AXIOMATIC CONSTRUCTION OF HIERARCHICAL CLUSTERING IN ASYMMETRIC NETWORKS

Gunnar Carlsson[‡] *Facundo Mémoli*[⋆] *Alejandro Ribeiro*[†] *Santiago Segarra*[†]

[‡] Department of Mathematics, Stanford University
[⋆] School of Computer Science, University of Adelaide
[†] Department of Electrical and Systems Engineering, University of Pennsylvania

ABSTRACT

We present an axiomatic construction of hierarchical clustering in asymmetric networks where the dissimilarity from node a to node b is not necessarily equal to the dissimilarity from node b to node a. The theory is built on the axioms of value and transformation which encode desirable properties common to any clustering method. Two hierarchical clustering methods that abide to these axioms are derived: reciprocal and nonreciprocal clustering. We further show that any clustering method that satisfies the axioms of value and transformation lies between reciprocal and nonreciprocal clustering in a well defined sense. We apply this theory to the formation of circles of trust in social networks.

Index Terms— Clustering, asymmetric networks.

1. INTRODUCTION

There are literally hundreds of methods, techniques, and heuristics that can be applied to the determination of hierarchical and non-hierarchical clusters in finite metric (thus symmetric) spaces – see, e.g., [1]. Methods to identify clusters in a network of asymmetric dissimilarities, however, are rarer. A number of approaches reduce the problem to symmetric clustering by introducing symmetrizations that are justified by a variety of heuristics; e.g., [2]. An idea that is more honest to the asymmetry in the dissimilarity matrix is the adaptation of spectral clustering [3–5] to asymmetric graphs by using a random walk perspective to define the clustering algorithm [6] or through the minimization of a weighted cut [7]. This relative rarity is expected because the intuition of clusters as groups of nodes that are closer to each other than to the rest is difficult to generalize when nodes are close in one direction but far apart in the other.

To overcome this generic difficulty we postulate two particular intuitive statements in the form of the axioms of value and transformation that have to be satisfied by allowable hierarchical clustering methods. The Axiom of Value states that for a network with two nodes the nodes are clustered together at the maximum of the two dissimilarities between them. The Axiom of Transformation states that if we consider a network and reduce all pairwise dissimilarities, the level at which two nodes become part of the same cluster is not larger than the level at which they were clustered together in the original network. In this paper we study the space of methods that satisfy the axioms of value and transformation.

Although the theoretical foundations of clustering are not as well developed as its practice [8–10], the foundations of clustering in metric spaces have been developed over the past decade [11–14]. Of particular relevance to our work is the case of hierarchical clustering where, instead of a single partition, we look for a family of partitions indexed by a resolution parameter; see e.g., [15], [16, Ch. 4], and [17]. In this context, it has been shown in [18] that single linkage [16, Ch. 4] is the unique hierarchical clustering method that satisfies symmetric versions of the axioms considered here and a third axiom stating that no clusters can be formed at resolutions smaller than the smallest distance in the given data. One may think of the work presented here as a generalization of [18] to the case of asymmetric (non-metric) data.

Our first contribution is the derivation of two hierarchical clustering methods that abide to the axioms of value and transformation. In reciprocal clustering closeness is propagated through links that are close in both directions, whereas in nonreciprocal clustering closeness is allowed to propagate through loops (Section 4). We further prove that any clustering method that satisfies the value and transformation axioms lies between reciprocal and nonreciprocal clustering in a well defined sense (Section 5).

Thanks to AFOSR MURI FA9550-10-1-0567 and NSF CCF-1217963.

2. PRELIMINARIES

Consider a finite set of points X jointly specified with a dissimilarity function A_X to define the network $N = (X, A_X)$. The set X represent the nodes in the network. The dissimilarity $A_X(x, x')$ between nodes $x \in X$ and $x' \in X$ is assumed to be non negative for all pairs (x, x') and null if and only if $x = x'$. However, dissimilarity values $A_X(x, x')$ need not satisfy the triangle inequality and, more consequential for the problem considered here, may be asymmetric in that it is possible to have $A_X(x, x') \neq A_X(x', x)$. We further define $\mathcal{N}$ as the set of all possible networks N.

