined by their predominant coauthors, i.e., the scholars that collaborated most prolifically with GG or VM during the period of interest. The distances $d^0_{\mathcal{P}}$ between GG and VM networks are large because these predominant collaborations are different. In GG networks there are usually groups of 3 to 5 predominant collaborators, whereas in MV networks there are usually one or two that concentrate a larger fraction of the total number of publications.

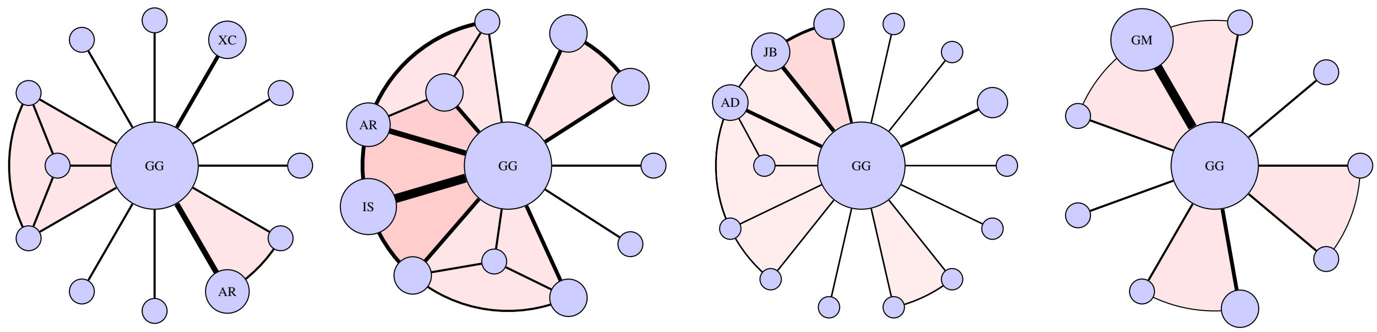
Similarly, high first order proximity distances between networks are likely due to one of the following situations: (i) Large differences between the numbers of papers authored by the predominant collaborators. (ii) Different patterns in the formation of communities – defined here as clusters of pairwise collaboration. In the latter case large distances arise because it is impossible to match the communities in one network to communities in the other. The distances $d^1_{\mathcal{P}}$ between GG and MV networks are large because the latter contain a smaller number of communities, which are also more strongly connected than the communities in GG networks.

In second order distances the bottleneck pair of triplets may reflect one of the following scenarios: (i) One network has collaboration between four or more authors while the



MV, 2004 - 2008 MV, 2004 - 2008 GG, 2004 - 2008 GG, 2009-2013

Fig. 4. Quinquennial coauthorship networks representing research communities centered at Prof. Georgios Giannakis (GG) or Prof. Martin Vetterli (MV). The size of the nodes is proportional to the zeroth order proximities, and the width of the links to the first order proximities. Second order proximities are represented by shading the triangle enclosed by the coauthor triplet. Color intensity is proportional to the second order proximities.



GG, 2006 - 2007 GG, 2008 - 2009 GG, 2010 - 2011 GG, 2012 - 2013

Fig. 5. Biennial coauthorship networks representing research communities centered at Prof. Georgios Giannakis (GG).

other doesn't (ii) There exist three authors with a strong collaboration between them in one network whereas in the other network there does not exist collaboration between three authors or, if such collaboration exists, it is weak. Many papers written by MV are collaborations of three or four scholars and the predominant coauthor in MV networks appears in at least one collaboration of four scholars. For GG, his 2004-2008 network has a few collaborations consisting of four scholars however all such collaborations are weak. His 2009-2013 network has no publications written by four authors.

B. Biennial networks

The networks GG0408 and GG0913 have more nodes than the networks MV0408 and MV0913 prompting the possibility that the differences in distances discussed in Section IV-B are just due their different number of publications. This is part of the reason, but not all. To see that this true we consider the biennial GG collaboration networks. Each of these networks contain numbers of papers that are comparable to the number of papers in the quinquennial MV networks.

The individual k-order distances $d_{\mathcal{P}}^k$ for $k \in \{0, 1, 2\}$ and the aggregate distance $d_{\mathcal{P},1}$ between the 4 quinquennial networks and the 5 biennial networks are represented in Figure 7 (top). Two dimensional Euclidean embeddings of these distances are shown in Figure 7 (bottom). An unsupervised classification run across all four distances would assign 6 networks correctly to GG and the other three networks to MV – one of them incorrectly.

We expect more variation in biennial networks because the time for averaging behavior is reduced. E.g., we may see deviations from usual collaboration patterns due to the presence of exceptional doctoral students. Still, three of the biennial networks, GG0405, GG0607, GG1011, (up triangles) and the two quinquennial networks GG0408, GG0913 (diamonds) are close to each other in every metric used and form a cluster clearly separate from the two five-year networks MV0408 and MV0913 (circles). This is due to the fact that the distinctive features of GG coauthorship are well reflected in GG0405, GG0607, GG1011. These features include: (i) Multiple predominant coauthors, each of whose collaboration with GG does not comprise a dominant portion of GG's scholarship during the period. (ii) Multiple small coauthorship communities in which strong collaborations within each community are rare. (iii) The number of publications with four or more authors is low. These features contrast with the rather opposite properties of the MV networks.

