Admissible Hierarchical Clustering Methods and Algorithms for Asymmetric Networks

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

Abstract—This paper characterizes hierarchical clustering methods that abide by two previously introduced axioms—thus, denominated admissible methods—and proposes tractable algorithms for their implementation. We leverage the fact that, for asymmetric networks, every admissible method must be contained between reciprocal and nonreciprocal clustering, and describe three families of intermediate methods. Grafting methods exchange branches between dendrograms generated by different admissible methods. The convex combination family combines admissible methods through a convex operation in the space of dendrograms, and third, the semireciprocal family clusters nodes that are related by strong cyclic influences in the network. An algorithmic framework for the computation of hierarchical clusters generated by reciprocal and nonreciprocal clustering as well as the grafting, convex combination and semireciprocal families is presented via matrix operations in a dioid algebra. Finally, the introduced clustering methods and algorithms are exemplified through their application to a network describing the interrelation between sectors of the United States (U.S.) economy.

Index Terms—Hierarchical clustering, asymmetric network, directed graph, dioid matrix algebra, axiomatic framework.

I. INTRODUCTION

THE relevance of clustering in modern data analysis is indubitable given its usage in multiple fields of knowledge such as genetics [2], computer vision [3], and sociology [4]. There are literally hundreds of methods that can be applied to the determination of hierarchical [5], [6] and non-hierarchical clusters in finite metric (thus symmetric) spaces—see, e.g., [7]. Even in the case of asymmetric networks [8], where the distance or dissimilarity from a node \( x \) to another node \( x' \) need not coincide with the dissimilarity from \( x' \) to \( x \), multiple methods have been developed to extend the notion of clustering into this less intuitive domain [9]–[14]. Although not as developed as its practice [15], the theoretical framework for clustering has been developed over the last decade for non-hierarchical [16]–[22] and hierarchical clustering [23]–[26]. Of special interest to us is this last direction where it has been shown in [23] that single linkage [6, Ch. 4] is the unique hierarchical clustering method for finite metric spaces that satisfies three reasonable axiomatic statements.

The output of a hierarchical clustering method is called a dendrogram, which can be understood as a nested collection of partitions indexed by a non-negative resolution parameter. Intuitively, at a resolution equal to zero, every node in the network is deemed as a singleton cluster whereas for a large enough resolution the whole network is seen as one large cluster. For intermediate resolutions, we have a successive agglomeration of the nodes into coarser and coarser clusters. Regarding hierarchical clustering of asymmetric networks, our work in [24] introduces the axioms of value—in a network with two nodes, the nodes cluster together at resolutions larger than the dissimilarities in both directions—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 clustering methods. Although weak in appearance, these axioms lead to the stringent result that all methods that abide by them—denominated admissible methods—must lie between two particular clustering methods in a well-defined sense. The first method, reciprocal clustering, requires clusters to form through edges exhibiting low dissimilarity in both directions whereas the second method, nonreciprocal clustering, allows clusters to form through cycles of small dissimilarity. When restricted to symmetric networks, reciprocal and nonreciprocal clustering yield equivalent outputs, which coincide with the output of single linkage.

The difference between reciprocal and nonreciprocal clustering for general asymmetric networks allows the existence of intermediate admissible methods. Hence, the contribution of this paper is twofold. First, we characterize intermediate clustering methods and study their properties. Second, we propose an algorithmic framework based on an alternative matrix dioid...
algebra to implement the intermediate methods introduced as well as reciprocal and nonreciprocal clustering. From an application-driven perspective, our motivation is to show that if one is interested in selecting a clustering method and considers the axioms of value and transformation to be reasonable properties, then one is not constrained to a binary choice between reciprocal and nonreciprocal clustering but rather there is a wide gamut of possible methods.

In Section III we unveil three families of intermediate clustering methods. The grafting methods consist of attaching the clustering output structures of the reciprocal and nonreciprocal methods in a way such that admissibility is guaranteed (Section III-A). We further present a construction that can be regarded as a convex combination in the space of clustering methods. This operation is shown to preserve admissibility therefore giving rise to a second family of admissible methods (Section III-B). A third family of admissible clustering methods is defined in the form of semi-reciprocal methods that allow the formation of cyclic influences in a more restrictive sense than nonreciprocal clustering but more permissive than reciprocal clustering (Section III-C).

In Section IV, we develop algorithms to compute the dendrograms associated with the methods introduced throughout the paper. The determination of algorithms for all of the methods introduced is given by the computation of matrix powers in a min-max dioid algebra [27]. In this algebra we operate in the field of positive reals and define the addition operation between two scalars to be their minimum and the product operation of two scalars to be their maximum. From this definition it follows that the \((i, j)\)-th entry of the \(l\)-th dioid power of a matrix of network dissimilarities represents the minimum cost of a chain linking node \(i\) to node \(j\) with at most \(l\) edges. Since reciprocal and nonreciprocal clustering require the determination of chains of minimax cost, their implementation can be framed in terms of dioid matrix powers. Similarly, other clustering methods introduced in this paper can be interpreted as minimax chain costs of a previously modified matrix of dissimilarities. Dioid matrices provide a unified framework to represent all the methods studied in Section III and to uncover the existence of additional admissible clustering methods.

Clustering methods are exemplified through their application to a real-world network representing the interactions between economic sectors of the U.S. economy (Section V). The purpose of this application is to understand which information can be extracted by performing hierarchical clustering analyses based on the different methods proposed. While the bidirectional influence required for cluster formation in reciprocal clustering might be too restrictive, nonreciprocal clustering propagates influence through arbitrarily large cycles, a feature which might be undesirable in practice. We illustrate how intermediate behaviors can be obtained by implementing the clustering methods here developed. Concluding remarks in Section VI close the paper.

II. PRELIMINARIES

We define a network \(N = (X, A_X)\) as a set of \(n\) points or nodes \(X\) jointly specified with a real-valued dissimilarity function \(A_X : X \times X \to \mathbb{R}_+\). Dissimilarities \(A_X(x, x')\) from \(x\) to \(x'\) are non-negative, 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 does not lead to confusion, we also denote by \(A_X \in \mathbb{R}^{n \times n}\). 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., \(D_X(0) = \{\{x\}, x \in X\}\), and for some sufficiently coarse resolution \(\delta_0\) all nodes are in the same cluster, i.e., \(D_X(\delta_0) = \{X\}\).

The requirement of nested partitions means that if \(x\) and \(x'\) are in the same cluster at resolution \(\delta\) they stay co-clustered for all larger resolutions \(\delta' > \delta\). From these requirements it follows that dendrograms can be represented as trees [23]; see, e.g., Fig. 6(a). When \(x\) and \(x'\) are co-clustered at resolution \(\delta\) in \(D_X\) we say that they are equivalent at that resolution and write \(x \sim_{D_X(\delta)} x'\).

An ultrametric \(u_X : X \times X \to \mathbb{R}_+\) on the set \(X\) 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

\[ u_X(x, x') \leq \max \{u_X(x, x''), u_X(x'', x')\}, \tag{1} \]

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

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

It can be shown that the function \(u_X\) in (2) satisfies (1). Conversely, given an ultrametric \(u_X\), one can define a collection of equivalence relations \(\sim_\delta\) indexed by \(\delta \geq 0\) as \(x \sim_\delta x'\) if and only if \(u_X(x, x') \leq \delta\). It can also be shown that the collection of partitions \(D_X(\delta)\) associated to the equivalence relations \(\sim_\delta\) constitutes a valid dendrogram, thus proving an equivalence between dendrograms and finite ultrametrics [23, Theorem 9]. While dendrograms are useful graphical representations, ultrametrics are more convenient to present the results contained in this paper.

In the description of hierarchical clustering methods the concepts of chain and chain cost are important. Given a network \((X, A_X)\) and \(x, x' \in X\), a chain from \(x\) to \(x'\) is any ordered sequence of nodes \([x = x_0, \ldots, x_{l-1}, x_l = x']\) starting at \(x\) and finishing at \(x'\). We use the notation \(C(x, x')\) to denote one such chain. We define the cost of a chain as the maximum dissimilarity encountered when traversing its links in order. Thus, the directed minimum chain cost \(\tilde{u}_X^*(x, x')\) between \(x\) and \(x'\) is then defined as the minimum cost among all chains connecting \(x\) to \(x'\),

\[ \tilde{u}_X^*(x, x') := \min_{C(x, x')} \max_{i \mid i \in C(x, x')} A_X(x_i, x_{i+1}). \tag{3} \]

We further define a loop as a chain of the form \(C(x, x)\) for some \(x \in X\) such that \(C(x, x)\) contains at least one node other than \(x\). Furthermore, we define the minimum loop cost \(\text{mlc}(X, A_X)\) of a network \((X, A_X)\) as the minimum across all possible loops
of each individual loop cost,

$$\text{mlc}(X, A_X) := \min_x \min_{C(x, x)} \max_{x \in C(x, x)} A_X(x, x_{i+1}),$$  \hspace{1cm} (4)$$

where \(C(x, x)\) contains at least one node different from \(x\).

A hierarchical clustering method is a map \(H : N \to D\) from the set of networks \(N\) to the set of dendromaps \(D\), or, equivalently, a map \(H : N \to U\) mapping each network \(N\) into the set \(U\) of networks with ultrametrics as dissimilarity functions, i.e., \(H(N) = (X, u_X)\). Our goal is to find methods \(H\) that satisfy the following intuitive restrictions:

1. **Axiom of Value.** Given a two-node network \(N = ((p, q), \Lambda_{p,q})\) with \(\Lambda_{p,q}(p, q) = \alpha\) and \(\Lambda_{p,q}(q, p) = \beta\), the ultrametric \((X, u_{p,q}) = H(N)\) output by \(H\) satisfies

$$u_{p,q}(p, q) = \max(\alpha, \beta).$$  \hspace{1cm} (5)$$

2. **Axiom of Transformation.** Given networks \(N_Y = (X, A_X)\) and \(N_Y = (Y, A_Y)\) and a dissimilarity reducing map \(\phi : X \to Y\), i.e., a map \(\phi\) such that for all \(x, x' \in X\) it holds \(A_X(x, x') \geq A_Y(\phi(x), \phi(x'))\), the outputs \((X, u_X) = H(N_X)\) and \((Y, u_Y) = H(N_Y)\) satisfy

$$u_X(x, x') \geq u_Y(\phi(x), \phi(x')).$$  \hspace{1cm} (6)$$

We say that node \(x\) is able to influence node \(x'\) at resolution \(\delta\) if the dissimilarity from \(x\) to \(x'\) is not greater than \(\delta\). Notice that the interpretation of the influence relation depends on the meaning encoded in the dissimilarities. E.g., if the dissimilarities represent some measure of distrust in a social network, then the meaning of the influence relation is closer to our usual understanding of influence whereas in the economic example in Section V, influence is more related with dependence in the productive process. In two-node networks, our intuition dictates that a cluster is formed if nodes \(p\) and \(q\) are able to influence each other. Thus, axiom (A1) states that in a network with two nodes, the dendrogram \(D_X\) has them merging at the maximum value of the two dissimilarities between them. Axiom (A2) captures the intuition that if a network is transformed such that some nodes become more similar but no pair of nodes increases its dissimilarity, then the transformed network should cluster at lower resolutions than the original one. Formally, (A2) states that a contraction of the dissimilarity function \(A_X\) entails a contraction of the associated ultrametric \(u_X\). Notice, however, that map \(\phi\) in (A2) need not be neither surjective nor injective.

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

$$u^R_X(x; x') := \min_{C(x, x')} \max_{x \in C(x, x')} A_X(x, x_{i+1}),$$  \hspace{1cm} (7)$$

where \(A_X(x, x') := \max(A_X(x, x'), A_X(x', x))\). Definition (7) is illustrated in Fig. 1. Intuitively, we search for chains \(C(x, x')\) linking nodes \(x\) and \(x'\). For a given chain we walk from \(x\) to \(x'\) and for every link, connecting say \(x_i\) with \(x_{i+1}\), we determine the maximum dissimilarity in both directions, i.e. the value of \(A_X(x_i, x_{i+1})\). We then determine the maximum across all the links in the chain. The reciprocal ultrametric \(u^R_X(x; x')\) between \(x\) and \(x'\) is the minimum of this value across all possible chains.

