Axiomatic Hierarchical Clustering for Intervals of Metric Distances

Weiyu Huang and Alejandro Ribeiro

***Abstract*—This paper considers metric spaces where distances between a pair of nodes are represented by distance intervals. The goal is to study methods for the determination of hierarchical clusters, i.e., a family of nested partitions indexed by a resolution parameter, induced from the given distance intervals of the metric spaces. Our construction is based on defining admissible methods to be those methods that abide to the axioms of value and transformation. Two admissible methods are constructed and are shown to provide upper and lower bounds in the space of admissible methods. Practical implications are explored by clustering networks representing brain structural connectivity using the lower and upper bounds of the network distance. The proposed clustering methods succeed in differentiating brain connectivity networks of patients from those of healthy controls.**

***Index Terms*—Clustering, hierarchical clustering, axiomatic clustering, network theory, metric spaces, network comparison**

I. Introduction

The problem of identifying clusters of metric data has been studied extensively. There are literally hundreds of methods, techniques, and heuristics that can be utilized to the identification of clusters in finite metric spaces [2]. While theoretical development of clustering methods is not as well studied as practical counterparts [3], the past decade saw the evolvement of theories [4]. Of strong relevance to our work is the hierarchical clustering, where, instead of a single partition, we search for a family of partitions indexed by a connectivity parameter [5]. In this context, it has been proved in [4] and that single linkage [5, Ch. 4] is the unique hierarchical clustering method that abides to the three reasonable properties. These properties expect that the hierarchical clustering of a metric space with two points is the same metric space, that no pairs are clustered at resolutions smaller than the minimum distance in the space, and that when distances reduce, the metric space may only be clustered more but not less.

When we consider the distance between any pair of nodes given in a distance interval, we move into the realm of clustering with uncertain observations where the unpredictability is given by the distance intervals. Of particular relevance is the construction of models to replicate the key properties of uncertainties in the data [6] as well as the consideration of multiple observations of points given in a Euclidean space [7]. Compared to these previous work, we investigate situations where the only available information is the upper and lower bounds of the actual metric between any pairs. This can be considered as a more crude observation and a generalization of the other two approaches. Another related problem is the clustering of asymmetric relationships between nodes [8], [9]. The difference between the two is that in metric spaces with distance given by intervals, both upper and lower bounds represent the symmetric relationship between the pair however involve the uncertainty, whereas in asymmetric networks, all observation is certain however the relationship from a to b may not be the same as the relationship from b to a.

Work in this paper is 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. A journal version of this paper can be found in [1].

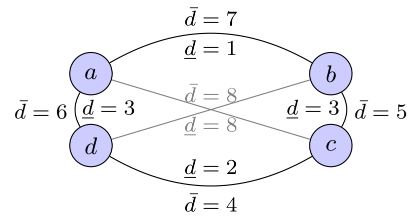


Fig. 1: An example of metric space where distances between pairs of nodes are given in lower and upper bounds.

The goal of this paper is to extend the axiomatic hierarchical clustering previously explored in [4], [9] for clustering of metric spaces where distance are given by intervals and to impose desired properties that we should demand for reasonable methods. With these properties, we proceed to characterize the space of methods that are admissible with respect to them.

II. Preliminaries

We consider a metric space M_X to be a pair (X, d_X) where X is a finite set of nodes and $d_X : X \times X \to \mathbb{R}_+$ is a metric distance. In specific, $d_X(x, x')$ between nodes $x \in X$ and $x' \in X$ is assumed to be nonnegative for all pairs x, x', is symmetric such that $d_X(x, x') = d_X(x', x)$, and is 0 if and only if the nodes coincide with $x = x'$; d_X also satisfies triangle inequality with $d_X(x, x'') \leq d_X(x, x') + d_X(x', x'')$ for any triplets $x, x', x'' \in X$. The interest of study in this paper is not on the metric space M_X, but in scenarios where observation of $d_X(x, x')$ is not exact but given in a confidence interval. Formally, we consider I_X as the triplet $(X, \bar{d}_X, \underline{d}_X)$ where given a pair of nodes $x, x' \in X$, we therefore have the relationship $0 < \underline{d}_X(x, x') \leq d_X(x, x') \leq \bar{d}_X(x, x')$. The bounds $\underline{d}(x, x')$ as well as $\bar{d}(x, x')$ between nodes $x, x' \in X$ are nonnegative for all pairs and 0 if and only if $x = x'$; moreover, they are symmetric, i.e. $\underline{d}(x, x')$ is the same as $\underline{d}(x', x)$ and similarly for $\bar{d}(x, x')$. However, they they may not necessarily satisfy the triangle inequality. We define $\mathcal{I}$ as the set of all metric spaces where the actual distance is observed in a confidence intervals. Entities in $\mathcal{I}$ may have different node sets X as well as different distance lower or upper bounds.