The networks GG0809 and GG1213 (down triangles) do not cluster with the other five GG networks. Depending on which distance we consider they may be closest to some of the other GG networks or to one of the two MV networks. This is because, likely due to random variation, GG0809 and GG1213 have some features that resemble GG networks and some other features that resemble MV networks. Fundamentally this

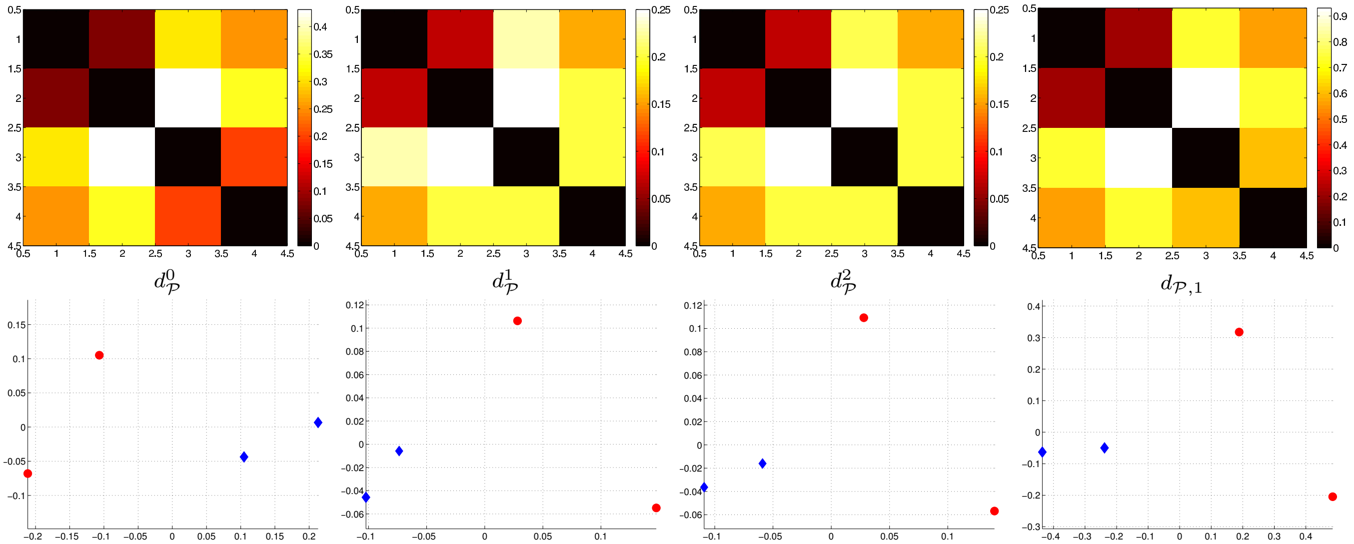


Fig. 6. Top: Heat map representations of the k-order proximity network distance $d_{\mathcal{P}}^0, d_{\mathcal{P}}^1, d_{\mathcal{P}}^2$ and the proximity network distance with respect to the 1-norm, $d_{\mathcal{P},1}$, for the quinquennial networks where the networks are ordered as GG0408, GG0913, MV0408, MV0913. Bottom: two dimensional Euclidean embeddings of the distances between quinquennial networks. In the embeddings, denote MV0408, MV0913 as circles, GG0408, GG0913 as diamonds.

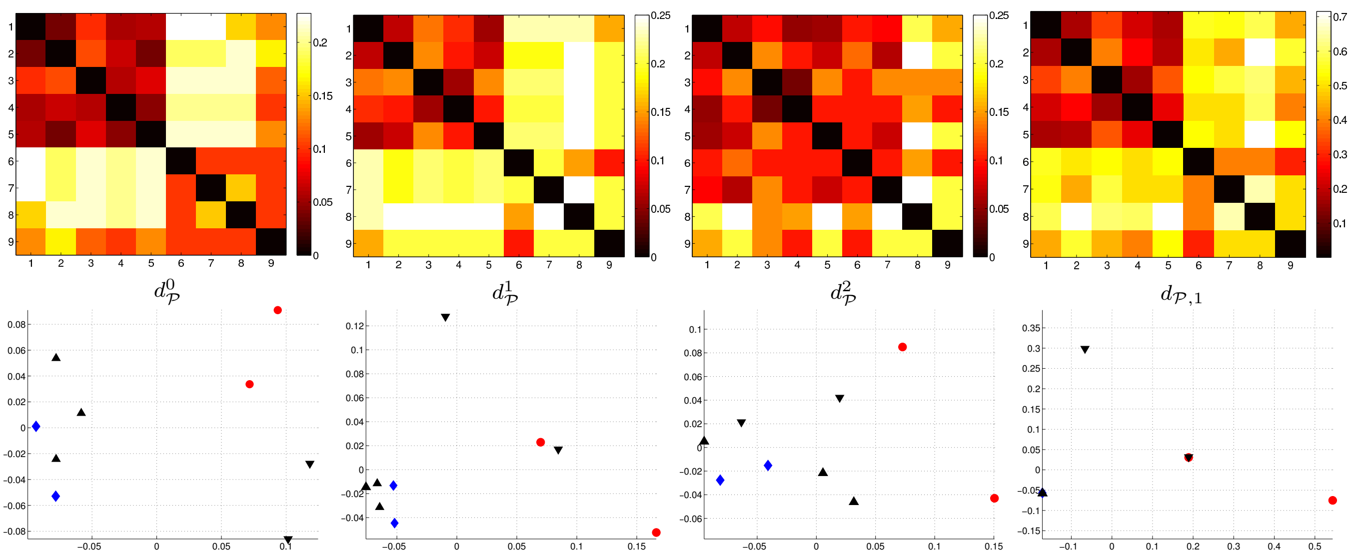


Fig. 7. Top: Heat map representations of the k-order proximity network distance $d_{\mathcal{P}}^0, d_{\mathcal{P}}^1, d_{\mathcal{P}}^2$ and the proximity network distance with respect to the 1-norm, $d_{\mathcal{P},1}$, for the biennial and quinquennial networks where the networks are ordered as GG0408, GG0913, GG0405, GG0607, GG1011, GG0809, GG1213, MV0408, MV0913. Bottom: two dimensional Euclidean embeddings of the distances between networks. In the embeddings, denote MV0408, MV0913 as circles, GG0408, GG0913 as diamonds, GG0405, GG0607, GG1011 as up triangles and GG0809, GG1212 as down triangles.