A clustering of the set X is a partition P_X defined as a collection of sets $P_X = \{B_1, \ldots, B_J\}$ that are nonintersecting, $B_i \cap B_j = \emptyset$ for $i \neq j$, and are required to cover X, $\cup_{i=1}^{J} B_i = X$. In this paper we focus on hierarchical clustering methods whose outcomes are not single partitions P_X but nested collections D_X of partitions $D_X(\delta)$ indexed by a resolution parameter $\delta \geq 0$. For a given D_X, whenever at resolution δ nodes x and x' are in the same cluster of $D_X(\delta)$, we say that they are equivalent at resolution δ and write $x \sim_{D_X(\delta)} x'$. The nested collection D_X is termed a *dendrogram* and is required to satisfy the following two properties plus an additional technical property (see [18]):

(D1) Boundary conditions. For $\delta = 0$ the partition $D_X(0)$ clusters each $x \in X$ into a separate singleton and for some δ_0 sufficiently large $D_X(\delta_0)$ clusters all elements of X into a single set,

$$D_X(0) = \Big\{\{x\},\, x \in X\Big\}, \;\; D_X(\delta_0) = \Big\{X\Big\} \text{ for some } \delta_0.$$

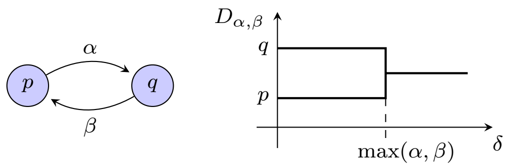

Fig. 1. Axiom of value. Nodes in a two-node network cluster at the minimum resolution at which both can influence each other.

(D2) Resolution. As δ increases clusters can be combined but not separated. I.e., for any $\delta_1 < \delta_2$ if we have $x \sim_{D_X(\delta_1)} x'$ for some pair of points we must have $x \sim_{D_X(\delta_2)} x'$.

As the resolution δ increases, partitions $D_X(\delta)$ become coarser implying that dendrograms are equivalent to trees; see figs. 1 and 2.

Denoting by $\mathcal{D}$ the space of all dendrograms we define a hierarchical clustering method as a function $\mathcal{H} : \mathcal{N} \to \mathcal{D}$ from the space of networks $\mathcal{N}$ to the space of dendrograms $\mathcal{D}$. For the network $N_X = (X, A_X)$ we denote by $D_X = \mathcal{H}(X, A_X)$ the output of clustering method $\mathcal{H}$. When dissimilarities A_X conform to the definition of a finite metric space, it is possible to show that there exists a hierarchical clustering method satisfying axioms similar to the ones proposed in this paper [18]. Furthermore, this method is unique and corresponds to single linkage. For resolution δ, single linkage makes x and x' part of the same cluster if they can be linked through a path of cost not exceeding δ. Formally, equivalence classes at resolution δ in the single linkage dendrogram SL_X are defined as

$$x \sim_{SL_X(\delta)} x' \iff \min_{C(x,x')} \max_{i | x_i \in C(x,x')} A_X(x_i, x_{i+1}) \leq \delta. \quad (1)$$

In (1), $C(x, x')$ denotes a *chain* between x and x', i.e., an ordered sequence of nodes connecting x and x'. We interpret $\max_{i|x_i \in C(x,x')} A_X(x_i, x_{i+1})$ as the maximum dissimilarity cost we need to pay when traversing the chain $C(x, x')$. The right hand side of (1) is this maximum cost for the best selection of the chain $C(x, x')$. Recall that in (1) we are assuming metric data, which in particular implies $A_X(x_i, x_{i+1}) = A_X(x_{i+1}, x_i)$.