An example metric space with distance given by intervals is shown in Fig. 1. The smallest nontrivial case contains two

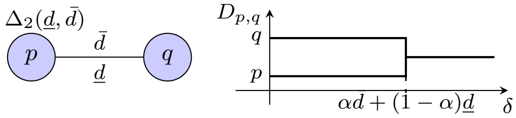


Fig. 2: Two-node space $\Delta_2(\bar{d}, \underline{d})$ and the Axiom of Value: nodes are clustered at the combination of the distance bounds.

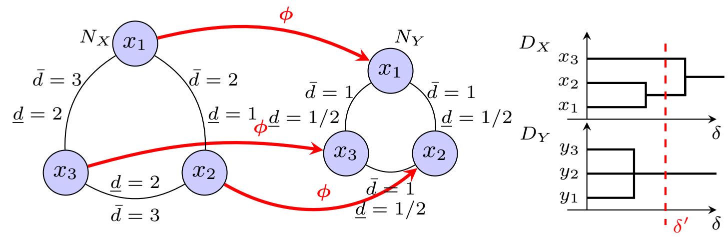


Fig. 3: Axiom of Transformation. If I_X can be mapped to I_Y using a α-distance-reducing map ϕ, then for every δ nodes clustered together in $D_X(\delta)$ must also be clustered in $D_Y(\delta)$.

nodes p and q with distance lower bound $\underline{d}(p,q) = \underline{d}$ as well as upper bound $\bar{d}(p,q) = \bar{d} \geq \underline{d} > 0$ is described in Fig. 2. We define the two-node space $\Delta_2(\underline{d}, \bar{d})$ with $\underline{d}$ and $\bar{d}$ as

$$\Delta_2(\underline{d}, \bar{d}) := (\{p, q\}, \underline{d}, \bar{d}). \tag{1}$$

A clustering of the set X denotes a partition P_X of X, i.e. a collection of pairwise disjoint sets $P_X = \{S_1, S_2, \dots, S_J\}$ with $S_i \cap S_j = \emptyset$ for any $i \neq j$ are required to cover X, $\cup_{j=1}^{J} S_j = X$. The sets $S_1, \dots, S_J$ are named the clusters of P_X. An equivalence relation $\sim$ on X is a binary relation such that for all $x, x', x'' \in X$ we have that (1) $x \sim x$, (2) $x \sim x'$ if and only if $x' \sim x$, and (3) $x \sim x'$ and $x' \sim x''$ would imply $x \sim x''$. A partition $P_X = \{S_1, S_2, \dots, S_J\}$ of X always induces and is induced by an equivalence relation $\sim_{P_X}$ on X where for all $x, x' \in X$ we have that $x \sim_{P_X} x'$ if and only if x and x' is cluttered to the same set S_j for some j.

In this paper we focus on hierarchical clustering methods [4], [9]. The output of hierarchical clustering methods is not a single partition P_X but a nested collection D_X of partitions $D_X(\delta)$ of X indexed by the resolution parameter $\delta \geq 0$. In the language of equivalence relation defined previously, for a given D_X, we say that two nodes x and x' are equivalent at resolution δ with notation $x \sim_{D_X(\delta)} x'$ if and only if nodes x and x' are in the same cluster of $D_X(\delta)$.

Dendrograms are difficult to analyze. A more convenient representation is acquired when dendrograms are identified with finite ultrametric spaces. An ultrametric on the space X is a metric $u_X : X \times X \to \mathbb{R}_+$ satisfying the stronger triangle inequality such that any points $x, x', x'' \in X$, the ultrametrics $u_X(x, x'')$, $u_X(x, x')$, and $u_X(x', x'')$ abide to the relationship

$$u_X(x, x'') \leq \max\big(u_X(x, x'), u_X(x', x'')\big). \tag{2}$$

We investigate ultrametrics because it preserves the structure of dendrograms [4]. Given a dendrogram, its corresponding ultrametric is $u_X(x, x') := \min\big\{\delta > 0 : x \sim_{D_X(\delta)} x'\big\}$; given an ultrametric, its associated equivalence relationship $\sim_{u_X(\delta)}$ is $x \sim_{u_X(\delta)} x' \Leftrightarrow u_X(x, x') \leq \delta$. From this equivalence, we consider hierarchical clustering method $\mathcal{H}$ as a map $\mathcal{H} : \mathcal{I} \to \mathcal{U}$ to the space of ultrametrics. Two clustering methods $\mathcal{H}_1$ and $\mathcal{H}_2$ are said to be equivalent with $\mathcal{H}_1 \equiv \mathcal{H}_2$ if and only if $\mathcal{H}_1(I) = \mathcal{H}_2(I)$ for any $I \in \mathcal{I}$.

