Hierarchical Clustering Methods and Algorithms for Asymmetric Networks

Gunnar Carlsson, Facundo Mémoli, Alejandro Ribeiro, and Santiago Segarra

Abstract—**Three different families of hierarchical clustering methods satisfying the axioms of value – in a network with two nodes the nodes cluster together at resolutions at which both can influence each other – and transformation – when we reduce some pairwise dissimilarities and increase none, the resolutions at which nodes cluster together may decrease but not increase – are introduced. The grafting family exchanges branches between dendrograms generated by different admissible methods. The convex combination family combines admissible methods using a convex operation in the space of dendrograms. The semi-reciprocal family is related to the reciprocal and nonreciprocal clustering methods introduced in [1]. Algorithms for the computation of hierarchical clusters generated by reciprocal and nonreciprocal clustering as well as the grafting, convex combination, and semi-reciprocal families are derived using matrix operations in a dioid algebra.**

I. Introduction

The output of hierarchical clustering methods is a dendrogram consisting of a nested set of partitions indexed by a resolution parameter [2]. We consider the problem of devising methods to construct dendrograms associated with a given network of asymmetric dissimilarities. While a large number of methods for determining hierarchical and nonhierarchical clusters in finite metric spaces exists – see, e.g., [3] –, methods to identify clusters in a network of asymmetric dissimilarities are rarer [4]–[6]. This relative scarcity is expected because the intuition for clusters as groups of nodes that are closer to each other than to the rest is difficult to generalize when nodes are close in one direction but far apart in the other. To overcome this generic difficulty we can draw inspiration from the fundamental underpinnings of clustering, which, although not as well developed as its practice [7], [8], are by now quite well established in the case of finite metric spaces [9]–[11]. Of particular relevance to our work is the case of hierarchical clustering [12]. In this context, it has been shown in [13] that single linkage [2, Ch. 4] is the unique hierarchical clustering method that satisfies three reasonable axiomatic statements.

In the context of asymmetric networks, our work in [1] introduces the axioms of value – in a network with two nodes the nodes cluster together at resolutions at which both can influence each other – and transformation – reducing some pairwise dissimilarities and increasing none cannot increase the resolution at which clusters form – as reasonable behaviors that we should expect to see in hierarchical clustering methods for asymmetric networks. These axioms are apparently not stringent but they do result in the strong conclusion that all methods that abide to these axioms lie between two particular cases in a well defined sense. The first method requires that clusters form through arcs in which both dissimilarities are small and is therefore termed reciprocal clustering. The second method, termed nonreciprocal clustering, allows clustering if loops of proximity can be formed. One of our results (Theorem 1) implies that any clustering method that satisfies the value and transformation axioms forms clusters at resolutions coarser than those of nonreciprocal clustering and finer than those of reciprocal clustering. For symmetric networks, reciprocal and nonreciprocal clustering coincide, recovering the uniqueness result in [13].

In the context of asymmetric networks, the difference between reciprocal and nonreciprocal clustering allows the existence of intermediate clustering methods. This paper introduces three families of intermediate clustering methods (Section III). The grafting family is built by exchanging branches between dendrograms generated by different admissible methods (Section III-A). Convex combinations of dendrograms generated by admissible methods yield a second family (Section III-B). The semi-reciprocal family requires part of the influence to be reciprocal and allows the rest to propagate through loops (Section III-C). We further derive algorithms to compute dendrograms generated by reciprocal and nonreciprocal clustering as well as the grafting, convex combination, and semi-reciprocal families using matrix operations in a dioid algebra (Section IV).

II. Preliminaries

Define the network $N = (X, A_X)$ as a set of n points or nodes X jointly specified with a real valued dissimilarity function $A_X : X \times X \to \mathrm{R}_+$ defined for all pairs $x, x' \in X$. Dissimilarities $A_X(x, x')$ from x to x' are nonnegative, and null if and only if $x = x'$, but may not satisfy the triangle inequality and may be asymmetric, i.e. $A_X(x, x') \neq A_X(x', x)$ for some $x, x' \in X$. The values $A_X(x, x')$ can be grouped in a matrix which, as it doesn't lead to confusion, we also denote as $A_X \in \mathbb{R}^{n\times n}$. A hierarchical clustering of the network $N = (X, A_X)$ is a dendrogram D_X which by definition is a nested set of partitions $D_X(\delta)$ indexed by the resolution parameter $\delta \geq 0$. Partitions in D_X are such that for $\delta = 0$ each point x is in a separate cluster, i.e., $D_X(0) = \big\{\{x\},\, x \in X\big\}$, and for some sufficiently coarse resolution δ_0 all nodes are in the same partition, i.e., $D_X(\delta_0) = \big\{X\big\}$. The requirement of nested partitions means that if x and x' are in the same partition at resolution δ_0 they stay co-clustered for all larger resolution $\delta > \delta_0$. From these requirements and a technical condition it follows that dendrograms can be represented as trees [13]. When x and x' are co-clustered at resolution δ in D_X we say that they are equivalent at that resolution and write $x \sim_{D_X(\delta)} x'$.

An ultrametric u_X on the space X is a function that satisfies the symmetry $u_X(x, x') = u_X(x', x)$ and identity $u_X(x, x') = 0 \iff x = x'$ properties as well as the strong triangle inequality

$$u_X(x, x') \leq \max\big(u_X(x, x''), u_X(x'', x')\big), \tag{1}$$

for all $x, x', x'' \in X$. For a given dendrogram D_X consider the minimum resolution δ at which x and x' are clustered together and define

$$u_X(x, x') := \min\big\{\delta \geq 0, x \sim_{D_X(\delta)} x'\big\}. \tag{2}$$

It can be shown that the function u_X satisfies (1) proving that dendrograms and finite ultrametric spaces are equivalent, [13, Theorem 9]. While dendrograms are useful graphical representations, ultrametrics are more convenient to present the results contained in this paper.