happens because of exceptionally prolific collaborations with Ioannis Schizas (IS) in the 2008-2009 period and Gonzalo Mateos (GM) in the 2012-2013 period. In the network GG0809 the IS node commands a significant fraction of GG publications and creates strong links between collaboration clusters that would be otherwise separate. Both of these features are more characteristic of MV networks. In the GG1213 network the node GM accounts for half of the publications in which GG is an author. This is, also, a feature more representative of MV networks than of GG networks.

In summary, proximity network distances capture features of scholar collaboration that permit discerning networks of different authors even when we consider networks that have very different numbers of nodes. The zeroth order distance $d_{\mathcal{P}}^0$ responds primarily to the number of predominant coauthors and the proportion of collaboration between predominant coauthors and the central scholar. The first order distance $d_{\mathcal{P}}^1$ is mostly determined by the fraction of collaborations that involve predominant coauthors and the central scholar as well as the level and number of strong collaborations within each community in the group. The second order distance $d_{\mathcal{P}}^2$ is largely given by the existence, level, and number of collaborations between four or more scholars and the appearance of predominant coauthors in a collaboration between four or more scholars.

V. Conclusion

High order networks, as a generalization of conventional pair-wise networks, was introduced. Restricting the relationship functions between members of tuples to the most commonly used ones, dissimilarities and proximities, yields two specific subspaces of high order networks – dissimilarity networks and proximity networks. Properties arisen from such restrictions were discussed. We defined two families of distances measuring differences between dissimilarity networks and between proximity networks. These distances are valid metrics in the corresponding subspace of high order networks modulo isomorphism. We use these distances to successfully identify collaboration patterns of Prof. Georgios B. Giannakis and Prof. Martin Vetterli. Tractable approximations will be provided in forthcoming contributions.

Appendix A
Proof of Proposition 1

To prove that $d_{\mathcal{N}}^{k}$ for any integer $0 \leq k \leq K$ is a pseudometric in the space of K-order networks modulo k-isomorphism we prove the (i) nonnegativity, (ii) symmetry, (iii') relaxed identity, and (iv) triangle inequality properties in Definition 3.

Proof of nonnegativity property: For any integers $0 \leq k \leq K$, since the function $|d_X^k(x_{0:k}) - d_Y^k(y_{0:k})|$ is nonnegative the k-order network difference with respect to C as defined in (6) also is. The network distance must then satisfy $d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) \geq 0$ because it is a minimum of nonnegative numbers. ■

Proof of symmetry property: A correspondence $C \subset X \times Y$ with elements $c_i = (x_i, y_i)$ results in the same associations as the correspondence $\tilde{C} \subset Y \times X$ with element $\tilde{c}_i = (y_i, x_i)$. Thus, for any correspondence C and integers $0 \leq k \leq K$, we have a correspondence $\tilde{C}$ such that $\Gamma_{X,Y}^k(C) = \Gamma_{Y,X}^k(\tilde{C})$. It follows that the minima in (7) must coincide from where it follows that $d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) = d_{\mathcal{N}}^{k}(N_Y^K, N_X^K)$. ■

Proof of relaxed identity property: We need to show that for any integers $0 \leq k \leq K$ if N_X^K and N_Y^K are k-isomorphic we must have $d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) = 0$. To see that this is true recall that for k-isomorphic networks there exists a bijection $\phi : X \to Y$ that preserves distance functions at order k [cf. (5)]. Consider then the particular correspondence $C_\phi = \{(x, \phi(x)), x \in X\}$. For all $x_0 \in X$ there is an element $c = (x_0, y) \in C_\phi$ and for all $y_0 \in Y$ there is an element $c' = (x, y_0) \in C_\phi$ since ϕ is bijective. Thus C_ϕ is a valid correspondence between X and Y for which (5) indicates that it must be

$$d_Y^k(y_{0:k}) = d_Y^k(\phi(x_{0:k}) = d_X^k(x_{0:k}), \tag{31}$$

for any $(x_{0:k}, y_{0:k}) \in C_\phi$. This implies $\Gamma_{X,Y}(C) = \left|d_X^k(x_{0:k}) - d_Y^k(y_{0:k})\right| = 0$ for any $(x_{0:k}, y_{0:k}) \in C_\phi$. Since C_ϕ is a particular correspondence, taking a minimum over all correspondence in (7) yields

$$d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) \leq \Gamma_{X,Y}^k(C) = 0. \tag{32}$$

Since $d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) \geq 0$, as already shown, it must be that $d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) = 0$ when two dissimilarity networks N_X^K and N_Y^K are k-isomorphic. ■

Proof of triangle inequality: To show that the triangle inequality holds, let the correspondence C_1 between X and Z and the correspondence C_2 between Z and Y be the minimizing correspondences in (7). We can then write

$$\begin{aligned} d_{\mathcal{N}}^{k}(N_X^K, N_Z^K) = \Gamma_{X,Z}^k(C_1). \\ d_{\mathcal{N}}^{k}(N_Z^K, N_Y^K) = \Gamma_{Z,Y}^k(C_2). \end{aligned} \tag{33}$$

