Hierarchical Quasi-Clustering Methods for Asymmetric Networks

Gunnar Carlsson GUNNAR@MATH.STANFORD.EDU
Department of Mathematics, Stanford University

Facundo Mémoli MEMOLI@MATH.OSU.EDU
Department of Mathematics and Department of Computer Science and Engineering, Ohio State University

Alejandro Ribeiro, Santiago Segarra {ARIBEIRO, SSEGARRA}@SEAS.UPENN.EDU
Department of Electrical and Systems Engineering, University of Pennsylvania

Abstract

This paper introduces hierarchical quasi-clustering methods, a generalization of hierarchical clustering for asymmetric networks where the output structure preserves the asymmetry of the input data. We show that this output structure is equivalent to a finite quasi-ultrametric space and study admissibility with respect to two desirable properties. We prove that a modified version of single linkage is the only admissible quasi-clustering method. Moreover, we show stability of the proposed method and we establish invariance properties fulfilled by it. Algorithms are further developed and the value of quasi-clustering analysis is illustrated with a study of internal migration within United States.

1. Introduction

Given a network of interactions, hierarchical clustering methods determine a dendrogram, i.e. a family of nested partitions indexed by a resolution parameter. Clusters that arise correspond to sets of nodes that are more similar to each other than to the rest and, as such, can be used to study the formation of communities (Shi & Malik, 2000; Newman & Girvan, 2002; 2004; Von Luxburg, 2007; Ng et al., 2002; Lance & Williams, 1967; Jain & Dubes, 1988). For asymmetric networks, in which the dissimilarity from node x to node x' may differ from the one from x' to x (Saito & Yadohisa, 2004), the determination of said clusters is not a straightforward generalization of the methods used to cluster symmetric datasets (Hubert, 1973; Slater, 1976; Boyd, 1980; Tarjan, 1983; Slater, 1984; Murtagh, 1985; Pentney & Meila, 2005; Meila & Pentney, 2007; Zhao & Karypis, 2005).

Proceedings of the 31^{st} International Conference on Machine Learning, Beijing, China, 2014. JMLR: W&CP volume 32.

This difficulty motivates formal developments whereby hierarchical clustering methods are constructed as those that are admissible with respect to some reasonable properties (Carlsson & Mémoli, 2010; 2013; Carlsson et al., 2013). A fundamental distinction between symmetric and asymmetric networks is that while it is easy to obtain uniqueness results for the former (Carlsson & Mémoli, 2010), there are a variety of methods that are admissible for the latter (Carlsson et al., 2013). Although one could conceive of imposing further restrictions to winnow the space of admissible methods for clustering asymmetric networks, it is actually reasonable that multiple methods should exist. Since dendrograms are symmetric structures one has to make a decision as to how to derive symmetry from an asymmetric dataset and there are different stages of the clustering process at which such symmetrization can be carried out (Carlsson et al., 2013). In a sense, there is a fundamental mismatch between having a network of *asymmetric* relations as input and a *symmetric* dendrogram as output.

This paper develops a generalization of dendrograms and hierarchical clustering methods to allow for asymmetric output structures. We refer to these asymmetric structures as quasi-dendrograms and to the procedures that generate them as hierarchical quasi-clustering methods. Since the symmetry in dendrograms can be traced back to the symmetry of equivalence relations we start by defining a quasi-equivalence relation as one that is reflexive and transitive but not necessarily symmetric (Section 3). We then define a quasi-partition as the structure induced by a quasi-equivalence relation, a quasi-dendrogram as a nested collection of quasi-partitions, and a hierarchical quasi-clustering method as a map from the space of networks to the space of quasi-dendrograms (Section 3.1). Quasi-partitions are similar to regular partitions in that they contain disjoint blocks of nodes but they also include an influence structure between the blocks derived from the asymmetry in the original network. This influence structure defines a partial order over the blocks (Harzheim, 2005).

We proceed to study admissibility of quasi-clustering methods with respect to the directed axioms of value and transformation. The Directed Axiom of Value states that the quasi-clustering of a network of two nodes is the network itself. The Directed Axiom of Transformation states that reducing dissimilarities cannot lead to looser quasi-clusters. We show that there is a unique quasi-clustering method admissible with respect to these axioms and that this method is an asymmetric version of the single linkage clustering method (Section 3.4). The analysis in this section hinges upon an equivalence between quasi-dendrograms and quasi-ultrametrics (Section 3.2) that generalizes the well-known equivalence between dendrograms and ultrametrics (Jardine & Sibson, 1971).

Exploiting the fact that quasi-dendrograms can be represented by quasi-ultrametrics, we propose a quantitative notion of stability of quasi-clustering methods (Section 3.5). We prove that the unique method from Section 3.4 is stable in the sense that we propose. We also establish several invariance properties enjoyed by this method.

In order to apply the quasi-clustering method to real data, we derive an algorithm based on matrix powers in a dioid algebra (Gondran & Minoux, 2008) (Section 3.6). As an example, we cluster a network that contains information about the internal migration between states of the United States for the year 2011 (Section 4). The quasi-clustering output unveils that migration is dominated by geographical proximity. Moreover, by exploiting the asymmetric influence between clusters, one can show the migrational influence of California over the West Coast.

Proofs of results in this paper not contained in the main body can be found in the supplementary material.

2. Preliminaries

A network N is a pair (X, A_X) where X is a finite set of points or nodes and $A_X : X \times X \to \mathbb{R}_+$ is a dissimilarity function. The value $A_X(x, x')$ is assumed to be nonnegative for all pairs $(x, x') \in X \times X$ and 0 if and only if $x = x'$. However, A_X need not satisfy the triangle inequality and may be asymmetric in that it is possible to have $A_X(x, x') \neq A_X(x', x)$ for some $x \neq x'$. We further define $\mathcal{N}$ as the set of all networks. Networks $N \in \mathcal{N}$ can have different node sets X and different dissimilarities A_X.

