

# Representable Hierarchical Clustering Methods for Asymmetric Networks

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

**Abstract** This paper introduces the generative model of representability for hierarchical clustering methods in asymmetric networks, i.e., the possibility to describe a method through its action on a collection of networks called representers. We characterize the necessary and sufficient structural conditions needed on these representers in order to generate a method which is scale preserving and admissible with respect to two known axioms and, based on this result, we construct the family of cyclic clustering methods.

## 1 Introduction

The relevance of clustering in modern data analysis is indubitable given its usage in multiple fields of knowledge from, e.g., medicine [22] to marketing [16]. There are literally hundreds of methods that can be applied to the determination of hierarchical [8, 11] and non-hierarchical clusters in finite metric (thus symmetric) spaces—see, e.g., [17]. Even in the case of asymmetric networks [18], multiple methods have been developed to extend the notion of clustering into this less intuitive domain [1, 7, 14, 15, 19, 20]. Although not as developed as its practice [6], the theoretical framework for clustering has been developed over the last decade for non-hierarchical [2, 3, 10, 12, 13, 21, 24] and hierarchical clustering [2, 4, 5]. Of special interest to us is this last direction, where two axioms were used to determine a bounded family of admissible clustering methods for asymmetric

---

G. Carlsson

Department of Mathematics, Stanford University, Stanford, CA, USA

e-mail: [gunnar@math.stanford.edu](mailto:gunnar@math.stanford.edu)

F. Mémoli (✉)

Department of Mathematics, The Ohio State University, Columbus, OH, USA

Department of Computer Science, The Ohio State University, Columbus, OH, USA

e-mail: [memoli@math.osu.edu](mailto:memoli@math.osu.edu)

A. Ribeiro • S. Segarra

Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA

e-mail: [aribeiro@seas.upenn.edu](mailto:aribeiro@seas.upenn.edu); [ssegarra@seas.upenn.edu](mailto:ssegarra@seas.upenn.edu)

networks. However, we consider that admissibility is an insufficient characterization criterion to identify clustering methods of practical importance, thus, we introduce additional features with the objective to further winnow the space of admissible methods.

In the current paper, we introduce and analyze the notion of representability for hierarchical clustering methods. A representable method is one that can be specified by determining its action on a collection of networks called representers. Our main characterization result [cf. Theorem 2] states the conditions needed on these representers to obtain an associated clustering method with desirable practical properties. In Sect. 3, we present the notion of scale preservation and highlight its practical utility. In Sect. 4 we introduce the concept of representability through a reinterpretation of reciprocal clustering and build upon this to develop a more general theory. We also connect representability to the more practical concepts of scale preservation and admissibility (Sect. 4.2), and we introduce the family of cyclic clustering methods (Sect. 4.3).

Proofs, algorithmic details, and computational results are omitted due to space constraints.

## 2 Preliminaries

We define a network  $N$  as a pair  $(X, A_X)$  where  $X$  is a set of  $n$  points or nodes and  $A_X : X \times X \rightarrow \mathbb{R}_+$  is a dissimilarity function. Dissimilarities  $A_X(x, x')$  from  $x$  to  $x'$  are nonnegative, null if and only if  $x = x'$ , need not satisfy the triangle inequality and may be asymmetric, i.e.,  $A_X(x, x') \neq A_X(x', x)$  for some  $x, x' \in X$ . We focus our study on asymmetric networks since these general structures include, as particular cases, symmetric networks and finite metric spaces.

The output of hierarchically clustering a network  $N = (X, A_X)$  is a dendrogram  $D_X$ , i.e., a nested collection of partitions  $D_X(\delta)$  indexed by a resolution parameter  $\delta \geq 0$  [9]. Dendrograms can be represented as trees. The interpretation of a dendrogram is that of a structure which yields different clusterings at different resolutions.

Given a network  $(X, A_X)$  and  $x, x' \in X$ , a chain  $C(x, x')$  is an ordered sequence of nodes,  $C(x, x') = [x = x_0, x_1, \dots, x_{l-1}, x_l = x']$ , which starts at  $x$  and finishes at  $x'$ . The *cost* of chain  $C(x, x')$  is defined as the maximum dissimilarity  $\max_{i|x_i, x_{i+1} \in C(x, x')} A_X(x_i, x_{i+1})$  encountered when traversing its links in order. We define the directed minimum chain cost  $\tilde{u}_X^*(x, x')$  between  $x$  and  $x'$  as the minimum cost among all the chains connecting  $x$  to  $x'$ ,

$$\tilde{u}_X^*(x, x') = \min_{C(x, x')} \max_{i|x_i, x_{i+1} \in C(x, x')} A_X(x_i, x_{i+1}). \quad (1)$$

