METRICS IN THE SPACE OF HIGH ORDER PROXIMITY NETWORKS

Weiyu Huang *Alejandro Ribeiro*

Department of Electrical and Systems Engineering, University of Pennsylvania

ABSTRACT

This paper presents two families of distances in the space of high order proximity networks. The distances measure differences between networks and are shown to be valid metrics in the space of high order 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.

Index Terms— High order networks, network distances.

1. INTRODUCTION

We consider high order proximity 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 [1, 2]. 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 were 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 [3–10].

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 [11–14] but very challenging otherwise, 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 [15], neighborhood topology [16], betweenness [17], motifs [18], wavelets [19], and graphlet-based heuristics [20–22]. 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 [23]. 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. These distances are build as generalizations of the pairwise distances in [23] which are themselves generalizations of the Gromov-Hausdorff distance between metric spaces [24,25]. 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. 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.

Supported by NSF CCF-1217963 and AFOSR MURI FA9550-10-1-0567.

2. 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 $d_X^1 : X \times X \to \mathbb{R}_+$ is a function encoding dissimilarity. We assume identity $d_X^1(x, x') = 0$ if and only if $x = x'$ and symmetry $d_X^1(x, x') = d_X^1(x', x)$ for all $x, x' \in X$. 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 two networks N_X and N_Y are isomorphic if there exists a bijection $\phi : X \to Y$ such that for all $x, x' \in X$,

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

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 [23]. The definition of this distance requires introducing the prerequisite notion of correspondence [26].

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. 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.

Definition 2 *Given two networks N_X and N_Y 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 N_X and N_Y is then defined as

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

For a given correspondence C 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. The distance in (3) is defined by selecting the correspondence that minimizes these maximal differences. 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.

For Definition 2 to be a valid distance it must define a metric in the space of networks modulo isomorphism. 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 $\mathcal{N} \mod \cong$; see [23]. The goal of this paper is to devise generalizations of Definition 2 to high order networks and to prove that they define valid metrics on the space of high order networks modulo isomorphism.

3. HIGH ORDER NETWORKS

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

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

For point collections $x_{0:k} := (x_0, x_1, \ldots, x_k) \in X^{k+1}$, $d_X^k(x_{0:k})$ are intended to represent a measure of similarity or dissimilarity for members of the group. Observe that pairwise networks are not particular cases of networks of order 1 because a network of order K requires relationships between $(k + 1)$-tuples for all integers $0 \leq k \leq K$. A 0-order network is one in which only node weights are given, an 1-order network is one in which weights and pairwise relationships are defined, a 2-order network 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 in which for all the $K+1$ functions d_X^k in (4) and $x_{0:k}$, $d_X^k(x_{[0:k]}) = d_X^k(x_{0:k})$ where $x_{[0:k]} = ([x_0], [x_1], \ldots, [x_k])$ is a reordering of $x_{0:k}$. 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 functions if d_X^k is a permutation of d_Y^k given integer $0 \leq k \leq K$. 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

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

for all $x_{0:k} \in X^{k+1}$ where $d_Y^k(\phi(x_{0:k})) := d_Y^k(\phi(x_0), \ldots, \phi(x_k))$. 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 $\mathcal{N}^K \mod \cong_k$. A stricter version of isomorphism is to consider K-order networks being equivalent if for all integers $0 \leq k \leq K$, d_X^k is a permutation of d_X^K. Formally, 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 $N_X^K \cong N_Y^K$. $N_X^K \cong N_Y^K$ implies $N_X^K \cong_k N_Y^K$ for every $0 \leq k \leq K$ but the converse is not true. The space of K-order networks modulo isomorphism is denoted by $\mathcal{N}^K \mod \cong$.

While different order functions d_X^k and d_X^l of a given network N_X^K need not be related, it is common to observe that adding nodes to a tuple results in decreasing relationships. This motivates the consideration of proximity networks that we undertake in the following section.