2.1. Dendrograms as ultrametrics

Dendrograms are convenient graphical representations but otherwise cumbersome to handle. A mathematically more convenient representation is to identify dendrograms with finite *ultrametric* spaces. An ultrametric defined on the space X is a function $u_X : X \times X \to \mathbb{R}$ that satisfies the strong triangle inequality

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

on top of the reflexivity $u_X(x, x') = u_X(x', x)$, non negativity and identity properties $u_X(x, x') = 0 \Leftrightarrow x = x'$. Hence, an ultrametric is a metric that satisfies (2), a stronger version of the triangle inequality.

As shown in [18], it is possible to establish a bijective mapping between dendrograms and ultrametrics.

Theorem 1 ([18]) *For a given dendrogram D_X define $u_X(x, x')$ as the smallest resolution at which x and x' are clustered together*

$$u_X(x, x') := \min \Big\{ \delta \geq 0, x \sim_{D_X(\delta)} x' \Big\}. \quad (3)$$

The function u_X is an ultrametric in the space X. Conversely, for a given ultrametric u_X define the relation $\sim_{U_X(\delta)}$ as

$$x \sim_{U_X(\delta)} x' \iff u_X(x, x') \leq \delta. \quad (4)$$

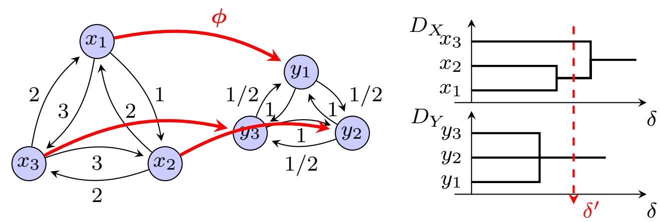

Fig. 2. Axiom of transformation. If network N_X can be mapped to network N_Y using a dissimilarity reducing map ϕ, nodes clustered together in $D_X(\delta)$ at arbitrary resolution δ must also be clustered in $D_Y(\delta)$. For example, x_1 and x_2 are clustered together at resolution δ', therefore y_1 and y_2 must also be clustered at this resolution.

The relation $\sim_{U_X(\delta)}$ is an equivalence relation and the collection of partitions of equivalence classes induced by $\sim_{U_X(\delta)}$, i.e. $U_X(\delta) := \big\{ X \mod \sim_{U_X(\delta)} \big\}$, is a dendrogram.

Given the equivalence between dendrograms and ultrametrics established by Theorem 1 we can think of hierarchical clustering methods $\mathcal{H}$ as inducing ultrametrics in the set of nodes X based on dissimilarity functions A_X. The distance $u_X(x, x')$ induced by $\mathcal{H}$ is the minimum resolution at which x and x' are co-clustered by $\mathcal{H}$.

3. VALUE AND TRANSFORMATION

To study hierarchical clustering algorithms in the context of asymmetric networks, we start from two intuitive notions that we translate into the axioms of value and transformation. The Axiom of Value is obtained from considering a two-node network with dissimilarities α and β; see Fig. 1. In this case, it makes sense for nodes p and q to be in separate clusters at resolutions $\delta < \max(\alpha, \beta)$. For these resolutions we have either no influence between the nodes, if $\delta < \min(\alpha, \beta)$, or unilateral influence from one node over the other, when $\min(\alpha, \beta) \leq \delta < \max(\alpha, \beta)$. In either case both nodes are different in nature. E.g., if we think of the network as a Markov chain, nodes p and q form separate classes. We thus require nodes p and q to cluster at resolution $\delta = \max(\alpha, \beta)$. This is somewhat arbitrary, as any number larger than $\max(\alpha, \beta)$ would work. As a value claim, however, it means that the clustering resolution parameter δ is expressed in the same units as the elements of the dissimilarity matrix. A formal statement in terms of ultrametric distances follows.