An ultrametric  $u_X$  on the set  $X$  is a function  $u_X : X \times X \rightarrow \mathbb{R}_+$  that satisfies symmetry  $u_X(x, x') = u_X(x', x)$ , identity  $u_X(x, x') = 0 \iff x = x'$ , and the strong triangle inequality

$$u_X(x, x') \leq \max(u_X(x, x''), u_X(x'', x')), \quad (2)$$

for all  $x, x', x'' \in X$ . It can be shown that dendrograms and finite ultrametric spaces are equivalent [2], where the ultrametric value between two points is given by the resolution at which they first merge in the dendrogram. However, ultrametries are more convenient than dendrograms for the results developed in this paper.

*Remark 1* For the particular case when  $N$  is a finite metric space, the minimum chain cost (1) is an ultrametric and coincides with the ultrametric output of the single linkage clustering method [5].

A hierarchical clustering method is defined as a map  $\mathcal{H} : \mathcal{N} \rightarrow \mathcal{D}$  from the space of networks  $\mathcal{N}$  to the space of dendrograms  $\mathcal{D}$ , or, equivalently, as a map  $\mathcal{H} : \mathcal{N} \rightarrow \mathcal{U}$  from  $\mathcal{N}$  to the space  $\mathcal{U}$  of networks with ultrametries as dissimilarity functions. This loose definition permits the existence of an ample diversity of methods, many of them carrying limited practical utility. Thus, we recall an axiomatic construction formulated to determine a subclass of admissible clustering methods.

## 2.1 Admissible Hierarchical Clustering Methods

In [4, 5], the authors impose the two following requirements on clustering methods:

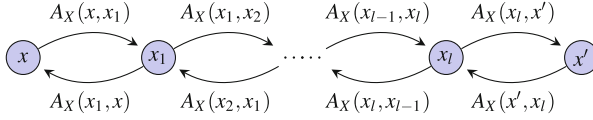
### (A1) Axiom of Value

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

### (A2) Axiom of Transformation

Given networks  $N_X = (X, A_X)$  and  $N_Y = (Y, A_Y)$  and a dissimilarity reducing map  $\phi : X \rightarrow Y$ , that is 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 output ultrametries  $(X, u_X) = \mathcal{H}(X, A_X)$  and  $(Y, u_Y) = \mathcal{H}(Y, A_Y)$  satisfy  $u_X(x, x') \geq u_Y(\phi(x), \phi(x'))$ .

Axiom (A1) states that in a network with two nodes  $p$  and  $q$ , the dendrogram  $D_X$  has them merging at the maximum value of the two dissimilarities  $\alpha$  and  $\beta$ . This is reasonable because at resolutions  $\delta < \max(\alpha, \beta)$  one node can influence



**Fig. 1** Reciprocal clustering.  $x, x'$  cluster at resolution  $\delta$  if they are joined by a bidirectional chain of max. dissimilarity  $\delta$  [cf. (3)]

the other but not vice versa, which in most situations means that the nodes are not alike. Axiom (A2) states that a contraction of the dissimilarity function  $A_X$  entails a contraction of the ultrametric  $u_X$ .

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

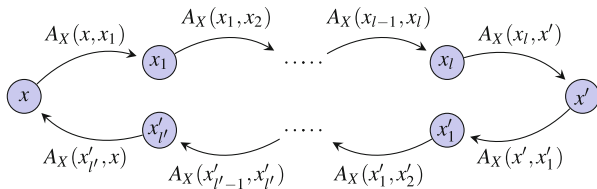
$$u_X^R(x, x') = \min_{C(x, x')} \max_{i: x_i \in C(x, x')} \bar{A}_X(x_i, x_{i+1}), \quad (3)$$

where  $\bar{A}_X(x, x') = \max(A_X(x, x'), A_X(x', x))$  for all  $x, x' \in X$ . Intuitively, in (3) we search for chains  $C(x, x')$  linking nodes  $x$  and  $x'$ . Then, for a given chain, walk from  $x$  to  $x'$  and determine the maximum dissimilarity, in either the forward or backward direction, across all links in the chain. The reciprocal ultrametric  $u_X^R(x, x')$  is the minimum of this value across all possible chains; see Fig. 1. We define the *nonreciprocal* clustering method  $\mathcal{H}^{NR}$  as one where the ultrametric between  $x$  and  $x'$  is given by the maximum of the directed minimum chain costs (1) in both directions

$$u_X^{NR}(x, x') = \max\left(\tilde{u}_X^*(x, x'), \tilde{u}_X^*(x', x)\right). \quad (4)$$

In (4) we implicitly consider forward chains  $C(x, x')$  going from  $x$  to  $x'$  and backward chains  $C(x', x)$  from  $x'$  to  $x$ . We then determine the respective maximum dissimilarities and search independently for the forward and backward chains that minimize the respective maximum dissimilarities. The nonreciprocal ultrametric  $u_X^{NR}(x, x')$  is the maximum of these two minimum values; see Fig. 2. Reciprocal and nonreciprocal clustering bound the range of ultrametries generated by any other admissible method  $\mathcal{H}$ .