A conventional non-hierarchical clustering of the set X is a partition P, i.e., a collection of sets $P = \{B_1, \ldots, B_J\}$ which are pairwise disjoint and required to cover X. The sets $B_1, B_2, \ldots B_J$ are called the *blocks* of P and represent *clusters*. A partition $P = \{B_1, \ldots, B_J\}$ of X induces and is induced by an equivalence relation $\sim$ on X such that for all $x, x', x'' \in X$ we have that $x \sim x$, $x \sim x'$ if and only if $x' \sim x$, and $x \sim x'$ combined with $x' \sim x''$ implies $x \sim x''$. In hierarchical clustering, the output is not a single partition P but a nested collection D_X of partitions $D_X(\delta)$ of X indexed by a resolution parameter $\delta \geq 0$. For a given D_X, we say that two nodes x and x' are equivalent at resolution $\delta \geq 0$ and write $x \sim_{D_X(\delta)} x'$ if and only if nodes x and x' are in the same cluster of $D_X(\delta)$. The nested collection D_X is termed a *dendrogram* (Jardine & Sibson, 1971). The interpretation of a dendrogram is that of a structure which yields different clusterings at different resolutions. At resolution $\delta = 0$ each point is in a cluster of its own and as the resolution parameter δ increases, nodes start forming clusters. We denote by $[x]_\delta$ the equivalence class to which the node $x \in X$ belongs at resolution δ, i.e. $[x]_\delta := \{x' \in X \,|\, x \sim_{D_X(\delta)} x'\}$.

Given a network (X, A_X) and $x, x' \in X$, a *chain* $C(x, x')$ is an *ordered* sequence of nodes in X,

$$C(x, x') = [x = x_0, x_1, \ldots, x_{l-1}, x_l = x'], \qquad (1)$$

which starts at x and ends at x'. We say that $C(x, x')$ links or connects x to x'. The *links* of a chain are the edges connecting consecutive nodes of the chain in the direction given by the chain. We define the *cost* of a chain (1) as the maximum dissimilarity $\max_{i | x_i \in C(x,x')} A_X(x_i, x_{i+1})$ encountered when traversing its links in order.

3. Quasi-Clustering methods

A partition $P = \{B_1, \ldots, B_J\}$ of a set X can be interpreted as a reduction in data complexity in which variations between elements of a group are neglected in favor of the larger dissimilarities between elements of different groups. This is natural when clustering datasets endowed with symmetric dissimilarities because the concepts of a node $x \in X$ being close to another node $x' \in X$ and x' being close to x are equivalent. In an asymmetric network these concepts are different and this difference motivates the definition of structures more general than partitions.

Considering that a partition $P = \{B_1, \ldots, B_J\}$ of X is induced by an equivalence relation $\sim$ on X we search for the equivalent of an asymmetric partition by removing the symmetry property in the definition of the equivalence relation. Thus, we define a *quasi-equivalence* $\rightsquigarrow$ as a binary relation that satisfies the reflexivity and transitivity properties but is not necessarily symmetric as stated next.

Definition 1 *A binary relation $\rightsquigarrow$ between elements of a set X is a quasi-equivalence if and only if the following properties hold true for all $x, x', x'' \in X$:*

(i) Reflexivity. Points are quasi-equivalent to themselves, $x \rightsquigarrow x$.

(ii) Transitivity. If $x \rightsquigarrow x'$ and $x' \rightsquigarrow x''$ then $x \rightsquigarrow x''$.

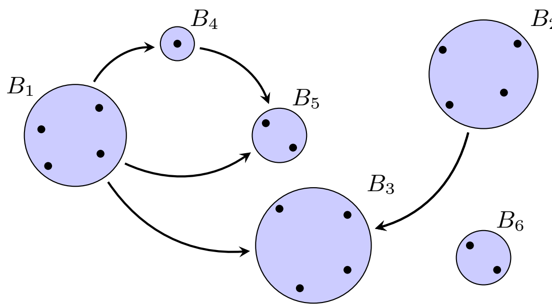

Figure 1. A quasi-partition $\tilde{P} = (P, E)$ on a set of nodes. The vertex set P of the quasi-partition is given by a partition of the nodes $P = \{B_1, B_2, \ldots, B_6\}$. The edges of the directed graph $\tilde{P} = (P, E)$ represent unidirectional influence between the blocks of the partition.

Quasi-equivalence relations are more often termed preorders or quasi-orders in the literature (Harzheim, 2005). We choose the term quasi-equivalence to emphasize that they are a modified version of an equivalence relation.

We define a *quasi-partition* of the set X as a directed, unweighted graph $\tilde{P} = (P, E)$ with no self-loops where the vertex set P is a partition $P = \{B_1, \ldots, B_J\}$ of the space X and the edge set $E \subseteq P \times P$ is such that the following properties are satisfied (see Fig. 1):

(QP1) Unidirectionality. For any given pair of distinct blocks B_i, $B_j \in P$ we have at most one edge between them. Thus, if for some $i \neq j$ we have $(B_i, B_j) \in E$ then $(B_j, B_i) \notin E$.

(QP2) Transitivity. If there are edges between blocks B_i and B_j and between blocks B_j and B_k, then there is an edge between blocks B_i and B_k.

The vertex set P of a quasi-partition $\tilde{P} = (P, E)$ represents sets of nodes that can influence each other, whereas the edges in E capture the notion of directed influence from one group to the next. In the example in Fig. 1, nodes which are drawn together can exert influence on each other. This gives rise to the blocks B_i which form the vertex set P of the quasi-partition. Additionally, some blocks have influence over others in only one direction. E.g., block B_1 can influence B_4 but not vice versa. This latter fact motivates keeping B_1 and B_4 as separate blocks in the partition whereas the former motivates the addition of the directed influence edge (B_1, B_4). Likewise, B_1 can influence B_3, B_2 can influence B_3 and B_4 can influence B_5 but none of these influences are true in the opposite direction. Block B_1 need not be able to directly influence B_5, but can influence it through B_4, hence the edge from B_1 to B_5, in accordance with (QP2). All other influence relations are not meaningful, justifying the lack of connections between the other blocks.

Requirements (QP1) and (QP2) in the definition of quasi-partition represent the relational structure that emerges from quasi-equivalence relations as we state in the following proposition.