Reciprocal clustering joins \(x\) to \(x'\) by going back and forth at maximum cost \(\delta\) through the same chain. **Nonreciprocal clustering** \(H^{NR}\) permits different chains and is defined as the maximum of the two minimum directed costs [cf. (3)] from \(x\) to \(x'\) and \(x'\) to \(x\):

$$u^{NR}_X(x; x') := \max\left(\bar{u}^{NR}_X(x; x'), \bar{u}^{NR}_X(x'; x)\right).$$  \hspace{1cm} (8)$$

Definition (8) 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 these maximum dissimilarities. The nonreciprocal ultrametric \(u^{NR}_X(x; x')\) is the maximum of these two minimum values. Observe that since reciprocal chains are particular cases of nonreciprocal chains we must have \(u^{NR}_X(x; x') \leq u^R_X(x; x')\) 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 \(H\) in the sense stated next.

**Theorem 1 (24], [28]):** Consider an arbitrary network \(N = (X, A_X)\) and let \(u^R_X\) and \(u^{NR}_X\) be the associated reciprocal and nonreciprocal ultrametrics as defined in (7) and (8). Then, for any admissible method \(H\) the output ultrametric \((X, u_X) = H(X, A_X)\) is such that for all pairs \(x, x'\),

$$u^{NR}_X(x; x') \leq u_X(x; x') \leq u^R_X(x; x').$$  \hspace{1cm} (9)$$

In particular, \(u^{NR}_X = u^R_X\) whenever \((X, A_X)\) is symmetric.
Proof: For completeness, a short version of the proof is included in Appendix A.

According to Theorem 1, nonreciprocal clustering yields uniformly minimal ultrametrics while reciprocal clustering yields uniformly maximal ultrametrics among all methods satisfying (A1)-(A2). For symmetric networks, reciprocal and nonreciprocal clustering coincide, implying that there is a unique admissible method which is equivalent to the well-known single linkage hierarchical clustering method [6, Ch. 4].

III. INTERMEDIATE CLUSTERING METHODS

Reciprocal and nonreciprocal clustering bound the range of methods satisfying axioms (A1)-(A2) in the sense specified by Theorem 1. Since \( \mathcal{H}^R \) and \( \mathcal{H}^{NR} \) are in general different, a question of interest is whether one can identify methods which are intermediate to \( \mathcal{H}^R \) and \( \mathcal{H}^{NR} \). We present three types of intermediate methods: grafting, convex combinations, and semi-reciprocal clustering. A summary of these methods along with their algebraic description introduced in Section IV is presented in Table I.

A. Grafting

A family of admissible methods can be constructed by grafting branches of the nonreciprocal dendrogram into corresponding branches of the reciprocal dendrogram. 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 \( \beta \). For each of these branches define the corresponding branch in the nonreciprocal tree as the one whose leaves are the same. Replacing the previously cut branches of the reciprocal tree by the corresponding branches of the nonreciprocal tree yields the \( \mathcal{H}^{R/NR}(\beta) \) method. Grafting is equivalent to providing the following piecewise definition of the output ultrametric

\[
\begin{align*}
    u_X^R(x, x'; \beta) &= \begin{cases} 
    u_X^{NR}(x, x'), & \text{if } u_X^R(x, x') \leq \beta, \\
    u_X^R(x, x'), & \text{if } u_X^R(x, x') > \beta.
    \end{cases}
\end{align*}
\]

For pairs \( x, x' \in X \) having large reciprocal ultrametric value we keep this value, whereas for pairs with small reciprocal ultrametric value, we replace it by the nonreciprocal one.

To prove admissibility, we need to show that (10) defines an ultrametric and that the method \( \mathcal{H}^{R/NR}(\beta) \) satisfies axioms (A1) and (A2). This is asserted in the following proposition.

Proposition 1: The hierarchical clustering method \( \mathcal{H}^{R/NR}(\beta) \) is valid and admissible. I.e., \( u_X^{R/NR}(\beta) \) defined in (10) is a valid ultrametric and \( \mathcal{H}^{R/NR}(\beta) \) satisfies axioms (A1)-(A2).

\( \blacksquare \)

Proof: See Appendix B.

Fig. 3. Dendrogram grafting. Reciprocal (\( \mathcal{H}^R \)), nonreciprocal (\( \mathcal{H}^{NR} \)), and grafting (\( \mathcal{H}^{R/NR}(\beta = 4) \)) dendrograms for the given network are shown – edges not drawn have dissimilarities greater than 5. To form the latter, branches of the reciprocal dendrogram are cut at resolution \( \beta = 4 \) and replaced by the corresponding branches of the nonreciprocal dendrogram.

However, the method \( \mathcal{H}^{R/R}(\beta) \) is not valid because for some networks the function \( u_X^{R/R}(\beta) \) is not an ultrametric as it violates the strong triangle inequality in (1). As a counterexample consider again the network in Fig. 3. Applying the definition in (11) we
obtain that \( u_X^{R/R}(a, b; 4) = u_X^R(a, b) = 3 \) while \( u_X^{R/R}(a, c; 4) = u_X^{R/R}(a, c) = 1 \) and similarly \( u_X^{R/R}(b, c; 4) = 1 \). In turn, this implies that \( u_X^{R/R}(a, b; 4) \geq \max(u_X^{R/R}(a, c; 4), u_X^{R/R}(c, b; 4)) \) violating the strong triangle inequality. Analogously, \( u_{NR}^{R/R}(\theta) \) and \( u_{NR}^{R/R}(\theta) \) can also be shown to be invalid clustering methods.

A second valid grafting alternative can be obtained as a modification of \( u_{R/R}(\theta) \) in which reciprocal ultrametrics are kept for pairs having small reciprocal ultrametrics, nonreciprocal ultrametrics are used for pairs having large reciprocal ultrametrics, but whenever a nonreciprocal ultrametric is smaller than \( \beta \), this value is used instead. Denoting the method by \( u_{NR}^{R/R}(\theta) \) the output ultrametrics are thereby given as

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

(12)

This alternative definition outputs a valid ultrametric and \( u_{NR}^{R/R}(\theta) \) satisfies axioms (A1)-(A2).

**Proposition 2:** The method \( u_{NR}^{R/R}(\theta) \) is valid and admissible, i.e., \( u_{NR}^{R/R} \) defined in (12) is a valid ultrametric and \( u_{NR}^{R/R} \) satisfies axioms (A1)-(A2).

**Proof:** This proof follows from a reasoning analogous to that in the proof of Proposition 1. In particular, by definition we have that [cf. (48)]

\[
u_X^{NR}(x, x') \leq \nu_X^{R/R}(x, x'; \beta) \leq \nu_X^R(x, x'),
\]

(13)

which immediately implies fulfillment of (A1). Also, as done for Proposition 1, the strong triangle inequality and the fulfillment of (A2) can be shown by dividing the proofs into the two cases \( \nu_X^R(x, x') \leq \beta \) and \( \nu_X^R(x, x') > \beta. \)

**Remark 1:** Intuitively, the grafting combination \( u_{NR}^{R/R}(\theta) \) allows nonreciprocal propagation of influence for resolutions smaller than \( \beta \) while requiring reciprocal propagation for higher resolutions. This is of interest if we want 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 \( u_{NR}^{R/R}(\theta) \) requires reciprocal influence within tight clusters of resolution smaller than \( \beta \) 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**

A different family of intermediate admissible methods can be constructed by performing a convex combination of methods known to satisfy axioms (A1) and (A2). Indeed, consider two admissible clustering methods \( u_1 \) and \( u_2 \) and a given parameter \( 0 \leq \theta \leq 1 \). For an arbitrary network \( N = (X, A_X) \) denote by \( (X, u_1^2) \) the \( u_1 \) and \( (X, u_2^2) \) the \( u_2 \) the respective outcomes of methods \( u_1 \) and \( u_2 \). Construct then the dissimilarity function \( \Omega_{u_1, u_2}(\theta) \) as the convex combination of \( u_1^2 \) and \( u_2^2 \), for all \( x, x' \in X \)

\[
\Omega_{u_1, u_2}(x, x'; \theta) := \theta u_1^2(x, x') + (1 - \theta) u_2^2(x, x').
\]

(14)

Although \( \Omega_{u_1, u_2}(\theta) \) is a well-defined dissimilarity function, it is not an ultrametric in general because it may violate the strong triangle inequality. Nevertheless, we can recover the ultrametric structure by applying any admissible clustering method \( u \) to the symmetric network \( N^2 = (X, A_X^2) \). Moreover, as explained after Theorem 1, single linkage is the unique admissible clustering method for symmetric networks. Thus, we define the convex combination method \( \Omega_{u_2, u_1}(\theta) \) as the application of single linkage on \( N_2^2 \). Formally, we define \( \Omega_{u_2, u_1}(\theta) \) as a method whose output \( (X, u_2^2(\theta)) = \min_{c(x, x')} A_X^2(x, x_i + 1; \theta), \)

for all \( x, x' \in X \) and \( A_X^2(\theta) \) as given in (14). We show that (15) defines a valid ultrametric and that \( \Omega_{u_1, u_2}(\theta) \) fulfills axioms (A1) and (A2) in the following proposition.

**Proposition 3:** Given two admissible hierarchical clustering methods \( u_1 \) and \( u_2 \), the convex combination method \( \Omega_{u_1, u_2}(\theta) \) is valid and admissible. I.e., \( u_{u_1}^2(\theta) \) defined in (15) is a valid ultrametric and \( \Omega_{u_1, u_2}(\theta) \) satisfies axioms (A1)-(A2).

**Proof:** As discussed in the paragraph preceding the statement of this proposition, \( u_{u_1}^2(\theta) \) is the output of applying single linkage to the symmetric network \( N_2^2 \), immediately implying that \( u_{u_1}^2(\theta) \) is a well-defined ultrametric.

To see that axiom (A1) is fulfilled, pick an arbitrary two-node network \((p, q) A_{p,q} \) with \( A_{p,q}(p, q) = \alpha \) and \( A_{p,q}(q, p) = \beta \). Since methods \( u_1 \) and \( u_2 \) are admissible, in particular they satisfy (A1), hence \( u_{p,q}^2(\theta) = u_{q,p}^2(\theta) = \max(\alpha, \beta) \). It then follows from (14) that \( A_{u_1}^2(\theta) = \max(\alpha, \beta) \) for all possible values of \( \theta \). Moreover, since in (15) all possible chains joining \( p \) and \( q \) must contain these two nodes as consecutive elements, we have that

\[
u_{p,q}^{u_1}(\theta) = A_{u_1}^2(\theta) = \max(\alpha, \beta),
\]

(16)

for all \( \theta \), satisfying axiom (A1).

Fulfillment of axiom (A2) also follows from admissibility of \( u_1 \) and \( u_2 \). Suppose there are two networks \( N_X = (X, A_X) \) and \( N_Y = (Y, A_Y) \) and a dissimilarity reducing map \( \phi : X \rightarrow Y \). From the facts that \( u_1 \) and \( u_2 \) satisfy (A2) we have

\[
u_\phi^X(x, x') \geq u_{1}^2(\phi(x), \phi(x')) \quad \text{and} \quad u_{\phi}^X(x, x') \geq u_{2}^2(\phi(x), \phi(x')); \]

(17)

By multiplying the left inequality by \( \theta \) and the right one by \( (1 - \theta) \), and adding both inequalities we obtain [cf. (14)]

\[
A_{u_1}^2(x, x'; \theta) \geq A_{u_1}^2(\phi(x), \phi(x'); \theta),
\]

(18)

for all \( 0 \leq \theta \leq 1 \). This implies that the map \( \phi \) is also dissimilarity reducing between the networks \((X, A_{u_1}^2(\theta)) \) and \((Y, A_{u_2}^2(\theta)) \). Combining this with the fact that we apply an admissible method (single linkage) to the previous networks to obtain the ultrametric outputs, it follows that

\[
u_{u_1}^2(x, x'; \theta) \geq A_{u_1}^2(\phi(x), \phi(x'); \theta),
\]

(19)
for all $\theta$, showing that axiom (A2) is satisfied by the convex combination method.