(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 $u_X = \mathcal{H}(X, A_X)$ produced by $\mathcal{H}$ is

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

The second restriction on the space of allowable methods $\mathcal{H}$ formalizes the expected behavior upon a modification of the dissimilarity matrix; see Fig. 2. Consider networks $N_X = (X, A_X)$ and $N_Y = (Y, A_Y)$ and denote by $D_X = \mathcal{H}(X, A_X)$ and $D_Y = \mathcal{H}(Y, A_Y)$ the corresponding dendrogram outputs. If we map all the nodes of the network $N_X = (X, A_X)$ into nodes of the network $N_Y = (Y, A_Y)$ in such a way that no pairwise dissimilarity is increased we expect the network to become more clustered. In terms of the respective clustering dendrograms we expect that nodes co-clustered at resolution δ in D_X are mapped to nodes that are also co-clustered at this resolution in D_Y. The Axiom of Transformation is a formal statement of this requirement as we introduce next.

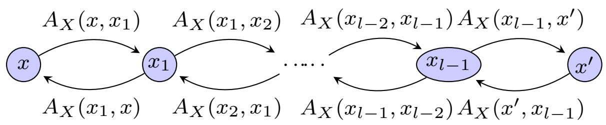

Fig. 3. Reciprocal clustering. Nodes x and x' are clustered together at resolution δ if they can be joined with a (reciprocal) chain whose maximum dissimilarity is smaller than δ in both directions [cf. (7)]. Of all methods that satisfy the axioms of value and transformation, the reciprocal ultrametric is the largest between any pair of nodes.

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

$$u_X(x, x') \geq u_Y(\phi(x), \phi(x')). \tag{6}$$

A hierarchical clustering method $\mathcal{H}$ is admissible if it satisfies Axioms (A1) and (A2). Axiom (A1) states that units of the clustering resolution parameter δ are the same units of the elements of the dissimilarity matrix. Axiom (A2) states that if we reduce dissimilarities, clusters may be combined but cannot be separated.

4. RECIPROCAL AND NONRECIPROCAL CLUSTERING

An admissible clustering method satisfying axioms (A1)-(A2) can be constructed by considering the symmetric dissimilarity $\bar{A}_X(x, x') = \max\big(A_X(x, x'), A_X(x', x)\big)$, for all $x, x' \in X$. This effectively reduces the problem to clustering of symmetric data, a scenario in which single linkage (1) is known to satisfy axioms similar to (A1)-(A2), [18]. Drawing upon this connection we define the *reciprocal* clustering method $\mathcal{H}^R$ with ultrametric outputs $u_X^R = \mathcal{H}^R(X, A_X)$ as the one for which the ultrametric $u_X^R(x, x')$ between nodes x and x' is given by

$$u_X^R(x, x') = \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \bar{A}_X(x_i, x_{i+1}). \tag{7}$$

An illustration of the definition in (7) is shown in Fig. 3. We search for chains $C(x, x')$ linking nodes x and x'. For a given chain we walk from x to x' and determine the maximum dissimilarity, in either the forward or backward direction, across all the links in the chain. The reciprocal ultrametric $u_X^R(x, x')$ between nodes x and x' is the minimum of this value across all possible chains. Recalling the equivalence of dendrograms and ultrametrics in Theorem 1 we know that the dendrogram produced by reciprocal clustering clusters x and x' together for resolutions $\delta \geq u_X^R(x, x')$. Combining this latter observation with (7) and denoting by R_X the reciprocal dendrogram we write the reciprocal equivalence classes as

$$x \sim_{R_X(\delta)} x' \iff \min_{C(x,x')} \max_{i | x_i \in C(x,x')} \bar{A}_X(x_i, x_{i+1}) \leq \delta. \tag{8}$$

Comparing (8) with the definition in (1), we see that reciprocal clustering is equivalent to single linkage for the network $N = (X, \bar{A}_X)$.

For the method $\mathcal{H}^R$ specified in (7) to be a proper hierarchical clustering method we need to show that u^R is an ultrametric. For $\mathcal{H}^R$ to be admissible it needs to satisfy axioms (A1)-(A2). Both of these properties are true as stated in the following proposition.