A hierarchical clustering method is a map $\mathcal{H} : \mathcal{N} \to \mathcal{D}$ from the space of networks $\mathcal{N}$ to the space of dendrograms $\mathcal{D}$, or, equivalently, a map $\mathcal{H} : \mathcal{N} \to \mathcal{U}$ mapping a network $\mathcal{H}(X, A_X) = (X, u_X)$ into the space $\mathcal{U}$ of networks with ultrametrics. Our goal here is to find methods $\mathcal{H}$ that abide to the following intuitive restrictions:

(A1) Axiom of Value. Consider a two-node network $N = (X, A_X)$ with $X = \{p, q\}$, $A_X(p, q) = \alpha$, and $A_X(q, p) = \beta$. The ultrametric $(X, u_X) = \mathcal{H}(X, A_X)$ produced by $\mathcal{H}$ satisfies

$$u_X(p, q) = \max(\alpha, \beta). \tag{3}$$

Supported by AFOSR MURI FA9550-10-1-0567 and NSF CCF-1217963.
G. Carlsson is with Dept. of Mathematics, Stanford University. F. Mémoli is with School of Computer Science, University of Adelaide. A. Ribeiro and S. Segarra are with Dept. of Electrical & Systems Eng., University of Pennsylvania. Email at gunnar@math.stanford.edu, facundo.memoli@adelaide.edu.au, aribeiro@seas.upenn.edu, and ssegarra@seas.upenn.edu.

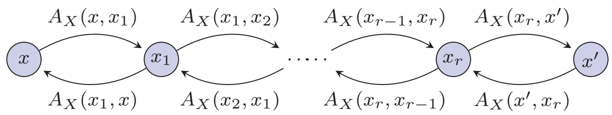

Fig. 1. Reciprocal clustering. Nodes x, x' cluster at resolution δ if they can be joined with a bidirectional chain of maximum dissimilarity δ [cf. (5)]. Reciprocal ultrametrics are largest among those produced by clustering methods satisfying the value and transformation axioms.

(A2) Axiom of Transformation. Given networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and a dissimilarity reducing map $\phi : X \to Y$, that is a map ϕ such that for all $x, x' \in X$ it holds $A_X(x, x') \geq A_Y(\phi(x), \phi(x'))$, the outputs $(X, u_X) = \mathcal{H}(X, A_X)$ and $(Y, u_Y) = \mathcal{H}(Y, A_Y)$ satisfy

$$u_X(x, x') \geq u_Y(\phi(x), \phi(x')). \tag{4}$$

Axiom (A1) says that in a network with two nodes p and q, the dendrogram D_X has them merging at the maximum value of the two dissimilarities $A_X(p, q) = \alpha$ and $A_X(q, p) = \beta$. This is reasonable because at resolutions $\delta < \max(\alpha, \beta)$ one node can influence the other but not vice versa, which in most situations means that the nodes are not alike. Axiom (A2) states that a contraction of the dissimilarity matrix A_X entails a contraction of the ultrametric u_X.

A hierarchical clustering method $\mathcal{H}$ is *admissible* if it satisfies axioms (A1) and (A2). Two admissible methods of interest are reciprocal and nonreciprocal clustering. The *reciprocal* clustering method $\mathcal{H}^{\text{R}}$ with output $(X, u_X^{\text{R}}) = \mathcal{H}^{\text{R}}(X, A_X)$ is the one for which the ultrametric $u_X^{\text{R}}(x, x')$ between points x and x' is given by

$$u_X^{\text{R}}(x, x') := \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \bar{A}_X(x_i, x_{i+1}), \tag{5}$$

where $\bar{A}_X(x, x') := \max(A_X(x, x'), A_X(x', x))$ for all $x, x' \in X$. In (5), the chain $C(x, x') = [x = x_0, x_1, \ldots, x_{r+1} = x']$ is defined as an ordered sequence of nodes linking x and x'. Definition (5) is illustrated in Fig. 1. Intuitively, search for chains $C(x, x')$ linking nodes x and x'. Then, for a given chain, walk from x to x' and determine the maximum dissimilarity, in either the forward or backward direction, across all links in the chain. The reciprocal ultrametric $u_X^{\text{R}}(x, x')$ is the minimum of this value across all possible chains.

Reciprocal clustering joins x to x' by going back and forth at maximum cost δ through the same chain. *Nonreciprocal* clustering $\mathcal{H}^{\text{NR}}$ permits different chains. Define the minimum directed cost as

$$\tilde{u}_X^{\text{NR}}(x, x') := \min_{C(x,x')} \max_{i | x_i \in C(x,x')} A_X(x_i, x_{i+1}), \tag{6}$$

and the nonreciprocal ultrametric as the maximum of the two minimum directed costs from x to x' and x' to x

$$u_X^{\text{NR}}(x, x') := \max\Big(\tilde{u}_X^{\text{NR}}(x, x'),\ \tilde{u}_X^{\text{NR}}(x', x)\Big). \tag{7}$$

Definition (7) is illustrated in Fig. 2. We consider forward chains $C(x, x')$ going from x to x' and backward chains $C(x', x)$ going from x' to x. We then determine the respective maximum dissimilarities and search independently for the best forward and backward chains that minimize the respective maximum dissimilarities. The nonreciprocal ultrametric $u_X^{\text{NR}}(x, x')$ is the maximum of these two minimum values. Observe that since reciprocal chains are particular cases of nonreciprocal chains we must have $u_X^{\text{NR}}(x, x') \leq u_X^{\text{R}}(x, x')$ for all pairs of nodes $x, x' \in X$.

Reciprocal and nonreciprocal clustering are of importance because they bound the range of ultrametrics generated by any other admissible method $\mathcal{H}$ in the sense stated in the following theorem.