The notions of chain and chain cost are substantial in hierarchical clustering. Given a metric space with distance intervals $(X, \underline{d}, \bar{d})$ and a pair of nodes $x, x' \in X$, a *chain* from x to x' is any ordered sequence of nodes in X,

$$[x = x_0, x_1, \dots, x_{l-1}, x_l = x'], \tag{3}$$

which begins with x and ends at x'. We denote $C(x, x')$ as one such chain and say $C(x, x')$ connects x to x'. Given two chains $C(x, x')$ and $C(x', x'')$ such that the end point x' of the first chain is the same as the starting point of the second, we define the *concatenated chain* $C(x, x') \uplus C(x', x)$ as

$$[x = x_0, x_1, \dots, x_{l-1}, x_l = x' = x'_0, x'_1, \dots, x'_{l'} = x'']. \tag{4}$$

For $C(x, x')$, we define its *upper cost* and *lower cost* as

$$\max_{i | x_i \in C(x,x')} \bar{d}_X(x_i, x_{i+1}), \qquad \max_{i | x_i \in C(x,x')} \underline{d}_X(x_i, x_{i+1}), \tag{5}$$

The *minimum upper chain cost* $\bar{c}(x, x')$ and the *minimum lower chain cost* $\underline{c}(x, x')$ between x and x' is then defined respectively as the minimum upper and lower cost among all chains connecting x to x',

$$\bar{c}_X(x, x') := \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \bar{d}_X(x_i, x_{i+1}), \tag{6}$$

$$\underline{c}_X(x, x') := \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \underline{d}_X(x_i, x_{i+1}). \tag{7}$$

The minimum upper chain cost $\bar{c}_X(x, x')$ and lower chain cost $\underline{c}_X(x, x')$ are different in general, however they are equal in the degenerate case where distance lower bounds and upper bounds coincide with $\underline{d}_X(x, x') = \bar{d}_X(x, x') := d_X(x, x')$ for any $x, x' \in X$. In this case, the minimum cost $\bar{c}_X(x, x') = \underline{c}_X(x, x')$ are important in the construction of the single linkage [4]. In specific, single linkage ultrametric $u_X^{\mathrm{SL}}(x, x')$ between x and x' is

$$u_X^{\mathrm{SL}}(x, x') = \min_{C(x,x')} \max_{i | x_i \in C(x,x')} d_X(x_i, x_{i+1}). \tag{8}$$

It can be seen that $\bar{c}_X$ is the result of applying single linkage towards the node set X equipped with dissimilarity $\bar{d}_X$ despite the fact that $\bar{d}_X$ may not be a valid metric; similar result holds for $\underline{c}_X$. In the degenerative case where distance lower bounds and upper bounds coincide, it is equivalent to consider metric spaces (X, d_X). It has been shown [4] that single linkage is the unique hierarchical clustering method fulfilling axioms (A1) and (A2) discussed in Section III. In the case when the dissimilarity $d_X(x, x')$ is unknown but given in an interval $[\underline{d}_X(x, x'), \bar{d}_X(x, x')]$ instead, the space of methods satisfying axioms (A1) and (A2) and their analogous ones becomes richer, as we explain throughout the paper.

III. Axioms of Value and Transformation

To study hierarchical clustering methods on metric spaces where dissimilarities are given in distance intervals, we translate natural concepts into the axioms of value and transformation, described in this section. We say a hierarchical clustering method $\mathcal{H}$ is *admissible* if and only if it satisfies both the the axioms of transformation and value.

The Axiom of Value is achieved by considering the two-node space $\Delta_2(\underline{d},\bar{d})$. In the degenerate special case where $\underline{d}=\bar{d}:=d(p,q)$, it is apparent that the resolution at which nodes p and q are first clustered together should be $d(p,q)$. In general scenarios where the dissimilarity $d(p,q)$ is given in an interval $[\underline{d},\bar{d}]$ with $\underline{d}<\bar{d}$, we say that nodes p and q form a single cluster first at resolution $\delta:=\alpha\bar{d}+(1-\alpha)\underline{d}$, the convex combination of the upper and lower bounds $\bar{d}$ and $\underline{d}$. Property of hierarchical clustering then indicates nodes p and q are clustered together at any resolution $\delta\geq\alpha\bar{d}+(1-\alpha)\underline{d}$. The parameter α controls the level of confidence in examining the distance intervals. A higher value of α implies a more conservative consideration, where in the extreme case with $\alpha=1$, nodes p and q are clustered together at the distance upper bound $\bar{d}$. We formalize this intuition as next.