Proposition 1 *The reciprocal clustering method $\mathcal{H}^R$ is valid and admissible. I.e., u^R as defined by* (7) *is a valid ultrametric and the method satisfies axioms (A1)-(A2).*

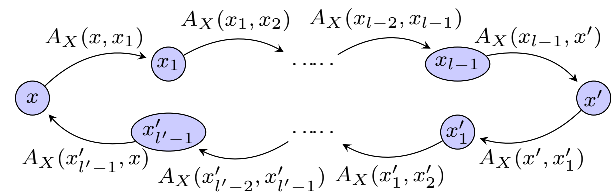

Fig. 4. Nonreciprocal clustering. Nodes x and x' are co-clustered at resolution δ if they can be joined in both directions with possibly different (nonreciprocal) chains of maximum dissimilarity not greater than δ [cf. (10)]. The nonreciprocal ultrametric is the smallest among all that abide to the value and transformation axioms.

Proof: See [19]. ■

In reciprocal clustering, nodes x and x' are joined together if we can go back and forth from x to x' at a maximum cost δ through the same chain. In *nonreciprocal* clustering we relax the restriction that the chain achieving the minimum cost must be the same in both directions and cluster nodes x and x' together if there are, possibly different, chains linking x to x' and x' to x. To state this definition in terms of ultrametrics, consider a given network $N_X = (X, A_X)$ and define the unidirectional minimum chain cost

$$\tilde{u}_X^{NR}(x, x') = \min_{C(x,x')} \max_{i | x_i \in C(x,x')} A_X(x_i, x_{i+1}). \tag{9}$$

We define the nonreciprocal clustering method $\mathcal{H}^{NR}$ with ultrametric outputs $u_X^{NR} = \mathcal{H}^{NR}(X, A_X)$ as the one for which the ultrametric $u_X^{NR}(x, x')$ between nodes x and x' is given by the maximum of the unidirectional minimum chain costs $\tilde{u}_X^{NR}(x, x')$ and $\tilde{u}_X^{NR}(x', x)$ in each direction,

$$u_X^{NR}(x, x') = \max\Big(\tilde{u}_X^{NR}(x, x'),\ \tilde{u}_X^{NR}(x', x)\Big). \tag{10}$$

An illustration of the definition in (10) is shown in Fig. 4. We consider forward chains $C(x, x')$ going from x to x' and backward chains $C(x', x)$ going from x' to x. For each of these chains we determine the maximum dissimilarity across all the links in the chain. We then search independently for the best forward chain $C(x, x')$ and the best backward chain $C(x', x)$ that minimize the respective maximum dissimilarities across all possible chains. The nonreciprocal ultrametric $u_X^{NR}(x, x')$ between nodes x and x' is the maximum of these two minimum values.

As is the case with reciprocal clustering we can verify that u^{NR} is a properly defined ultrametric. We also show that $\mathcal{H}^{NR}$ is admissible in the following proposition.

Proposition 2 *The nonreciprocal clustering method $\mathcal{H}^{NR}$ is valid and admissible. That is, u^{NR} as defined by* (10) *is a valid ultrametric and the method satisfies axioms (A1)-(A2).*

Proof: See [19]. ■

Remark 1 Reciprocal and nonreciprocal clustering are different in general. However, for symmetric networks, they are equivalent and coincide with single linkage as defined by (1). To see this, note that in the symmetric case $\tilde{u}_X^{NR}(x, x') = \tilde{u}_X^{NR}(x', x)$. Therefore, from (10), $u_X^{NR}(x, x') = \tilde{u}_X^{NR}(x, x')$. Comparing (9) and (7) we get the equivalence of nonreciprocal and reciprocal clustering by noting that dissimilarities $A_X = \bar{A}_X$ for the symmetric case. By further comparison with (1) the equivalence with single linkage follows.