Define a correspondence C between X and Y as the one induced by pairs (x, z) and (z, y) sharing a common node $z \in Z$,

$$C := \{(x, y) \mid \exists z \in Z \text{ with } (x, z) \in C_1, (z, y) \in C_2\}. \tag{34}$$

To show that C is a well defined correspondence we need to show that for every $x \in X$ there exists $y_0 \in Y$ such that $(x, y_0) \in C$ and by symmetry for every $y \in Y$ there exists $x_0 \in Y$ such that $(x_0, y) \in C$. To see this, first pick an arbitrary $x \in X$. Because C_1 is a correspondence between X and Z there must exist $z_0 \in Z$ such that $(x, z_0) \in C_1$. There must exist $y_0 \in Y$ such that $(z_0, y_0) \in C_2$ since C_2 is also a correspondence between Y and Z. Therefore, there exists a pair $(x, y_0) \in T$ with $y_0 \in Y$ for any $x \in X$. The second part follows by symmetry and C is a well defined correspondence. The correspondence C may not be the minimizing correspondence for the distance $d_{\mathcal{N}}^{k}(N_X^K, N_Y^K)$. However since it is a valid correspondence with the definition in (7) we can write

$$d_{\mathcal{N}}^{k}(N_X^K, N_Y^K) \leq \Gamma_{X,Y}^k(C) \tag{35}$$

By the definition of C in (34), the requirement $(x_{0:k}, y_{0:k}) \in C$ is equivalent as $(x_{0:k}, z_{0:k}) \in C_1$ and $(z_{0:k}, y_{0:k}) \in C_2$ for any $0 \leq k \leq K$. Further adding and subtracting $d_Z^k(z_{0:k})$ in the absolute value of $\Gamma_{X,Y}^k(C) = \left|d_X^k(x_{0:k}) - d_Y^k(y_{0:k})\right|$ and using the triangle inequality of the absolute value yields

$$\begin{aligned} \Gamma_{X,Y}^k(C) \leq \max_{\substack{(x_{0:k}, z_{0:k}) \in C_1 \\ (z_{0:k}, y_{0:k}) \in C_2}} & \Big\{ \left|d_X^k(x_{0:k}) - d_Z^k(z_{0:k})\right| \\ & + \left|d_Z^k(z_{0:k}) - d_Y^k(y_{0:k})\right| \Big\}. \end{aligned} \tag{36}$$

We can further bound (36) by taking maximum over each summand,

$$\begin{aligned} \Gamma_{X,Y}^k(C) \leq & \max_{(x_{0:k}, z_{0:k}) \in C_1} \left|d_X^k(x_{0:k}) - d_Z^k(z_{0:k})\right| \\ & + \max_{(z_{0:k}, y_{0:k}) \in C_2} \left|d_Z^k(z_{0:k}) - d_Y^k(y_{0:k})\right| \\ & = \Gamma_{X,Z}^k(C_1) + \Gamma_{Z,Y}^k(C_2). \end{aligned} \tag{37}$$

Substituting (35) and (33) back into (37) yields the triangle inequality. ■

APPENDIX B
PROOF OF PROPOSITION 2

To prove that $d_{\mathcal{N},p}$ is a distance in the space of K-order dissimilarity networks modulo isomorphism we prove the (i) nonnegativity, (ii) symmetry, (iii') relaxed identity, and (iv) triangle inequality properties in Definition 3.

Proof of nonnegativity property: Since the norm $\|\mathbf{\Gamma}^K_{X,Y}(C)\|_p$ is nonnegative, the network distance must then satisfy $d_{\mathcal{N},p}(N^K_X, N^K_Y) \geq 0$ because it is a minimum of nonnegative numbers. ■

Proof of symmetry property: A correspondence $C \subset X \times Y$ with elements $c_i = (x_i, y_i)$ results in the same associations as the correspondence $\tilde{C} \subset Y \times X$ with element $\tilde{c}_i = (y_i, x_i)$. Thus, for any correspondence C we have a correspondence $\tilde{C}$ such that $\mathbf{\Gamma}^K_{X,Y}(C) = \mathbf{\Gamma}^K_{Y,X}(\tilde{C})$. This implies their p-norms are also the same $\|\mathbf{\Gamma}^K_{X,Y}(C)\|_p = \|\mathbf{\Gamma}^K_{Y,X}(\tilde{C})\|_p$. It follows that the minima in (10) must coincide from where it follows that $d_{\mathcal{N},p}(N^K_X, N^K_Y) = d_{\mathcal{N},p}(N^K_Y, N^K_X)$. ■

Proof of relaxed identity property: We need to show that if N^K_X and N^K_Y are isomorphic we must have $d_{\mathcal{N},p}(N^K_X, N^K_Y) = 0$. To see that this is true recall that for isomorphic networks there exists a bijection $\phi : X \to Y$ that preserves distance functions at every order [cf. (5)]. Consider then the particular correspondence $C_\phi = \{(x, \phi(x)), x \in X\}$. We have demonstrated in Appendix A that C_ϕ is a valid correspondence between X and Y. The definition of isomorphism indicates that it must be (31) holds true for all $0 \leq k \leq K$ and $(x_{0:k}, y_{0:k}) \in C_\phi$. Since C_ϕ is a particular correspondence, taking a minimum over all correspondence in (7) yields

$$d_{\mathcal{N},p}(N^K_X, N^K_Y) \leq \left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p. \tag{38}$$