Proposition 1 *Given a node set X and a quasi-equivalence relation $\rightsquigarrow$ on X [cf. Definition 1] define the relation $\leftrightarrow$ on X as*

$$x \leftrightarrow x' \quad \iff \quad x \rightsquigarrow x' \text{ and } x' \rightsquigarrow x, \tag{2}$$

for all $x, x' \in X$. Then, $\leftrightarrow$ is an equivalence relation. Let $P = \{B_1, \ldots, B_J\}$ be the partition of X induced by $\leftrightarrow$. Define $E \subseteq P \times P$ such that for all distinct $B_i, B_j \in P$

$$(B_i, B_j) \in E \quad \iff \quad x_i \rightsquigarrow x_j, \tag{3}$$

for some $x_i \in B_i$ and $x_j \in B_j$. Then, $\tilde{P} = (P, E)$ is a quasi-partition of X. Conversely, given a quasi-partition $\tilde{P} = (P, E)$ of X, define the binary relation $\rightsquigarrow$ on X so that for all $x, x' \in X$

$$x \rightsquigarrow x' \iff [x] = [x'] \text{ or } ([x], [x']) \in E, \tag{4}$$

where $[x] \in P$ is the block of the partition P that contains the node x and similarly for $[x']$. Then, $\rightsquigarrow$ is a quasi-equivalence on X.

Proof: See Theorem 4.9, Ch. 1.4 in (Harzheim, 2005). ■

In the same way that an equivalence relation induces and is induced by a partition on a given node set X, Proposition 1 shows that a quasi-equivalence relation induces and is induced by a quasi-partition on X. We can then adopt the construction of quasi-partitions as the natural generalization of clustering problems when given asymmetric data. Further, observe that if the edge set E contains no edges, $\tilde{P} = (P, E)$ is equivalent to the regular partition P when ignoring the empty edge set. In this sense, partitions are particular cases of quasi-partitions having the generic form $\tilde{P} = (P, \emptyset)$. To allow generalizations of hierarchical clustering methods with asymmetric outputs we introduce the notion of *quasi-dendrogram* in the following section.

3.1. Quasi-dendrograms

Given that a dendrogram is defined as a nested set of partitions, we define a *quasi-dendrogram* $\tilde{D}_X$ of the set X as a nested set of quasi-partitions $\tilde{D}_X(\delta) = (D_X(\delta), E_X(\delta))$ indexed by a resolution parameter $\delta \geq 0$. Recall the definition of $[x]_\delta$ from Section 2. Formally, for $\tilde{D}_X$ to be a quasi-dendrogram we require the following conditions:

($\tilde{D}$1) Boundary conditions. At resolution $\delta = 0$ all nodes are in separate clusters with no edges between them and

for some δ_0 sufficiently large all elements of X are in a single cluster,

$$\tilde{D}_X(0) = \Big(\big\{\{x\},\, x \in X\big\},\, \emptyset\Big),$$
$$\tilde{D}_X(\delta_0) = \Big(\{X\}, \emptyset\Big) \quad \text{for some } \delta_0 \geq 0. \tag{5}$$

($\tilde{\text{D}}$2) Equivalence hierarchy. For any pair of points x, x' for which $x \sim_{D_X(\delta_1)} x'$ at resolution δ_1 we must have $x \sim_{D_X(\delta_2)} x'$ for all resolutions $\delta_2 > \delta_1$.

($\tilde{\text{D}}$3) Influence hierarchy. If there is an edge $([x]_{\delta_1}, [x']_{\delta_1}) \in E_X(\delta_1)$ between the equivalence classes $[x]_{\delta_1}$ and $[x']_{\delta_1}$ of nodes x and x' at resolution δ_1, at any resolution $\delta_2 > \delta_1$ we either have $([x]_{\delta_2}, [x']_{\delta_2}) \in E_X(\delta_2)$ or $[x]_{\delta_2} = [x']_{\delta_2}$.

($\tilde{\text{D}}$4) Right continuity. For all $\delta \geq 0$ there exists $\epsilon > 0$ such that $\tilde{D}_X(\delta) = \tilde{D}_X(\delta')$ for all $\delta' \in [\delta, \delta + \epsilon]$.

Requirement ($\tilde{\text{D}}$1) states that for resolution $\delta = 0$ there should be no influence between any pair of nodes and that, for a large enough resolution $\delta = \delta_0$, there should be enough influence between the nodes for all of them to belong to the same cluster. According to ($\tilde{\text{D}}$2), nodes become ever more clustered since once they join together in a cluster, they stay together in the same cluster for all larger resolutions. Condition ($\tilde{\text{D}}$3) states for the edge set the analogous requirement that ($\tilde{\text{D}}$2) states for the node set. If there is an edge present at a given resolution δ_1, that edge should persist at coarser resolutions $\delta_2 > \delta_1$ except if the groups linked by the edge merge in a single cluster. Requirement ($\tilde{\text{D}}$4) is a technical condition that ensures the correct definition of a hierarchical structure [cf. (8) below].

Comparison of ($\tilde{\text{D}}$1), ($\tilde{\text{D}}$2), and ($\tilde{\text{D}}$4) with the three properties defining a dendrogram (Carlsson & Mémoli, 2010) implies that given a quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ on a node set X, the component D_X is a dendrogram on X. I.e, the vertex sets $D_X(\delta)$ of the quasi-partitions $(D_X(\delta), E_X(\delta))$ for varying δ form a nested set of partitions. Hence, if the edge set $E_X(\delta) = \emptyset$ for every resolution parameter, $\tilde{D}_X$ recovers the structure of the dendrogram D_X. Thus, quasi-dendrograms are a generalization of dendrograms, or, equivalently, dendrograms are particular cases of quasi-dendrograms with empty edge sets. Regarding dendrograms D_X as quasi-dendrograms $(D_X, \emptyset)$ with empty edge sets, we have that the set of all dendrograms $\mathcal{D}$ is a subset of $\tilde{\mathcal{D}}$, the set of all quasi-dendrograms.

A hierarchical clustering method $\mathcal{H} : \mathcal{N} \to \mathcal{D}$ is defined as a map from the space of networks $\mathcal{N}$ to the space of dendrograms $\mathcal{D}$. This motivates the definition of a hierarchical *quasi-clustering* method as follows.