Theorem 1 ([1]) *Consider an arbitrary network $N = (X, A_X)$ and let $u_X^{\text{R}}(x, x')$ and $u_X^{\text{NR}}(x, x')$ be the associated reciprocal and nonreciprocal ultrametrics as defined in* (5) *and* (7)*. Then, for any admissible method $\mathcal{H}$ the output ultrametric $(X, u_X) = \mathcal{H}(X, A_X)$ is such that for all pairs x, x',*

$$u_X^{\text{NR}}(x, x') \leq u_X(x, x') \leq u_X^{\text{R}}(x, x'). \tag{8}$$

In particular, $u_X^{NR} = u_X^{R}$ whenever $N = (X, A_X)$ is symmetric.

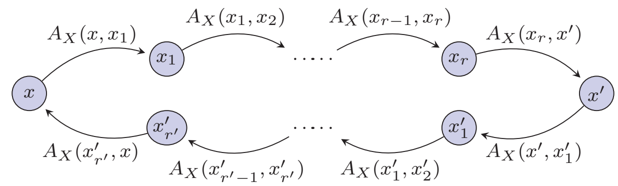

Fig. 2. Nonreciprocal clustering. Nodes x, x' cluster at resolution δ if they can be joined in both directions with possibly different chains of maximum dissimilarity δ [cf. (7)]. Nonreciprocal ultrametrics are smallest among those produced by clustering methods that satisfy the value and transformation axioms.

According to Theorem 1, nonreciprocal clustering yields uniformly minimal ultrametrics while reciprocal clustering yields uniformly maximal ultrametrics among all methods satisfying (A1)-(A2). Section III presents intermediate methods lying in the space between $\mathcal{H}^{\text{NR}}$ and $\mathcal{H}^{\text{R}}$. Section IV develops algorithms for the computation of u_X^{NR}, u_X^{R}, and the intermediate output ultrametrics of the methods derived in Section III.

III. Intermediate Clustering Methods

Fig. 3 shows an example network as well as its corresponding reciprocal and nonreciprocal dendrograms. Since these two are different, Theorem 1 allows the existence of intermediate admissible methods, which we study in this section.

A. Grafting and related constructions

A family of admissible methods can be constructed by grafting branches of the nonreciprocal dendrogram into corresponding branches of the reciprocal dendrogram; see Fig. 3. To be precise consider a given positive constant $\beta > 0$. For any given network $N = (X, A_X)$ compute the reciprocal and nonreciprocal dendrograms and cut all branches of the reciprocal dendrogram at resolution β. For each of these branches define the corresponding branch in the nonreciprocal tree as the one whose leaves are the same. Replacing the severed branches of the reciprocal tree by the corresponding branches of the nonreciprocal tree yields the $\mathcal{H}^{\text{R/NR}}(\beta)$ method. Grafting is equivalent to providing the following piecewise definition of the output ultrametric. For $x, x' \in X$ let

$$u_X^{\text{R/NR}}(x, x'; \beta) := \begin{cases} u_X^{\text{NR}}(x, x'), & \text{if } u_X^{\text{R}}(x, x') \leq \beta, \\ u_X^{\text{R}}(x, x'), & \text{if } u_X^{\text{R}}(x, x') > \beta. \end{cases} \tag{9}$$

For pairs x, x' having large reciprocal ultrametric $u_X^{\text{R}}(x, x') > \beta$ we keep the reciprocal ultrametric value $u_X^{\text{R/NR}}(x, x'; \beta) = u_X^{\text{R}}(x, x')$. For pairs x, x' with small reciprocal ultrametric $u_X^{\text{R}}(x, x') \leq \beta$ we replace the reciprocal by the nonreciprocal ultrametric and make $u_X^{\text{R/NR}}(x, x'; \beta) = u_X^{\text{NR}}(x, x')$.

To show that $\mathcal{H}^{\text{R/NR}}(\beta)$ is an admissible method we need to show that (9) defines an ultrametric in the space X and that $\mathcal{H}^{\text{R/NR}}(\beta)$ satisfies axioms (A1) and (A2). This is asserted in the following proposition.

Proposition 1 *The hierarchical clustering method $\mathcal{H}^{\text{R/NR}}(\beta)$ with ultrametrics as in* (9) *satisfies axioms (A1) and (A2).*

Proof: See [14]. ■

Since $u_X^{\text{R/NR}}(x, x'; \beta)$ coincides with either $u_X^{\text{NR}}(x, x')$ or $u_X^{\text{R}}(x, x')$, it then satisfies (8) as it should by the combination of Theorem 1 and Proposition 1.

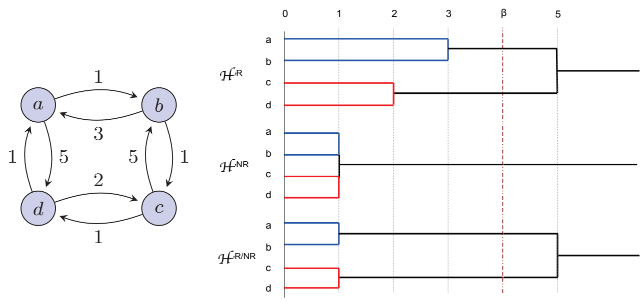

Fig. 3. Dendrogram grafting. Dendrograms resulting from three different clustering methods applied on the network on the left. Undrawn edges have dissimilarities greater than 5. The first two dendrograms correspond to the reciprocal and the nonreciprocal methods respectively. The third dendrogram is the result of grafting the first two as defined in (9) with $\beta = 4$. The third dendrogram is constructed by replacing the colored branches of the reciprocal dendrogram with the corresponding branches of the nonreciprocal dendrogram.