**Theorem 1 [4]** *Given any network  $N = (X, A_X)$  and the associated reciprocal  $u_X^R$  and nonreciprocal  $u_X^{NR}$  ultrametries as defined in (3) and (4). Then, for any admissible method  $\mathcal{H}$ , the output ultrametric  $(X, u_X) = \mathcal{H}(X, A_X)$  is such that  $u_X^{NR}(x, x') \leq u_X(x, x') \leq u_X^R(x, x')$  for all  $x, x'$ .*



**Fig. 2** Nonreciprocal clustering.  $x, x'$  cluster at resolution  $\delta$  if they are joined in both directions by chains of max. dissimilarity  $\delta$  [cf. (4)]

### 3 Scale Preservation

In this section we begin to explore additional conditions to winnow the space of admissible methods. One such condition is the concept of a scale preserving hierarchical clustering method.

#### (P1) Scale Preservation

Consider a network  $N_X = (X, A_X)$  and a nondecreasing function  $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  with  $\psi(0) = 0, \psi(z) > 0$  for all  $z > 0$  and  $\lim_{z \rightarrow \infty} \psi(z) = \infty$ . Define the network  $N_X^\psi = (X, \psi(A_X))$  with the same set of nodes and dissimilarities transformed by  $\psi$ . A hierarchical clustering method  $\mathcal{H}$  is said to be *scale preserving* if for an arbitrary network  $N_X = (X, A_X)$  and a function  $\psi$  satisfying the above requirements, the outputs  $(X, u_X) = \mathcal{H}(N_X)$  and  $(X, u_X^\psi) = \mathcal{H}(N_X^\psi)$  satisfy  $u_X^\psi = \psi \circ u_X$ .

Scale preservation entails two consequences of practical relevance. First, since ultrametric outcomes vary according to the same function that transforms the dissimilarities, scale preserving methods are invariant with respect to units. Second, in some applications we may be interested in clustering a network of which we are given a similarity function instead of dissimilarities. In such case, in order to apply a dissimilarity based method, we must apply a decreasing function to the similarity information to transform it into dissimilarities. Scale preservation ensures that the structure of the clustering hierarchy does not depend on the decreasing function used to turn similarities into dissimilarities. This property is helpful in many practical instances. In, e.g., trust networks it is more natural for subjects to express their trust of neighbors rather than their distrust. In social networks, proximity indicators—like number of exchanged messages—are more common than distance indicators.

## 4 Representability

We build upon the notion of representable methods—introduced for non-hierarchical clustering in [3]—to specify the hierarchical clustering of arbitrary networks through the clustering of particular examples that we call *representers*. To explain the concept of a representable method we first present an alternative definition for the reciprocal ultrametric (3). Start by considering a given asymmetric network  $N = (X, A_X)$  and define  $\mathcal{U}_2 = (\{p, q\}, A_{p,q})$  as a two-node network with both dissimilarities equal to 1. Define the  $\lambda$ -multiple of the network  $\mathcal{U}_2$  as the network  $\lambda * \mathcal{U}_2 = (\{p, q\}, \lambda A_{p,q})$  whose underlying set is the same and its dissimilarities are linearly scaled by a given  $\lambda > 0$ . Further define the Lipschitz constant of a map  $\phi : \{p, q\} \rightarrow X$  from  $\mathcal{U}_2$  to  $N$  as

$$L(\phi; \mathcal{U}_2, N) = \max(A_X(\phi(p), \phi(q)), A_X(\phi(q), \phi(p))), \quad (5)$$

i.e., the maximum dissimilarity into which one of the unit dissimilarities in  $\mathcal{U}_2$  is mapped. For any nodes  $x, x' \in X$  we define the optimal multiple  $\lambda_X^{\mathcal{U}_2}(x, x')$  between  $x$  and  $x'$  with respect to  $\mathcal{U}_2$  as

$$\lambda_X^{\mathcal{U}_2}(x, x') = \{L(\phi; \mathcal{U}_2, N) \mid \phi : \{p, q\} \rightarrow X, x, x' \in \text{Im}(\phi)\}. \quad (6)$$

Notice that  $\lambda_X^{\mathcal{U}_2}(x, x')$  is the minimum multiple needed for the existence of a dissimilarity reducing map with  $x$  and  $x'$  in its image between a multiple of  $\mathcal{U}_2$  and  $N$ .