Definition 2 *A hierarchical quasi-clustering method $\tilde{\mathcal{H}}$ is defined as a map from the space of networks $\mathcal{N}$ to the space of quasi-dendrograms $\tilde{\mathcal{D}}$,*

$$\tilde{\mathcal{H}} : \mathcal{N} \to \tilde{\mathcal{D}}. \tag{6}$$

Since $\mathcal{D} \subset \tilde{\mathcal{D}}$ we have that every clustering method is a quasi-clustering method but not vice versa. Our goal here is to study quasi-clustering methods satisfying desirable axioms that define the concept of admissibility. In order to facilitate this analysis, we introduce quasi-ultrametrics as asymmetric versions of ultrametrics and show their equivalence to quasi-dendrograms in the following section.

Remark 1 Unidirectionality (QP1) ensures that no cycles containing exactly two nodes can exist in any quasi-partition $\tilde{P} = (P, E)$. If there were longer cycles, transitivity (QP2) would imply that every two distinct nodes in a longer cycle would have to form a two-node cycle, contradicting (QP1). Thus, conditions (QP1) and (QP2) imply that every quasi-partition $\tilde{P} = (P, E)$ is a directed acyclic graph (DAG). The fact that a DAG represents a partial order shows that our construction of a quasi-partition from a quasi-equivalence relation is consistent with the known set theoretic construction of a partial order on a partition of a set given a preorder on the set (Harzheim, 2005).

3.2. Quasi-ultrametrics

We define a *quasi-ultrametric* $\tilde{u}_X$ on a given node set X as follows.

Definition 3 *Given a node set X, a quasi-ultrametric $\tilde{u}_X$ is a non-negative function $\tilde{u}_X : X \times X \to \mathbb{R}_+$ satisfying the following properties for all $x, x', x'' \in X$:*

(i) Identity. $\tilde{u}_X(x, x') = 0$ if and only if $x = x'$.

(ii) Strong triangle inequality. $\tilde{u}_X$ satisfies

$$\tilde{u}_X(x, x') \leq \max(\tilde{u}_X(x, x''), \tilde{u}_X(x'', x')). \tag{7}$$

Quasi-ultrametrics may be regarded as ultrametrics where the symmetry property is not imposed. In particular, the space $\tilde{\mathcal{U}}$ of quasi-ultrametric networks, i.e. networks with quasi-ultrametrics as dissimilarities, is a superset of the space of ultrametric networks $\mathcal{U} \subset \tilde{\mathcal{U}}$. See (Gurvich & Vyalyi, 2012) for structural properties of quasi-ultrametrics.

The following constructions and theorem establish a structure preserving equivalence between quasi-dendrograms and quasi-ultrametrics. Consider the map $\Psi : \tilde{\mathcal{D}} \to \tilde{\mathcal{U}}$ defined as follows: for a given quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ over the set X write $\Psi(\tilde{D}_X) = (X, \tilde{u}_X)$, where we define $\tilde{u}_X(x, x')$ for each $x, x' \in X$ as the smallest resolution δ at which either both nodes belong to the same equivalence class $[x]_\delta = [x']_\delta$, i.e. $x \sim_{D_X(\delta)} x'$, or there exists an edge in $E_X(\delta)$ from the equivalence class $[x]_\delta$ to the equivalence class $[x']_\delta$,

$$\tilde{u}_X(x,x') := \min\Big\{\delta \geq 0 \,\Big|\, [x]_\delta = [x']_\delta \quad \text{or} \quad ([x]_\delta, [x']_\delta) \in E_X(\delta)\Big\}. \tag{8}$$

We also consider the map $\Upsilon : \tilde{\mathcal{U}} \to \tilde{\mathcal{D}}$ constructed as follows: for a given quasi-ultrametric $\tilde{u}_X$ on the set X and each $\delta \geq 0$ define the relation $\sim_{\tilde{u}_X(\delta)}$ on X as

$$x \sim_{\tilde{u}_X(\delta)} x' \iff \max\big(\tilde{u}_X(x,x'), \tilde{u}_X(x',x)\big) \leq \delta. \tag{9}$$

Define further $D_X(\delta) := \big\{X \mod \sim_{\tilde{u}_X(\delta)}\big\}$ and the edge set $E_X(\delta)$ for every $\delta \geq 0$ as follows: $B_1 \neq B_2 \in D_X(\delta)$ are such that

$$(B_1, B_2) \in E_X(\delta) \iff \min_{\substack{x_1 \in B_1 \\ x_2 \in B_2}} \tilde{u}_X(x_1, x_2) \leq \delta. \tag{10}$$

Finally, $\Upsilon(X, \tilde{u}_X) := \tilde{D}_X$, where $\tilde{D}_X := (D_X, E_X)$.

Theorem 1 *The maps $\Psi : \tilde{\mathcal{D}} \to \tilde{\mathcal{U}}$ and $\Upsilon : \tilde{\mathcal{U}} \to \tilde{\mathcal{D}}$ are both well defined. Furthermore, $\Psi \circ \Upsilon$ is the identity on $\tilde{\mathcal{U}}$ and $\Upsilon \circ \Psi$ is the identity on $\tilde{\mathcal{D}}$.*

Theorem 1 implies that quasi-dendrograms $\tilde{D}_X$ can be represented as quasi-ultrametric networks defined on the same underlying node set X. This allows us to reinterpret hierarchical quasi-clustering methods [cf. (6)] as maps

$$\tilde{\mathcal{H}} : \mathcal{N} \to \tilde{\mathcal{U}}, \tag{11}$$

from the space of networks to the space of quasi-ultrametric networks. Apart from the theoretical importance of Theorem 1, this equivalence result is of practical importance since quasi-ultrametrics are mathematically more convenient to handle than quasi-dendrograms. However, quasi-dendrograms are more convenient for representing data as illustrated in Section 4.

Given a quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$, the value $\tilde{u}_X(x,x')$ of the associated quasi-ultrametric for $x, x' \in X$ is given by the minimum resolution δ at which x can influence x'. This may occur when x and x' belong to the same block of $D_X(\delta)$ or when they belong to different blocks $B, B' \in D_X(\delta)$, but there is an edge from the block containing x to the block containing x', i.e. $(B, B') \in E_X(\delta)$. Conversely, given a quasi-ultrametric network $(X, \tilde{u}_X)$, for a given resolution δ the graph $\tilde{D}_X(\delta)$ has as a vertex set the classes of nodes whose quasi-ultrametric is less than δ in both directions. Furthermore, $\tilde{D}_X(\delta)$ contains a directed edge between two distinct equivalence classes if the quasi-ultrametric from some node in the first class to some node in the second is not greater than δ.