Because $d^k_X(x_{0:k}) - d^k_Y(y_{0:k}) = 0$ for any $0 \leq k \leq K$ and any $(x_{0:k}, y_{0:k}) \in C_\phi$ by the first equality in (31),

$$\mathbf{\Gamma}^K_{X,Y}(C) = \mathbf{0}. \tag{39}$$

$\|\cdot\|_p$ is a proper norm implies $\|\mathbf{\Gamma}^K_{X,Y}(C)\|_p = 0$. Substituting this back into (38) shows $d_{\mathcal{N},p}(N^K_X, N^K_Y) \leq 0$. Since $d_{\mathcal{N},p}(N^K_X, N^K_Y) \geq 0$, as already shown, it must be that $d_{\mathcal{N},p}(N^K_X, N^K_Y) = 0$ when two dissimilarity networks N^K_X and N^K_Y are isomorphic. ■

Proof of triangle inequality: To show that the triangle inequality holds, let the correspondence C_1 between X and Z and the correspondence C_2 between Z and Y be the minimizing correspondences in (10). We can then write

$$\begin{aligned} d_{\mathcal{N},p}(N^K_X, N^K_Z) &= \left\|\mathbf{\Gamma}^K_{X,Z}(C_1)\right\|_p. \\ d_{\mathcal{N},p}(N^K_Z, N^K_Y) &= \left\|\mathbf{\Gamma}^K_{Z,Y}(C_2)\right\|_p. \end{aligned} \tag{40}$$

Define a correspondence C between X and Y in the same way as (34). We have demonstrated in Appendix A that C is a well defined correspondence. Therefore with the definition in (10) we can write

$$d_{\mathcal{N},p}(N^K_X, N^K_Y) \leq \left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p \tag{41}$$

Moreover, in Appendix A we also showed for any $0 \leq k \leq K$,

$$\Gamma^k_{X,Y}(C) \leq \Gamma^k_{X,Z}(C_1) + \Gamma^k_{Z,Y}(C_2). \tag{42}$$

This implies the vector $\mathbf{\Gamma}^K_{X,Z}(C_1) + \mathbf{\Gamma}^K_{Z,Y}(C_2)$ is elementwise no smaller than the vector $\mathbf{\Gamma}^K_{X,Y}(C)$. The definition of p-norm $\|\mathbf{x}\|_p = \left(\sum_{i=0}^n |x_i|^p\right)^{1/p}$ guarantees that the value of $\|\mathbf{x}\|_p$ is monotonically nondecreasing on each element x_i in $\mathbf{x} = (x_0, x_1, \ldots, x_n)^T$. Therefore,

$$\left\|\mathbf{\Gamma}^k_{X,Y}(C)\right\|_p \leq \left\|\mathbf{\Gamma}^k_{X,Z}(C_1) + \mathbf{\Gamma}^k_{Z,Y}(C_2)\right\|_p. \tag{43}$$

We can further bound (43) by using the triangle inequality of the p-norm,

$$\left\|\mathbf{\Gamma}^k_{X,Y}(C)\right\|_p \leq \left\|\mathbf{\Gamma}^k_{X,Z}(C_1)\right\|_p + \left\|\mathbf{\Gamma}^k_{Z,Y}(C_2)\right\|_p. \tag{44}$$

Substituting (41) and (40) back into (44) yields the triangle inequality. ■

APPENDIX C
PROOF OF THEOREMS IN SECTION III-A

Proof of Theorem 1: The proof in Appendix A has demonstrated that $d^k_{\mathcal{D}}$ for any integer $0 \leq k \leq K$ is a pseudometric in the space of K-order dissimilarity networks modulo k-isomorphism. To prove that $d^k_{\mathcal{D}}$ for any integer $1 \leq k \leq K$ is a metric in the space of K-order dissimilarity networks modulo k-isomorphism we need to show the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: We want to prove that having $d^k_{\mathcal{D}}(D^K_X, D^K_Y) = 0$ must imply that D^K_X and D^K_Y are k-isomorphic. If $d^k_{\mathcal{D}}(D^K_X, D^K_Y) = 0$, there exists a correspondence C_0 such that $d^k_X(x_{0:k}) = d^k_Y(y_{0:k})$ for any $(x_{0:k}, y_{0:k}) \in C_0$. Define a function $\phi : X \to Y$ that associates x with an arbitrary y chosen from the set that form a pair with x in C_0,

$$\phi : x \mapsto y_0 \in \{y \mid (x, y) \in C_0\}. \tag{45}$$

Since C_0 is a correspondence the set $\{y \mid (x, y) \in C_0\}$ is nonempty for any x implying that ϕ is well-defined for any $x \in X$. Therefore $d^k_X(x_{0:k}) = d^k_Y(\phi(x_{0:k}))$ for any $x_{0:k} \in C$. This implies the function ϕ must be injective. If it were not, there would be a pair of nodes $x_l \neq x_{l'}$ with $\phi(x_l) = \phi(x_{l'}) = y \in Y$. Hence the k-order dissimilarity between $x_{l,l,\ldots,l,l'}$ where the first k nodes in the tuple are x_l and the last node is $x_{l'}$ would satisfy

$$d^k_X(x_{l,l,\ldots,l'}) = d^k_Y(\phi(x_{l,l,\ldots,l,l'})) = 0, \tag{46}$$

where the first equality follows from the definition of ϕ and the second equality is because of $\phi(x_l) = \phi(x_{l'}) = y$ and the identity property of dissimilarity networks that $d^k_Y(y, y, \ldots, y) = 0$. However, this is inconsistent with the identity property of dissimilarity networks which requires $d^k_X(x_{l,l,\ldots,l'}) = 0$ if and only if all the members in the tuple $x_{l,l,\ldots,l,l'}$ are identical. It then must be $\phi(x_l) = \phi(x_{l'})$ if and only if $x_l = x_{l'}$ implying that ϕ is an injection.