We define the representable clustering method  $\mathcal{H}^{\mathcal{U}_2}$  associated with the representer network  $\mathcal{U}_2$  as the method with ultrametric output  $(X, u_X^{\mathcal{U}_2}) = \mathcal{H}^{\mathcal{U}_2}(N)$  given by

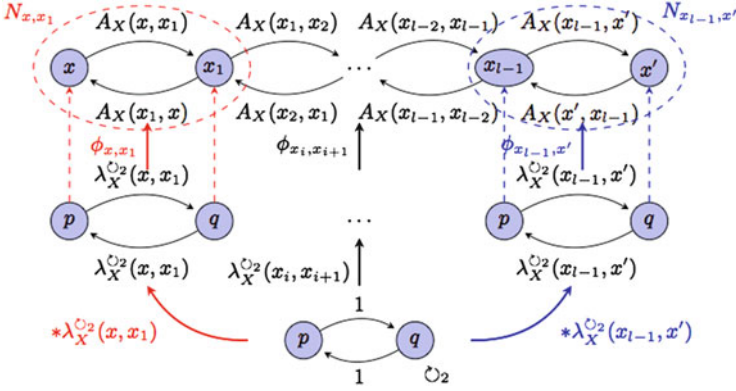
$$u_X^{\mathcal{U}_2}(x, x') = \min_{C(x, x')} \max_{i, x_i \in C(x, x')} \lambda_X^{\mathcal{U}_2}(x_i, x_{i+1}), \quad (7)$$

for all  $x, x' \in X$ . Definition (7) is illustrated in Fig. 3.

It is immediate that the method  $\mathcal{H}^{\mathcal{U}_2}$  with output ultrametrics as in (7) is equivalent to reciprocal clustering  $\mathcal{H}^R$  with output ultrametrics as in (3). Indeed, given a network  $N = (X, A_X)$  and points  $x, x' \in X$ , there are only two possible maps from  $\lambda * \mathcal{U}_2$  to  $N$  containing nodes  $x$  and  $x'$  in their images. One map takes  $p$  to  $x$  and  $q$  to  $x'$  and the other reverses the images and takes  $p$  to  $x'$  and  $q$  to  $x$ . However, either maps have the same Lipschitz constant as defined in (5), ensuring that the optimal multiple in (6) is well defined. Consequently, we obtain that

$$\lambda_X^{\mathcal{U}_2}(x, x') = \max(A_X(x, x'), A_X(x', x)) = \bar{A}_X(x, x'). \quad (8)$$

Comparing (7) with (3) and using the observation in (8) the equivalence  $\mathcal{H}^R \equiv \mathcal{H}^{\mathcal{U}_2}$  follows, i.e.,  $\mathcal{H}^R(N) = \mathcal{H}^{\mathcal{U}_2}(N)$  for all networks  $N$ . Thus, we say that



**Fig. 3** Representable method  $\mathcal{H}^{C_2}$  with ultrametric output as in (7). For every pair  $x_i, x_{i+1}$  of consecutive nodes in the chain  $C(x, x')$  we multiply the network  $\mathcal{C}_2$  by the minimum multiple  $\lambda_X^{C_2}(x_i, x_{i+1})$  that allows the existence of a dissimilarity reducing map  $\phi_{x_i, x_{i+1}}$  containing nodes  $x_i$  and  $x_{i+1}$  in its image. The maximum among all these multiples determines the cost of the chain  $C(x, x')$ . The value  $u_X^{C_2}(x, x')$  is the minimum chain cost over all possible chains linking  $x$  to  $x'$

the reciprocal clustering method is represented by the network  $\mathcal{C}_2$ . That is, if we consider  $\mathcal{C}_2$  as an interaction modality defining a basic clustering unit, the hierarchical clustering of a generic network follows from application of (7).

The definition in (7) is certainly more cumbersome than (3). However, the former can be generalized to cases in which we consider arbitrary representers in lieu of  $\mathcal{C}_2$  as we explain next.