The construction in (15) can be generalized to produce intermediate clustering methods generated by convex combinations of any number (i.e. not necessarily two) of admissible methods. These convex combinations can be seen to satisfy axioms (A1) and (A2) through recursive applications of Proposition 3.

**Remark 2:** Since (15) is equivalent to single linkage applied to the symmetric network $N_{\theta}^{12}$, it follows [18], [23] that $u_{X}^{c}(\theta)$ is the largest ultrametric bounded above by $A_{X}^{\mathcal{F}}(\theta)$, i.e., the largest ultrametric for which $u_{X}^{c}(x, x'; \theta) \leq A_{X}(x, x'; \theta)$ for all $x, x'$. We can then think of (15) as an operation ensuring a valid ultrametric definition while deviating as little as possible from $A_{X}^{\mathcal{F}}(\theta)$, thus, retaining as much information as possible in the convex combination of $u_{X}$ and $u_{X}^{c}$.

C. Semi-reciprocal

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-length two.

**Semi-reciprocal** clustering is a generalization of this concept where loops consisting of at most $t$ nodes in each direction are allowed. Given $t \in \mathbb{N}$ such that $t \geq 2$, we use the notation $C_{t}(x, x')$ to denote any chain $[x = x_{0}, x_{1}, \ldots, x_{t} = x']$ joining $x$ to $x'$ where $l \leq t - 1$. That is, $C_{t}(x, x')$ is a chain starting at $x$ and finishing at $x'$ with at most $t$ nodes. We reserve the notation $C(x, x')$ to represent a chain from $x$ to $x'$ where no maximum is imposed on the number of nodes. Given an arbitrary network $N = (X, A_{X})$, define as $A_{X}^{SR}(x, x')$ the minimum cost incurred when traveling from node $x$ to node $x'$ using a chain of at most $t$ nodes. I.e.,

$$A_{X}^{SR}(x, x') := \min_{C_{t}(x, x')} \max_{i_{t}, \in C_{t}(x, x')} A_{X}(x_{i}, x_{i+1}). \quad (20)$$

We define the family of semi-reciprocal clustering methods $\mathcal{H}_{SR}(t)$ with output $(X, u_{X}^{SR}(t)) = \mathcal{H}_{SR}(t)(X, A_{X})$ as the one for which the ultrametric $u_{X}^{SR}(t)(x, x')$ between $x$ and $x'$ is

$$u_{X}^{SR}(t)(x, x') := \min_{C(x, x')} \max_{i_{t}, \in C(x, x')} A_{X}^{SR}(x_{i}, x_{i+1}). \quad (21)$$

where the function $A_{X}^{SR}(t)$ is defined as

$$A_{X}^{SR}(x_{i}, x_{i+1}) := \max\{A_{X}^{SR}(x_{i}, x_{i+1}), A_{X}^{SR}(x_{i+1}, x_{i})\}. \quad (22)$$

The chain $C(x, x')$ of unconstrained length in (21) is called the main chain, represented by $[x = x_{0}, x_{1}, \ldots, x_{t} = x']$ in Fig. 4. Between consecutive nodes $x_{i}$ and $x_{i+1}$ of the main chain, we build loops consisting of secondary chains in each direction, represented in Fig. 4 by $[x_{i}, y_{i1}, \ldots, y_{ik_{i}}, x_{i+1}]$ and $[x_{i+1}, y'_{i1}, \ldots, y'_{k_{i}}'x_{i}]$ for all $i$. For the computation of $u_{X}^{SR}(t)(x, x')$, the maximum allowed length of secondary chains is equal to $t$ nodes, i.e., $k_{i}, k_{i}' \leq t - 2$ for all $i$. In particular, for $t = 2$ we recover the reciprocal chain; see Fig. 1.

We can reinterpret (21) as the application of reciprocal clustering [cf. (7)] to a network with dissimilarities $A_{X}^{SR}(t)$ as in (20), 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 $\mathcal{H}_{SR}(t)$ is valid and admissible for all integers $t \geq 2$. I.e., $u_{X}^{SR}(t)$ is a valid ultrametric and $\mathcal{H}_{SR}(t)$ satisfies axioms (A1)-(A2).

**Proof:** See Appendix B.

The semi-reciprocal family is a countable family of clustering methods parameterized by integer $t \geq 2$ representing the allowed maximum node-length of secondary chains. Reciprocal and nonreciprocal ultrametrics are equivalent to semi-reciprocal ultrametrics for specific values of $t$. For $t = 2$ we have $u_{X}^{SR}(2) = u_{X}^{NR}$ meaning that we recover reciprocal clustering. To see this formally, note that $A_{X}^{SR}(2)(x, x') = A_{X}(x, x')$ [cf. (20)] since the only chain of length two joining $x$ and $x'$ is $[x, x']$. Hence, for $t = 2$, (21) reduces to

$$u_{X}^{SR}(2)(x, x') = \min_{C(x, x')} \max_{i_{t}, \in C(x, x')} A_{X}(x_{i}, x_{i+1}), \quad (23)$$

which is the definition of the reciprocal ultrametric [cf. (7)]. Nonreciprocal ultrametrics can be obtained as $u_{X}^{SR}(t) = u_{X}^{NR}$ for any parameter $t$ exceeding the number of nodes in the network analyzed. To see this, notice that minimizing over $C(x, x')$ is equivalent to minimizing over $C_{t}(x, x')$ for all $t \geq n$, since we are looking for minimizing chains in a network with non-negative dissimilarities. 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 $x'$, i.e., all chains of interest have at at most $n$ nodes. Hence, by inspecting (20), $A_{X}^{SR}(t)(x, x') = A_{X}(x, x')$ [cf. (3)] for all $t \geq n$. Furthermore, when $t \geq n$, 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 (21) reduces to

$$u_{X}^{SR}(t)(x, x') = \max\{\tilde{u}_{X}^{*}(x, x'), \tilde{u}_{X}(x, x')\}, \quad (24)$$

for all $x, x' \in X$ and for all $t \geq n$. The right hand side of (24) is the definition of the nonreciprocal ultrametric [cf. (8)].

For the network in Fig. 5, we compute the semi-reciprocal ultrametrics between $x$ and $x'$ for different values of $t$. The edges which are not delineated are assigned dissimilarity values greater than 4. Since the only bidirectional chain between
$x$ and $x'$ uses $x_3$ as the intermediate node, we conclude that $u_x^R(x, x) = u_x^{SR(2)}(x, x') = 4$. Furthermore, by constructing a path through the outermost clockwise cycle in the network, we conclude that $u_x^{SR}(x, x') = 1$. Since the longest secondary chain in the minimizing chain for the nonreciprocal case, $[x, x_1, x_2, x_4, x']$, has node-length 5, we may conclude that $u_x^{SR(t)}(x, x') = 1$ for all $t \geq 5$. For intermediate values of $t$, if e.g., we fix $t = 3$, the minimizing chain is given by the main chain $[x, x_3, x']$ and the secondary chains $[x, x_1, x_3]$, $[x_3, x_4, x']$, $[x', x_3, x_3]$ 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, $u_x^{SR(3)}(x, x') = 3$. The minimizing chain for $t = 4$ is similar to the minimizing one for $t = 3$ but replacing the secondary chain $[x, x_1, x_3]$ by $[x, x_1, x_2, x_3]$. In this way, we obtain $u_x^{SR(4)}(x, x') = 2$.

**Remark 3:** Intuitively, 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 where the parameter $t$ represents the maximum length of chains through which influence can propagate in a nonreciprocal manner.

**Remark 4:** Although a wide variety of admissible methods were introduced in this section, Theorem 1 forces all of them to coincide with single linkage when the input network is symmetric. This implies that natural extensions of other clustering methods – such as complete or Ward’s linkage [6] – to asymmetric networks will not be admissible. Notice that this is not a declaration of the practical validity of these alternative methods but rather a clear-cut classification that can be useful for the practitioner. More precisely, if the proposed axioms are reasonable properties for the practitioner, then admissible methods must be chosen whereas if a non-admissible method is used then at least one of the axioms will be violated.

**IV. DIOID MATRIX REPRESENTATIONS**

Recall that, for convenience, we can interpret the dissimilarity function $A_X$ as an $n \times n$ matrix and, similarly, $u_X$ can be regarded as a matrix of ultrametrics. By (7), reciprocal clustering searches for chains that minimize their maximum dissimilarity in the symmetric matrix $A_X := \max(A_X, A_X^*)$, where the $\max$ is applied element-wise. This is equivalent to finding chains in $A_X$ that have minimum cost in a $\ell_\infty$ sense. Likewise, nonreciprocal clustering searches for directed chains of minimum cost in $A_X$ to construct the matrix $\tilde{u}_X^*$ [cf. (3)] and selects the maximum of the directed costs by performing the operation $u_X^{NR} = \max(\tilde{u}_X, \tilde{u}_X^T)$ [cf. (8)]. These operations can be performed algorithmically using matrix powers in the dioid algebra $\mathcal{A} := (\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ [27].

In $\mathcal{A}$, the regular sum is replaced by the minimization operator and the regular product by maximization. Indeed, using $\oplus$ and $\otimes$ to denote sum and product, respectively, on this dioid algebra we have $a \oplus b := \min(a, b)$ and $a \otimes b := \max(a, b)$ for all $a, b \in \mathbb{R}^+ \cup \{+\infty\}$. In the algebra $\mathcal{A}$, the matrix product $A \otimes B$ of two real valued matrices of compatible sizes is therefore given by the matrix with entries

$$[A \otimes B]_{ij} := \bigoplus_{k=1}^n (A_{ik} \oplus B_{kj}) = \min_{k=1}^n \max_{k=1}^n (A_{ik}, B_{kj}).$$

(25)

For integers $k \geq 2$ dioid matrix powers $A_X^k := A_X \otimes A_X^{k-1}$ with $A_X := A_X$ of a dissimilarity matrix are related to ultrametric matrices $u_X$. We delve into this relationship in the next section.

**A. Diod Powers and Ultrametrics**

Notice that the elements of the dioid power $u_X^2$ of a given ultrametric matrix $u_X$ are given by

$$[u_X^2]_{ij} = \min_{k \in \{1, \ldots, n\}} \max\left([u_X]_{ik}, [u_X]_{kj}\right).$$

(26)

Since $u_X$ satisfies the strong triangle inequality we have that $[u_X]_{ij} \leq \max([u_X]_{ik}, [u_X]_{kj})$ for all $k \in \{1, \ldots, n\}$. And for $k = j$ in particular we further have that $\max([u_X]_{ij}, [u_X]_{jj}) = \max([u_X]_{ij}, 0) = [u_X]_{ij}$. Combining these two observations it follows that the result of the minimization in (26) is $[u_X^2]_{ij} = [u_X]_{ij}$ since none of its arguments is smaller that $[u_X]_{ij}$ and one of them is exactly $[u_X]_{ij}$. This being valid for all $i, j$ implies

$$u_X^2 = u_X.$$  

(27)

Furthermore, a matrix having the property in (27) is such that $[u_X^2]_{ij} = \min_{k \in \{1, \ldots, n\}} \max([u_X]_{ik}, [u_X]_{kj}) \leq \max([u_X]_{ij}, [u_X]_{ij})$ for all $l$, which is just a restatement of the strong triangle inequality. Therefore, a non-negative matrix $u_X$ represents a finite ultrametric if and only if (27) is true, has null diagonal elements and positive off-diagonal elements, and is symmetric, $u_X = u_X^T$. From definition (25) it follows that the $l$-th dioid power $A_X^l$ is such that its entry $[A_X^l]_{ij}$ represents the minimum cost of a chain from node $i$ to $j$ containing at most $l$ hops. We then expect dioid powers to play a key role in the construction of ultrametrics.