Likewise, define the function $\psi : Y \to X$ that associates y with an arbitrary x chosen from the set that form a pair with y in C_0,

$$\psi : y \mapsto x_0 \in \{x | (x, y) \in C_0\}. \tag{47}$$

It follows by similar arguments that ψ must be injective. By applying the Cantor-Bernstein-Schroeder theorem [29, Section 2.6] to the reciprocal injections $\phi : X \to Y$ and $\psi : Y \to X$, the existence of a bijection between X and Y is guaranteed. This forces X and Y to have same cardinality and ϕ and ψ being bijections. Pick the bijection ϕ and it follows $d_X^k(x_{0:k}) = d_Y^k(\phi(x_{0:k})$ for all nodes $(k+1)$-tuples $x_{0:k} \in X$. This shows that $D_X^K \cong_k D_Y^K$ and completes the proof of the identity statement. ■

Having demonstrated all four properties in Theorem 1, the global proof completes. ■

Proof of Theorem 2: The proof in Appendix B has demonstrated that $d_{\mathcal{D},p}$ is a pseudometric in the space of K-order dissimilarity networks modulo isomorphism. To prove that $d_{\mathcal{D},p}$ is a metric in the space of K-order dissimilarity networks modulo isomorphism we further demonstrate the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: We want to show that having $d_{\mathcal{D},p}(D_X^K, D_Y^K) = 0$ must imply that D_X^K and D_Y^K are isomorphic. If $d_{\mathcal{D},p}(D_X^K, D_Y^K) = \min_{C \in \mathcal{C}(X,Y)} \|\mathbf{\Gamma}_{X,Y}^K(C)\|_p = 0$, there exists a correspondence C_0 such that

$$\|\mathbf{\Gamma}_{X,Y}^K(C_0)\|_p = 0. \tag{48}$$

The property of p-norm implies that this correspondence C_0 satisfies $\Gamma_{X,Y}^k(C_0) = 0$ for any integers $0 \leq k \leq K$, i.e. $d_X^k(x_{0:k}) = d_Y^k(y_{0:k})$ for any integers $0 \leq k \leq K$ and any $(x_{0:k}, y_{0:k}) \in C_0$. Define functions $\phi : X \to Y$ as in (45) and $\psi : Y \to X$ as in (47). The analysis in Appendix C Proof of Theorem 1 for any integers $0 \leq k \leq K$ demonstrated that ϕ and ψ are bijections and that X and Y have same cardinality. Pick the bijection ϕ and it follows $d_X^k(x_{0:k}) = d_Y^k(\phi(x_{0:k}))$ for any integers $0 \leq k \leq K$ and all nodes $(k+1)$-tuples $x_{0:k} \in X$. This shows that $D_X^K \cong D_Y^K$ and completes the proof of the identity statement. ■

Having demonstrated all four properties in Theorem 2, the global proof completes. ■

Appendix D
Proof of Theorems in Section III-B

Proof of Theorem 3 : The proof in Appendix A has demonstrated that $d_{\mathcal{P}}^k$ for any integer $0 \leq k \leq K$ is a pseudometric in the space of K-order dissimilarity networks modulo k-isomorphism. To prove that $d_{\mathcal{P}}^k$ for any integer $1 \leq k \leq K$ is a metric in the space of K-order dissimilarity networks modulo k-isomorphism we need to show the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: Most parts of the proof follow from the proof of the second part of the identity property for Theorem 1 in Appendix C. The only difference is in demonstrating the function ϕ constructed in (45) is injective. Under the same setup where there exist a pair of nodes $x_l \neq x_{l'}$ such that $\phi(x_l) = \phi(x_{l'}) = y \in Y$. The k-order proximity between $x_{l,l,\ldots,l,l'}$ where the first k nodes in the tuple are x_l and the last node is $x_{l'}$ would satisfy

$$d_X^k(x_{l,l,\ldots,l,l'}) = d_Y^k(\phi(x_{l,l,\ldots,l'})) = 1, \tag{49}$$

where the first equality comes from the definition of ϕ and the second equality is due to the fact that $\phi(x_l) = \phi(x_{l'}) = y$ and the identity property of proximity networks that $d_Y^k(y, y, \ldots, y) = 1$. However, this is consistent with the identity property of proximity networks which requires $d_X^k(x_{l,l,\ldots,l'}) = 1$ if and only if all nodes in the tuple $x_{l,l\ldots,l,l'}$ are identical. It then must be $\phi(x_l) = \phi(x_{l'})$ if and only if $x_l = x_{l'}$ implying that ϕ is an injection. The rest of the proof follows. ■

Having demonstrated all four properties in Theorem 3, the global proof completes. ■

Proof of Theorem 4: The proof in Appendix B has demonstrated that $d_{\mathcal{P},p}$ is a pseudometric in the space of K-order proximity networks modulo isomorphism. To prove that $d_{\mathcal{P},p}$ is a metric in the space of K-order proximity networks modulo isomorphism we further demonstrate the missing part in the (iii) identity property in Definition 3.