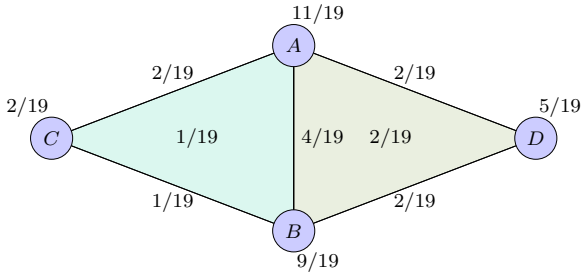

Fig. 1. Collaboration between authors of a research community. The k-order proximity function marks the number of publications between members of $(k + 1)$-tuples normalized by the total number of papers. E.g., author A published 11 papers which implies $d_X^0(A) = 11/19$. The number of papers jointly published by A and B results in $d_X^1(A, B) = 4/19$ and the number of papers written by A, B, and C implies that $d_X^2(A, B, C) = 1/19$. The order decreasing property in Definition 4 is satisfied because a paper written by a $(k + 1)$-tuple is also written by members of each of the included k-tuples.

3.1. Proximity Networks

In proximity networks the relationship functions $d_X^k(x_{0:k})$ denote similarity between elements of a tuple. Thus, large values of the proximity functions represent strong relationship whereas small values denote weak relationships. In this framework it is reasonable 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 that follows.

Definition 4 *The K-order network $P_X^K = \left(X, d_X^0, d_X^1, \ldots, d_X^K\right)$ is said to be a proximity network if the following two properties holds:*

Identity. *For any $0 \leq k \leq K$, $p_X^k(x_{0:k}) = 1$ 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*

$$d_X^k(x_{0:k}) \leq d_X^{k-1}(x_{0:k-1}). \tag{6}$$

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

In pairwise networks we required $d_X^k(x, x') = 0$ if and only if $x = x'$. The identity property in Definition 4 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 proximity networks the highest proximity $p_X^k(x_{0:k}) = 1$ is reserved to represent the closeness of a node to itself. Further note that since we restricted attention to symmetric networks a relationship as in (6) 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.

To see that the order decreasing property in Definition 4 is reasonable consider a 2-order network where the k-order proximity function records the normalized number of papers between members of a given $(k + 1)$-tuple. Proximities $d_X^0(x)$ are the numbers of papers published by author x, proximities $d_X^1(x, x')$ are the total number of papers in which x and x' are coauthors, and $d_X^2(x, x', x'')$ the number of papers jointly written by x, x', and x''. In all three cases we divide proximities by the total number of papers. Since a paper for a pair is also a paper for each of the individuals, $d_X^1(x, x') \leq d_X^0(x)$ and $d_X^1(x, x') \leq d_X^0(x')$ for all x and x'. Likewise, a paper of a triplet is also a paper of each of the three pairs, which implies that $d_X^2(x, x', x'') \leq d_X^1(x, x')$, $d_X^2(x, x', x'') \leq d_X^1(x, x'')$, and $d_X^2(x, x', x'') \leq d_X^1(x', x'')$ for all x, x', and x''. The order decreasing property is satisfied in both cases; see also Figure 4.

For each integer $0 \leq k \leq K$, the space $\mathcal{P}^K \mod \cong_k$ can be endowed with a proper metric akin to the pairwise network distance in Definition 2 as we formally specify next.

Definition 5 *Given proximity networks P_X^K and P_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*

$$\Gamma_{X,Y}^k(C) := \max_{(x_{0:k},y_{0:k})\in C} \left| d_X^k(x_{0:k}) - d_Y^k(y_{0:k}) \right|. \quad (7)$$

The k-order network distance between P_X^K and P_Y^K is then defined as

$$d_{\mathcal{P}}^k(P_X^K, P_Y^K) := \min_{C\in\mathcal{C}(X,Y)} \left\{ \Gamma_{X,Y}^k(C) \right\}. \quad (8)$$

The distance vector between P_X^K and P_Y^K is defined as

$$\mathbf{d}_{\mathcal{P}}^K(P_X^K, P_Y^K) = \left(d_{\mathcal{P}}^0(P_X^K, P_Y^K), \ldots, d_{\mathcal{P}}^K(P_X^K, P_Y^K) \right)^T. \quad (9)$$

Both, Definition 2 and Definition 5 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 (8) in Definition 5 is that $d_{\mathcal{P}}^k$ only considers one out of $K+1$ relationship functions. Other than that the definition is not much different since $\Gamma_{X,Y}^k(C)$ selects the maximum k-order difference $|d_X^k(x_{0:k}) - d_Y^k(y_{0:k})|$ among all tuples of correspondents. The distance $d_{\mathcal{P}}^k$ is defined by selecting the correspondence that minimizes these maximal differences. The distance vector $\mathbf{d}_{\mathcal{P}}^K$ defined in (9) is a vector with each element measuring the dissimilarity between functions of a specific order. Similar as in Definition 2, $d_{\mathcal{P}}^k$ and $\mathbf{d}_{\mathcal{P}}^K$ are defined even if networks have different number of nodes. The function $d_{\mathcal{P}}^k$ is a valid metric in $\mathcal{P}^K \mod \cong_k$ for any integer $1 \leq k \leq K$ (see [27] for proofs in this paper).