The quasi-inverse of a matrix in a dioid algebra is a useful concept that simplifies the proofs within this section. In any dioid algebra we call quasi-inverse of $A$, denoted by $A^!$, to the limit, when it exists, of the sequence of matrices [27, Ch.4, Def. 3.1.2]

$$A^! := \lim_{k \to \infty} I \oplus A \oplus A^2 \oplus \ldots \oplus A^k,$$  

(28)
where \( I \) has zeros in the diagonal and \(+\infty\) in the off-diagonal elements. The utility of the quasi-inverse resides in the fact that, given a dissimilarity matrix \( A_X \), then \cite[Ch. 6, Sec. 6.1]{27}

\[
[A_X^-]_{ij} = \min_{C(x, x_j)} \max_{k | x_k \in C(x, x_j)} A_X (x_k, x_{k+1}).
\]

(29)

I.e., the elements of the quasi-inverse \( A_X^- \) correspond to the directed minimum chain costs \( \tilde{u}_X^\ast \) of the associated network \((X, A_X^-)\) as defined in (3).

B. Algorithms for Admissible Clustering Methods

The reciprocal and nonreciprocal ultrametrics can be obtained via simple dioid matrix operations, as stated next.

Theorem 2: For any network \( N = (X, A_X) \) with \( n \) nodes the reciprocal ultrametric \( u_X^R \) defined in (7) can be computed as

\[
u_X^R = \left( \max \left( A_X, A_X^T \right) \right)^{n-1},
\]

(30)

where the matrix operations are in the dioid algebra \( \mathfrak{M} \). Similarly, the nonreciprocal ultrametric \( u_X^{NR} \) defined in (8) can be computed as

\[
u_X^{NR} = \max \left( A_X^{-1}, (A_X^T)^{n-1} \right).
\]

(31)

Proof: By comparing (29) with (3), we can see that \( A_X^- \) is the quasi-inverse of \( A_X \) from where it follows [cf. (8)].

Similarly, if we consider the quasi-inverse of the symmetrized matrix \( \tilde{A}_X := \max(A_X, A_X^T) \), expression (29) becomes

\[
[\tilde{A}_X^-]_{ij} = \min_{C(x, x_j)} \max_{k | x_k \in C(x, x_j)} \tilde{A}_X (x_k, x_{k+1}).
\]

(33)

From comparing (33) and (7) it is immediate that

\[
u_X^R = \tilde{A}_X^- = \left( \max(A_X, A_X^T) \right)^{1/2}.
\]

(34)

If we show that \( A_X^{-1} = A_X^{(t-1)} \), then (34) and (32) imply equations (30) and (31) respectively, completing the proof.

Notice that in \( \mathfrak{M} \), the \( \min \) or \( \oplus \) operation is idempotent, i.e., \( a \oplus a = a \) for all \( a \). In this case, it can be shown that for any \( k \geq 1 \),

\[
I \oplus A_X \oplus A_X^2 \oplus \ldots \oplus A_X^k = (I \oplus A_X)^k,
\]

(35)

for all \( k \geq 1 \). Recalling that \( I \) has zeros in the diagonal and \(+\infty\) in the off-diagonal elements, it is immediate that \( \oplus \tilde{A}_X = A_X \). Consequently, (35) becomes

\[
\oplus \tilde{A}_X \oplus A_X \oplus \ldots \oplus A_X^k = A_X^k.
\]

(36)

Taking the limit to infinity in both sides of equality (36) and invoking the definition of the quasi-inverse in (28), we obtain

\[
A_X^- = \lim_{k \to \infty} A_X^k.
\]

(37)

Finally, it can be shown \cite[Ch. 4, Sec. 3.3, Th. 1]{27} that \( A_X^{-1} = A_X^R \), proving that the limit in (37) exists and, more importantly, that \( A_X^- = A_X^{R^{-1}} \), as desired.

For the reciprocal ultrametric we symmetrize dissimilarities with a maximization operation and take the \((n-1)\)-th power of the resulting matrix on the dioid algebra \( \mathfrak{M} \). For the nonreciprocal ultrametric we revert the order of these two operations. We first consider matrix powers \( A_X^{-1} \) and \( (A_X^T)^{n-1} \) of the dissimilarity matrix and its transpose which we then symmetrize with a maximization operator. Besides emphasizing the extremal nature (cf. Theorem 1) of reciprocal and nonreciprocal clustering, Theorem 2 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 applying matrix powers. 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 5: For any network \( N = (X, A_X) \) with \( n \) nodes the \( t \)-th semi-reciprocal ultrametric \( u_X^{SR(t)} \) in (21) for every natural \( t \geq 2 \) can be computed as

\[
\nu_X^{SR(t)} = \left( \max \left( A_X^{-1}, (A_X^T)^{t-1} \right) \right)^{n-1},
\]

(38)

where the matrix operations are in the dioid algebra \( \mathfrak{M} \).

Proof: By comparison with (30), in (38) we have that reciprocal clustering is equivalent to applying reciprocal clustering to network \((X, A_X^{-1}) \). Furthermore, from the definition of matrix multiplication (25) in \( \mathfrak{M} \), the \((t-1)\)-th dioid power \( A_X^{t-1} \) is such that its entry \( [A_X^{t-1}]_{ij} \) represents the minimum cost of a chain containing at most \( t \) nodes, i.e.

\[
[A_X^{t-1}]_{ij} = \min_{C(x, x_j)} \max_{k | x_k \in C(x, x_j)} A_X (x_k, x_{k+1}).
\]

(39)

Observe that we recover (30) by making \( t = 1 \) and (20) to see that \( A_X^{-1} = A_X^{SR(1)} \). Since semi-reciprocal clustering is equivalent to applying reciprocal clustering to network \((X, A_X^{SR(t)}) \), the proof concludes.

The result in (38) is intuitively clear. The powers \( A_X^{-1} \) and \( (A_X^T)^{t-1} \) represent the minimum cost among directed chains of at most \( t - 1 \) links. In the terminology of Section III-C these are the costs of optimal secondary chains containing at most \( t \) nodes. Therefore, the maximization \( \max \left( A_X^{-1}, (A_X^T)^{t-1} \right) \) computes the cost of joining two nodes with secondary chains of at most \( t \) nodes in each direction. This is the definition of \( A_X^{SR(t)} \) in (21). Applying the \((n-1)\)-th dioid power to this new matrix is equivalent to looking for minimizing chains in the network with costs given by the secondary chains. Thus, the outermost dioid power computes the costs of the optimal main chains that achieve the ultrametric values in (21).

Observe that we recover (20) by making \( t = 2 \) in (38) and that we recover (31) when \( t = n \). For this latter case note that when \( t = n \) in (38), comparison with (31) shows that \( \max \left( A_X^{-1}, (A_X^T)^{t-1} \right) = \max \left( A_X^{-1}, (A_X^T)^{n-1} \right) = u_X^{NR} \). However, since \( u_X^{NR} \) is an ultrametric it is idempotent in the dioid algebra [cf. (27)] and the outermost dioid power in (38) is moot. This recovery is consistent with the observations in (23) and (24) that reciprocal and nonreciprocal clustering are particular cases of semi-reciprocal clustering \( H^{SR(t)} \) such that for \( t = 2 \) we have \( u_X^{SR(2)} = u_X^R \) and for \( t = n \) it holds that \( u_X^{SR(t)} = u_X^{NR} \). The
results in Theorem 2 and Proposition 5 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 \( A^T_X \) are different. To be precise consider positive integers \( t, t' > 0 \) and define the algorithmic intermediate clustering method \( H^{t, t'} \) with parameters \( t, t' \) as the one that maps the given network \( N = (X, A_X) \) to the ultrametric set \((X, u_X^{t, t'}) = H^{t, t'}(N)\) given by

\[
u_X^{t, t'} := \left( \max \left( A_X^t, (A^T_X)^{t'} \right) \right)^{n-1}.
\]

The ultrametric (40) 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 \( t + 1 \) nodes whereas backward secondary chains may have at most \( t' + 1 \) nodes. The algorithmic intermediate family \( H^{t, t'} \) encapsulates the semi-reciprocal family since \( H^{t, t'} = H^{R(t), R(t')} \) as well as the reciprocal method since \( H^R = H^{1, 1} \) as it follows from comparison of (40) with (38) and (30), respectively. We also have that \( H^{NR}(N) = H^{NR}(N) \) for all networks \( N = (X, A_X) \) such that \( |X| \leq n \). This follows from the comparison of (40) with (31) and the idempotency of \( u_X^{NR} = \max(A_X^{t}, (A^{T}_X)^{t'}) \) with respect to the dioid algebra. The intermediate algorithmic methods \( H^{t, t'} \) are admissible as we claim in the following proposition.

**Proposition 6:** The hierarchical clustering method \( H^{t, t'} \) is valid and admissible. I.e., \( u_X^{t, t'} \) defined in (40) is a valid ultrametric and \( H^{t, t'} \) satisfies axioms (A1)-(A2).

**Proof:** See Appendix C. □

Algorithms to compute ultrametrics associated with the grafting families in Section III-A entail simple combinations of matrices \( u_X^R \) and \( u_X^{NR} \). E.g., the ultrametrics in (10) corresponding to the grafting method \( H^{R, NR}(\beta) \) can be computed as

\[
u_X^{R, NR}(\beta) = u_X^{NR} \circ I \{ u_X^R \leq \beta \} + u_X^R \circ I \{ u_X^R > \beta \},
\]

where \( \circ \) denotes the Hadamard matrix product and \( I \{ \cdot \} \) is an element-wise indicator function.

In symmetric networks, Theorem 1 states that any admissible method must output an ultrametric equal to the single linkage ultrametric, that we can denote by \( u_X^{SL} \). Thus, all algorithms in this section yield the same output \( u_X^{SL} \) when restricted to symmetric matrices \( A_X \). Considering, e.g., the algorithm for the reciprocal ultrametric in (30) and noting that for a symmetric network \( A_X = \max(A_X, A^T_X) \) we conclude that single linkage can be computed as

\[
u_X^{SL} = A_X^{n-1}.
\]

Observe that, from an algebraic perspective, single linkage hierarchical clustering is closely related to two classical graph-theoretical problems. Indeed, if one considers the alternative dioid \( \mathcal{R}_1 := \{ \mathbb{R}^+ \cup \{ \infty \}, \min, + \} \), then \( A_X^{n-1} \) can be shown to solve the all pairs shortest paths problem, where element \((i, j)\) of the resulting matrix contains the length of the shortest path going from node \( i \) to node \( j \) [29], [30]. Similarly, when considering the dioid \( \mathcal{R}_2 := \{ \mathbb{R}^+ \cup \{ \infty \}, \min, \max \} \), (42) solves the all pairs bottleneck paths problem, where paths of maximum capacity between every pair of nodes are found [31]. It follows from the above discussion that the dioid framework here introduced is versatile in the sense that different problems can be formulated in these terms but, at the same time, specific features of the dioid \( \mathcal{R} \) are not being leveraged in the design of the algorithm; see Remark 5.

Algorithms for the convex combination family in Section III-B involve computing dioid algebra powers of a convex combination of ultrametrics. Given two admissible methods \( H^1 \) and \( H^2 \) with outputs \( (X, u_X^1) = H^1(N) \) and \( (X, u_X^2) = H^2(N) \), and \( \theta \in [0, 1] \), the ultrametric in (15) corresponding to the method \( H^\theta \) can be computed as

\[
u_X^\theta(\theta) = \left( \theta u_X^1 + (1 - \theta) u_X^2 \right)^{n-1}.
\]

The operation \( \theta u_X^1 + (1 - \theta) u_X^2 \) is just the regular convex combination in (14) and the dioid power in (43) implements the single linkage operation in (15) as it follows from (42).

Before illustrating the implementation of the described methods, a summary of these is presented in Table I.