(A1) Axiom of Value. Given $0\leq\alpha\leq 1$, the ultrametric output $(\{p,q\},u_{p,q})=\mathcal{H}(\Delta_2(\underline{d},\bar{d}))$ resulted from applying $\mathcal{H}$ upon the two-node space $\Delta_2(\underline{d},\bar{d})$ satisfies that $u_{p,q}(p,q)=\alpha\bar{d}+(1-\alpha)\underline{d}$.

The second requirement on the space of desired methods $\mathcal{H}$ formalizes the intuition for the behavior of $\mathcal{H}$ when considering a transformation on the distance bounds on the underlying space X; see Fig. 3. Consider $I_X=(X,\underline{d}_X,\bar{d}_X)$ and $I_Y=(Y,\underline{d}_Y,\bar{d}_Y)$ and denote $D_X=\mathcal{H}(X,\underline{d}_X,\bar{d}_X)$ and $D_Y=\mathcal{H}(Y,\underline{d}_Y,\bar{d}_Y)$ as the corresponding dendrogram outputs. If we can map all the nodes of the triplet $(X,\underline{d}_X,\bar{d}_X)$ into nodes of $(Y,\underline{d}_Y,\bar{d}_Y)$ such that the combination of lower and upper bounds for any pair of nodes is not increased, we expect the latter metric distance intervals to be more clustered than the former one at any given resolution. Intuitively, nodes in I_Y are less dissimilar with respect to each other, and therefore at any resolution δ in the respective dendrograms, we expect that for nodes that are clustered in I_X, their corresponding nodes in Y are also clustered in I_Y. In order to formalize this intuition, we introduce the following notion that given $I_X=(X,\underline{d}_X,\bar{d}_X)$, $I_Y=(Y,\underline{d}_Y,\bar{d}_Y)$, and a value $0\leq\alpha\leq 1$, the map $\phi:X\to Y$ is called *α-distance-reducing* if for any $x,x'\in X$, it holds that

$$\hat{d}_X(x,x')\geq\hat{d}_Y(\phi(x),\phi(x')),\ \hat{c}_X(x,x')\geq\hat{c}_Y(\phi(x),\phi(x')),\quad (9)$$

where we define $\hat{d}_X(x,x'):=\alpha\bar{d}_X(x,x')+(1-\alpha)\underline{d}_X(x,x')$ and similarly $\hat{c}_X(x,x'):=\alpha\bar{c}_X(x,x')+(1-\alpha)\underline{c}_X(x,x')$. A mapping is α-distance-reducing if both the combinations of distance bounds and chain costs are non-increasing. In the degenerate case where distance lower and upper bounds coincide, $u_X^{\mathrm{SL}}(x,x'):=\bar{c}_X(x,x')=\underline{c}_X(x,x')$ is the output of applying single linkage upon the metric space. Therefore the first inequality in (9) becomes identical as $d_X(x,x')\geq d_Y(\phi(x),\phi(x'))$, from which $c_X(x,x')\geq c_Y(\phi(x),\phi(x'))$ follows directly. In general cases, $\hat{c}_X(x,x')\geq\hat{c}_Y(\phi(x),\phi(x'))$ does not follow from $\hat{d}_X(x,x')\geq\hat{d}_Y(\phi(x),\phi(x'))$. The Axiom of Transformation is a formal statement of the intinction.

(A2) Axiom of Transformation. Consider $I_X=(X,\underline{d}_X,\bar{d}_X)$ and $I_Y=(Y,\underline{d}_Y,\bar{d}_Y)$ and a given α-distance-reducing map $\phi:X\to Y$. For any pair of nodes $x,x'\in X$, the output ultrametrics $u_X=\mathcal{H}(X,\underline{d}_X,\bar{d}_X)$ and $u_Y=\mathcal{H}(Y,\underline{d}_Y,\bar{d}_Y)$ satisfy $u_X(x,x')\geq u_Y(\phi(x),\phi(x'))$.

In summary, Axiom (A1) specifics our tendency in believing lower or upper bounds. Axiom (A2) states that if we reduce both the distance bounds, clusters may be combined but cannot be separated. These axioms are an adaption of the axioms proposed in [4] for the degenerate case of $\underline{d}_X=\bar{d}_X$, and the axioms proposed in [9] for asymmetric networks.