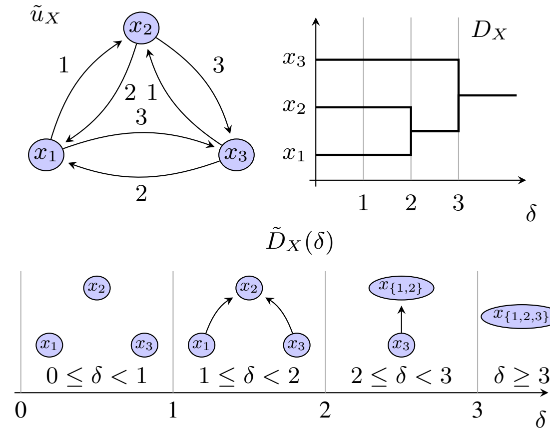

Figure 2. Equivalence between quasi-dendrograms $\tilde{D}_X = (D_X, E_X)$ and quasi-ultrametrics $\tilde{u}_X$.

In Fig. 2 we present an example of the equivalence between quasi-dendrograms and quasi-ultrametric networks stated by Theorem 1. At the top left of the figure, we present a quasi-ultrametric $\tilde{u}_X$ defined on a three-node set $X = \{x_1, x_2, x_3\}$. At the top right, we depict the dendrogram component D_X of the quasi-dendrogram $\tilde{D}_X = (D_X, E_X)$ equivalent to $(X, \tilde{u}_X)$ as given by Theorem 1. At the bottom of the figure, we present graphs $\tilde{D}_X(\delta)$ for a range of resolutions $\delta \geq 0$. To obtain $\tilde{D}_X$ from $\tilde{u}_X$, we first obtain the dendrogram component D_X by symmetrizing $\tilde{u}_X$ to the maximum [cf. (9)], nodes x_1 and x_2 merge at resolution 2 and x_3 merges with $\{x_1, x_2\}$ at resolution 3. To see how the edges in $\tilde{D}_X$ are obtained, at resolutions $0 \leq \delta < 1$, there are no edges since there is no quasi-ultrametric value between distinct nodes in this range [cf. (10)]. At resolution $\delta = 1$, we reach the first non-zero values of $\tilde{u}_X$ and hence the corresponding edges appear in $\tilde{D}_X(1)$. At resolution $\delta = 2$, nodes x_1 and x_2 merge and become the same vertex in graph $\tilde{D}_X(2)$. Finally, at resolution $\delta = 3$ all the nodes belong to the same equivalence class and hence $\tilde{D}_X(3)$ contains only one vertex. Conversely, to obtain $\tilde{u}_X$ from $\tilde{D}_X$ as depicted in the figure, note that at resolution $\delta = 1$ two edges $([x_1]_1, [x_2]_1)$ and $([x_3]_1, [x_2]_1)$ appear in $\tilde{D}_X(1)$, thus the corresponding values of the quasi-ultrametric are fixed to be $\tilde{u}_X(x_1, x_2) = \tilde{u}(x_3, x_2) = 1$. At resolution $\delta = 2$, when x_1 and x_2 merge into the same vertex in $\tilde{D}_X(2)$, an edge is generated from $[x_3]_2$ to $[x_1]_2$ the equivalence class of x_1 at resolution $\delta = 2$ which did not exist before, implying that $\tilde{u}_X(x_3, x_1) = 2$. Moreover, we have that $[x_2]_2 = [x_1]_2$, hence $\tilde{u}_X(x_2, x_1) = 2$. Finally, at $\tilde{D}_X(3)$ there is only one equivalence class, thus the values of $\tilde{u}_X$ that have not been defined so far must equal 3.

3.3. Admissible quasi-clustering methods

We encode desirable properties of quasi-clustering methods into axioms which we use as a criterion for admissibility. The Directed Axiom of Value (Ã1) and the Directed Axiom of Transformation (Ã2) winnow the space of quasi-clustering methods by imposing conditions on their output quasi-ultrametrics which, by Theorem 1, is equivalent to imposing conditions on the output quasi-dendrograms. Defining an arbitrary two-node network $\vec{\Delta}_2(\alpha,\beta) := (\{p,q\}, A_{p,q})$ with $A_{p,q}(p,q) = \alpha$ and $A_{p,q}(q,p) = \beta$ for some $\alpha, \beta > 0$,

(Ã1) Directed Axiom of Value. $\tilde{\mathcal{H}}(\vec{\Delta}_2(\alpha,\beta)) = \vec{\Delta}_2(\alpha,\beta)$ for every two-node network $\vec{\Delta}_2(\alpha,\beta)$.

(Ã2) Directed 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, for all $x, x' \in X$, the outputs $(X, \tilde{u}_X) = \tilde{\mathcal{H}}(X, A_X)$ and $(Y, \tilde{u}_Y) = \tilde{\mathcal{H}}(Y, A_Y)$ satisfy

$$\tilde{u}_X(x,x') \geq \tilde{u}_Y(\phi(x), \phi(x')). \tag{12}$$

The Directed Axiom of Transformation (Ã2) states that no influence relation can be weakened by a dissimilarity reducing transformation. That is, if relations in the network are strengthened, the tendency of nodes to cluster cannot decrease. The Directed Axiom of Value (Ã1) simply recognizes that in any two-node network, the dissimilarity function is itself a quasi-ultrametric and that there is no valid justification to output a different quasi-ultrametric.

3.4. Existence and uniqueness of admissible quasi-clustering methods: directed single linkage

We call a quasi-clustering method $\tilde{\mathcal{H}}$ admissible if it satisfies axioms (Ã1) and (Ã2) and we want to find methods that are admissible with respect to these axioms. This is not difficult. Define the directed minimum chain cost $\tilde{u}^*_X(x,x')$ between nodes x and x' as the minimum chain cost among all chains connecting x to x'. Formally, for all $x, x' \in X$,

$$\tilde{u}^*_X(x,x') = \min_{C(x,x')} \max_{i | x_i \in C(x,x')} A_X(x_i, x_{i+1}). \tag{13}$$

Define the *directed single linkage* (DSL) hierarchical quasi-clustering method $\tilde{\mathcal{H}}^*$ as the one with output quasi-ultrametrics $(X, \tilde{u}^*_X) = \tilde{\mathcal{H}}^*(X, A_X)$ given by the directed minimum chain cost function $\tilde{u}^*_X$. The DSL method is valid and admissible as we show in the following proposition.

