A Density-Sensitive Hierarchical Clustering Method

Álvaro Martínez-Pérez

Universidad de Castilla-La Mancha, Spain

Abstract: We define a hierarchical clustering method: $\alpha$-unchaining single linkage or $SL(\alpha)$. The input of this algorithm is a finite space with a distance function and a certain parameter $\alpha$. This method is sensitive to the density of the distribution and offers some solution to the so-called chaining effect. We also define a modified version, $SL^*(\alpha)$, to treat the chaining through points or small blocks. We study the theoretical properties of these methods and offer some theoretical background for the treatment of chaining effects.

Keywords: Hierarchical clustering; Single linkage; Chaining effect; Weakly un-unchaining; $\alpha$-bridge-unchaining.

MSC: 62H30, 68T10

1. Introduction

A clustering method is an algorithm that takes as input a finite space with a distance function (typically, a finite metric space) $(X, d)$ and gives as output a partition of $X$.

Kleinberg (2002) proposed an axiomatic approach to clustering defining a few basic properties that a clustering function should satisfy. Then, he proved that no clustering function can satisfy these conditions simultaneously. This does not imply the impossibility of defining a consistent standard
clustering algorithm. Kleinberg’s impossibility only holds when the unique input in the algorithm is the space and the set of distances. It can be avoided including, for example, the number of clusters to be obtained as part of the input. See Ackerman, Ben-David and Locker (2010) and Zadeh and Ben-David (2009). Also, Ackerman and Ben-David (2008) showed that these axioms are consistent in the setting of clustering quality measures. Carlsson and Mémoli (2010b) study the analogous problem for clustering schemes that yield hierarchical decompositions instead of a certain partition of the space. See also Carlsson and Mémoli (2008; 2010a). Hierarchical clustering methods also take as input a finite metric space but the output is a hierarchical family of partitions of $X$. They approach the subject focusing on a theoretical basis for the study of hierarchical clustering ($HC$). In the spirit of Kleinberg’s result, they define a few reasonable conditions that a $HC$ method should hold. They prove that the unique $HC$ method satisfying three basic conditions is (the well-known) single linkage hierarchical clustering, $SLHC$ (see Jardine and Sibson (1971) for a seminal study of its theoretical properties). Ackerman and Ben-David (2011) proved also a characterization of the class of linkage-based algorithms, including $SL$. See also Ackerman, Brown and Locker (2012). In the setting of partitional (standard) clustering, there is another characterization by Zadeh and Ben-David (2009).

Carlsson and Mémoli (2010b) also study the theoretical properties of $SLHC$ obtaining some interesting results. The main advantage seems to be that this method enjoys some sort of stability which is defined by means of the Gromov-Hausdorff distance. However, the main weakness of $SLHC$ is the so-called chaining effect which may merge clusters that, in practice, “should be detected” by the algorithm as independent clusters. One way to address this difficulty is to take into account the density. In Carlsson and Mémoli (2010c), the authors do this by including in the input of the algorithm a function that provides that information. Other approach is using minimal spanning trees as in Zahn (1971) or Narendra and Goldberg (1980).

The first challenge is that what we may consider the “undesired” chaining effect depends on the characteristics of the problem under study. The definition from Lance and Williams (1967) makes reference to the higher tendency of the points to add to a pre-existing group rather than defining the nucleus of a new group or joining to another single point. Our algorithm is oriented to prevent the tendency to merge two clusters when the minimal distance between them is small. These are called touching clusters in Zahn (1971). This is typically the problem of $SLHC$ and it might be a problem in many practical situations. Consider, for example, two clusters in $\mathbb{R}^n$ following a multivariate normal distribution where the minimal distance between the clusters is small.
We also include as an undesired chaining effect the case of two big clusters joined by a chain of points or small clusters. This idea is closer to the type of chaining effect considered by Wishart (1969).

There exist other linkage-based methods that offer some resistance to these chaining effects as average linkage, AL, or complete linkage, CL. See Rohlf (1982) for a comparative survey of linkage-based clustering. However, although the main problem of the chaining effect of SL HC is reduced, these methods have a tendency to merge isolated points before joining them to pre-existing big clusters and this might be unwanted too. Also, these methods are proved to be extremely unstable in the sense that small perturbations on the data yield very different dendrograms. See Carlsson and Mémoli (2010b).