IV. Admissible Ultrametrics

Consider a specific metric space with distances given by intervals $I_X=(X,\underline{d}_X,\bar{d}_X)\in\mathcal{I}$. Given a value $0\leq\alpha\leq 1$, one particular clustering method satisfying axioms (A1) and (A2) can be established by examining the *α-combined* dissimilarity $\hat{d}_X(x,x')$ for all pair of nodes $x,x'\in X$. Though $\hat{d}_X$ does not necessarily satisfy the triangle inequality as the original metric distance d_X, it is symmetric; therefore the α-combined dissimilarity effectively reduces the problem to clustering of symmetric data, a case where the single linkage method defined in (8) is shown to abide to axioms analogous to (A1) and (A2) [4]. Based on this observation, we define the *combine-and-cluster* method $\mathcal{H}^{\mathrm{CO}}$ with output $(X,u_X^{\mathrm{CO}})=\mathcal{H}^{\mathrm{CO}}(X,A_X)$ between a pair x and x' as

$$u_X^{\mathrm{CO}}(x,x'):=\min_{C(x,x')}\ \max_{i|x_i\in C(x,x')}\hat{d}_X(x_i,x_{i+1}).\qquad (10)$$

An illustration of the combine-and-cluster clustering method is shown in Fig. 4. For a given pair of nodes x and x', we look for chains $C(x,x')$ connecting them. For a considered chain we examine each of its link, connecting say x_i with x_{i+1}, and investigate the convex combination of the distance bounds, i.e. the value of $\hat{d}_X(x_i,x_{i+1})=\alpha\bar{d}_X(x_i,x_{i+1})+(1-\alpha)\underline{d}_X(x_i,x_{i+1})$. The maximum value across all links in this chain is then recorded. The combine-and-cluster ultrametric $u_X^{\mathrm{CO}}(x,x')$ between points x and x' is the minimum of this value across all possible chains connecting x and x'.

In combine-and-cluster clustering, nodes x and x' belong to the same cluster at resolution δ whenever we can find a single chain such that the maximum convex combination of distance bounds is no greater than δ. In *cluster-and-combine* clustering, we switch the order of operations and investigate chains, potentially different, connecting x and x', with one chain focusing on the distance upper bounds and the other chain examining the distance lower bounds, before combining the upper and lower estimations. To state this definition regarding ultrametrics, consider $I_X=(X,\underline{d}_X,\bar{d}_X)$ and $0\leq\alpha\leq 1$. We define the cluster-and-combine method $\mathcal{H}^{\mathrm{CL}}$ with output $(X,u_X^{\mathrm{CL}})=\mathcal{H}^{\mathrm{CL}}(X,\underline{d}_X,\bar{d}_X)$ as

$$u_X^{\mathrm{CL}}(x,x'):=\min_{C(x,x')}\ \max_{i|x_i\in C(x,x')}\hat{c}_X(x_i,x_{i+1})),\qquad (11)$$

An illustration of the cluster-and-combine clustering method is described in Fig. 5. For any pair of nodes, we consider the convex combination $\hat{c}_X(x,x')$ of minimum chain costs. The output of the cluster-and-combine clustering method is the result by applying single linkage $\mathcal{H}^{\mathrm{SL}}$ [cf. (8)] onto the convex combination $\hat{c}_X(x,x')$. The single linkage is applied towards $\hat{c}_X(x,x')$ because convex combination of ultrametrics is a metric but not necessarily an ultrametric. We demonstrate that the output u_X^{CO} and u_X^{CL} are valid ultrametrics and the methods $\mathcal{H}^{\mathrm{CL}}$ and $\mathcal{H}^{\mathrm{CO}}$ abide to axioms (A1) and (A2).

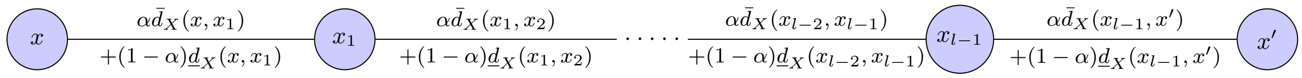


Fig. 4: Combine-and-cluster clustering. Nodes x and x' are clustered together at resolution δ if there exists a chain such that the maximum convex combination of distance bounds $\hat{d}_X(x_i, x_{i+1})$ is no greater than δ [cf. (10)].