Proposition 2 *The hierarchical quasi-clustering method $\tilde{\mathcal{H}}^*$ is valid and admissible. I.e., $\tilde{u}^*_X$ defined by (13) is a quasi-ultrametric and $\tilde{\mathcal{H}}^*$ satisfies axioms (Ã1)-(Ã2).*

We next ask which other methods satisfy (Ã1)-(Ã2) and what special properties DSL has. As it turns out, DSL is the unique quasi-clustering method that is admissible with respect to (Ã1)-(Ã2) as we assert in the following theorem.

Theorem 2 *Let $\tilde{\mathcal{H}}$ be a valid hierarchical quasi-clustering method satisfying axioms (Ã1) and (Ã2). Then, $\tilde{\mathcal{H}} \equiv \tilde{\mathcal{H}}^*$ where $\tilde{\mathcal{H}}^*$ is the DSL method with output quasi-ultrametrics as in (13).*

In (Carlsson & Mémoli, 2010), it was shown that single linkage is the only admissible hierarchical clustering method for finite metric spaces. Admissibility was defined by three axioms, two of which are undirected versions of (Ã1) and (Ã2). In (Carlsson et al., 2013), they show that when replacing metric spaces by more general asymmetric networks, the uniqueness result is lost and an infinite number of methods satisfy the admissibility axioms. In our paper, by considering the more general framework of quasi-clustering methods, we recover the uniqueness result even for asymmetric networks. Moreover, Theorem 2 shows that the only admissible method is a directed version of single linkage. In this way, it becomes clear that the non-uniqueness result for asymmetric networks in (Carlsson et al., 2013) is originated in the symmetry mismatch between the input asymmetric network and the output symmetric dendrogram. When we allow the more general asymmetric quasi-dendrogram as output, the uniqueness result is recovered.

DSL was identified as a natural extension of single linkage hierarchical clustering to asymmetric networks in (Boyd, 1980). In our paper, by developing a framework to study hierarchical quasi-clustering methods and leveraging the equivalence result in Theorem 1, we show that DSL is the *unique* admissible way of quasi-clustering asymmetric networks. Furthermore, stability and invariance properties are established in the following section.

Remark 2 (Axiomatic strength and chaining effect) DSL, having a strong resemblance to single linkage hierarchical clustering on finite metric spaces, is likely to be sensitive to a directed version of the so called chaining effect (Jain & Dubes, 1988). By requiring a weaker version of (Ã2), the most stringent of our two axioms, the uniqueness result in Theorem 2 is lost and density aware methods, that do not suffer from the chaining effect, become admissible. This direction, shown to be successful for finite metric spaces (Carlsson & Mémoli, 2013), appears to be an interesting research avenue.

3.5. Stability and invariance properties of DSL

DSL is stable in the sense that if it is applied to similar networks then it outputs similar quasi-dendrograms. This notion has been used to study stability of clustering methods for finite metric spaces (Carlsson & Mémoli, 2010). More precisely, we define a notion of distance between networks

$d_{\mathcal{N}}$ which is an analogue to the Gromov-Hausdorff distance (Gromov, 2007) between metric spaces and defines a legitimate metric on $\mathcal{N}$ (see A.4 in supplementary material for details). Since we may regard DSL as a map $\mathcal{N} \longrightarrow \tilde{\mathcal{U}}$ and $\tilde{\mathcal{U}}$ is a subset of $\mathcal{N}$, we are in a position in which we can use $d_{\mathcal{N}}$ to express the stability of $\tilde{\mathcal{H}}^*$.

Theorem 3 *For all $N_X, N_Y \in \mathcal{N}$,*

$$d_{\mathcal{N}}\big(\tilde{\mathcal{H}}^*(N_X), \tilde{\mathcal{H}}^*(N_Y)\big) \leq d_{\mathcal{N}}(N_X, N_Y).$$

Theorem 3 states that the distance between the output quasi-ultrametrics is upper bounded by the distance between the input networks. Thus, for DSL, nearby networks yield nearby quasi-ultrametrics. In particular, Theorem 3 ensures that noisy dissimilarity data has limited effect on quasi-dendrograms. Also, the theorem implies that DSL is permutation invariant; see A.7 in supplementary material.

For a non-decreasing function $\psi : [0,\infty) \to [0,\infty)$ such that $\psi(a) = 0$ if and only if $a = 0$, and $N_X = (X, A_X) \in \mathcal{N}$ we write $\psi(N_X)$ to denote the network $(X, \psi(A_X))$. Any such ψ will be referred to as a *change of scale function*. Then, DSL is a scale invariant method as the following proposition asserts.

Proposition 3 *For all $N_X \in \mathcal{N}$ and all change of scale functions ψ one has $\psi\big(\tilde{\mathcal{H}}^*(N_X)\big) = \tilde{\mathcal{H}}^*\big(\psi(N_X)\big)$.*

Since Proposition 3 asserts that the quasi-ultrametric outcome is transformed by the same function ψ that alters the dissimilarity function in the original network, DSL is invariant to change of units. More precisely, in terms of quasi-dendrograms, a transformation of dissimilarities through ψ results in a transformed quasi-dendrogram where the order in which influences between nodes arise is the same as in the original one while the resolution at which they appear changes according to ψ. For further invariances of DSL, see A.7 in the supplementary materials.

3.6. Algorithms

In this section we interpret A_X as a matrix of dissimilarities and $\tilde{u}^*_X$ as a symmetric matrix with entries corresponding to the quasi-ultrametric values $\tilde{u}^*_X(x,x')$ for all $x, x' \in X$. By (13), DSL quasi-clustering searches for directed chains of minimum infinity norm cost in A_X to construct the matrix $\tilde{u}^*_X$. This operation can be performed algorithmically using matrix powers in the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ (Gondran & Minoux, 2008).