**Remark 5:** It follows from (30), (31), (38), (40), (41), and (43) that all methods presented in this paper can be computed in a number of operations of order \( O(n^4) \) which coincides with the time it takes to compute \( n \) matrix products of matrices of
size \( n \times n \). This complexity can be reduced to \( O(n^3 \log n) \) by noting that the dioid matrix power \( A^n \) can be computed via the sequence \( A, A^2, A^3, \ldots \) which requires \( O(\log n) \) matrix products at a cost of \( O(n^3) \) each. Notice however that the algebraic framework here presented is valid for any dioid algebra and does not leverage the specific min-max structure in \( \mathbb{R} \). One can exploit the structure of \( \mathcal{A} \) and resort to sub cubic dioid matrix multiplication algorithms similar to those in [31], [32] that have complexity \( O(n^{2.688}) \), reducing the total complexity to \( O(n^2 \log n) \). It should be noted that there exist related algorithms that depart from the dioid framework here presented that attain lower complexity in some cases. For the case of reciprocal clustering, complexity of order \( O(n^2) \) can be achieved by leveraging an equivalence between single linkage and a minimum spanning tree problem [33], [34]. For the case of nonreciprocal clustering, Tarjan’s method [12] can be implemented to reduce complexity to \( O(n^2 \log n) \). Further scalability analysis of the methods here presented is left as a future research avenue.

V. NUMERICAL RESULTS

The U.S. Department of Commerce publishes a yearly table of input and outputs organized by economic sectors\(^2\). We focus on a particular section of this table, called input sectors, that corresponds to the inputs to production for year 2011. More precisely, we are given a set \( I \) of 61 industrial sectors as defined by the North American Industry Classification System and a similarity function \( U : I \times I \rightarrow \mathbb{R}_+ \) where \( U(i, j) \) represents how much of the production of sector \( i \), expressed in dollars, is used as an input of sector \( j \). Based on this, we define the network \( N_I = (I, A_I) \) where the dissimilarity function \( A_I \) satisfies \( A_I(i, i) = 0 \) for all \( i \in I \) and, for \( i \neq j \in I \), is given by

\[
A_I(i, j) := 1 - \frac{U(i, j)}{\sum_{j} U(j, j)}.
\]

The normalized similarity \( \tilde{U}(i, j) := U(i, j) / \sum_{j} U(j, j) \) in (44) can be interpreted as the proportion of the input in dollars to productive sector \( j \) that comes from sector \( i \). In this way, we focus on the combination of inputs of a sector rather than the size of the economic sector itself. That is, a small dissimilarity \( A_I(i, j) \) from sector \( i \) to sector \( j \) implies that sector \( j \) highly relies on the output of sector \( i \) as input for its own production. Network \( N_I \) is markedly asymmetric, hence, well-suited for illustrating the clustering methods here developed. Indeed, if one decomposes the asymmetric matrix \( \tilde{U} \) into the following symmetric and skew-symmetric components, \( \tilde{U} = (\tilde{U} + \tilde{U}^T)/2 + (\tilde{U} - \tilde{U}^T)/2 \), then the relative norm of the skew-symmetric component is \( \|\tilde{U} - \tilde{U}^T\|_F / \|\tilde{U}\|_F = 0.658 \), where \( \|\cdot\|_F \) stands for the matrix Frobenius norm.

Reciprocal clustering: The outcome of applying the reciprocal clustering method \( \mathcal{H}_R^\delta \) defined in (7) to the network \( N_I \) is computed with the formula in (30). A partial view of the resulting dendrogram is shown in Fig. 6(a) where two clusters appearing at resolutions \( \delta_1^R = 0.959 \) and \( \delta_2^R = 0.969 \) are highlighted in blue and red, respectively. We also depict in Fig. 6(b) the nodes in the blue cluster with edges representing bidirectional influence between industrial sectors at the corresponding resolution. That is, a double arrow is drawn between two nodes if and only if the dissimilarity between these nodes in both directions is less than or equal to \( \delta_1^R \). In particular, it shows the bidirectional chains of minimum cost between two nodes. E.g., the first two sectors to be merged by \( \mathcal{H}_R^\delta \) are ‘Administrative and support services’ (AS) and ‘Miscellaneous professional, scientific and technical services’ (MP) at resolution \( \delta = 0.887 \). This occurs because 13.2% of the input of AS comes from MP – corresponding to \( A_I(MP, AS) = 0.868 \) – and 11.3% of MP’s input comes from AS, both influences being similar in magnitude. It is reasonable that these two sectors hire services from each other in order to better deliver their own services. This balanced behavior is more frequent among service sectors than between raw material extraction (primary) or manufacturing (secondary) sectors. Indeed, the blue cluster in Fig. 6(b) is mainly composed of services.

Requiring direct bidirectional influence generates some clusters which are counter-intuitive. E.g., at resolution \( \delta = 0.971 \) when the blue and red clusters merge together we have that the ‘Oil and gas extraction’ sector (OG) in the red cluster joins, e.g., the insurance sector IC in the blue cluster. However, OG does not merge with ‘Petroleum and coal products’ (PC), a sector that one would expect to be more closely related, until resolution \( \delta = 0.975 \). In order to avoid this situation, we may allow nonreciprocal influence as we do next.

---

\(^2\)Available at http://www.bea.gov/industry/io_annual.htm

---

Fig. 6. (a) Partial view of the reciprocal dendrogram output by \( \mathcal{H}_R^\delta \) when applied to \( N_I \), (b) Detail of blue cluster. Edges represent bidirectional influence between adjacent sectors.
Nonreciprocal clustering: The outcome of applying the nonreciprocal clustering method \( \mathcal{H}^{NR} \) defined in (8) to \( \mathcal{N}_I \) is computed via (31) and is partially depicted in Fig. 7(a). Let us first observe that the nonreciprocal ultrametrics (merging resolutions) in Fig. 7(a) are not larger than the reciprocal ultrametrics in Fig. 6(a), as it should be the case given the inequality in (9). As a test case we have that the ‘Mining, except oil and gas’ (MI) and the ‘Utilities’ (UT) sectors become part of the same cluster in the reciprocal dendrogram at a resolution \( \delta = 0.943 \) whereas they merge in the nonreciprocal dendrogram at resolution \( \delta' = 0.912 < 0.943 \).

A more interesting observation is that the nonreciprocal dendrogram is qualitatively very different from the reciprocal dendrogram. In the reciprocal dendrogram we tended to see the formation of definite clusters that then merged into larger clusters at coarser resolutions. In the nonreciprocal dendrogram, in contrast, we see the progressive agglutination of economic sectors into a central cluster. Indeed, the first non-singleton cluster to arise is formed at resolution \( \delta = 0.885 \) by the sectors of oil and gas extraction OG, petroleum and coal products PC, and ‘Construction’ (CO). In Fig. 7(b) we see that this cluster forms due to the influence cycle [OG, PC, CO, OG]. Of all the economic input to PC, 82.6\% comes from the OG sector – which is represented by the dissimilarity \( A_I(\text{OG, PC}) = 0.174 \) – in the form of raw material for its productive processes of which oil refining is the dominant one. In the input to CO a total of 11.5\% comes from PC as fuel and lubricating oil for heavy machinery as well as asphalt coating, and 12.3\% of OG’s input comes from CO mainly from engineering projects to enable extraction such as perforation and the construction of pipelines and their maintenance.

Nonreciprocal clustering \( \mathcal{H}^{NR} \) merges the oil and gas OG and petroleum products PC sectors at resolution \( \delta = 0.885 \) before they merge with the insurance sector IC at resolution \( \delta = 0.923 \).

By contrast, as has been already stated, \( \mathcal{H}^R \) merges OG with IC before their common joining with PC. However, the preponderance of cyclic influences in the network of economic interactions \( \mathcal{N}_I \) leads to the formation of clusters that look more like artifacts than fundamental features. E.g., the cluster that forms at resolution \( \delta = 0.887 \) has AS and MP joining the three-node cluster CO-PC-OG because of an influence cycle of five nodes. It is thus apparent that allowing clusters to be formed by arbitrarily long cycles might overlook important bidirectional influences between co-clustered nodes. If we wanted a clustering method which at resolution \( \delta = 0.887 \) would cluster the nodes PC, CO, and OG into one cluster and AS and MP into another cluster, we should allow influence to propagate through cycles of at most three or four nodes. Semi-reciprocal methods \( \mathcal{H}^{SR(1)} \) permit this degree of flexibility as we exemplify next.

Semi-reciprocal clustering: The outcome of applying the semi-reciprocal clustering method \( \mathcal{H}^{SR(1)} \) defined in Section III-C to \( \mathcal{N}_I \) is computed with the formula in (38). A partial view of the resulting dendrogram is shown in Fig. 8(a). Two clusters generated at resolutions \( \delta^{SR}_1 = 0.909 \) and \( \delta^{SR}_2 = 0.917 \) are highlighted in red and blue, respectively. These clusters are depicted in Fig. 8(b) with directed edges between the nodes representing dissimilarities less than or equal to the corresponding resolution. E.g., for the cluster generated at resolution \( \delta^{SR}_1 = 0.909 \), we draw an edge from sector \( i \) to sector \( i' \) if and only if \( A_I(i, i') \leq \delta^{SR}_1 \). Comparing the semi-reciprocal dendrogram in Fig. 8(a) with the reciprocal and nonreciprocal dendrograms in Figs. 6(a) and 7(a), we observe that semi-reciprocal clustering merges any pair of sectors at a resolution not higher than the resolution at which they are co-clustered by reciprocal clustering and not lower than the one at which they are co-clustered by nonreciprocal clustering. E.g., the financial sectors FR and SC become part of the same cluster at resolu-
tions \( \delta_R = 0.925 \) in the reciprocal dendrogram, \( \delta_{SR} = 0.909 \) in the semi-reciprocal dendrogram and \( \delta_{NR} = 0.900 \) in the nonreciprocal dendrogram, satisfying \( \delta_{NR} \leq \delta_{SR} \leq \delta_R \). This ordering of the merging resolutions is as it should be since the reciprocal and nonreciprocal ultrametrics uniformly bound the output ultrametric of any admissible clustering method such as semi-reciprocal clustering [cf. (9)].

The clustering method \( \mathcal{H}^{SR,(3)} \) allows reasonable cyclic influences and is insensitive to intricate influences described by long cycles. Indeed, as can be seen from Fig. 8(a), \( \mathcal{H}^{SR,(3)} \) recognizes the heavy industry cluster OG-PC-CO since these three sectors are the first to merge at resolution \( \delta = 0.885 \). However, the service sectors MP and AS form a cluster of their own before merging with the heavy industry cluster. To be more precise, MP and AS merge at resolution \( \delta = 0.887 \) due to the bidirectional influence between them. When we increase the resolution, at \( \delta_{SR} = 0.917 \) the ‘Rental and leasing services’ (RL) sector acts as an intermediary merging the OG-PC-CO cluster with the MP-AS cluster forming the blue cluster in Fig. 8(b). The cycle containing RL with secondary chains of length at most 3 nodes is \{RL, OG, PC, AS, RL\}. The sector RL uses administrative and support services from AS to provide their own leasing services, and leasing is a common practice in the OG sector. Thus, we obtain the influences depicted in the blue cluster. At resolution \( \delta_{SR} = 0.909 \) the credit intermediation sector FR, the investment sector SC and the real estate sector RA form a three-node cluster given by the influence cycle \{RA, SC, FR, RA\} and depicted in red in Fig. 8(b). Notice that in the nonreciprocal dendrogram in Fig. 7(a), these three sectors join the main blue cluster separately due to the formation of intricate influence loops. The semi-reciprocal method, by not allowing the formation of long loops, distinguishes the more reasonable cluster formed by FR-RA-SC.

Given that \( \mathcal{H}^{SR,(2)} = \mathcal{H}^R \) and \( \mathcal{H}^{SR,(1)} = \mathcal{H}^{NR} \) for large enough \( t \), it is interesting to see how the output of \( \mathcal{H}^{SR,(t)} \) varies with increasing \( t \). In Fig. 11(a) we plot the relative difference with the nonreciprocal ultrametric, i.e., \( \| u^{SR,(t)}_R - u_{NR}^{R} \|_\infty / \| u_{R}^R - u_{NR}^{R} \|_F \) as a function of \( t \). Given that \( u^{SR,(2)}_R = u^{R}_{R} \), the initial relative difference is equal to 1. More importantly, the fact that for this dataset \( u^{SR,(5)}_R = u_{NR}^{R} \) implies that no optimal nonreciprocal chain contains loops with more than 8 nodes, i.e., the concatenation of two secondary chains of length \( t = 5 \) nodes each. Moreover, notice that the relative difference attained by \( \mathcal{H}^{SR,(3)} \) is already smaller than 0.4, implying that more than half of the discrepancy between the reciprocal and nonreciprocal outputs can be explained by the existence of relatively short influence loops.