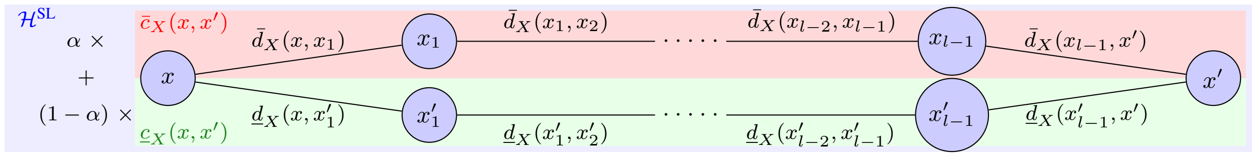


Fig. 5: Cluster-and-combine clustering. Nodes x and x' are clustered together at resolution δ if there exists a chain such that the maximum convex combination $\hat{c}_X(x,x')$ of minimum upper and lower chain costs is no greater than δ [cf. (11)].

Proposition 1 *The combine-and-cluster method $\mathcal{H}^{CO}$ and cluster-and-combine method $\mathcal{H}^{CL}$ is valid and admissible. I.e., the ultrametrics u_X^{CO} defined by* (11) *and u_X^{CL} defined by* (10) *are ultrametrics for all $I_X = (X, \underline{d}_X, \bar{d}_X)$; moreover, the methods $\mathcal{H}^{CO}$ and $\mathcal{H}^{CL}$ satisfy axioms (A1) and (A2).*

Proof: The proofs for Prop. 1 and Thm. 1 are both technically rich and therefore we refer readers to [1] for details. ■

Given that we have constructed two admissible methods satisfying axioms (A1)-(A2), it is natural to ask whether these two constructions are the only possible ones, and if not, whether they are special with respect to other satisfying methods. We prove the important characterization that any method $\mathcal{H}$ satisfying axioms (A1)-(A2) yields ultrametrics that lie between u_X^{CL} and u_X^{CO}.

Theorem 1 *Consider an admissible clustering method $\mathcal{H}$ satisfying axioms (A1)-(A2). For an arbitrary $I_X = (X, \bar{d}_X, \underline{d}_X)$ and $0 \leq \alpha \leq 1$, denote $(X, u_X) = \mathcal{H}(I_X)$ the output of applying $\mathcal{H}$ onto I_X. Then for any pair of nodes $x, x' \in X$,*

$$u_X^{CL}(x,x') \leq u_X(x,x') \leq u_X^{CO}(x,x'). \tag{12}$$

V. Clustering of networks via distance bounds

In this section, we apply clustering methods to bounds on metrics in the space of networks [10]. We have defined network differences derived from an underlying metric in the space of networks [11]. We briefly review definitions here. We consider network in the form of $N_Z = (Z, r_Z)$ with Z denotes the set of points in the network and $r_Z : Z \times Z \to \mathbb{R}^+$ a relationship function. For points z and z', the value $r_Z(z,z')$ is intended to represent similarity between the pair. The function r_Z is nonnegative and symmetric, satisfies $r_Z(z,z') = 0$ if and only if $z = z'$, however, does not necessarily satisfy the triangle inequality. The set of all networks is denoted as $\mathcal{N}$. Two networks N_Z and N_W are said isomorphic if there exists a bijection $\pi : Z \to W$ such that for all $x_{0:k} \in X^{k+1}$ we have $r_W(\pi(z), \pi(z')) = r_Z(z,z')$ for any $z \neq z' \in Z$. Since the map π is bijective, isomophism can only be satisfied when Z is a permutation of W. When networks N_Z and N_W are isomorphic we write $N_Z \cong N_W$. The space of networks where isomorphic networks are represented by the same element is termed the set of networks modulo isomorphism and denoted by $\mathcal{N} \mod \cong$. The space $\mathcal{N}^K \mod \cong_k$ can be endowed with a valid metric [11]. The definition of this distance requires introducing the notion of correspondence [12, Def. 7.3.17]:

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

A correspondence in Definition 1 connects node sets X and Y so that every element of one set has at least one correspondent in the other. Network distance is defined by selecting the correspondence that makes them the most similar.