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

$$\big[A \otimes B\big]_{ij} = \bigoplus_{k=1}^{n} \big(A_{ik} \otimes B_{kj}\big) = \min_{k\in[1,n]} \max\big(A_{ik}, B_{kj}\big). \tag{14}$$

Dioid powers $A_X^{(k)} := A_X \otimes A_X^{(k-1)}$ with $A_X^{(1)} = A_X$ of a dissimilarity matrix are related to quasi-ultrametric matrices $\tilde{u}$. For instance, the elements of the dioid power $\tilde{u}^{(2)}$ of a given quasi-ultrametric matrix $\tilde{u}$ are given by

$$\big[\tilde{u}^{(2)}\big]_{ij} = \min_{k\in[1,n]} \max\big(\tilde{u}_{ik}, \tilde{u}_{kj}\big). \tag{15}$$

Since $\tilde{u}$ satisfies the strong triangle inequality we have that $\tilde{u}_{ij} \leq \max(\tilde{u}_{ik}, \tilde{u}_{kj})$ for all k. In particular, for $k = j$ we have that $\max(\tilde{u}_{ij}, \tilde{u}_{jj}) = \max(\tilde{u}_{ij}, 0) = \tilde{u}_{ij}$. Combining these two observations it follows that the result of the minimization in (15) is $[\tilde{u}^{(2)}]_{ij} = \tilde{u}_{ij}$. This being valid for all i, j implies $\tilde{u}^{(2)} = \tilde{u}$. Furthermore, a matrix satisfying $\tilde{u}^{(2)} = \tilde{u}$ is such that $\tilde{u}_{ij} = [\tilde{u}^{(2)}]_{ij} = \min_{k\in[1,n]} \max(\tilde{u}_{ik}, \tilde{u}_{kj}) \leq \max(\tilde{u}_{ik}, \tilde{u}_{kj})$ for all k, which is just a restatement of the strong triangle inequality. Therefore, a non-negative matrix $\tilde{u}$ represents a finite quasi-ultrametric space if and only if $\tilde{u}^{(2)} = \tilde{u}$ and only the diagonal elements are null. Building on this fact, we develop the following algorithm to implement DSL.

Proposition 4 *For every network (X, A_X) with $|X| = n$, the quasi-ultrametric $\tilde{u}^*_X$ is given by*

$$\tilde{u}^*_X = A_X^{(n-1)}, \tag{16}$$

where the operation $(\cdot)^{(n-1)}$ denotes the $(n-1)$st matrix power in the dioid algebra $(\mathbb{R}^+ \cup \{+\infty\}, \min, \max)$ with matrix product as defined in (14).

Matrix powers in dioid algebras are tractable operations. Indeed, there exist sub cubic dioid power algorithms (Vassilevska et al., 2009; Duan & Pettie, 2009) of complexity $O(n^{2.688})$. Thus, Proposition 4 shows computational tractability of DSL. There exist related methods with lower complexity. For instance, Tarjan's method (Tarjan, 1983), which takes as input an asymmetric network but in contrast to our method enforces symmetry in its output, runs in time $O(n^2 \log n)$ for complete networks. It seems of interest to ascertain whether Tarjan's method can be modified to suit our (asymmetric) output construction. In the following section we use (16) to quasi-cluster a real-world network.

4. Applications

The number of migrants from state to state in the U.S. is published yearly (United States Census Bureau, 2011). We denote by S the set of all states and by $A_S : S \times S \to \mathbb{R}_+$ a migrational dissimilarity such that $A_S(s,s) = 0$ for all

$s \in S$ and $A_S(s, s')$ for all $s \neq s' \in S$ is a monotonically decreasing function of the fraction of immigrants to s' that come from s (see A.9 in supplementary material for details). A small dissimilarity from state s to s' implies that, among all the immigrants into s', a high percentage comes from s. We then construct the network $N_S = (S, A_S)$ with node set S and dissimilarities A_S. The application of hierarchical clustering to migration data has been extensively investigated by Slater, see (Slater, 1976; 1984).

The outcome of applying DSL to the migration network N_S is computed via (16). By Theorem 1, the output quasi-ultrametric is equivalent to a quasi-dendrogram $\tilde{D}^*_S = (D^*_S, E^*_S)$. By analyzing the dendrogram component D^*_S of the quasi-dendrogram $\tilde{D}^*_S$, the influence of geographical proximity in migrational preference is evident; see Fig. 4 in Section A.9 of the supplementary material.

To facilitate display and understanding, we do not present quasi-partitions for all the nodes and resolutions. Instead, we restrict the quasi-ultrametric to a subset of states representing an extended West Coast including Arizona and Nevada. In Fig. 3, we depict quasi-partitions at four relevant resolutions of the quasi-dendrogram equivalent to the restricted quasi-ultrametric. States represented with the same color in the maps in Fig. 3 are part of the same cluster at the given resolution and states in white form singleton clusters. Arrows between clusters for a given resolution δ represent the edge set $E^*_S(\delta)$ which we interpret as a migrational influence relation between the blocks of states.

The DSL quasi-clustering method $\tilde{\mathcal{H}}^*$ captures not only the formation of clusters but also the asymmetric influence between them. E.g. the quasi-partition in Fig. 3 for resolution $\delta = 0.859$ is of little interest since every state forms a singleton cluster. The influence structure, however, reveals a highly asymmetric migration pattern. At this resolution California has migrational influence over every other state in the region as depicted by the four arrows leaving California and entering each of the other states. This influence can be explained by the fact that California contains the largest urban areas of the region such as Los Angeles. Hence, these urban areas attract immigrants from all over the country, reducing the proportional immigration into California from its neighbors and generating the asymmetric influence structure observed. Since this influence structure defines a partial order over the clusters, the quasi-partition at resolution $\delta = 0.859$ permits asserting the reasonable fact that California is the dominant migration force in the region.