In the method $\mathcal{H}^{\text{R/NR}}(\beta)$ we use the reciprocal ultrametric as a decision variable and use nonreciprocal ultrametrics for nodes having small reciprocal ultrametrics. There are three other grafting combinations $\mathcal{H}^{\text{R/R}}(\beta)$, $\mathcal{H}^{\text{NR/R}}(\beta)$ and $\mathcal{H}^{\text{NR/NR}}(\beta)$ but none of them outputs a valid ultrametric. In $\mathcal{H}^{\text{R/R}}(\beta)$, we use reciprocal ultrametrics as decision variables and as the choice for small values of reciprocal ultrametrics,

$$u_X^{\text{R/R}}(x,x';\beta) := \begin{cases} u_X^{\text{R}}(x,x'), & \text{if } u_X^{\text{R}}(x,x') \leq \beta, \\ u_X^{\text{NR}}(x,x'), & \text{if } u_X^{\text{R}}(x,x') > \beta. \end{cases} \quad (10)$$

The method $\mathcal{H}^{\text{R/R}}(\beta)$ as defined in (10) is not valid, however, because the function $u_X^{\text{R/R}}(\beta)$ is not an ultrametric as it violates the strong triangle inequality in (1). E.g., focusing on the network in Figure 3, we use (10) to compute $u_X^{\text{R/R}}(a,b;4) = 3$, since $u_X^{\text{R}}(a,b) = 3$ and the reciprocal ultrametric between a and b is $u_X^{\text{R}}(a,b) \leq \beta = 4$. Similarly, $u_X^{\text{R/R}}(b,c;4) = 1$ and $u_X^{\text{R/R}}(a,c;4) = 1$. Hence, we obtain $u_X^{\text{R/R}}(a,b;4) > \max(u_X^{\text{R/R}}(a,c;4), u_X^{\text{R/R}}(b,c;4))$, violating (1).

In $\mathcal{H}^{\text{NR/NR}}(\beta)$ we use nonreciprocal ultrametrics as decision variables and as the choice for small values of nonreciprocal ultrametrics. In $\mathcal{H}^{\text{NR/R}}(\beta)$ nonreciprocal ultrametrics are used as decision variables and reciprocal ultrametrics are used for small values of nonreciprocal ultrametrics. Both of these methods can be seen to also violate the strong triangle inequality.

A second valid grafting alternative can be obtained as a modification of $\mathcal{H}^{\text{R/R}}(\beta)$ in which reciprocal ultrametrics are kept for pairs having small reciprocal ultrametrics, nonreciprocal ultrametrics are used for pairs having large reciprocal ultrametrics, but all nonreciprocal ultrametrics smaller than β are saturated to this value. Denoting the method as $\mathcal{H}^{\text{R/R}_{\max}}(\beta)$ the output ultrametrics are thereby given as

$$u_X^{\text{R/R}_{\max}}(x,x';\beta) := \begin{cases} u_X^{\text{R}}(x,x'), & \text{if } u_X^{\text{R}}(x,x') \leq \beta, \\ \max\left(\beta\,,\, u_X^{\text{NR}}(x,x')\right), & \text{if } u_X^{\text{R}}(x,x') > \beta. \end{cases} \quad (11)$$

This alternative definition entails a valid clustering method satisfying axioms (A1)-(A2) as we claim in the following proposition.

Proposition 2 *The hierarchical clustering method $\mathcal{H}^{\text{R/R}_{\max}}(\beta)$ with ultrametrics as in* (11) *satisfies axioms (A1) and (A2).*

Proof: See [14]. ■

Remark 1 The grafting combination $\mathcal{H}^{\text{R/NR}}(\beta)$ allows nonreciprocal propagation of influence for resolutions smaller than β while requiring reciprocal propagation for higher resolutions. This is of interest if we want tight clusters of small dissimilarity to be formed through loops of influence while looser clusters of higher dissimilarity are required to form through links of bidirectional influence. Conversely, the clustering method $\mathcal{H}^{\text{R/R}_{\max}}(\beta)$ requires reciprocal influence within tight clusters of resolution smaller than β but allows nonreciprocal influence in clusters of higher resolutions. This latter behavior is desirable in, e.g., trust propagation in social interactions, where we want tight clusters to be formed through links of mutual trust but allow looser clusters to be formed through unidirectional trust loops.

B. Convex combinations

Intermediate admissible methods can also be obtained by performing a convex combination of methods known to satisfy axioms (A1) and (A2). Indeed, consider two admissible clustering methods $\mathcal{H}^1$ and $\mathcal{H}^2$ and a given parameter $0 \leq \theta \leq 1$. For arbitrary network $N = (X, A_X)$ denote as $(X, u_X^1) = \mathcal{H}^1(N)$ and $(X, u_X^2) = \mathcal{H}^2(N)$ the respective outcome ultrametrics of methods $\mathcal{H}^1$ and $\mathcal{H}^2$. Construct then the dissimilarity function $A_X^{12}(\theta)$ as the convex combination of ultrametrics u_X^1 and u_X^2,

$$A_X^{12}(x,x';\theta) := \theta\, u_X^1(x,x') + (1-\theta)\, u_X^2(x,x'), \quad (12)$$

for all $x, x' \in X$. While the dissimilarity function $A_X^{12}(\theta)$ is not an ultrametric in general because it may violate the strong triangle inequality, we can recover the ultrametric structure by applying an admissible clustering method $\mathcal{H}$ to the network $N_\theta^{12} = (X, A_X^{12})$ to obtain $(X, u_X) = \mathcal{H}(N_\theta^{12})$. Notice however that the network N_θ^{12} is symmetric because the ultrametrics u_X^1 and u_X^2 are symmetric and that, in such case, by Theorem 1 reciprocal and nonreciprocal clustering yield the same outcome [1]. It then follows from (8) that the ultrametric u_X is independent of the admissible method $\mathcal{H}$ applied to N_θ^{12}. Thus, we define the convex combination method $\mathcal{H}_\theta^{12}$ as the one where the ultrametric $(X, u_X^{12}(\theta)) = \mathcal{H}_\theta^{12}(N)$ corresponding to $N = (X, A_X)$ is given by

$$u_X^{12}(x,x';\theta) := \min_{C(x,x')} \max_{i | x_i \in C(x,x')} A_X^{12}(x_i, x_{i+1}; \theta), \quad (13)$$

for all $x, x' \in X$ and A_X^{12} as given in (12). The operation in (13) is equivalent to the definition of single linkage applied to the symmetric network N_θ^{12}. It can be shown that (13) defines a valid ultrametric and fulfills axioms (A1) and (A2) as stated in the following proposition.