Theorem 1 *Given any nonnegative integer K, for any integers $1 \leq k \leq K$, the function $d_{\mathcal{P}}^k : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ defined in* (8) *is a metric in $\mathcal{P}^K \mod \cong_k$. The function $d_{\mathcal{P}}^0 : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ is a pseudometric in $\mathcal{P}^K \mod \cong_0$.*

The caveat for $d_{\mathcal{P}}^0$ is because two networks may own different number of nodes and identical zero order proximities for any nodes. Networks are not isomorphic however their 0-order distance is zero. A family of valid metrics measuring the difference between networks over all order functions can be endowed on the space $\mathcal{P}^K \mod \cong$.

Definition 6 *Given networks P_X^K and P_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 and the p-norm $\|\cdot\|_p$ as*

$$\left\| \mathbf{\Gamma}_{X,Y}^K(C) \right\|_p := \left\| \left(\Gamma_{X,Y}^0(C), \Gamma_{X,Y}^1(C), \ldots, \Gamma_{X,Y}^K(C) \right)^T \right\|_p, \quad (10)$$

where $\Gamma_{X,Y}^k(C)$ is defined in (7)*. The network distance respect to the p-norm $\|\cdot\|_p$ between P_X^K and P_Y^K is then defined as*

$$d_{\mathcal{P},p}(P_X^K, P_Y^K) := \min_{C\in\mathcal{C}(X,Y)} \left\{ \left\| \mathbf{\Gamma}_{X,Y}^K(C) \right\|_p \right\}. \quad (11)$$

The difference between Definition 2, Definition 5 and Definition 6 is that in $d_{\mathcal{P},p}$ we compare not only one functions but also all functions defined on networks. The norm $\|\mathbf{\Gamma}_{X,Y}^K(C)\|_p$ is assigned as the difference between P_X^K and P_Y^K measured by the correspondence C. The distance $d_{\mathcal{P},p}(P_X^K, P_Y^K)$ is then defined as the minimum of these differences achieved by some correspondence. Similarly $d_{\mathcal{P},p}$ is defined even if the numbers of nodes in networks are different. The function $d_{\mathcal{P},p} : \mathcal{P}^K \times \mathcal{P}^K \to \mathbb{R}_+$ is a proper metric in $\mathcal{P}^K \mod \cong$.

Theorem 2 *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 (11) is a metric in $\mathcal{P}^K \mod \cong$.*

In (11) we are only allowed to pick one correspondence minimizing $\|\mathbf{\Gamma}_{X,Y}^K(C)\|_p$ whereas in (8) for each k we are able to pick one correspondence minimizing the order specific $\Gamma_{X,Y}^k(C)$. This establishes a relationship between $d_{\mathcal{P},p}$ and $\|\mathbf{d}_{\mathcal{P}}^K\|_p$ as we formally state next.

Proposition 1 *Given some p-norm $\|\cdot\|_p$, for any nonnegative integer K and any proximity networks P_X^K, P_Y^K,*

$$d_{\mathcal{P},p}(P_X^K, P_Y^K) \geq \left\| \mathbf{d}_{\mathcal{P}}^K(P_X^K, P_Y^K) \right\|_p. \quad (12)$$

4. COMPARISON OF COAUTHORSHIP NETWORKS

We apply the metrics of Section 3.1 to compare 2-order coauthorship networks. Since Definitions 5 and 6 require searching over all correspondences, 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) but restrict attention to the collaboration networks of Prof. Georgios B. Giannakis (GG) and Prof. Martin Vetterli (MV). For each of them we construct networks for the 2004-2008 and 2009-2013 quinquennia (GG0408, GG0913, MV0408, and MV0913). For GG we also define networks for the five biennia between 2004 and 2013 (GG0405 through GG1213). Publication lists are queried from [28].