Herein, we define a new HC method on the basis of SL: \( \alpha \)-unchaining single linkage or \( SL(\alpha) \). The definition of \( SL(\alpha) \) uses the dimension of the Vietoris-Rips complexes defined by the points of \( X \). These complexes reflect the density distribution of the sample inducing high dimensional simplices where points are closely packed together. Thus, \( SL(\alpha) \) is a density-sensitive algorithm where the input is just the set of distances between the points and a fixed parameter \( \alpha \in \mathbb{N} \). The parameter determines how sensitive the method will be to the chaining effect.

To treat the chaining through single points or smaller blocks (being a block an element of the partition at some level of the dendrogram as in Carlsson and Mémoli (2010b)), we define another version of the method, \( SL^*(\alpha) \), by adding an extra condition on \( SL(\alpha) \).

There exist many different approaches to clustering that provide useful and efficient algorithms such as Ward’s method, Baire ultrametric, Zahn’s minimal spanning trees, etc. Of course, different approaches provide different strengths and weaknesses. Among the variety of clustering and hierarchical clustering methods some of them exhibit some sensitivity to the density of the distribution. It is not possible to analyze them all herein to provide a good comparison with our method so we focus on linkage-based methods. However, it is worth mentioning at least one of the most popular algorithms, DBSCAN, defined by Ester et al. (1996).

This paper intends to give a theoretical basis to the study of the problem. So, instead of checking the algorithm on examples of real data we rather try to find general properties characterizing what would be an undesired chaining of two blocks and how good is the algorithm detecting and unchaining them. However, we include several examples where the unchaining properties can be explicitly seen in the resulting dendrogram.

We introduce the concepts of chained subsets and subsets chained by smaller blocks so that they contain what we consider the problematic examples. Nevertheless, there may be many examples of chained subsets
which should be merged by any reasonable algorithm. In such property-based context, we say that a $HC$ method is strongly chaining if every pair of chained subsets are always merged before they appear contained in different clusters. A $HC$ method is completely chaining if, in addition, every pair of subsets chained through smaller blocks are merged before they appear contained in different clusters. Thus, strongly chaining methods and completely chaining methods are extremely sensitive to these chaining effects. This is the case, for example, of $SLHC$. In Section 4 we first introduce the idea of what are we considering undesired chaining effect. Formal definitions and theoretical results are left to the last sections to enhance readability.

We say that a $HC$ method is weakly unchaining if, at least, it is able to detect two clusters where points accumulate in the interior, and such that the minimal distance between the clusters is small only because of a single pair of points. Then, we prove that $SL(\alpha)$ satisfies this condition while other methods which are not strongly chaining as $AL$ and $CLHC$ fail to be weakly unchaining. See Section 6.

We also define a minimal condition for two subsets chained by single points that should be detected. We say that a $HC$ method is $\alpha$-bridge-unchaining if it is able to separate two blocks in that situation. $SL^*(\alpha)$ is proved to be more sensitive than that. It also detects some classes of chaining through smaller blocks. In particular, $SL^*(\alpha)$ is $\alpha$-bridge-unchaining. See Section 7.

Thus, this work has two objectives. The first one is to offer an alternative approach to deal with the chaining effect providing new algorithms, and the second one is to give some theoretical background for the study of this effect. So, herein we check the chaining properties of $SL(\alpha)$, $SL^*(\alpha)$, $SL$, $AL$ and $CL$, this is, if they are strongly chaining, completely chaining, weakly unchaining and $\alpha$-bridge unchaining. This work is complemented in Martínez-Pérez (2016) by studying some other abstract significant properties (permutation invariant, rich, weight-robust, faithful...) of $SL(\alpha)$ and $SL^*(\alpha)$ and comparing them with $SL$, $AL$ or $CL$. This offers the practitioner the opportunity to choose the most appropriate (linkage-based) algorithm based on which are the most relevant properties for the problem. Stability properties are studied in Martínez-Pérez (2015). Computational efficiency and effective performance of the algorithm are yet to be studied.