**Grafting:** A partial view of the dendrogram obtained from applying the grafting clustering method \( \mathcal{H}^{R/NR,\beta} \) for \( \beta = 0.96 \) to \( N_f \) – computed via (41) – is depicted in Fig. 9. Notice that the main differences between the dendrogram here shown and the reciprocal one in Fig. 6 are concentrated in the blue clusters highlighted in both figures. To be more precise, given that the blue reciprocal cluster forms at resolution \( \delta^R = 0.959 < \beta \), in constructing the grafted dendrogram we cut this branch from the reciprocal one and replace it by the corresponding branch in the nonreciprocal dendrogram. Notice however that a cluster containing exactly the same sectors as the blue cluster does not exist in the nonreciprocal dendrogram [cf. Fig. 7(a)]. Nonetheless, what occurs is that the贸ed dendrogram sees any two sectors in the blue cluster – e.g., AS and RL – merging at the resolution that they would have merged in the nonreciprocal dendrogram. This occurs irrespectively of the fact that in the nonreciprocal dendrogram the loop merging these two sectors might also contain other sectors – e.g., OG – outside the blue cluster under consideration. Putting it differently, for sectors that achieve certain level of reciprocal influence, the grafting method allows clusters to form via nonreciprocal influence.

As explained for semi-reciprocal clustering, in Fig. 11(b) we illustrate the relative difference between \( u^{R/NR,\beta} \) and \( u_{NR}^{R} \) as a function of \( \beta \). Given that the dissimilarities \( A_I \) in (44) are upper bounded by 1, the equality \( u^{R/NR,\beta} = u_{NR}^{R} \) is attained. Also notice that even for \( \beta = 0.97 \) the grafting outcome is very similar to the reciprocal one, which is not surprising since most of the reciprocal mergings (those not shown in the partial view in Fig. 6) occur for resolutions larger \( \delta = 0.97 \).

**Convex combinations:** The output obtained from \( \mathcal{H}_\theta^{1.2} \) for \( \theta = 0.5 \) which convexly combines the reciprocal and nonreciprocal methods in equal proportions – computed using (43) – is partially portrayed in Fig. 10. To illustrate the behavior of the convex combination, focus on the blue cluster and compare it with the five leftmost sectors of the reciprocal dendrogram in Fig. 6(a). Notice that there is a difference in the merging order,
namely, RA merges with the rest of the sectors last whereas in the reciprocal case it merges before the cluster FR-SC. To understand this distinction, we have to look at the nonreciprocal dendrogram in Fig. 7(a) and observe that RA merges the main cluster after both FR and SC. Thus, the larger merging resolution in the nonreciprocal dendrogram outweighs the earlier merging in the reciprocal dendrogram giving rise to the observed convex combination. Intuitively, for two sectors to merge at a low resolution in the convex combination method, there must exist strong reciprocal and nonreciprocal influences between them.

If we quantify the difference between $u_{12}^R(\theta)$ and $u_{12}^{NR}$ as a function of $\theta$ we see that the variation is almost linear; see Fig. 11(c) where a linear variation is depicted for reference (dashed line). The departure from linearity is given by the application of single linkage as a necessary means to ensure the validity of the ultrametric output as explained in Section III-B. Moreover, since the ultrametric obtained from single linkage is upper bounded by the input dissimilarities [23], it follows that the dashed line being an upper bound of the curve of interest is not an artifact of $N_7$ but holds in general.

VI. CONCLUSION

We identified and described three families of hierarchical clustering methods that, by satisfying the axioms of value and transformation, are contained between reciprocal and nonreciprocal clustering: i) The grafting methods are defined by exchanging branches between the reciprocal and nonreciprocal dendrograms; ii) The convex combination methods are built around the definition of a convex operation in the space of dendrograms; and iii) The semi-reciprocal clustering methods allow the generation of clusters via cyclic influence of a fixed maximum length. Algorithms for the application of the methods described throughout the paper were developed via matrix operations in a min-max dioid algebra. The reciprocal ultrametric was computed by first symmetrizing directed dissimilarities to their maximum and then computing increasing powers of the symmetrized dissimilarity matrix until stabilization whereas, for the nonreciprocal case, the opposite was shown to be true. In a similar fashion, algorithms for the remaining clustering methods presented throughout the paper were developed in terms of finite matrix powers, thus exhibiting computational tractability of our clustering constructions. Finally, we applied the derived clustering methods and algorithms to study the relationship between economic sectors in the United States. As a future research avenue, we seek to further winnow the set of admissible methods by requiring additional properties such as stability – when clustering similar networks we should obtain similar dendrograms -- and scale invariance – the formation of clusters should not depend on the scale used to measure dissimilarities. The objective will be to achieve a complete taxonomic description of the landscape of admissible methods so that the correct clustering method can be chosen based on a set of prescribed desirable features.

APPENDIX A

PROOF OF THEOREM 1

We prove the theorem by showing both inequalities in (9).

Proof of $u_X(x, x') \leq u^R_X(x, x')$: Consider points $x$ and $x'$ with reciprocal ultrametric $u^R_X(x, x') = \delta$. Let $C^*(x, x') = \{x = x_0, \ldots, x_i = x'\}$ be a chain achieving the minimum in (7) so that we can write

$$u^R_X(x, x') = \max_{i|x_i \in C^*(x, x')} \max \left( A_X(x_i, x_{i+1}), A_X(x_{i+1}, x_i) \right).$$

(45)

Turn attention to the symmetric two-node network $\Delta_2(\delta, \delta) := \{(p, q), A_{p,q}\}$ with $A_{p,q}(p, q) = A_{p,q}(q, p) = \delta$ and define $\{(p, q), u_{p,q}\} := \mathcal{H}(\Delta_2(\delta, \delta))$. Notice that according to Axiom (A1) we have $u_{p,q}(p, q) = \max(\Delta_2(\delta, \delta)) = \delta$.

Focus now on transformations $\phi_i : \{p, q\} \to X$ given by $\phi_i(p) = x_i, \phi_i(q) = x_{i+1}$ so as to map $p$ and $q$ to subsequent points in the chain $C^*(x, x')$ used in (45). Since it follows from (45) that $A_X(x_i, x_{i+1}) \leq \delta$ and $A_X(x_{i+1}, x_i) \leq \delta$ for all $i$, it is just a simple matter of notation to observe that

$$A_X(\phi_i(p), \phi_i(q)) \leq A_{p,q}(p, q) = \delta,$$

$$A_X(\phi_i(q), \phi_i(p)) \leq A_{p,q}(q, p) = \delta.$$  

(46)

Given that according to (46) transformations $\phi_i$ are dissimilarity reducing, it follows from Axiom (A2) that $u_X(x_i, x_{i+1}) = u_X(\phi_i(p), \phi_i(q)) \leq u_{p,q}(p, q) = \delta$, for all $i$. To complete the proof we use the fact that since $u_X$ is an ultrametric and $C^*(x, x') = \{x = x_0, \ldots, x_i = x'\}$ is a chain joining $x$ and $x'$ the strong triangle inequality dictates [cf. (1)] that $u_X(x, x') \leq \max u_X(x_i, x_{i+1}) \leq \delta$. The proof of the second inequality in (9) follows by substituting $\delta = u^R_X(x, x')$ [cf. (45)].
Proof of $u_{X}^{NR}(x,x') \leq u_{X}(x,x')$: We first state the following result, that is instrumental in showing the required bound.

Claim 1: For any admissible method $\mathcal{H}$, the output ultrametric $(X, u_{X}) = \mathcal{H}(X, A_{X})$ is such that for all pairs $x \neq x'$, $u_{X}(x,x') \geq \text{mcl}(X, A_{X})$, where $\text{mcl}(X, A_{X})$ is the minimum loop cost of $(X, A_{X})$ [cf. (4)].

Proof: See [28, Th. 3].

Consider the nonreciprocal clustering equivalence relation $\sim_{\text{NR}, \phi}(\delta)$ at resolution $\delta$ according to which $x \sim_{\text{NR}, \phi}(\delta) x'$ if and only if $x$ and $x'$ belong to the same nonreciprocal cluster at resolution $\delta$. Notice that this is true if and only if $u_{X}^{\text{NR}}(x,x') \leq \delta$. Further consider the set $Z := X \mod \sim_{\text{NR}, \phi}(\delta)$ of corresponding equivalence classes and the map $\phi_{\delta} : X \rightarrow Z$ that maps each point of $X$ to its equivalence class. Notice that $x$ and $x'$ are mapped to the same point $z$ if they belong to the same cluster at resolution $\delta$.

We define the network $N_{Z} := (Z, A_{Z})$ by endowing $Z$ with the dissimilarity $A_{Z}$ derived from the dissimilarity $A_{X}$ as

$$A_{Z}(z,z') := \min_{x \in \phi_{\delta}^{-1}(z), x' \in \phi_{\delta}^{-1}(z')} A_{X}(x,x').$$

(47)

The dissimilarity $A_{Z}(z,z')$ compares all the dissimilarities $A_{X}(x,x')$ between a member of the equivalence class $z$ and a member of the equivalence class $z'$ and sets $A_{Z}(z,z')$ to the value corresponding to the least dissimilar pair. Notice that according to construction, the map $\phi_{\delta}$ is dissimilarity reducing $A_{X}(x,x') \geq A_{Z}(\phi_{\delta}(x), \phi_{\delta}(x'))$ as stated in Axiom (A2), because we either have $A_{Z}(\phi_{\delta}(x), \phi_{\delta}(x')) = 0$ if $x$ and $x'$ are co-clustered at resolution $\delta$, or $A_{X}(x,x') \geq \min_{x \in \phi_{\delta}^{-1}(z), x' \in \phi_{\delta}^{-1}(z')} A_{X}(x,x') = A_{Z}(\phi_{\delta}(x), \phi_{\delta}(x'))$ if they are mapped to different equivalence classes.

Consider now an arbitrary method $\mathcal{H}$ satisfying axioms (A1)-(A2) and define $(Z, u_{Z}) = \mathcal{H}(N_{Z})$ We determine the minimum loop cost [cf. (4)] of $N_{Z}$ in the following claim.

Claim 2: The minimum loop cost of the network $N_{Z}$ is $\text{mcl}(N_{Z}) > \delta$.

Proof: Assume that Claim 2 is not true, denote by $C(z,z) = [z, z', \ldots, z^{(l)}, z]$ a loop of cost smaller than $\delta$ and consider arbitrary nodes $x \in \phi_{\delta}^{-1}(z)$ and $x' \in \phi_{\delta}^{-1}(z')$. By definition, given two nodes in the same equivalence class, we can always find a chain from one to the other of cost not larger than $\delta$. Moreover, since we are assuming that $A_{Z}(z,z') \leq \delta$, this implies that there exists at least one node $x_{1}$ belonging to class $z$ and another node $x_{2}$ belonging to $z'$ such that $A_{X}(x_{1}, x_{2}) \leq \delta$. Combining these two facts, we can guarantee the existence of a chain from $x$ to $x'$ of cost not larger than $\delta$, since we can go first from $x$ to $x_{1}$ then from $x_{1}$ to $x_{2}$ and finally from $x_{2}$ to $x'$ without encountering dissimilarities greater than $\delta$. In a similar way, we can go from $x'$ to $x$ by constructing a chain that goes through all the equivalence classes in $C(z,z)$, i.e., from $z'$ to $z^{(3)}$ and so on until we reach $z$. Since we can go from $x$ to $x'$ and back with chains of cost not exceeding $\delta$, it follows that $u_{X}^{\text{NR}}(x,x') \leq \delta$ contradicting the assumption that $x$ and $x'$ belong to different equivalent classes. Therefore, the assumption that Claim 2 is false cannot hold.