#### 4.1 Representable Hierarchical Clustering Methods

Generalizing  $\mathcal{H}^{C_2}$  entails redefining the Lipschitz constant of a map and the optimal multiples so that they are calculated with respect to an arbitrary representer network  $\omega = (X_\omega, A_\omega)$  instead of  $\mathcal{C}_2$ . In representer networks  $\omega$ , we allow the domain  $\text{dom}(A_\omega)$  of the dissimilarity function  $A_\omega$  to be a proper subset of the product space, i.e., we may have  $\text{dom}(A_\omega) \neq X_\omega \times X_\omega$ . This is a technical modification that allows representer networks to have some dissimilarities that can be interpreted as arbitrarily large. Generalizing (5), given an arbitrary network  $N = (X, A_X)$  we define the *Lipschitz constant* of a map  $\phi : X_\omega \rightarrow X$  from  $\omega$  to  $N$  as

$$L(\phi; \omega, N) = \max_{\substack{(z, z') \in \text{dom}(A_\omega) \\ z \neq z'}} \frac{A_X(\phi(z), \phi(z'))}{A_\omega(z, z')}. \quad (9)$$

Notice that  $L(\phi; \omega, N)$  is the minimum multiple of the network  $\omega$  such that the considered map  $\phi$  is dissimilarity reducing from  $L(\phi; \omega, N) * \omega$  to  $N$ . Further,

observe that (9) reduces to (5) when  $\omega = \mathcal{U}_2$ . Notice as well that the maximum in (9) is computed for pairs  $(z, z')$  in the domain of  $A_\omega$ . Pairs not belonging to the domain could be mapped to any dissimilarity without modifying the value of the Lipschitz constant. Mimicking (6), for arbitrary nodes  $x, x' \in X$  we define the optimal multiple  $\lambda_X^\omega(x, x')$  between  $x$  and  $x'$  with respect to  $\omega$  as

$$\lambda_X^\omega(x, x') = \min\{L(\phi; \omega, N) \mid \phi: X_\omega \rightarrow X, x, x' \in \text{Im}(\phi)\}. \quad (10)$$

This means that  $\lambda_X^\omega(x, x')$  is the minimum Lipschitz constant among those maps that have  $x$  and  $x'$  in its image. Observe that (10) reduces to (6) when  $\omega = \mathcal{U}_2$ .

Representable methods are generalized to cases in which we are given a nonempty set  $\Omega$  of representer networks  $\omega$ . In such case we define the function  $\lambda_X^\Omega$  by considering the infimum across all representer  $\omega \in \Omega$ ,

$$\lambda_X^\Omega(x, x') = \inf_{\omega \in \Omega} \lambda_X^\omega(x, x'), \quad (11)$$

for all  $x, x' \in X$ . The value  $\lambda_X^\Omega(x, x')$  is the infimum across all multiples  $\lambda > 0$  such that there exists a dissimilarity reducing map from  $\lambda * \omega$  to the network containing  $x$  and  $x'$  in its image for some representer  $\omega \in \Omega$ . For a given network  $N = (X, A_X)$ , the representable clustering method  $\mathcal{H}^\Omega$  associated with the collection of representer  $\Omega$  is the one with outputs  $(X, u_X^\Omega) = \mathcal{H}^\Omega(X, A_X)$  such that the ultrametric  $u_X^\Omega$  is given by

$$u_X^\Omega(x, x') = \min_{C(x, x')} \max_{i, x_i \in C(x, x')} \lambda_X^\Omega(x_i, x_{i+1}), \quad (12)$$

for all  $x, x' \in X$ . The definition in (12) is interpreted in Fig. 4.

We say that  $\Omega$  is *uniformly bounded* if there exists a finite  $M$  such that for all  $\omega = (X_\omega, A_\omega) \in \Omega$  we have that  $\max_{(z, z') \in \text{dom}(A_\omega)} A_\omega(z, z') \leq M$ . We now formally define the notion of representability.