2. Background and Notation

A hierarchical clustering scheme over a finite set is a nested family of partitions (see Johnson, 1967) which is usually represented by a dendrogram.
Let \( \mathcal{P}(X) \) denote the collection of all partitions of a finite set \( X = \{x_1, \ldots, x_n\} \). Then, a dendrogram can be described as a rooted tree (with some labeling) where the leaves are the singletons or as a map \( \theta: [0, \infty) \to \mathcal{P}(X) \) such that:

1. \( \theta(0) = \{\{x_1\}, \{x_2\}, \ldots, \{x_n\}\} \),
2. there exists \( T \) such that \( \theta(t) = X \) for every \( t \geq T \),
3. if \( r \leq s \) then \( \theta(r) \) refines \( \theta(s) \),
4. for all \( r \) there exists \( \varepsilon > 0 \) such that \( \theta(r) = \theta(t) \) for \( t \in [r, r + \varepsilon] \).

Notice that conditions 2 and 4 imply that there exist \( t_0 < t_1 < \ldots < t_m \) such that \( \theta(r) = \theta(t_{i-1}) \) for every \( r \in [t_{i-1}, t_i) \), \( i = 0, 1, \ldots, m \) and \( \theta(r) = \theta(t_m) = \{X\} \) for every \( r \in [t_m, \infty) \).

For any partition \( \{B_1, ..., B_k\} \in \mathcal{P}(X) \), the subsets \( B_i \) are called blocks.

Let \( D(X) \) denote the collection of all possible dendrograms over a finite set \( X \). Given some \( \theta \in D(X) \), let us denote \( \theta(t) = \{B_1^t, ..., B_k^t\} \). Therefore, the nested family of partitions is given by the corresponding partitions at \( t_0, ..., t_m \), this is, \( \{B_1^{t_i}, ..., B_k^{t_i}\} \), \( i = 0, ..., m \).

An ultrametric space is a metric space \((X, d)\) such that \( d(x, y) \leq \max\{d(x, z), d(z, y)\} \) for all \( x, y, z \in X \). Given a finite metric space \( X \) let \( \mathcal{U}(X) \) denote the set of all ultrametrics over \( X \).

There is a well known equivalence between trees and ultrametrics. Hughes (2004) or Martínez-Pérez and Morón (2009) show how this correspondence can be seen as an equivalence of categories. In particular, this may be translated into an equivalence between dendrograms and ultrametrics.

Thus, a hierarchical clustering method \( \Xi \) can be presented as an algorithm whose output is a dendrogram or an ultrametric space. Let \( \Xi_D(X, d) \) denote the dendrogram obtained by applying \( \Xi \) to a metric space \((X, d)\) and \( \Xi_U(X, d) \) denote the corresponding ultrametric space.

Let us define the map \( \eta: D(X) \to U(X) \) as follows:

Given a dendrogram \( \theta \in D(X) \), let \( \eta(\theta) = u_\theta \) be such that \( u_\theta(x, x') = \min\{r \geq 0 \mid x, x' \text{ belong to the same block of } \theta(r)\} \).

**Proposition 2.1** (Carlsson and Mémoli 2010b, Theorem 9) \( \eta \) is a bijection such that \( \eta \circ \Xi_D = \Xi_U \).

**Notation:** For any HC method \( \Xi \) and any finite metric space \((X, d)\), let us denote \( \Xi_D(X, d) = \theta_X \) and \( \Xi_U(X, d) = (X, u_X) \). If there is no ambiguity on which is the metric space we shall just write \( \Xi_D(X, d) = \theta \).
3. Hierarchical Clustering Methods

Let us recall the definition of some well-known hierarchical clustering methods. For a detailed description with multiple examples we refer the reader to Sneath and Sokal (1973). We include here the description of single linkage based on its t-connected components and the recursive description of single linkage, complete linkage and average linkage as presented in Carlsson and Mémoli (2010b). Based on the latter one, we introduce also an alternative description of these methods which will be the key to build our new method, $SL^* (\alpha)$. The flexibility of our approach might be useful to define other algorithms adapted to other specific problems since it allows to define complex conditions to prevent two blocks from being merged.