Since the minimum loop cost of $N_{Z}$ satisfies Claim 2, it follows from Claim 1 that $u_{Z}(z,z') > \delta$ for all pairs of distinct equivalence classes $z, z'$. Further note that, since $\phi_{\delta}$ is dissimilarity reducing, Axiom (A2) implies that $u_{X}(x,x') \geq u_{Z}(z,z')$. Combining these facts, we can conclude that when $x$ and $x'$ map to different equivalence classes it holds that $u_{X}(x,x') \geq u_{Z}(z,z') > \delta$. Recall that $x$ and $x'$ mapping to different equivalence classes is equivalent to $u_{X}^{\text{NR}}(x,x') > \delta$. Consequently, we can claim that $u_{X}^{\text{NR}}(x,x') \geq \delta$ implies $u_{X}(x,x') > \delta$, or, in set notation that $\{x, x' : u_{X}^{\text{NR}}(x,x') > \delta\} \subseteq \{x, x': u_{X}(x,x') > \delta\}$. Since the previous expression is true for arbitrary $\delta > 0$ it implies that $u_{X}^{\text{NR}}(x,x') \leq u_{X}(x,x')$ for all $x, x' \in X$ as in the first inequality in (9).

Having showed both inequalities in (9), the global proof concludes.

APPENDIX B
PROOFS IN SECTION III

Proof of Proposition 1: The function $u_{X}^{R/\text{NR}}(\beta)$ fulfills the symmetry and identity properties of ultrametrics because $u_{X}^{\text{NR}}$ and $u_{X}^{R}$ fulfill them separately. Hence, to show that $u_{X}^{R/\text{NR}}(\beta)$ is a properly defined ultrametric, we need to show that it satisfies the strong triangle inequality (1). To show this, we split the proof into two cases: $u_{X}^{R}(x,x') \leq \beta$ and $u_{X}^{R}(x,x') > \beta$. Note that, by definition (10),

$$u_{X}^{R/\text{NR}}(x,x') \leq u_{X}^{R/\text{NR}}(x,x'; \beta) \leq u_{X}^{R}(x,x').$$

(48)

Starting with the case where $u_{X}^{R}(x,x') \leq \beta$, since $u_{X}^{\text{NR}}$ satisfies (1) we can state that,

$$u_{X}^{R/\text{NR}}(x,x'; \beta) = u_{X}^{\text{NR}}(x,x') \leq \max \left( u_{X}^{\text{NR}}(x,x''), u_{X}^{\text{NR}}(x'',x') \right).$$

(49)

Using the lower bound inequality in (48) we can write

$$\max \left( u_{X}^{\text{NR}}(x,x''), u_{X}^{\text{NR}}(x'',x') \right) \leq \max \left( u_{X}^{R/\text{NR}}(x,x'; \beta), u_{X}^{R/\text{NR}}(x',x'') \right).$$

(50)

Combining (49) and (50), we see that $u_{X}^{R/\text{NR}}(\beta)$ fulfills the strong triangle inequality in this case. As a second case, suppose that $u_{X}^{R}(x,x') > \beta$, from the validity of the strong triangle inequality (1) for $u_{X}^{R}$, we can write

$$\beta < u_{X}^{R/\text{NR}}(x,x'; \beta) = u_{X}^{R}(x,x') \leq \max \left( u_{X}^{R}(x,x''), u_{X}^{R}(x'',x') \right).$$

(51)

This implies that at least one of $u_{X}^{R}(x,x'')$ and $u_{X}^{R}(x'',x')$ is greater than $\beta$. When this occurs, $u_{X}^{R/\text{NR}}(\beta) = u_{X}^{R}$ Hence,

$$\max \left( u_{X}^{R}(x,x''), u_{X}^{R}(x'',x') \right) = \max \left( u_{X}^{R/\text{NR}}(x,x'; \beta), u_{X}^{R/\text{NR}}(x'',x'; \beta) \right).$$

(52)

By substituting (52) into (51), we see that for this second case the strong triangle inequality is also satisfied.

To show that $u_{X}^{R/\text{NR}}(\beta)$ satisfies Axiom (A1) it suffices to see that in a two-node network $u_{X}^{\text{NR}}$ and $u_{X}^{R}$ coincide, meaning
that we must have \( u^{R/NR}(\beta) = u^{NR}_{A_{11}} = u^R_X \). Since \( H^R \) and \( H^{NR} \) fulfill (A1), the method \( H^{R/NR}(\beta) \) must satisfy (A1) as well.

To prove (A2) consider a dissimilarity reducing map \( \phi : X \rightarrow Y \) and split consideration with regards to whether the reciprocal ultrametric is \( u^R_X(x, x') \leq \beta \) or \( u^{R/NR}_X(x, x') > \beta \). When \( u^R_X(x, x') \leq \beta \) we must have \( u^R_{A_{11}}(\phi(x), \phi(x')) \leq \beta \) because \( H^R \) satisfies (A2) and \( \phi \) is a dissimilarity reducing map. Hence, according to the definition in (10) we must have both \( u^{R/NR}_X(x, x'; \beta) \) and \( u^{R/NR}_X(\phi(x), \phi(x'); \beta) \) coincide with the nonreciprocal ultrametric and, since \( H^{NR} \) satisfies (A2), it immediately follows that \( u^{R/NR}_X(x, x'; \beta) \leq u^{R/NR}_X(\phi(x), \phi(x'); \beta) \), showing that \( H^{R/NR}(\beta) \) satisfies (A2) when \( u^R_X(x, x') \leq \beta \).

In the second case, when \( u^R_X(x, x') > \beta \), the validity of (A2) for the reciprocal ultrametric \( u^R_X \) allows us to write
\[
\frac{u^{R/NR}_X(x, x'; \beta)}{\beta} = \frac{u^R_X(x, x')}{\beta} \geq u^R_{A_{11}}(\phi(x), \phi(x')).
\] (53)
Combining this with the fact that \( u^R_Y \) is an upper bound on \( u^{R/NR}_X(\beta) \) [cf. (48)], we see that \( H^{R/NR}(\beta) \) satisfies (A2) also for this second case.

**Proof of Proposition 4:** We begin the proof by showing that (21) outputs a valid ultrametric where the only non-trivial property to be shown is the strong triangle inequality (1). For a fixed \( t \), pick an arbitrary pair of nodes \( x \) and \( x' \) and an arbitrary intermediate node \( x'' \). Let us denote by \( C^*(x, x') \) and \( C^*(x', x'') \) a pair of main chains that satisfy definition (21) for \( u^{SR}_X(x, x') \) and \( u^{SR}_X(x', x'') \) respectively. Construct \( C(x, x') \) by concatenating the aforementioned minimizing chains \( C^*(x, x') \) and \( C^*(x', x'') \). However, \( C(x, x') \) is a particular chain for computing \( u^{SR}_X(x, x') \) and need not be the minimizing one. This implies that
\[
u^{SR}_X(x, x') \leq \max \left( u^{SR}_X(x, x'), u^{SR}_X(x', x'') \right),
\] (54)
proving the strong triangle inequality.

To show fulfillment of (A1), consider the network \((p, q); A_{pq}\) with \( A_{pq}(p, q) = \alpha \) and \( A_{pq}(q, p) = \beta \). Note that in this situation, \( A_{pq}(p, q) = \alpha \) and \( A_{pq}(q, p) = \beta \) for all \( t \geq 2 \) [cf. (20)], since there is only one possible chain between them and contains only two nodes. Hence, from (21),
\[
u^{SR}_X(p, q) = \max(\alpha, \beta),
\] (55)
for all \( t \). Consequently, axiom (A1) is satisfied.

To show fulfillment of (A2), consider two arbitrary networks \((X, A_X)\) and \((Y, A_Y)\) and a dissimilarity reducing map \( \phi : X \rightarrow Y \) between them. Further, denote by \( C^*_X(x, x') = [x = x_0, \ldots, x = x'] \) a main chain that achieves the minimum semi-reciprocal cost in (21). Then, for a fixed \( t \), we can write
\[
u^{SR}_X(x, x') = \max_{i; t; x_i \in C^*_X(x, x')} A^{SR}_X(x_i, x_{i+1}).
\] (56)
Consider now a secondary chain \( C^*_Y(x_i, x_{i+1}) = [x_i = x^{(0)}, \ldots, x_i = x_t] \) between two consecutive nodes \( x_i \) and \( x_{i+1} \) of the minimizing chain \( C^*_X(x, x') \). Further, focus on the image of this secondary chain under the map \( \phi \), that is \( C^*_Y(\phi(x_i), \phi(x_{i+1})) := [\phi(x_i) = \phi(x^{(0)}), \ldots, \phi(x^{(t)}) = \phi(x_{i+1})] \) in the set \( Y \).

Since the map \( \phi : X \rightarrow Y \) is dissimilarity reducing, \( A_Y(\phi(x^{(t)}), \phi(x^{(t+1)})) \leq A_X(x_i, x_{i+1}) \) for all links in this chain. Analogously, we can bound the dissimilarities in secondary chains \( C^*_Y(x_i, x_{i+1}) \) from \( x_{i+1} \) back to \( x_i \). Thus, from (20) we can state that,
\[
A^{SR}_X(x_i, x_{i+1}) \geq \frac{A^{SR}_Y(\phi(x_i), \phi(x_{i+1}))}{\beta}.
\] (57)
Denote by \( C^*_Y(\phi(x), \phi(x')) \) the image of the main chain \( C^*_X(x, x') \) under the map \( \phi \). Notice that \( C^*_Y(\phi(x), \phi(x')) \) is a particular chain joining \( \phi(x) \) and \( \phi(x') \), whereas the semi-reciprocal ultrametric computes the minimum across all main chains. Therefore,
\[
u^{SR}_X(\phi(x), \phi(x')) \leq \max_{i; p \in C^*_Y(\phi(x), \phi(x'))} A^{SR}_Y(\phi(x_i), \phi(x_{i+1})),
\] (58)
By bounding the right-hand side of (58) using (57) and recalling (56), it follows that
\[
u^{SR}_X(\phi(x), \phi(x')) \leq u^{SR}_X(x, x').
\] This proves that (A2) is satisfied.

**APPENDIX C**

**Proof of Proposition 6**

Since method \( H^{l,t'} \) is a generalization of \( H^{SR}(t) \), the proof is almost identical to the one of Proposition 4. The only major difference is that showing symmetry of \( u^{l,t'}_X(x, x') \), i.e. \( u^{l,t'}_X(x', x) = u^{l,t'}_X(x', x) \) for all \( x, x' \in X \), is not immediate as in the case of \( u^{SR}_X(t) \).

In a fashion similar to (21), we rewrite the definition of \( u^{l,t'}_X \) given an arbitrary network \((X, A_X)\) in terms of minimizing chains,
\[
u^{l,t'}_X(x, x') = \min_{C(x, x')} \max_{i; x_i \in C(x, x')} A_X^{l,t'}(x_i, x_{i+1})
\] (59)
where the function \( A^{l,t'}_X \) is defined as
\[
A^{l,t'}_X(x, x') := \max \left( A^{SR(t+1)}_X(x', x), A^{SR(t+1)}_X(x', x) \right),
\] (60)
for all \( x, x' \in X \) and functions \( A^{SR}(t) \) as defined in (20). Notice that \( A^{l,t'}_X \) is not symmetric in general. Symmetry of \( u^{l,t'}_X \), however, follows from the following claim.

**Claim 3:** Given any network \((X, A_X)\) and a pair of nodes \( x, x' \in X \) such that \( u^{l,t'}_X(x, x') = \delta \), then \( u^{l,t'}_X(x', x) \leq \delta \).