At larger resolutions we can ascertain the relative importance of clusters. At resolution $\delta = 0.921$ we can say that California is more important than the cluster formed by Oregon and Washington as well as more important than Arizona and Nevada. We can also see that Arizona precedes Nevada in the migration ordering at this resolution

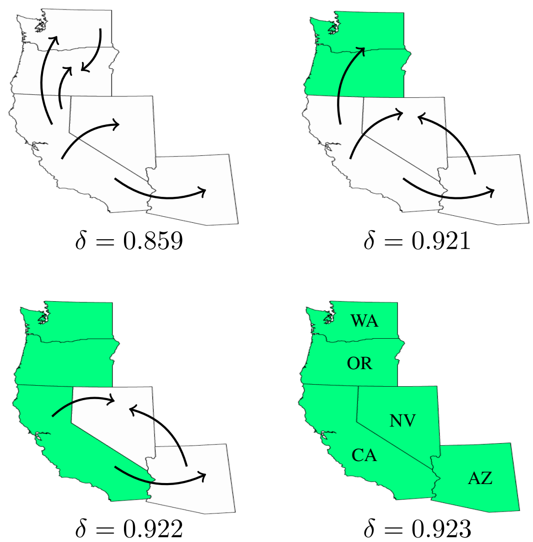

Figure 3. Directed single linkage quasi-clustering method applied to the extended West Coast migration flow.

while the remaining pairs of the ordering are undefined. At resolution $\delta = 0.922$ there is an interesting pattern as we can see the cluster formed by the three West Coast states preceding Arizona and Nevada in the partial order. At this resolution the partial order also happens to be a total order as Arizona is seen to precede Nevada.

For further applications, refer to Section A.9 in the supplementary material.

5. Conclusion

When clustering asymmetric networks, requiring the output to be symmetric – as in hierarchical clustering – might be undesirable. Hence, we defined quasi-dendrograms, a generalization of dendrograms that admits asymmetric relations, and developed a theory for quasi-clustering methods. We formalized the notion of admissibility by introducing two axioms. Under this framework, we showed that DSL is the unique admissible method. We pointed out that less stringent frameworks that give rise to new admissible methods can be explored by weakening the Directed Axiom of Transformation. Furthermore, we proved an equivalence between quasi-dendrograms and quasi-ultrametrics that generalizes the well-known equivalence between dendrograms and ultrametrics, and established the stability and invariance properties of the DSL method. Finally, we illustrated the application of DSL to a migration network.

Acknowledgments

Work in this paper is supported by NSF CCF-0952867, AFOSR MURI FA9550-10-1-0567, DARPA GRAPHS FA9550-12-1-0416, AFOSR FA9550-09-0-1-0531, AFOSR FA9550-09-1-0643, NSF DMS 0905823, and NSF DMS-0406992.

References

Boyd, J.P. Asymmetric clusters of internal migration regions of france. *Ieee Transactions on Systems Man and Cybernetics*, (2):101–104, 1980.

Bureau of Economic Analysis. Input-output accounts: the use of commodities by industries before redefinitions. *U.S. Department of Commerce*, 2011. URL http://www.bea.gov/iTable/index_industry.cfm.

Carlsson, G. and Mémoli, F. Characterization, stability and convergence of hierarchical clustering methods. *Journal of Machine Learning Research*, 11:1425–1470, 2010.

Carlsson, G. and Mémoli, F. Classifying clustering schemes. *Foundations of Computational Mathematics*, 13(2):221–252, 2013.

Carlsson, G., Memoli, F., Ribeiro, A., and Segarra, S. Axiomatic construction of hierarchical clustering in asymmetric networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 5219–5223, 2013.

Duan, R. and Pettie, S. Fast algorithms for (max, min)-matrix multiplication and bottleneck shortest paths. *Symposium on discrete algorithms*, 2009.

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

Gromov, M. *Metric structures for Riemannian and non-Riemannian spaces*. Birkhäuser Boston Inc., Boston, MA, 2007. ISBN 978-0-8176-4582-3; 0-8176-4582-9.

Gurvich, V. and Vyalyi, M. Characterizing (quasi-) ultrametric finite spaces in terms of (directed) graphs. *Discrete Applied Mathematics*, 160(12):1742–1756, 2012.

Harzheim, E. *Ordered sets*. Springer, 2005.

Hubert, L. Min and max hierarchical clustering using asymmetric similarity measures. *Psychometrika*, 38(1): 63–72, 1973.

Jain, A.K. and Dubes, R. C. *Algorithms for clustering data*. Prentice Hall Advanced Reference Series. Prentice Hall Inc., 1988.

Jardine, N. and Sibson, R. *Mathematical taxonomy*. John Wiley & Sons Ltd., London, 1971. Wiley Series in Probability and Mathematical Statistics.

Lance, G. N. and Williams, W. T. A general theory of classificatory sorting strategies 1. Hierarchical systems. *Computer Journal*, 9(4):373–380, 1967.

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

Murtagh, F. Multidimensional clustering algorithms. *Compstat Lectures, Vienna: Physika Verlag, 1985*, 1, 1985.

Newman, M. and Girvan, M. Community structure in social and biological networks. *Proc. Ntnl. Acad. Sci.*, 99 (12):7821–7826, 2002.

Newman, M. and Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69, 026113, 2004.

Ng, A., Jordan, M., and Weiss, Y. 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*, 2:849–856, 2002.

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

Saito, T. and Yadohisa, H. *Data analysis of asymmetric structures: advanced approaches in computational statistics*. CRC Press, 2004.

Shi, J. and Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

Slater, P.B. Hierarchical internal migration regions of france. *Systems, Man and Cybernetics, IEEE Transactions on*, (4):321–324, 1976.

Slater, P.B. A partial hierarchical regionalization of 3140 us counties on the basis of 1965-1970 intercounty migration. *Environment and Planning A*, 16(4):545–550, 1984.

Tarjan, R. E. An improved algorithm for hierarchical clustering using strong components. *Inf. Process. Lett.*, 17 (1):37–41, 1983.

United States Census Bureau. State-to-state migration flows. *U.S. Department of Commerce*, 2011. URL http://www.census.gov/hhes/migration/data/acs/state-to-state.html.

Vassilevska, V., Williams, R., and Yuster, R. All pairs bottleneck paths and max-min matrix products in truly subcubic time. *Theory of Computing*, 5:173–189, 2009.

Von Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 12 2007.

Zhao, Y. and Karypis, G. Hierarchical clustering algorithms for document datasets. *Data Mining and Knowledge Discovery*, 10:141–168, 2005.