We consider all TSP publications in the period of interest and construct proximity networks where the node space X is formed by the lead author and their respective set of coauthors; see Figure 2. 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 by pairs, and second order proximities as the number of papers coauthored by triplets. We then normalize all proximities by the total number of papers in the network. With this construction the zeroth order proximities 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 higher than 2.

There are clear differences in the collaboration patterns which the network distances succeed in identifying. Two dimensional Euclidean embeddings of the distances $d_{\mathcal{P}}^0$, $d_{\mathcal{P}}^1$, $d_{\mathcal{P}}^2$, and $d_{\mathcal{P},1}$ between all networks are shown in Figure 3. Five of the seven GG networks; the biennia GG0405, GG0607, GG1011 (up triangles) and the quinquennia GG0408, GG0913 (diamonds); cluster closely for all four distances shown. The other two biennia (down triangles) may be closest to some of the other GG networks or to one of the two MV networks (circles), depending on which distance we consider. 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 6 networks correctly to GG and the other three networks to MV – one of them incorrectly.

To further parse these results, recall that $d_{\mathcal{P}}^k$ is defined by searching for a correspondence such that the maximum k-order difference $|d_X^k(x_{0:k}) - d_Y^k(y_{0:k})|$ is minimized [cf. (7) and (8)]. For the optimal correspondence $C^\star = \operatorname{argmin}_{C\in\mathcal{C}(X,Y)} \Gamma_{X,Y}^k(C)$, define the pair of correspondent tuples achieving the maximum k-order difference as

$$(x_{0:k}^\star, y_{0:k}^\star) = \operatorname*{argmax}_{(x_{0:k},y_{0:k})\in C^\star} \left| d_X^k(x_{0:k}) - d_Y^k(y_{0:k}) \right|. \quad (13)$$

The tuple pair $(x_{0:k}^\star, y_{0:k}^\star)$ 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 $d_{\mathcal{P}}^k$ 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 networks of Figure 2 the bottleneck pair for 0-order distances $d_{\mathcal{P}}^0$ is formed by nodes with high zero order proximities and $d_{\mathcal{P}}^0$ reflects the difference between their zero order proximities. Since the networks are normalized so that the lead nodes have size 1, $d_{\mathcal{P}}^0$ is determined by their predominant coauthors, i.e., the scholars that collaborated most prolifically with GG or MV during the period of interest. Focusing first on the quinquennial networks, observe that the distances $d_{\mathcal{P}}^0$ between GG