**Proof:** Assuming \( u^{l,t'}_X(x, x') = \delta \), we denote by \( C(x, x') = [x = x_0, x_1, \ldots, x_i = x'] \) a minimizing main chain achieving the cost \( \delta \) in (59). Thus, we must show that there exists a main chain \( C(x', x) \) from \( x' \) back to \( x \) with cost not exceeding \( \delta \). From definition (60), there must exist secondary chains in both directions between every pair of consecutive nodes \( x_i, x_{i+1} \) in \((X, C(x, x'))\) with cost no greater than \( \delta \). This secondary chains \( C_{t_i}(x_i, x_{i+1}) \) and \( C_{t_i+1}(x_{i+1}, x_i) \) can have at most \( t + 1 \) nodes in the forward direction and at most \( t' + 1 \) nodes in the opposite direction. Moreover, without loss of generality we may consider the secondary chains as having exactly \( t + 1 \) nodes in one direction and \( t' + 1 \) in the other if we do not require consecutive nodes to be distinct.

Focus on a pair of consecutive nodes \( x_i, x_{i+1} \) of the main chain \( C(x, x') \). If we can construct a main chain from \( x_{i+1} \)
back to \( x_i \) with cost not greater than \( \delta \), then we can concatenate these chains for pairs \( x_{i+1}, x_i \) for all \( i \) and obtain the required chain \( C'(x', x) \) in the opposite direction.

Notice that the secondary chains \( C_{t+1}(x_{i+1}, x_i) \) and \( C_{t+1}(x_i, x_{i+1}) \) can be concatenated to form a loop \( L(x_{i+1}, x_i, x_{i+1}) \), i.e., a chain starting and ending at the same node, of length \( t' + t + 1 \) and cost not larger than \( \delta \). We rename the nodes in \( L(x_{i+1}, x_i) = x_{i+1} = x_0, x_1, \ldots, x_i' = x_i, x_i+1, \ldots, x_{i+t'+1} = x_{i+1} \) starting at \( x_i \) and following the direction of the loop. See Fig. 12(a) for an example when \( t = 2 \) and \( t' = 3 \).

Now we are going to construct a main chain \( C'(x_{i+1}, x_i) \) from \( x_{i+1} \) to \( x_i \). We may reinterpret the loop \( L(x_{i+1}, x_i, x_{i+1}) \) as the concatenation of two secondary chains \([x_0, x_1, \ldots, x_i']\) and \([x_i', x_i+1, \ldots, x_{i+t'+1} = x_0] \) of each of them having cost not exceeding \( \delta \). Thus, we may pick \( x_0 = x_{i+1} \) and \( x_i' \) as the first two nodes of the main chain \( C'(x_{i+1}, x_i) \). With the same reasoning, we may link \( x_i' \) with \( x_i^{2t \mod (t+t')} \) with cost not exceeding \( \delta \), and we may link \( x_i^{2t \mod (t+t')} \) with \( x_i^{3t \mod (t+t')} \) with cost not exceeding \( \delta \), and so on. Hence, we construct the main chain

\[
C(x_{i+1}, x_i) = [x_0, x_i', x_i^{2t \mod (t+t')}, \ldots, x_i^{(t+t'-1)t \mod (t+t')}],
\]

which, by construction, has cost not exceeding \( \delta \). In Fig. 12(b) we present an example of this construction.

In order to finish the proof, we need to verify that the last node in the chain in (61) is in fact \( x_i' \). To do so, we have to show that \( (t+t'-1) t \equiv t' \mod (t+t') \), which follows from rewriting the left-hand side as \( (t+t')(t'-1)+t' \).

Applying Claim 3 to an arbitrary pair of nodes \( x, x' \) and then to the pair \( x', x'' \) implies that \( u_X(x, x') = u_X(x', x) \), as needed to show Proposition 6.

REFERENCES

[1] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, “Hierarchical clustering methods and algorithms for asymmetric networks,” in Proc. Asilomar Conf. Signals, Syst., Comput., Nov. 2013, pp. 1773–1777.

[2] H. Caron et al., “The human transcriptome map: Clustering of highly expressed genes in chromosomal domains,” Science, vol. 291, no. 5507, pp. 1289–1292, 2001.

[3] H. Frigui and R. Krishnapuram, “A robust competitive clustering algorithm with applications in computer vision,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 5, pp. 450–465, May 1999.

[4] M. S. Handcock, A. E. Raftery, and J. M. Tantrum, “Model-based clustering for social networks,” J. Roy. Statist. Soc., Ser. A, vol. 170, no. 2, pp. 301–354, 2007.

[5] G. N. Lance and W. T. Williams, “A general theory of classification sorting strategies 1: Hierarchical systems,” Comput. J., vol. 9, no. 4, pp. 373–380, 1967.

[6] A. K. Jain and R. C. Dubes, Algorithms for Clustering Data (Ser. Prentice Hall Advanced Reference Series). Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

[7] R. Xu and D. Wunsch, “Survey of clustering algorithms,” IEEE Trans. Neural Netw. vol. 16, no. 3, pp. 645–678, May 2005.

[8] T. Saito and H. Yadohisa, Data Analysis of Asymmetric Structures: Advanced Approaches in Computational Statistics. Boca Raton, FL, USA: CRC Press, 2004.

[9] L. Hubert, “Min and max hierarchical clustering using asymmetric similarity measures,” Psychometrika, vol. 38, no. 1, pp. 63–72, 1973.

[10] P. B. Slater, “Hierarchical internal migration regions of France,” IEEE Trans. Syst. Man Cybern., vol. SMC-4, no. 4, pp. 321–324, Apr. 1976.

[11] J. P. Boyd, “Asymmetric clusters of internal migration regions of France,” Edinburgh, Scotland: Academic Press, 1980.

[12] R. E. Tarjan, “An improved algorithm for hierarchical clustering using strong components,” Inf. Process. Lett., vol. 7, no. 1, pp. 37–41, 1983.

[13] F. Murtagh, Multidimensional Clustering Algorithms (Computastar LectureNotes). Vienna, Austria: Physica Verlag, 1985.

[14] W. Pentney and M. Meila, “Spectral clustering of biological sequence data,” in Proc. Nat. Conf. Artif. Intell., 2005, pp. 845–850.

[15] I. Guyon, U. V. Luxburg, and R. C. Williamson, “Clustering: Science or art?” in Proc. NIPS Workshop Clustering Theory, 2009, pp. 1–11.

[16] M. Ackermann and S. Ben-David, “Measures of clustering quality: A working set of axioms for clustering,” in Proc. Adv. Neural Info. Process. Syst., 2008, pp. 121–128.

[17] R. B. Zadeh and S. Ben-David, “A uniqueness theorem for clustering,” in Proc. Conf. Uncertainty Artif. Intell., 2009, pp. 639–646.

[18] G. Carlsson and F. Mémoli, “Classifying clustering schemes,” Found. Comput. Math., vol. 13, no. 2, pp. 221–252, 2013.

[19] J. M. Kleinberg, “An impossibility theorem for clustering,” in Proc. Adv. Neural Info. Process. Syst., 2002, pp. 446–453.

[20] T. V. Laarhoven and E. Marchiori, “Axioms for graph clustering quality functions,” J. Mach. Learn. Res., vol. 15, no. 1, pp. 193–215, Jan. 2014.

[21] M. Meila, “Comparing clusterings: An axiomatic view,” in Proc. Int. Conf. Mach. Learn., 2005, pp. 577–584.

[22] M. Meila, “Comparing clusterings—An information based distance,” J. Multivariate Anal., vol. 98, no. 5, pp. 873–895, 2007.

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

[24] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, “Axiomatic construction of hierarchical clustering in asymmetric networks,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2011, pp. 5219–5223.

[25] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, “Alternative axiomatic constructions for hierarchical clustering of asymmetric networks,” in Proc. IEEE Global Conf. Signal Inf. Process., Dec. 2013, pp. 791–794.

[26] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, “Hierarchical quasi-clustering methods for asymmetric networks,” in Proc. Int. Conf. Mach. Learn., 2014, pp. 352–360.

[27] M. Gondran and M. Minoux, Graphs, Dioids and Semirings: New Models and Algorithms. Berlin, Germany: Springer, 2008.

[28] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, “Hierarchical clustering of asymmetric networks,” 2016, arXiv:1607.06294.

[29] R. W. Floyd, “Algorithm 97: shortest path,” Commun. ACM, vol. 5, no. 6, p. 345, 1962.

[30] S. Warshall, “A theorem on boolean matrices,” J. ACM, vol. 9, no. 1, pp. 11–12, 1962.

[31] V. Vassilevska, R. Williams, and R. Yuster, “All pairs bottleneck paths and max-min matrix products in truly subcubic time,” Theory Comput., vol. 5, pp. 173–189, 2009.

[32] R. Duan and S. Pettie, “Fast algorithms for (max, min)-matrix multiplication and bottleneck shortest paths,” in Proc. ACM/SIAM Symp. Discret. Algorithms, 2009, pp. 384–391.

[33] T. C. Hu, “The maximum capacity route problem,” Oper. Res., vol. 9, no. 6, pp. 898–900, 1961.

[34] D. Müller, “Modern hierarchical, agglomerative clustering algorithms,” 2011, arXiv:1109.2378.
Gunnar Carlsson received the Ph.D. degree in mathematics from Stanford University, Stanford, CA, USA, in 1976. He has taught at University of Chicago, University of California (San Diego), and Princeton University. Since 1991, he has been teaching at Stanford University. In 1986, he gave an invited address at the International Congress of Mathematicians, as well as numerous lectures and lecture series. He was the PI of a multiuniversity DARPA initiative on topological methods for the study of large and complex data sets. He is the founder of Ayasdi Inc., which commercializes the work on topological data analysis.

Facundo Mémoli received the B.Sc. and M.Sc. degrees from the Universidad de la República Oriental del Uruguay, Montevideo, Uruguay, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN, USA, in 2000, 2001, and 2005, respectively, all in electrical engineering. From 2005 to 2010, after the Ph.D degree, he was a Postdoctoral Scholar in the Mathematics Department, Stanford University. From 2010 to 2013, he was an Assistant Professor in the University of Adelaide, Adelaide, SA, Australia. He is currently an Associate Professor in the Departments of Mathematics and Computer Science and Engineering, Ohio State University, Columbus, OH, USA. His current research focuses on the applications of geometry and topology to the analysis of data and networks.

Alejandro Ribeiro (M’07) received the B.Sc. degree from the Universidad de la República Oriental del Uruguay, Montevideo, Uruguay, and the M.Sc. and Ph.D. degrees from the Department of Electrical and Computer Engineering, the University of Minnesota, Minneapolis, MN, USA, in 1998, 2005, and 2007, respectively, all in electrical engineering. From 1998 to 2003, he was a Member of the Technical Staff at BellSouth Montevideo. In 2008, after the M.Sc. and Ph.D. degrees, he joined the University of Pennsylvania (Penn), Philadelphia, PA, USA, where he is currently the Rosenbluth Associate Professor in the Department of Electrical and Systems Engineering. His research interests include the applications of statistical signal processing to the study of networks and networked phenomena, structured representations of networked data structures, graph signal processing, network optimization, robot teams, and networked control. He received the 2014 O. Hugo Schuck Best Paper Award, the 2012 S. Reid Warren, Jr., Award presented by Penn’s undergraduate student body for outstanding teaching, the NSF CAREER Award in 2010, and paper awards at the 2016 SSP Workshop, 2016 SAM Workshop, 2015 Asilomar SSC Conference, ACC 2013, ICASSP 2006, and ICASSP 2005. He is a Fulbright Scholar and a Penn Fellow.

Santiago Segarra (M’16) received the B.Sc. degree in industrial engineering with highest honors (Valedictorian) from the Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina, the M.Sc. degree in electrical engineering from the University of Pennsylvania (Penn), Philadelphia, PA, USA, and the Ph.D. degree in electrical and systems engineering from Penn, in 2011, 2014, and 2016, respectively. Since 2016, he has been a Postdoctoral Research Associate in the Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA. His research interests include network theory, data analysis, machine learning, and graph signal processing. He received the ITBA’s 2011 Best Undergraduate Thesis Award in industrial engineering, the 2011 Outstanding Graduate Award granted by the National Academy of Engineering of Argentina, the Best Student Paper Awards at the 2015 Asilomar Conference and the 2016 IEEE Statistical Signal Processing Workshop, and the Best Paper Award at the 2016 IEEE Sensor Array and Multichannel Signal Processing Workshop.