Definition 2 *Given N_Z and N_W and a correspondence C between Z and W, the network difference w.r.t C is*

$$\Gamma_{Z,W}^k(C) := \max_{(z,w),(z',w') \in C} |r_Z(z,z') - r_W(w,w')|, \tag{13}$$

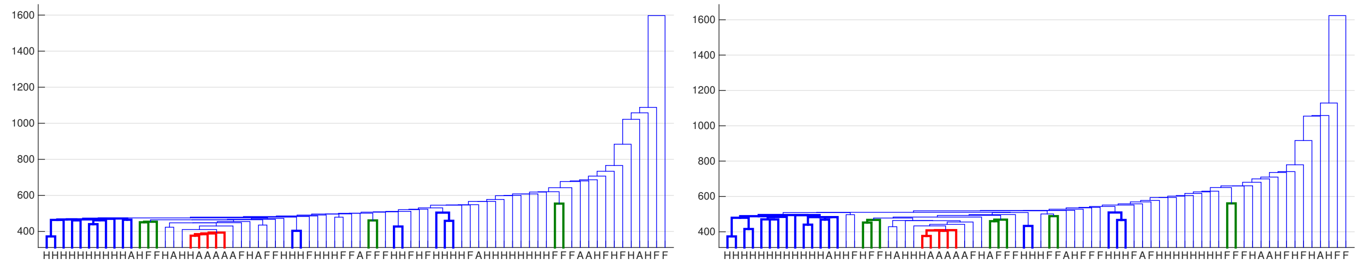
The network distance between N_Z and N_W is then defined as

$$d_{\mathcal{N}}(N_Z, N_W) := \min_{C \in \mathcal{C}(Z,W)} \{\Gamma_{Z,W}(C)\}. \tag{14}$$

For a given correspondence $C \in \mathcal{C}(Z,W)$ the network difference $\Gamma_{Z,W}(C)$ selects the maximum distance difference $|r_Z(z,z') - r_W(w,w')|$ among all pairs of correspondents. The distance in (14) is defined by selecting the correspondence that minimizes these maximal differences. The metric distances defined here have been applied to compare networks with small number of nodes and have succeeded in identifying collaboration patterns of coauthorship networks [11]. However, because they have to consider all possible correspondences, network distances are difficult to compute when the number of nodes in the networks is large. To solve such problem, we mapped networks to filtrations of simplicial complexes and demonstrated that the difference between the homological features of their respective filtration can be used as a lower bound of $d_{\mathcal{N}}$ [13]. Computation of homological features is fast and we have applied these lower bounds in comparing the coauthorship networks of academic journals.

On the other hand, $\Gamma_{Z,W}^k(C)$ in (13) for any correspondence C witnesses an upper bound on the distance $d_{\mathcal{N}}$. Therefore, given a set of networks X where the i-th element x_i denotes a network N_i, we can evaluate the upper and lower bounds of network distance $d_{\mathcal{N}}(N_i, N_j)$ for any pair N_i and N_j in X to yield a metric in the space of networks where distances are given by intervals $(X, \underline{d}_X, \bar{d}_X)$. Clustering methods can then be applied towards the triplet to categorize networks.

As an example of network classification, we consider real-world brain networks of patients diagnosed with Fronto-Temporal Dementia (FTD) or Alzheimer's disease (AD) as



(a) Cluster-and-combine dendrogram

(b) Combine-and-cluster dendrogram

Fig. 6: Resulting dendrograms of cluster-and-combine method (a) and combine-and-cluster method (b) applied upon the triplet $(X, \underline{d}_X, \bar{d}_X)$ in the brain network application. Each node in the dendrograms represents the brain structural connectivity network of a participant, where healthy controls are labeled as 'H', patients with FTD as 'F', and patients with AD as 'A'.

well as healthy controls [14]. For each network $N_Z = (Z, r_Z)$, Z denotes the set of 119 brain regions, where each region z represents brain areas that are anatomically similar. The relationship function $r_Z(z, z')$ denotes the number of connecting neurons between brain regions z and z'. There is an underlying labeling of brain regions, and therefore given a pair of networks $N_Z = (Z, r_Z)$ and $N_{\tilde{Z}} = (Z, r_{\tilde{Z}})$, it is reasonable to utilize the sum of differences on all connections

$$d_b(N_Z, N_{\tilde{Z}}) := \sum_{z \neq z'} |r_Z(z, z') - r_{\tilde{Z}}(z, z')|, \tag{15}$$

as the difference between the pair; we use this as a benchmark with the methods proposed in the paper. Nonetheless, using the same labeling assumes that same brain region has exactly identical functionality across all subjects, while in reality brain region z of one subject might be the most similar with brain region $z' \neq z$ of another subject. Motivated by this, we consider brain networks as unlabeled entities and evaluate the bounds of network distance $d_{\mathcal{N}}(N_Z, N_{\tilde{Z}})$ defined in Definition 2. In specific, the lower bound $\underline{d}_X(N_Z, N_{\tilde{Z}})$ is established via the difference in the homological features. The upper bound $\bar{d}_X(N_Z, N_{\tilde{Z}})$ is constructed using the correspondence C

$$\bar{d}_X(N_Z, N_{\tilde{Z}}) := \max_{z \neq z'} |r_Z(z, z') - r_{\tilde{Z}}(z, z')|. \tag{16}$$