Proposition 3 *Given two admissible hierarchical clustering methods $\mathcal{H}^1$ and $\mathcal{H}^2$, the convex combination method $\mathcal{H}_\theta^{12}$ with ultrametrics as in* (13) *satisfies axioms (A1) and (A2).*

Proof: See [14]. ■

The construction in (13) can be generalized to a family of intermediate clustering methods generated by arbitrary convex combinations of reciprocal, nonreciprocal, members of the grafting family of Section III-A, members of the semi-reciprocal family to be introduced in Section III-C, or any other admissible method. These arbitrary combinations can be seen to satisfy axioms (A1) and (A2) through recursive application of Proposition 3.

Remark 2 Since (13) is equivalent to single linkage applied to the symmetric network N_θ^{12}, it follows that the ultrametric $u_X^{12}(\theta)$ in (13) is the largest ultrametric uniformly bounded by $A_X^{12}(\theta)$, i.e., the largest ultrametric for which $u_X^{12}(x,x';\theta) \leq A_X^{12}(x,x';\theta)$ for all pairs x, x'. We can then think of (13) as an operation ensuring a valid ultrametric definition while retaining as much information as possible in the convex combination of u_X^1 and u_X^2.

C. Semi-reciprocal ultrametrics

In reciprocal clustering we require influence to propagate through bidirectional chains; see Fig. 1. We could reinterpret bidirectional propagation as allowing loops of node length two in both directions. E.g., the bidirectional chain between x and x_1 in Fig. 1 can be interpreted as a loop between x and x_1 composed by two chains $[x, x_1]$ and $[x_1, x]$ of node

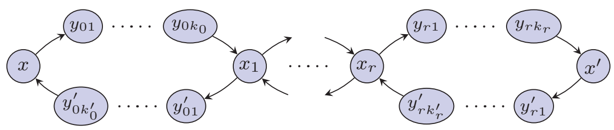

Fig. 4. Semi-reciprocal chains. The main chain joining x and x' is formed by $[x, x_1, ..., x_r, x']$. Between two consecutive nodes of the main chain x_i and x_{i+1}, we have a secondary chain in each direction $[x_i, y_{i1}, ..., y_{ik_i}, x_{i+1}]$ and $[x_{i+1}, y'_{i1}, ..., y'_{ik'_i}, x_i]$. For $u_X^{\mathrm{SR}(l)}(x,x')$, the maximum allowed node length of secondary chains is l, i.e., $k_i, k'_i \leq l-2$ for all i.

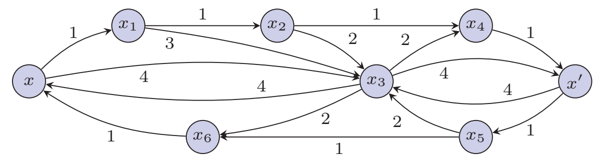

Fig. 5. Semi-reciprocal example. Computation of semi-reciprocal ultrametrics between nodes x and x' for different values of parameter l. $u_X^{\mathrm{SR}(2)}(x,x') = 4$, $u_X^{\mathrm{SR}(3)}(x,x') = 3$, $u_X^{\mathrm{SR}(4)}(x,x') = 2$ and $u_X^{\mathrm{SR}(l)}(x,x') = 1$ for all $l \geq 5$; see text for details.

length two. Semi-reciprocal clustering is a generalization of this concept where loops consisting of at most l nodes in each direction are allowed. Define as $C_l(x,x')$, $l \in \mathbb{N}$, a chain $[x = x_0, x_1, \ldots, x_{k-1} = x']$ where $k \in \mathbb{N}$, $2 \leq k \leq l$. In other words, $C_l(x,x')$ is a chain starting at x and finishing at x' with *at most* l nodes. For consistency, we require $l \geq 2$, since a chain joining two nodes must at least contain both extremes. We reserve the notation $C(x,x')$ to represent a chain linking x with x' where no maximum is imposed on the amount of nodes in the chain. Given an arbitrary network $N = (X, A_X)$, define as $A_X^{\mathrm{SR}(l)}(x,x')$ the minimum cost of going from node x to node x' using a chain of at most l nodes,

$$A_X^{\mathrm{SR}(l)}(x,x') := \min_{C_l(x,x')} \max_{k | x_k \in C_l(x,x')} A_X(x_k, x_{k+1}). \quad (14)$$

We define the family of semi-reciprocal clustering methods $\mathcal{H}^{\mathrm{SR}(l)}$ with output $(X, u_X^{\mathrm{SR}(l)}) = \mathcal{H}^{\mathrm{SR}(l)}(X, A_X)$ as the one for which the ultrametric value $u_X^{\mathrm{SR}(l)}(x,x')$ between points x and x' is

$$u_X^{\mathrm{SR}(l)}(x,x') := \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \bar{A}_X^{\mathrm{SR}(l)}(x_i, x_{i+1}) \quad (15)$$

where the function $\bar{A}_X^{\mathrm{SR}(l)}(x_i, x_{i+1})$ is defined as

$$\bar{A}_X^{\mathrm{SR}(l)}(x_i, x_{i+1}) := \max\left(A_X^{\mathrm{SR}(l)}(x_i, x_{i+1}), A_X^{\mathrm{SR}(l)}(x_{i+1}, x_i)\right).$$