Proof of the second part of the identity property: We want to show that having $d_{\mathcal{P},p}(P_X^K, P_Y^K) = 0$ must imply that P_X^K and P_Y^K are isomorphic. If $d_{\mathcal{D},p}(P_X^K, P_Y^K) = 0$, there exists a correspondence C_0 such that $\|\mathbf{\Gamma}_{X,Y}^K(C_0)\|_p = 0$. The property of p-norm implies that this correspondence C_0 satisfies $d_X^k(x_{0:k}) = d_Y^k(y_{0:k})$ for any integers $0 \leq k \leq K$ and any $(x_{0:k}, y_{0:k}) \in C_0$. Define functions $\phi : X \to Y$ as in (45) and $\psi : Y \to X$ as in (47), the analysis in Appendix D Proof of Theorem 3 has demonstrated that ϕ and ψ are bijections and that X and Y have same cardinality. Pick the bijection ϕ and it follows $d_X^k(x_{0:k}) = d_Y^k(\phi(x_{0:k}))$ for any integers $0 \leq k \leq K$ and all nodes $(k+1)$-tuples $x_{0:k} \in X$. This shows that $P_X^K \cong P_Y^K$ and completes the proof of the identity statement. ■

Having demonstrated all four properties in Theorem 4, the global proof completes. ■

Appendix E
Proofs in Section III-C

Proof of Proposition 6 : We first prove that $d_{\mathcal{D}}^k(D_{X'}^K, D_{Y'}^K) = d_{\mathcal{P}}^k(P_X^K, P_Y^K)$ for any integer $0 \leq k \leq K$ where $D_{X'}^K$ and $D_{Y'}^K$ are the constructed dissimilarity networks built from P_X^K and P_Y^K using (24). Let the correspondence C between X and Y be the minimizing correspondence in (20) so that we can write

$$d_{\mathcal{P}}^k(P_X^K, P_Y^K) = \Gamma_{X,Y}^k(C). \tag{50}$$

Since $X = X', Y = Y'$, although C may not be the minimizing correspondence for the distance $d_{\mathcal{D}}^K(P_{X'}^K, P_{Y'}^K)$, it is a valid correspondence. With the definition in (15) we can write,

$$d_{\mathcal{D}}^k(D_{X'}^K, D_{Y'}^K) \leq \Gamma_{X',Y'}^k(C). \tag{51}$$

Substituting the definition of $d_{X'}^k$ and $d_{Y'}^k$ in (24) for the right hand side of the definition for $\Gamma_{X',Y'}^k(C)$ in (51),

$$\Gamma_{X',Y'}^k(C) = \max_{(x_{0:k}, y_{0:k}) \in C} \Big| \big(1 - d_X^k(x_{0:k})\big) - \big(1 - d_Y^k(y_{0:k})\big) \Big|. \tag{52}$$

The 1s in (52) cancels and therefore,

$$\Gamma^k_{X',Y'}(C) = \Gamma^k_{X,Y}(C). \tag{53}$$

Substituting (50) and (51) back to (53) implies

$$d^k_{\mathcal{P}}(P^K_X, P^K_Y) \geq d^k_{\mathcal{D}}(D^K_{X'}, D^K_{Y'}). \tag{54}$$

Let the correspondence C' between X' and Y' be the minimizing correspondence in (15). Then C' is also a valid correspondence for the distance $d^K_{\mathcal{P}}(P^K_X, P^K_Y)$ used in (20). By symmetry, we have

$$d^k_{\mathcal{D}}(D^K_{X'}, D^K_{Y'}) \geq d^k_{\mathcal{P}}(P^K_X, P^K_Y). \tag{55}$$

Combining (54) and (55) yields the desired result

$$d^k_{\mathcal{D}}(D^K_{X'}, D^K_{Y'}) = d^k_{\mathcal{P}}(P^K_X, P^K_Y). \tag{56}$$

The other part of the proposition in which $d^k_{\mathcal{P}}(P^K_{X'}, P^K_{Y'}) = d^k_{\mathcal{D}}(D^K_X, D^K_Y)$ for any integer $0 \leq k \leq K$ where $P^K_{X'}$ and $P^K_{Y'}$ are the constructed proximity networks built from D^K_X and D^K_Y using (25) follows by symmetry. ■

Proof of Proposition 7 : We first prove that $d_{\mathcal{D},p}(D^K_{X'}, D^K_{Y'}) = d_{\mathcal{P},p}(P^K_X, P^K_Y)$ for some p-norm $\|\cdot\|_p$ where $D^K_{X'}$ and $D^K_{Y'}$ are the constructed dissimilarity networks built from P^K_X and P^K_Y using (24). Let the correspondence C between X and Y be the minimizing correspondence in (22) so that we can write

$$d_{\mathcal{P},p}(P^K_X, P^K_Y) = \left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p. \tag{57}$$

Since $X = X', Y = Y'$, although C may not be the minimizing correspondence for the distance $d_{\mathcal{D},p}(P^K_{X'}, P^K_{Y'})$, it is a valid correspondence. With the definition in (17) we can write,

$$d_{\mathcal{D},p}(D^K_{X'}, D^K_{Y'}) \leq \left\|\mathbf{\Gamma}^K_{X',Y'}(C)\right\|_p. \tag{58}$$

We have demonstrated in the Proof of Proposition 6 in Appendix E that for any integers $0 \leq k \leq K$, $\Gamma^k_{X',Y'}(C) = \Gamma^k_{X,Y}(C)$. In vector form, this is $\mathbf{\Gamma}^K_{X',Y'}(C) = \mathbf{\Gamma}^K_{X,Y}(C)$. Therefore, the property of p-norm implies that

$$\left\|\mathbf{\Gamma}^K_{X',Y'}(C)\right\|_p = \left\|\mathbf{\Gamma}^K_{X,Y}(C)\right\|_p. \tag{59}$$