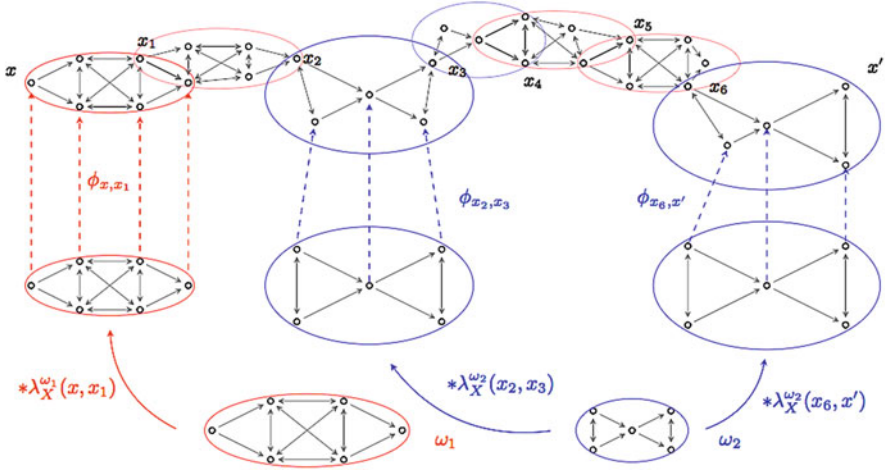
## (P2) Representability

We say that a clustering method  $\mathcal{H}$  is representable if there exists a uniformly bounded collection  $\Omega$  of weakly connected representer each with finite number of nodes such that  $\mathcal{H} \equiv \mathcal{H}^\Omega$  where  $\mathcal{H}^\Omega$  has output ultrametrics as in (12). If the collection  $\Omega$  is finite, we say that  $\mathcal{H}$  is finitely representable.

For every collection of representer  $\Omega$  satisfying the conditions in property (P2), (12) defines a valid ultrametric. Moreover, every representable clustering method abides by axiom (A2), as stated next.

**Proposition 1** *Given a collection of representer  $\Omega$  satisfying the conditions in (P2), the representable method  $\mathcal{H}^\Omega$  is valid, i.e.,  $u_X^\Omega$  defined in (12) is an ultrametric for all networks  $N = (X, A_X)$ , and satisfies the Axiom of Transformation (A2)*





**Fig. 4** A representable clustering method  $\mathcal{H}^\Omega$ . The collection of representers  $\Omega$  is composed by two representers  $\omega_1$  and  $\omega_2$  shown at the *bottom* of the figure. In order to compute the ultrametric value  $u_X^\Omega(x, x')$  we link  $x$  and  $x'$  through a chain, e.g.,  $[x, x_1, \dots, x_6, x']$  in the figure, and link pairs of consecutive nodes with multiples of the representers. We depict these multiples for pairs  $(x, x_1)$ ,  $(x_2, x_3)$ , and  $(x_6, x')$  and the corresponding dissimilarity reducing maps  $\phi_{x, x_1}$ ,  $\phi_{x_2, x_3}$ ,  $\phi_{x_6, x'}$  from the multiple of the representers to the network, containing the corresponding pair of nodes in their images. The ultrametric value  $u_X^\Omega(x, x')$  is given by minimizing over all chains joining  $x$  and  $x'$  the maximum multiple of a representer used to link consecutive nodes in the chain (12)

The condition in (P2) that a valid representable method is defined by a set of weakly connected [23] representers is necessary and sufficient. However, the condition in (P2) that  $\Omega$  be uniformly bounded is sufficient but not necessary for  $\mathcal{H}^\Omega$  to output a valid ultrametric. Although (A2) is guaranteed for every representable method, the Axiom of Value (A1) need not be satisfied. Thus, admissibility and representability are independent properties.

*Remark 2* Representability is a mechanism for defining universal hierarchical clustering methods from given representative examples. Each representer  $\omega \in \Omega$  can be interpreted as defining a particular structure that is to be considered a cluster unit. The scaling of this unit structure [cf. (10), (11)] and its replication through the network [cf. (12)] indicate the resolution at which nodes become part of a cluster. The interest in representability is that it is easier to state desirable clustering structures for particular networks rather than for arbitrary ones. We refer the reader to Sect. 4.3 for particular examples of representer networks that give rise to intuitively appealing clustering methods.

## 4.2 Representability, Scale Preservation, and Admissibility

Are all representable clustering methods relevant in practice? To answer this question we seek to characterize methods that satisfy some desired properties that we deem reasonable. In particular, we consider methods that are admissible with respect to the axioms of value and transformation (A1) and (A2) as well as scale preserving in the sense of (P1).

In characterizing admissible, representable, and scale preserving methods, the concept of structure representer appears naturally. We say that a representer  $\omega = (X_\omega, A_\omega)$  is a *structure representer* if and only if  $|X_\omega| \geq 2$  and

$$A_\omega(z, z') = 1, \text{ for all } z \neq z' \text{ s.t. } (z, z') \in \text{dom}(A_\omega). \quad (13)$$

The requirement in (13) implies that structure representers define the relationships that are necessary in a cluster unit but do not distinguish between different levels of influence. In the following theorem we claim that admissible, representable, and scale preserving hierarchical clustering methods are those represented by a collection  $\Omega$  of strongly connected, structure representers.