The chain $C(x,x')$ of unconstrained length in (15) is denoted as the *main chain*, represented by $[x = x_0, x_1, ..., x_r, x_{r+1} = x']$ in Fig. 4. Between consecutive nodes of the main chain x_i and x_{i+1}, we build loops consisting of *secondary chains* in each direction, represented in Fig. 4 by $[x_i, y_{i1}, ..., y_{ik_i}, x_{i+1}]$ and $[x_{i+1}, y'_{i1}, ..., y'_{ik'_i}, x_i]$ for all i. For the computation of $u_X^{\mathrm{SR}(l)}(x,x')$, the maximum allowed length of secondary chains is equal to l nodes, i.e., $k_i, k'_i \leq l-2$ for all i. In particular, for $l = 2$ we recover the reciprocal chain depicted in Fig. 1.

We can reinterpret (15) as the application of reciprocal clustering [cf. (5)] to a network with dissimilarities $A_X^{\mathrm{SR}(l)}$ as in (14), i.e., a network with dissimilarities given by the optimal choice of secondary chains. Semi-reciprocal clustering methods are valid and satisfy axioms (A1)-(A2) as shown in the following proposition.

Proposition 4 *The semi-reciprocal clustering method $\mathcal{H}^{\mathrm{SR}(l)}$ with ultrametrics as in* (15) *satisfies axioms (A1) and (A2) for all integers $l \geq 2$.*

Proof: See [14]. ■

The semi-reciprocal is a countable family of clustering methods parameterized by integer l representing the allowed maximum node length of secondary chains. Reciprocal and nonreciprocal ultrametrics are equivalent to semi-reciprocal ultrametrics for specific values of l. For $l = 2$ we have $u_X^{\mathrm{SR}(2)} = u_X^{\mathrm{R}}$ meaning that we recover reciprocal clustering. To see this formally, note that $\tilde{u}_X^{\mathrm{SR}(2)}(x,x') = A_X(x,x')$ [cf. (14)] since the only chain of length two joining x and x' is $[x,x']$. Hence, for the case where $l = 2$, (15) reduces to

$$u_X^{\mathrm{SR}(2)}(x,x') = \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \bar{A}_X(x_i, x_{i+1}), \quad (16)$$

which is the definition of the reciprocal ultrametric [cf. (5)]. Nonreciprocal ultrametrics can be obtained as $u_X^{\mathrm{SR}(l)} = u_X^{\mathrm{NR}}$ for any parameter $l \geq n$ exceeding the number of nodes in the network. To see this, notice that minimizing over $C(x,x')$ is equivalent to minimizing over $C_l(x,x')$ for all $l \geq n$, since we are looking for minimizing chains in a network with non negative weights. Therefore, visiting the same node twice is not an optimal choice. This implies that $C_n(x,x')$ contains all possible minimizing chains between x and x'. In other words, all chains of interest have at most n nodes or $n-1$ hops. Hence, by inspecting (14), $\tilde{u}_X^{\mathrm{SR}(l)}(x,x') = \tilde{u}_X^{\mathrm{NR}}(x,x')$ [cf. (6)] for all $l \geq n$. Furthermore, when $l \geq n$, the best main chain that can be picked is formed only by nodes x and x' because, in this way, no additional meeting point is enforced between the chains going from x to x' and vice versa. As a consequence, definition (15) reduces to

$$u_X^{\mathrm{SR}(l)}(x,x') = \max\left(\tilde{u}_X^{\mathrm{NR}}(x,x'), \tilde{u}_X^{\mathrm{NR}}(x',x)\right), \quad (17)$$

for all $l \geq n$. The right hand side of (17) is the definition of the nonreciprocal ultrametric [cf. (7)].

For the network in Fig. 5, we calculate the semi-reciprocal ultrametrics between x and x' for different values of l. Edges which have not been drawn have dissimilarity values greater than the ones depicted in the figure. Since the only bidirectional chain between x and x' uses x_3 as the intermediate node, we conclude that $u_X^{\mathrm{R}}(x,x') = u_X^{\mathrm{SR}(2)}(x,x') = 4$. Furthermore, by constructing a path through the outmost clockwise cycle in the network, we conclude that $u_X^{\mathrm{NR}}(x,x') = 1$. Since the longest secondary chain in the minimizing path for the nonreciprocal case, $[x, x_1, x_2, x_4, x']$, has length 5, we may conclude that $u_X^{\mathrm{SR}(l)}(x,x') = 1$ for all $l \geq 5$. For intermediate values of l, if e.g., we fix $l = 3$, the minimizing path is given by the main chain $[x, x_3, x']$ and the secondary chains $[x, x_1, x_3]$, $[x_3, x_4, x']$, $[x', x_5, x_3]$ and $[x_3, x_6, x]$ joining consecutive nodes in the main chain in both directions. The maximum cost among all dissimilarities in this path is $A_X(x_1, x_3) = 3$. Hence, $u_X^{\mathrm{SR}(3)}(x,x') = 3$. The minimizing path for $l = 4$ is similar to the minimizing one for $l = 3$ but replacing the secondary chain $[x, x_1, x_3]$ by $[x, x_1, x_2, x_3]$. In this way, we obtain $u_X^{\mathrm{SR}(4)}(x,x') = 2$.

Remark 3 When propagating influence through a network, reciprocal clustering requires bidirectional influence whereas nonreciprocal clustering allows arbitrarily large unidirectional cycles. In many applications, such as trust propagation in social networks, it is reasonable to look for an intermediate situation where influence can propagate through cycles but of limited length. Semi-reciprocal ultrametrics represent this intermediate situation with parameter l accounting for the size of the influence cycles permitted.