An $\varepsilon$-chain is a finite sequence of points $x_0,\ldots,x_N$ that are separated by distances less than or equal to $\varepsilon$: $d(x_i,x_{i+1}) < \varepsilon$. Two points are $\varepsilon$-connected if there is an $\varepsilon$-chain joining them. Any two points in an $\varepsilon$-connected set can be linked by an $\varepsilon$-chain. An $\varepsilon$-component is a maximal $\varepsilon$-connected subset.

Clearly, given a metric space and any $\varepsilon > 0$, there is a partition of $X$ into its $\varepsilon$-components $\{C_{1}^{\varepsilon},\ldots,C_{k(\varepsilon)}^{\varepsilon}\}$.

Let $X$ be a finite metric set. The single linkage HC is defined by the map $\theta_{SL}: [0, \infty) \rightarrow P(X)$ such that $\theta_{SL}(t)$ is the partition of $X$ into its $t$-components. See Figure 1.

In Carlsson and Mémoli (2010b), the authors use a recursive procedure to redefine $SLHC$, average linkage (AL) and complete linkage (CL) hierarchical clustering. The main advantage of this procedure is that it allows to merge more than two clusters at the same time and therefore, the hierarchical clustering does not depend on the order in which the points are introduced in the algorithm. We reproduce here, for completeness, their formulation.

Let $(X,d)$ be a finite metric space where $X = \{x_1,\ldots,x_n\}$ and let $L$ denote a family of linkage functions on $X$:

$$L := \{\ell: \mathcal{C}(X) \times \mathcal{C}(X) \rightarrow \mathbb{R}^+ \mid \ell \text{ is bounded and non-negative} \}$$

where $\mathcal{C}(X)$ denotes the collection of all non-empty subsets of $X$.

Some standard choices for $\ell$ are:

- Single linkage: $\ell^{SL}(B, B') = \min_{(x,x') \in B \times B'} d(x, x')$
- Complete linkage: $\ell^{CL}(B, B') = \max_{(x,x') \in B \times B'} d(x, x')$
- Average linkage: $\ell^{AL}(B, B') = \frac{\sum_{(x,x') \in B \times B'} d(x, x')}{\#(B) \cdot \#(B')}$ where $\#(X)$ denotes the cardinality of the set $X$. 

A Density-Sensitive Hierarchical Clustering Method

Figure 1. $\theta_{SL}(\varepsilon)$ is the partition of $X$ into its $\varepsilon$-components.

Fix some linkage function $\ell \in L$. Then, in the recursive formulation the algorithm starts with a partition where the blocks are the single points. Then, at every step, the minimal distance between blocks is computed using the linkage function and every pair of blocks minimizing this distance is merged. This process is repeated until there is a unique block left. Formally,

1. For each $R > 0$ consider the equivalence relation $\sim_{\ell,R}$ on blocks of a partition $\Pi \in \mathcal{P}(X)$, given by $B \sim_{\ell,R} B'$ if and only if there is a sequence of blocks $B = B_1, ..., B_s = B'$ in $\Pi$ with $\ell(B_k, B_{k+1}) \leq R$ for $k = 1, ..., s - 1$.

2. Consider the sequences $R_0, R_1, R_2, ... \in [0, \infty)$ and $\Theta_0, \Theta_1, \Theta_2, ... \in \mathcal{P}(X)$ given by $R_0 = 0$, $\Theta_0 := \{x_1, ..., x_n\}$, and recursively for $i \geq 1$ by $\Theta_i = \frac{\Theta_{i-1}}{\sim_{\ell,R_i}}$ where $R_i := \min\{\ell(B, B') \mid B, B' \in \Theta_{i-1}, B \neq B'\}$ until $\Theta_i = \{X\}$.

3. Finally, let $\theta_{\ell} : [0, \infty) \to \mathcal{P}(X)$ be such that $\theta_{\ell}(r) := \Theta_{i(r)}$ with $i(r) := \max\{i \mid R_i \leq r\}$.