**Theorem 2** *A representable clustering method  $\mathcal{H}$  satisfies axioms (A1)–(A2) and scale preservation (P1) if and only if  $\mathcal{H} \equiv \mathcal{H}^\Omega$  where  $\Omega$  is a collection of strongly connected, structure representers as defined by the condition in (13).*

Recalling the interpretation of representability as the extension of clustering defined for particular cases, Theorem 2 entails that the definitions of particular cases cannot present dissimilarity degrees if we require scale preservation. That is, the dissimilarity between every pair of distinct nodes in the representers must be either 1 or undefined. The edges with value 1 imply that the corresponding influence relations are required for the formation of a cluster whereas the influence relations associated with undefined edges are not required. Conversely, Theorem 2 states that encoding different degrees of required influence for different pairs of nodes within the representers is impossible if we want the resulting clustering method to be scale preserving.

## 4.3 Cyclic Clustering Methods

Let  $\mathcal{C}_t = (\{1, \dots, t\}, A_t)$  denote a cycle network with  $t$  nodes such that the domain of the dissimilarity function  $\text{dom}(A_t) = \{(i, i+1)\}_{i=1}^{t-1} \cup (t, 1)$  and every defined dissimilarity is equal to 1. In this section we study representable methods where the representer collections contain cycle networks. We first note that the method defined by a representer collection that contains a finite number of cycle networks is equivalent to the method represented by the longest cycle in the collection.

**Proposition 2** *Given a finite collection  $\Omega = \{\mathcal{U}_t | t \in \mathcal{T}\}$  of cyclic representers, we have that  $\mathcal{H}^\Omega \equiv \mathcal{H}^{\mathcal{U}_{t_{\max}}}$ , where  $t_{\max} = \max \mathcal{T}$ .*

The method  $\mathcal{H}^{\mathcal{U}_t}$  is referred to as the  $t$ th cyclic method. Cyclic methods  $\mathcal{H}^{\mathcal{U}_t}$  for all  $t \geq 2$  are admissible and scale preserving as stated in the following corollary of Theorem 2.

**Corollary 1** *Cyclic methods  $\mathcal{H}^{\mathcal{U}_t}$  satisfy axioms (A1)–(A2) and the scale preservation property (P1).*

The corollary follows from the fact that networks  $\mathcal{U}_t$  are strongly connected and structure representers. The second cyclic method  $\mathcal{H}^{\mathcal{U}_2}$  was used to introduce the concept of representable clustering in (5)–(7) and shown to coincide with the reciprocal clustering method  $\mathcal{H}^{\mathcal{R}}$  in (8). Interpreting  $\mathcal{U}_2$  as a basic cluster unit we can then think of reciprocal clustering  $\mathcal{H}^{\mathcal{R}} \equiv \mathcal{H}^{\mathcal{U}_2}$  as a method that allows propagation of influence through cycles that contain at most two nodes. Likewise, the method  $\mathcal{H}^{\mathcal{U}_3}$  can be interpreted as a method that allows propagation of influence through cycles that contain at most three nodes, and so on.

As we increase  $t$ , the output ultrametrics of methods  $\mathcal{H}^{\mathcal{U}_t}$  become smaller, in particular smaller than those output by  $\mathcal{H}^{\mathcal{U}_2} \equiv \mathcal{H}^{\mathcal{R}}$ . This is consistent with Theorem 1 and is indicative of the status of these methods as relaxations of the condition of direct mutual influence. As we increase the length of the cycles, the question arises of whether we recover nonreciprocal clustering. This is not true for any  $\mathcal{U}_t$  where  $t$  is finite. However, if we define  $\mathcal{C}_\infty = \{\mathcal{U}_t\}_{t=1}^\infty$  the following result holds.

**Proposition 3** *The clustering method  $\mathcal{H}^{\mathcal{C}_\infty}$  represented by the family of all cycle networks  $\mathcal{C}_\infty$  is equivalent to the nonreciprocal clustering method  $\mathcal{H}^{\mathcal{NR}}$  with output ultrametrics as defined in (4).*

Combining the results in Propositions 2 and 3, it follows that by considering methods  $\mathcal{H}^{\mathcal{U}_t}$  for finite  $t$  and method  $\mathcal{H}^{\mathcal{C}_\infty}$  we are considering every method that can be represented by a countable collection of cyclic representers. The reformulation in Proposition 3 expresses nonreciprocal clustering through the consideration of particular cases, namely cycles of arbitrary length. This not only uncovers a drawback of nonreciprocal clustering—propagating influence through cycles of arbitrary length is perhaps unrealistic—but also offers alternative formulations that mitigate this limitation—restrict the propagation of influence to cycles of certain length. In that sense, cyclic methods of length  $t$  can be interpreted as a tightening of nonreciprocal clustering. This interpretation is complementary of their interpretation as relaxations of reciprocal clustering that we discussed above. Given this dual interpretation, cyclic clustering methods are of practical importance.