5. EXTREMAL ULTRAMETRICS

Given that we have constructed two admissible methods satisfying axioms (A1)-(A2), the question arises of whether these two constructions are the only possible ones and if not whether they are special in some sense, if any. One can find constructions other than reciprocal and nonreciprocal clustering that satisfy axioms (A1)-(A2). However, we prove in this section that reciprocal and nonreciprocal clustering are a peculiar pair in that all possible admissible clustering methods are contained between them in a well defined sense. To explain this sense properly, observe that since reciprocal chains (see Fig. 3) are particular cases of nonreciprocal chains (see Fig. 4) we must have that for all pairs of nodes x, x'

$$u_X^{NR}(x,x') \leq u_X^{R}(x,x'). \tag{11}$$

I.e., nonreciprocal clustering distances do not exceed reciprocal clustering distances. An important characterization is that any method $\mathcal{H}$ satisfying axioms (A1)-(A2) yields ultrametrics that lie between $u_X^{NR}(x,x')$ and $u_X^{R}(x,x')$ as we formally state next.

Theorem 2 *Consider an admissible clustering method $\mathcal{H}$, that is a clustering method satisfying axioms (A1)-(A2). For arbitrary given network $N = (X, A_X)$ denote by $u_X = \mathcal{H}(X, A_X)$ the outcome of $\mathcal{H}$ applied to N. Then, for all pairs of nodes x, x'*

$$u_X^{NR}(x,x') \leq u_X(x,x') \leq u_X^{R}(x,x'), \tag{12}$$

where $u_X^{NR}(x,x')$ and $u_X^{R}(x,x')$ denote the nonreciprocal and reciprocal ultrametrics as defined by (10) *and* (7)*, respectively.*

Proof: See [19]. ■

According to Theorem 2, nonreciprocal clustering applied to the network $N = (X, A_X)$ yields a uniformly minimal ultrametric that satisfies axioms (A1)-(A2). Reciprocal clustering yields a uniformly maximal ultrametric. Any other clustering method abiding to (A1)-(A2) yields an ultrametric such that the distances $u_X(x,x')$ between any two pairs of nodes lie between the distances $u_X^{NR}(x,x')$ and $u_X^{R}(x,x')$ assigned by reciprocal and nonreciprocal clustering. In terms of dendrograms, (12) implies that among all possible clustering methods, the smallest possible resolution at which nodes are clustered together is the one corresponding to nonreciprocal clustering. The highest possible resolution is the one that corresponds to reciprocal clustering.

Remark 2 From Remark 1, the upper and lower bounds in (12) coincide with single linkage for symmetric networks. Thus, (12) becomes an equality in such context. Since metric spaces are particular cases of symmetric networks, Theorem 2 recovers the uniqueness result in [18] and extends it to symmetric – but not necessarily metric – data. Further, the result in [18] is based on three axioms, two of which are the symmetric particular cases of the axioms of value and transformation. It then follows that one of the three axioms in [18] is redundant. See [19] for details.

6. CIRCLES OF TRUST

We apply the theory developed to the formation of trust clusters – circles of trust – in social networks [20]. Recalling the equivalence between dendrograms and ultrametrics, it follows that we can think of trust propagation in a network as inducing a trust ultrametric T_X. The induced trust distance bound $T_X(x,x') \leq \delta$ signifies that, at resolution δ, individuals x and x' are part of a circle of trust. Since

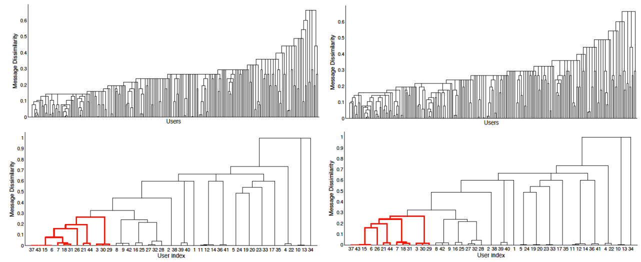

Fig. 5. Nonreciprocal (left) and reciprocal (right) clustering for an online social network [21]. Dissimilarities are inversely proportional to the number of messages sent between any two users. Dendrogram closeups shown in second row.

axioms (A1)-(A2) are reasonable requirements in the context of trust networks Theorem 2 implies that the trust ultrametric must satisfy

$$u_X^{NR}(x,x') \leq T_X(x,x') \leq u_X^{R}(x,x'), \tag{13}$$

which is just a reinterpretation of (12). While (13) does not give a value for trust ultrametrics, reciprocal and nonreciprocal clustering provide lower and upper bounds in the formation of circles of trust.