We apply cluster-and-combine as well as combine-and-cluster methods upon the constructed metric in the space of networks with distances given by intervals $(X, \underline{d}_X, \bar{d}_X)$ with confidence level $\alpha = 0.5$. Figure 6 shows the respective dendrogram, where healthy controls are labeled as 'H', patients with FTD as 'F', and patients with AD as 'A'. We note that from the resulting dendrograms, the difference between the resulting dendrogram of $\mathcal{H}^{\text{CL}}$ and $\mathcal{H}^{\text{CO}}$ is small. This observation combined with the guarantee established by Theorem 1 further demonstrates that the outputs of all feasible hierarchical clustering methods do not differ by much. Besides, networks corresponding to patients with same health status ('H' or 'F' or 'A') tend to be clustered together at lower resolution, which is highlighted by bold colored lines in the dendrograms. Finally, as a quantitative benchmark, we evaluate the error of unsupervised classification based on the output ultrametrics u_X^{CL} and u_X^{CO} in classifying healthy controls from patients with either FTD or AD, and compare its error with the similar investigation based on single linkage applied towards the benchmark difference $d_b(N_Z, N_{\tilde{Z}})$ defined in (15). The clustering methods $\mathcal{H}^{\text{CL}}$ and $\mathcal{H}^{\text{CO}}$ yield 28.38% unsupervised classification error, which is slightly better than the 29.73% error of the benchmark.

VI. Conclusions

We developed a theory for hierarchically clustering metric spaces where distances are given by intervals. We begin by identifying reasonable axioms. Two admissible methods were constructed and were proved as upper and lower bounds for all admissible methods. We explored the practical usefulness by clustering networks denoting brain structural connectivity via the upper and lower bounds of network distances. The proposed methods made a moderate success in associating brain connectivity networks to their neruroscitific behaviors.

References

[1] W. Huang and A. Ribeiro, "Axiomatic hierarchical clustering given confidence intervals of metric distances," *Signal Process., IEEE Trans. on*, vol. (submitted), 2016.

[2] R. Xu, D. Wunsch *et al.*, "Survey of clustering algorithms," *Neural Networks, IEEE Trans. on*, vol. 16, no. 3, pp. 645–678, 2005.

[3] U. Von Luxburg and S. Ben-David, "Towards a statistical theory of clustering," in *Pascal work. on stat. & opt. of clust.*, 2005, pp. 20–26.

[4] G. Carlsson and F. Mémoli, "Characterization, stability and convergence of hierarchical clustering methods," *J. Mach. Learn. Res*, vol. 11, pp. 1425–1470, 2010.

[5] A. K. Jain and R. C. Dubes, *Algorithms For Clustering Data*. Prentice-Hall, Inc., 1988.

[6] G. Cormode and A. McGregor, "Approximation algorithms for clustering uncertain data," in *Proc. of the 27th ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2008, pp. 191–200.

[7] R. M. de Souza and F. d. A. De Carvalho, "Clustering of interval data based on city–block distances," *Pattern Recognition Let.*, vol. 25, no. 3, pp. 353–365, 2004.

[8] W. Pentney and M. Meila, "Spectral clustering of biological sequence data," in *AAAI*, vol. 5, 2005, pp. 845–850.

[9] G. Carlsson, F. Memoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," *Trans. Inf. Theory*, vol. (submitted.), 2015.

[10] W. Ali, T. Rito, G. Reinert, F. Sun, and C. M. Deane, "Alignment-free protein interaction network comparison," *Bioinformatics*, vol. 30, no. 17, pp. i430–i437, 2014.

[11] W. Huang and A. Ribeiro, "Metrics in the space of high order networks," *Signal Process., IEEE Trans. on*, vol. 64, no. 3, pp. 615–629, 2016.

[12] D. Burago, Y. Burago, and S. Ivanov, *A Course In Metric Geometry*. American Mathematical Society Providence, 2001, vol. 33.

[13] W. Huang and A. Ribeiro, "Persistent homology lower bounds on network distances," in *In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE Int. Conf. on*, vol. (to appear), Shanghai, 2016.

[14] J. D. Medaglia, W. Huang, S. Segarra, C. Olm, J. Gee, M. Grossman, A. Ribeiro, C. T. McMillan, and D. S. Bassett, "Brain network efficiency is influenced by pathological source of corticobasal syndrome," *Neurology*, vol. (submitted), 2016. [Online]. Available: http://arxiv.org/pdf/1601.07867v1.pdf