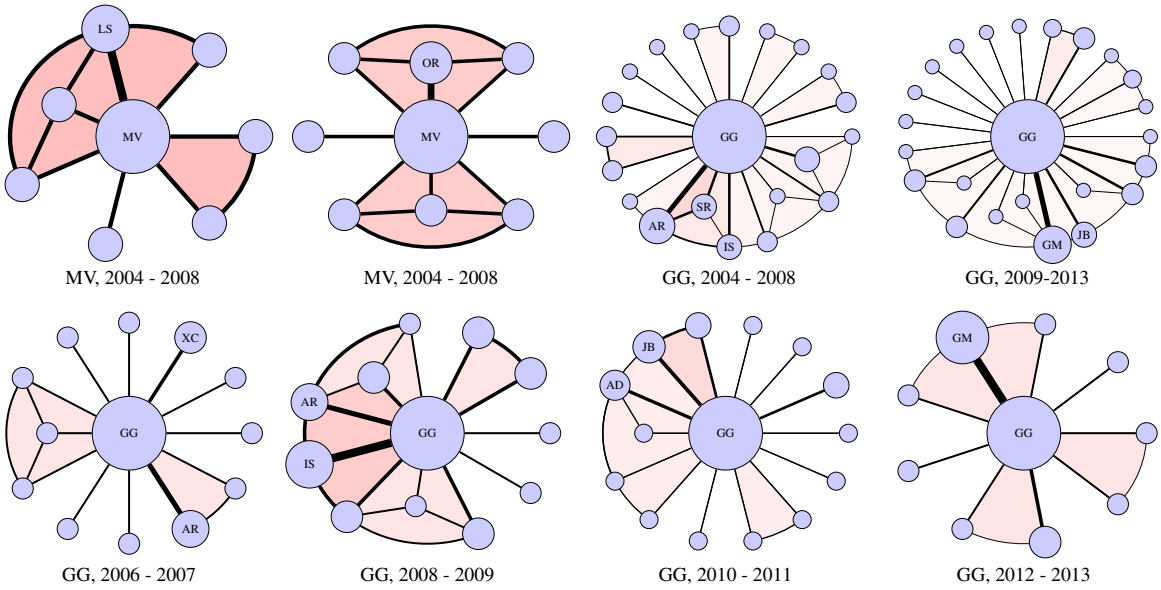

Fig. 2. Coauthorship networks representing research communities centered at Prof. Georgios Giannakis (GG) or Prof. Martin Vetterli (MV). The size of the nodes is proportional to the zeroth order proximities, and the width of the links to the first order proximities. Second order proximities are represented by shading the triangle enclosed by the coauthor triplet. Color intensity is proportional to the second order proximities.

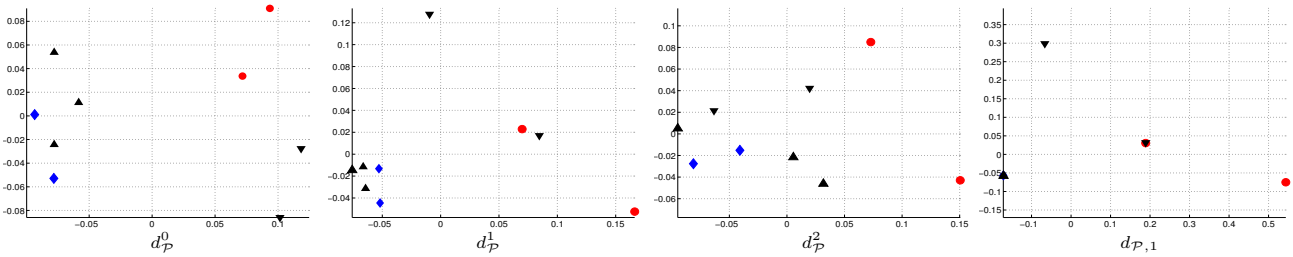

Fig. 3. Two dimensional Euclidean embeddings of the distances $d^0_{\mathcal{P}}$, $d^1_{\mathcal{P}}$, $d^2_{\mathcal{P}}$, and $d_{\mathcal{P},1}$ between networks. In the embeddings, denote MV0408, MV0913 as circles, GG0408, GG0913 as diamonds, GG0405, GG0607, GG1011 as up triangles and GG0809, GG1212 as down triangles.

and MV 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, first-order proximity distances between networks are likely due to: (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 quinquennial 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: (i) One network has collaboration between four or more authors while the other doesn't (ii) There exist three authors with a strong collaboration 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 0408 network has a few collaborations consisting of four scholars however all such collaborations are weak. GG0913 has no publications written by four authors.

In biennial networks we see more random variation. Still, the biennial networks, GG0405, GG0607, GG1011 (up triangles) are close to the quinquennial networks GG0408, GG0913 (diamonds) in every metric used because the distinctive features of GG coauthorship are well reflected in them. Indeed, these networks have: (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) Few publications with four or more authors. GG0809 and GG1213 do not cluster with other GG networks because they have features that resemble GG networks and some features that resemble MV networks. This happens because of prolific collaborations with Ioannis Schizas (IS) in the 08-09 period and Gonzalo Mateos (GM) in the 12-13 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. In the network GG1213 GM accounts for half of the publications in which GG is an author. Both of these features are more characteristic of MV networks.

5. CONCLUSION

We defined two families of distances measuring differences between proximity networks. These distances are valid metrics in the space of high order networks modulo isomorphism. We use this distances to successfully identify collaboration patters of Prof. Georgios B. Giannakis and Prof. Martin Vetterli. Tractable approximations will be provided in forthcoming contributions.

6. REFERENCES

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

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

[3] R. Ghrist and A. Muhammad, "Coverage and hole-detection in sensor networks via homology," in *International Symposium on Information Processing in Sensor Networks*, 2005, pp. 254–260.

[4] 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.

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

[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] 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.

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

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

[22] 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.

[23] 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

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

[25] 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.

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

[27] W. Huang and A. Ribeiro, "Metrics in the Space of High Order Networks," 2014. [Online]. Available: https://alliance.seas.upenn.edu/~aribeiro/wiki/index.php?n=Research.Publications

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