Remark 1. We can also reformulate the recursive algorithm above as follows. The algorithm starts with a partition where the blocks are the single points. Now, at every step we define a graph whose vertices are the blocks of the last partition and there is an edge joining a pair of blocks if and only if they minimize the distance, measured with the linkage function. Then, two blocks are merged if there is a path connecting them in the graph. Formally,

1. Let $\Theta_0 := \{x_1, ..., x_n\}$ and $R_0 = 0$.

2. For every $i \geq 1$, while $\Theta_{i-1} \neq \{X\}$, let $R_i := \min\{\ell(B, B') \mid B, B' \in \Theta_{i-1}, B \neq B'\}$. Then, let $G_{R_i}^\ell$ be a graph whose vertices are the
blocks of $\Theta_{i-1}$ and such that there is an edge joining $B$ and $B'$ if and only if $\ell(B, B') \leq R_i$.

3. Consider the equivalence relation $B \sim_{\ell, R} B'$ if and only if $B, B'$ are in the same connected component of $G_{R_i}^{\ell}$. Then, $\Theta_i = \frac{\Theta_{i-1}}{\sim_{\ell, R_i}}$.

4. Finally, let $\theta^\ell: [0, \infty) \to \mathcal{P}(X)$ be such that $\theta^\ell(r) := \Theta_i(r)$ with $i(r) := \max\{i \mid R_i \leq r\}$.

This formulation is equivalent to the previous one. Its advantage is that it allows us to consider extra properties to define the edges of the graph $G_{R_i}^{\ell}$, or other equivalence relations on the vertices different from the relation $B \sim_{\ell, R} B'$ defined in step 3 above. In particular, we may introduce further conditions to merge the blocks and, this way, reduce some undesired results as the chaining effect. See also Martínez-Pérez (2015). This approach will be used in Section 5 to define $SL^*(\alpha)$.

4. $\alpha$-unchaining Single Linkage Hierarchical Clustering: $SL(\alpha)$

The chaining effect is usually mentioned as one of the problems to solve in clustering. However, there are different approaches to define “chaining effects”.

Carlsson and Mémoli (2010b) refer to the chaining effect from Lance and Williams (1967) which is the one defined by Williams, Lambert and Lance (1966). This version of the “chaining effect” takes account of the tendency of a group to merge with single points or small groups rather than with other groups of comparable size. Thus, Williams, Lambert and Lance (1966) study and measure it by comparing the cardinality of the groups.

Here, we are focusing on another aspect. We want to deal with the tendency of two clusters to be merged when the minimal distance between them is small independently of their distribution. This can be a problem, for example, when the clusters are close and follow a multivariate normal distribution in $\mathbb{R}^n$ where each cluster has a point density which takes high values at the center and low values at the boundary. These are called “touching Gaussian clusters” in Zahn (1971). See Figure 2. This is, typically, the chaining effect one finds in $SLHC$.

Let us recall that a family $\Delta$ of non-empty finite subsets of a set $X$ is an abstract simplicial complex if, for every set $A$ in $\Delta$ and every non-empty subset $B \subset A$, $B$ also belongs to $\Delta$. Any set $A$ in $\Delta$ is called a simplex.

Given a finite metric space $(X, d)$, let $F_t(X, d)$ be the Vietoris-Rips complex of $(X, d)$. Let us recall that the Vietoris-Rips complex of a metric space $(X, d)$ is a simplicial complex whose vertices are the points of $X$ and $[v_0, \ldots, v_k]$ is a simplex of $F_t(X, d)$ if and only if $d(v_i, v_j) \leq t$ for every
Figure 2. The minimal distance between the blocks \( B_1 \) and \( B_2 \) is \( \varepsilon \). The clustering \( \{ B_1, B_2 \} \) will not appear (at any level) in the output of \( SLHC \).

Given any subset \( Y \subset X \), by \( F_t(Y) \) we refer to the subcomplex of \( F_t(X) \) defined by the vertices in \( Y \). A simplex \([v_0, ..., v_k]\) has dimension \( k \). The dimension of a simplicial complex (or any of its subsets) is the maximal dimension of its simplices.

Notice that densely packed points produce high-dimensional simplices in the Vietoris-Rips complex. We will consider this as a sign that the cluster is significant. Low dimensional simplices mean that there are few points located in close proximity to one or more neighbor points, so they might be interpreted as noise or poor measurements.

**Example 4.1** Let \((X, d)\) be the graph from Figure 3. The edges in \( N_1, N_2 \) have length 1 and the rest have length 3. The distances between vertices are measured as the minimal length of a path joining them.