IV. Algorithms

In this section we interpret A_X as a given matrix of dissimilarities and u_X as a symmetric matrix with entries corresponding to the ultrametric $u_X(x,x')$. As per (5), reciprocal clustering searches for chains that minimize the maximum dissimilarity in the symmetric matrix $\bar{A}_X := \max(A_X, A_X^T)$. This is equivalent to finding chains in $\bar{A}_X$ that have minimum cost as measured in the infinity norm. Likewise, nonreciprocal clustering searches for directed chains of minimum infinity norm cost in A_X to construct the matrix $\tilde{u}_X$ [cf. (6)] and selects the maximum of the directed costs by performing the operation $u_X^{\mathrm{NR}} = \max(\tilde{u}_X, \tilde{u}_X^T)$ [cf. (7)]. These operations can be performed algorithmically using matrix powers in the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ [15].

In the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ the regular sum is replaced by the minimization operator and the regular product by maximization. Using $\oplus$ and $\otimes$ to denote sum and product respectively on this dioid algebra we have $a \oplus b := \min(a,b)$ and $a \otimes b := \max(a,b)$. The matrix product $A \otimes B$ is therefore given by the matrix with entries

$$\left[A \otimes B\right]_{i,j} = \bigoplus_{k=1}^{n} \left(A_{i,k} \otimes B_{k,j}\right) = \min_{k \in [1,n]} \max\left(A_{i,k}, B_{k,j}\right). \quad (18)$$

From the definition in (18), it follows that for given matrix A the lth dioid power $A^{(l)}$ is such that its i,j entry $[A^{(l)}]_{i,j}$ represents the minimum infinity norm cost of a chain containing at most l hops. As discussed in Section III-C, we can restrict candidate minimizing chains to those with at most $n-1$ hops, entailing the following result.

Proposition 5 *For given network $N=(X,A_X)$ with n nodes the reciprocal ultrametric u_X^{R} defined in* (5) *can be computed as*

$$u_X^{\mathrm{R}} = \Big(\max\left(A_X, A_X^T\right) \Big)^{(n-1)}, \tag{19}$$

where the operation $(\cdot)^{(n-1)}$ denotes the $(n-1)$st matrix power in the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ with matrix product as defined in (18)*. The nonreciprocal ultrametric u_X^{NR} defined in* (7) *can be computed as*

$$u_X^{\mathrm{NR}} = \max\Big(A_X^{(n-1)}, \left(A_X^T\right)^{(n-1)}\Big). \tag{20}$$

Proof: See [14]. ■

For the reciprocal ultrametric we symmetrize dissimilarities with a maximization operation and take the $(n-1)$st power of the resulting matrix on the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$. For the nonreciprocal ultrametric we revert the order of these two operations. We first consider matrix powers $A_X^{(n-1)}$ and $\left(A_X^T\right)^{(n-1)}$ of the dissimilarity matrix and its transpose which we then symmetrize to the maximum. Besides emphasizing the relationship between reciprocal and nonreciprocal clustering, Proposition 5 suggests the existence of intermediate methods in which we raise dissimilarity matrices A_X and A_X^T to some power, perform a symmetrization, and then continue matrix multiplications. These procedures yield methods that are not only valid but coincide with the family of semi-reciprocal ultrametrics introduced in Section III-C as the following proposition asserts.

Proposition 6 *For a given network $N=(X,A_X)$ with n nodes the lth semi-reciprocal ultrametric $u_X^{\mathrm{SR}(l)}$ in* (15) *can be computed as*

$$u_X^{\mathrm{SR}(l)} = \Big(\max\left(A_X^{(l-1)}, (A_X^T)^{(l-1)}\right) \Big)^{(n-1)}. \tag{21}$$

where $(\cdot)^{(l-1)}$ and $(\cdot)^{(n-1)}$ denote matrix powers in the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ with matrix product as defined in (18)*.*

Proof: See [14]. ■

The result in (21) is intuitive. The powers $A_X^{(l-1)}$ and $\left(A_X^T\right)^{(l-1)}$ represent the minimum infinity norm cost among directed chains of at most $l-1$ hops. In terms of Section III-C, these are the cost of the optimal secondary chains of at most l nodes. Therefore the maximization $\max\left(A_X^{(l-1)}, \left(A_X^T\right)^{(l-1)}\right)$ computes the cost in both directions of joining two given nodes with secondary chains of at most l nodes, i.e. $\bar{A}_X^{\mathrm{SR}(l)}$ in (15). Applying the dioid power $(n-1)$ to this new matrix is equivalent to looking for minimizing chains in the network with costs given by the secondary chains, i.e., the dioid power computes the cost of the optimal main chain, as described in Section III-C. Observe that we recover (19) by making $l=2$ in (21). Also, it can be shown that (20) is equivalent to (21) when $l=n$. Thus, the results in propositions 5 and 6 further emphasize the extremal nature of the reciprocal and nonreciprocal methods and characterize the semi-reciprocal ultrametrics as natural intermediate clustering methods in an algorithmic sense.