As a numerical application consider an online social network of a community of students at the University of California at Irvine, [21]. In Fig. 5-top, we depict both clustering algorithms for a subset of the users of the social network. The dissimilarity between nodes has been normalized as a function of the messages sent between any two users where lower distances represent more intense exchange. Note that although the ultrametrics are lower for the nonreciprocal case – as they should (11) –, the overall structure is similar. The similarity between both dendrograms could be interpreted as an indicator of symmetry in the communication. Indeed, in a completely symmetric case both dendrograms would coincide. However, there is another source of similarity between the two proposed algorithms which can be interpreted as consistent asymmetry. For example, someone who rarely replies to a message regardless of the sender or someone who sends messages but gets few replies regardless of the receiver. The similarity between both dendrograms hints that answering all messages of some people but none of others is not ubiquitous.

Fig. 5-bottom presents a closeup of the dendrograms in Fig. 5-top and the major cluster at resolution $\delta = 0.3$ is highlighted in red. We see that this cluster has a different hierarchical genesis in both algorithms. I.e., the two clustering methods alter the clustering order between nodes, which in terms of ultrametrics corresponds to an inversion of the nearest neighbors ordering. Nonetheless, at $\delta = 0.3$ the red cluster contains the same nodes in both clustering methods. This implies that, for the given resolution, this cluster constitutes a circle of trust for any choice of admissible trust metric T_X.

7. CONCLUSION

An axiomatic construction of hierarchical clustering in asymmetric networks was presented. Based on two axioms proposed, the axioms of value and transformation, two particular clustering methods were developed: reciprocal and nonreciprocal clustering. Furthermore, these methods were shown to be well-defined extremes of all possible clustering methods satisfying the proposed axioms. Finally, the theoretical developments were applied to real data in order to study the formation of circles of trust in social networks.

8. REFERENCES

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

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

[3] F. Chung, *Spectral graph theory*, American Mathematical Society, 1997.

[4] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 12 2007.

[5] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *In T.K. Leen, T.G. Dietterich and V. Tresp (Eds.), Advances in neural information processing systems 14, MIT Press, Cambridge*, vol. 2, pp. 849–856, 2002.

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

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

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

[9] Shai Ben-David, Ulrike von Luxburg, and Dávid Pál, "A sober look at clustering stability," in *COLT*, Gábor Lugosi and Hans-Ulrich Simon, Eds. 2006, vol. 4005 of *Lecture Notes in Computer Science*, pp. 5–19, Springer.

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

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

[12] R. Zadeh and S. Ben-David, "A uniqueness theorem for clustering," *Proceedings of Uncertainty in Artificial Intelligence*, 2009.

[13] G. Carlsson and F. Mémoli, "Classifying clustering schemes," *arXiv:1011.5270v2*, 2010.

[14] Jon M. Kleinberg, "An impossibility theorem for clustering," in *NIPS*, Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, Eds. 2002, pp. 446–453, MIT Press.

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

[16] Anil K. Jain and Richard C. Dubes, *Algorithms for clustering data*, Prentice Hall Advanced Reference Series. Prentice Hall Inc., Englewood Cliffs, NJ, 1988.

[17] Ying Zhao and George Karypis, "Hierarchical clustering algorithms for document datasets," *Data Mining and Knowledge Discovery*, vol. 10, pp. 141–168, 2005.

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

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

[20] R. Cross and A. Parker, *The hidden power of social networks*, Harvard business school press, 2004.

[21] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social Networks*, vol. 31, pp. 155–163, 2009.