## 5 Conclusion

The notion of representability was introduced as the possibility of specifying a hierarchical clustering method through its action on a collection of representers. Moreover, the characteristics needed on the representers to obtain an admissible and scale preserving method were detailed. We then focused our attention on cyclic methods, a particular family within the representable methods.

## References

1. Boyd, J.: Asymmetric clusters of internal migration regions of France. *IEEE Trans. Syst. Man Cybern.* **10**(2), 101–104 (1980)
2. Carlsson, G., Mémoli, F.: Characterization, stability and convergence of hierarchical clustering methods. *J. Mach. Learn. Res.* **11**, 1425–1470 (2010)
3. Carlsson, G., Mémoli, F.: Classifying clustering schemes. *Found. Comput. Math.* **13**(2), 221–252 (2013)
4. Carlsson, G., Memoli, F., Ribeiro, A., Segarra, S.: Axiomatic construction of hierarchical clustering in asymmetric networks. In: *IEEE International Conference on Acoustics, Speech and Signal Process (ICASSP)*, pp. 5219–5223 (2013)
5. Carlsson, G., Memoli, F., Ribeiro, A., Segarra, S.: Axiomatic construction of hierarchical clustering in asymmetric networks (2014). [arXiv:1301.7724v2](https://arxiv.org/abs/1301.7724v2)
6. Guyon, I., von Luxburg, U., Williamson, R.: Clustering: science or art? Tech. rep. Paper presented at the NIPS 2009 Workshop Clustering: Science or Art? (2009)
7. Hubert, L.: Min and max hierarchical clustering using asymmetric similarity measures. *Psychometrika* **38**(1), 63–72 (1973)
8. Jain, A.K., Dubes, R.C.: *Algorithms for Clustering Data*. Prentice-Hall, Upper Saddle River, NJ (1988)
9. Jardine, N., Sibson, R.: *Mathematical Taxonomy*. Wiley Series in Probability and Mathematical Statistics. Wiley, London (1971)
10. Kleinberg, J.M.: An impossibility theorem for clustering. In: *NIPS*, pp. 446–453 (2002)
11. Lance, G.N., Williams, W.T.: A general theory of classificatory sorting strategies 1. Hierarchical systems. *Comput. J.* **9**(4), 373–380 (1967)
12. Meila, M.: Comparing clusterings: an axiomatic view. In: *Proceedings of the 22nd International Conference on Machine Learning*, pp. 577–584. ACM, New York (2005)
13. Meila, M.: Comparing clusterings – an information based distance. *J. Multivar. Anal.* **98**(5), 873–895 (2007)
14. Murtagh, F.: Multidimensional clustering algorithms. In: *Compstat Lectures*, vol. 1. Physika Verlag, Vienna (1985)
15. Pentney, W., Meila, M.: Spectral clustering of biological sequence data. In: *Proceedings of National Conference on Artificial Intelligence* (2005)
16. Punj, G., Stewart, D.W.: Cluster analysis in marketing research: review and suggestions for application. *J. Market. Res.* **20**(2), 134–148 (1983)
17. Rui, X., Wunsch-II, D.: Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **16**(3), 645–678 (2005)
18. Saito, T., Yadohisa, H.: *Data Analysis of Asymmetric Structures: Advanced Approaches in Computational Statistics*. CRC Press, Boca Raton, FL (2004)
19. Slater, P.: Hierarchical internal migration regions of France. *IEEE Trans. Syst. Man Cybern.* **6**(4), 321–324 (1976)

20. Tarjan, R.E.: An improved algorithm for hierarchical clustering using strong components. *Inf. Process. Lett.* **17**(1), 37–41 (1983)
21. Van Laarhoven, T., Marchiori, E.: Axioms for graph clustering quality functions. *J. Mach. Learn. Res.* **15**(1), 193–215 (2014)
22. Walsh, D., Rybicki, L.: Symptom clustering in advanced cancer. *Support. Care Cancer* **14**(8), 831–836 (2006)
23. West, D.B., et al.: *Introduction to Graph Theory*, vol. 2. Prentice Hall, Upper Saddle River (2001)
24. Zadeh, R., Ben-David, S.: A uniqueness theorem for clustering. In: *Proceedings of Uncertainty in Artificial Intelligence* (2009)