Substituting (57) and (58) back to (59) yields

$$d_{\mathcal{P},p}(P^K_X, P^K_Y) \geq d_{\mathcal{D},p}(D^K_{X'}, D^K_{Y'}). \tag{60}$$

Let the correspondence C' between X' and Y' be the minimizing correspondence in (17). Then C' is also a valid correspondence for the distance $d^K_{\mathcal{P}}(P^K_X, P^K_Y)$ used in (22). By symmetry, we have

$$d_{\mathcal{D},p}(D^K_{X'}, D^K_{Y'}) \geq d_{\mathcal{P},p}(P^K_X, P^K_Y). \tag{61}$$

Combining (60) and (61) yields the desired result

$$d_{\mathcal{D},p}(D^K_{X'}, D^K_{Y'}) = d_{\mathcal{P},p}(P^K_X, P^K_Y). \tag{62}$$

The other part of the proposition in which $d_{\mathcal{P},p}(P^K_{X'}, P^K_{Y'}) = d_{\mathcal{D},p}(D^K_X, D^K_Y)$ for some p-norm where $P^K_{X'}$ and $P^K_{Y'}$ are the constructed proximity networks built from D^K_X and D^K_Y using (25) follows by symmetry. ■

References

[1] W. Huang and A. Ribeiro, "Metrics in the Space of High Order Proximity Networks (submitted)," in *In Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, 2015.

[2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Inc., 1993.

[3] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, ser. Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[4] R. Ghrist and A. Muhammad, "Coverage and hole-detection in sensor networks via homology," in *International Symposium on Information Processing in Sensor Networks*, vol. 00, no. 1, 2005, pp. 254–260.

[5] V. de Silva and R. Ghrist, "Coordinate-free Coverage in Sensor Networks with Controlled Boundaries via Homology," *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1205–1222, Dec. 2006.

[6] H. Chintakunta and H. Krim, "Divide and Conquer: Localizing Coverage Holes in Sensor Networks," in *2010 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, Jun. 2010, pp. 1–8.

[7] B. Zhang, "Hypernetworks: A molecular evolutionary architecture for cognitive learning and memory," *Computational Intelligence Magazine, IEEE*, vol. 3, no. 3, pp. 49–63, 2008.

[8] W. Ren, Q. Zhao, R. Ramanathan, J. Gao, A. Swami, A. Bar-Noy, M. P. Johnson, and P. Basu, "Broadcasting in multi-radio multi-channel wireless networks using simplicial complexes," in *Wireless Networks*, vol. 19, no. 6, Nov. 2012, pp. 1121–1133.

[9] J. Xu and V. Singh, "Unified Hypergraph for Image Ranking in a Multi-modal Context," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 2333–2336.

[10] Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis." *IEEE Transactions on Image Processing*, vol. 21, no. 9, pp. 4290–303, Sep. 2012.

[11] A. Wilkerson, T. Moore, A. Swami, and H. Krim, "Simplifying the Homology Of Networks via Strong Collapses," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 5258–5262.

[12] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, "Structure and tie strengths in mobile communication networks." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 18, pp. 7332–6, May 2007.

[13] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network." *Science (New York, N.Y.)*, vol. 311, no. 5757, pp. 88–90, Jan. 2006.

[14] S. Segarra, M. Eisen, and A. Ribeiro, "Authorship attribution using function words adjacency networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, no. 2, 2013, pp. 5563–5567.

[15] D. Khmelev and F. Tweedie, "Using Markov Chains for Identification of Writer," *Literary and linguistic computing*, vol. 16, no. 3, 2001.

[16] T. Wang and H. Krim, "Statistical classification of social networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 3977–3980.

[17] 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*, vol. 105, no. 35, pp. 12 763–12 768, 2008.

[18] L. Peng, L. Liu, S. Chen, and Q. Sheng, "A network comparison algorithm for predicting the conservative interaction regions in protein-protein interaction network," in *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, Sep. 2010, pp. 34–39.

[19] S. Choobdar, P. Ribeiro, S. Bugla, and F. Silva, "Comparison of Co-authorship Networks across Scientific Fields Using Motifs," *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 147–152, Aug. 2012.

[20] L. Yong, Z. Yan, and C. Lei, "Protein-protein interaction network comparison based on wavelet and principal component analysis," in *2010 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW)*, 2010, pp. 430–437.

[21] N. Pržulj, "Biological network comparison using graphlet degree distribution." *Bioinformatics*, vol. 23, no. 2, pp. e177–183, Jan. 2007.

[22] T. Milenković and N. Pržulj, "Uncovering biological network function via graphlet degree signatures." *Cancer informatics*, vol. 6, p. 257, Jan. 2008.

[23] N. Shervashidze, S. Vishwanathan, T. H. Petri, K. Mehlhorn, and K. M. Borgwardt, "Efficient graphlet kernels for large graph comparison," in *International Conference on Artificial Intelligence and Statistics*, vol. 5, 2009, pp. 488–495.

[24] G. Carlsson, F. Memoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," 2014. [Online]. Available: http://arxiv.org/abs/1301.7724

[25] M. Gromov, *Metric structures for Riemannian and non- Riemannian spaces*. Birkha user Boston Inc., Boston, MA., 2007.

[26] F. Memoli, "Gromov-Hausdorff distances in Euclidean spaces," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2008, pp. 1–8.

[27] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*. American Mathematical Soc., 2001, vol. 33.

[28] "Engineering Village: the place to find answers to engineering questions." [Online]. Available: http://www.engineeringvillage.com/search/quick.url

[29] A. N. Kolmogorov, S. V. Fomine, and R. A. Silverman, *Introductory Real Analysis*. Dover, 1975.