This algorithmic perspective allows for a generalization in which the powers of the matrices A_X and A_X^T are different. To be precise, consider strictly positive integers $l, l' > 0$ and define the algorithmic intermediate clustering method $\mathcal{H}^{l,l'}$ with parameters l, l' as the one that maps the given network $N=(X,A_X)$ to the output ultrametric $(X, u_X^{l,l'}) = \mathcal{H}^{l,l'}(N)$ given by

$$u_X^{l,l'} = \Big(\max\left(A_X^{(l)}, (A_X^T)^{(l')}\right) \Big)^{(n-1)}. \tag{22}$$

The ultrametric (22) can be interpreted as a semi-reciprocal ultrametric where the allowed length of secondary chains varies with the direction. Forward secondary chains may have at most $l+1$ nodes whereas backward secondary chains may have at most $l'+1$ nodes. The algorithmic intermediate family $\mathcal{H}^{l,l'}$ encapsulates the semi-reciprocal family since $\mathcal{H}^{l,l} \equiv \mathcal{H}^{\mathrm{SR}(l+1)}$ as well as the reciprocal method since $\mathcal{H}^{\mathrm{R}} \equiv \mathcal{H}^{1,1}$ as it follows from comparison of (22) with (21) and (19), respectively. It can also be shown that $\mathcal{H}^{\mathrm{NR}}(X,A_X) = \mathcal{H}^{n-1,n-1}(X,A_X)$ for all networks with $|X| \leq n$. The intermediate algorithmic methods $\mathcal{H}^{l,l'}$ are admissible as we claim in the following proposition.

Proposition 7 *The hierarchical clustering method $\mathcal{H}^{l,l'}$ with ultrametrics as in* (22) *satisfies axioms (A1) and (A2).*

Proof: See [14]. ■

Algorithms to compute ultrametrics associated with the grafting families in Section III-A entail combinations of matrices u_X^{R} and u_X^{NR}. E.g., ultrametrics in (9) corresponding to the grafting method $\mathcal{H}^{R/NR}(\beta)$ can be computed as

$$u_X^{R/NR}(\beta) = u_X^{\mathrm{NR}} \circ \mathbb{I}\left\{u_X^{\mathrm{R}} \leq \beta\right\} + u_X^{\mathrm{R}} \circ \mathbb{I}\left\{u_X^{\mathrm{R}} > \beta\right\}, \tag{23}$$

where $A \circ B$ denotes the Hadamard product of matrices A and B and $\mathbb{I}\{\cdot\}$ is an element wise indicator function which outputs a matrix with a 1 in the positions satisfying the condition and a 0 otherwise.

Algorithms for the convex combination family in Section III-B involve computing dioid algebra powers of a convex combination of ultrametric matrices. Given two admissible methods with output ultrametrics $(X, u_X^1) = \mathcal{H}^1(N)$ and $(X, u_X^2) = \mathcal{H}^2(N)$, and a scalar $0 \leq \theta \leq 1$, the ultrametric in (13) corresponding to the method $\mathcal{H}^{12}(\theta)$ can be computed as

$$u_X^{12}(\theta) = \left(\theta\, u_X^1 + (1-\theta)\, u_X^2\right)^{(n-1)}. \tag{24}$$

Remark 4 It follows from (19), (20) and (21) that methods in this paper are computationally tractable as the total number of operations is of order n^4. This complexity can be reduced to $n^3 \log n$ by noting that the dioid matrix power A^n can be computed with the sequence $A, A^2, A^4, \ldots$, which requires $o(logn)$ matrix products at a cost of $o(n^3)$ each.

Remark 5 In the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$, it can be shown that a matrix A satisfies the strong triangle inequality if and only if $A = A^{(2)}$ [15]. Also, for a nonnegative matrix with null diagonal, $A^{(n-1)} = A^{(n)}$. Hence, the dioid matrix powers in (19)-(22) and (24) ensure a valid ultrametric definition.

REFERENCES

[1] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," *Proc. Int. Conf. Acoustics Speech Signal Process*, 2013.

[2] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*, ser. Prentice Hall Advanced Reference Series. Englewood Cliffs, NJ: Prentice Hall Inc., 1988.

[3] X. Rui and D. Wunsch-II, "Survey of clustering algorithms," *IEEE Trans. Neural Netw.*, vol. 16, no. 3, pp. 645–678, May 2005.

[4] D. Zhou, B. Scholkopf, and T. Hofmann, "Semi-supervised learning on directed graphs," *Advances in Neural Information Processing Systems*, 2005.

[5] W. Pentney and M. Meila, "Spectral clustering of biological sequence data," *Proc. Ntnl. Conf. Artificial Intel.*, 2005.

[6] M. Meila and W. Pentney, "Clustering by weighted cuts in directed graphs," *Proceedings of the 7th SIAM International Conference on Data Mining*, 2007.

[7] U. von Luxburg and S. Ben-David, "Towards a statistical theory of clustering." PASCAL workshop on clustering, London, Tech. Rep., 2005.

[8] I. Guyon, U. von Luxburg, and R. Williamson, "Clustering: Science or art?" Paper presented at the NIPS 2009 Workshop Clustering: Science or Art?, Tech. Rep., 2009.

[9] M. Ackerman and S. Ben-David, "Measures of clustering quality: A working set of axioms for clustering," *Proceedings of Neural Information Processing Systems*, 2008.

[10] G. Carlsson and F. Mémoli, "Classifying clustering schemes," *Foundations of Computational Mathematics*, vol. 13, no. 2, pp. 221–252, April 2013.

[11] J. M. Kleinberg, "An impossibility theorem for clustering," in *NIPS*, S. Becker, S. Thrun, and K. Obermayer, Eds. MIT Press, 2002, pp. 446–453.

[12] G. N. Lance and W. T. Williams, "A general theory of classificatory sorting strategies 1. Hierarchical systems," *Computer Journal*, vol. 9, no. 4, pp. 373–380, Feb. 1967.

[13] G. Carlsson and F. Mémoli, "Characterization, stability and convergence of hierarchical clustering methods," *Journal of Machine Learning Research*, vol. 11, pp. 1425–1470, 2010.

[14] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Axiomatic construction of hierarchical clustering in asymmetric networks," *https://fling.seas.upenn.edu/~ssegarra/wiki/index.php?n=Research.Publications*, 2012.

[15] M. Gondran and M. Minoux, *Graphs, dioids and semi rings: New models and algorithms*. Springer, 2008.