Consider \( F_1(X, d) \). Then, the vertices of \( N_1 \) and \( N_2 \) define 3-dimensional simplices. For any vertex \( v \in X \setminus \{N_1 \cup N_2\} \), there is no vertex \( w \) with \( d(v, w) \leq 1 \). Therefore, they define 0-dimensional simplices and they are not part of any 1-dimensional simplex of \( F_1(X, d) \).

We define a modified single linkage hierarchical clustering method, \( SL(\alpha) \), on the basis of \( SL \) introducing a parameter \( \alpha \in \mathbb{N} \). This method allows us to take into account density without having to provide any additional input to the algorithm apart from \( \alpha \) and the distances between the points.

Let \((X, d)\) be a finite metric space with \( X = \{x_1, ..., x_n\} \).

Notice that in Section 3 the recursive definition of \( SL, CL \) and \( AL \) computed the minimal distances, \( R_i \), between the blocks from the previous step. For technical reasons, to define our method we use the ordered set of distances in the data set instead.
Let $d_{ij} := d(x_i, x_j)$ and $D := \{d_{ij} : 1 \leq i, j \leq n\} = \{t_i : 0 \leq i \leq m\}$ with $t_i < t_j \forall i < j$ where “<” denotes the order of the real numbers $t_i$. Clearly, $t_0 = 0$.

Let the dendrogram defined by $SL(\alpha)$, $\Sigma^{SL(\alpha)}_D(X, d) = \theta_{X, \alpha}$ or simply $\theta_\alpha$, be as follows:

1) Let $\theta_\alpha(0) := \{\{x_1\},...,\{x_n\}\}$ and $\theta_\alpha(t) := \theta_\alpha(0) \forall t < t_1$. Now, for every $i$, given $\theta_\alpha[t_{i-1}, t_i) = \theta_\alpha(t_{i-1}) = \{B_1,...,B_m\}$, we define recursively $\theta_\alpha$ on the interval $[t_i, t_{i+1})$ as follows:

2) Let us define a relation, $\sim_{t_i, \alpha}$, in $\theta_\alpha(t_{i-1})$ as follows. Suppose $B_j \sim_{t_i, \alpha} B_k$ if the following conditions hold:
   
   i) $\min\{d(x, y) \mid x \in B_j, y \in B_k\} \leq t_i$.
   
   ii) there is a simplex $\Delta \in F_{t_i}(B_j \cup B_k)$ such that $\Delta \cap B_j \neq \emptyset$, $\Delta \cap B_k \neq \emptyset$ and $\alpha \cdot \text{dim}(\Delta) \geq \min\{\text{dim}(F_{t_i}(B_j)), \text{dim}(F_{t_i}(B_k))\}$.

Then, $\sim_{t_i, \alpha}$ induces an equivalence relation where $B_j \sim_{t_i, \alpha} B_k$ if and only if there is a sequence $B_j = B_{i_0},...,B_{i_s} = B_k$ such that for every pair $B_{i_l-1}, B_{i_l}$ with $1 \leq l \leq s$, $B_{i_{l-1}} \sim_{t_i, \alpha} B_{i_l}$.

3) For every $t \in [t_i, t_{i+1})$, $\theta_\alpha(t) := \theta_\alpha(t_{i-1})/\sim_{t_i, \alpha}$.

In this algorithm we start with the partition defined by the single blocks. Then, at every step, blocks are merged if two conditions are simultaneously satisfied:
Condition \( i \) states that the distance between the blocks is minimal which is just the condition used in SL HC to define the graph in Remark 1.

Condition \( ii \) is used to avoid the chaining effect. The minimal dimension of the simplices generated by the Vietoris-Rips complex in both blocks must be less than \( \alpha \) times the maximal dimension of some simplex intersecting both blocks. Suppose we have two adjacent blocks which are close to each other as sets but contain dense cores, this is, where the points are closely packed in the interior. Suppose also that these cores are a certain distance apart as in Figure 2. Then, the dense cores will produce high dimensional simplices in the Vietoris-Rips complex while the connection between the blocks might be a low dimensional simplex. In this case, condition \( (ii) \) will not be satisfied and these blocks are not directly related in step 2.

Repeating the process, we eventually obtain a partition with a single block.

**Remark 2.** Notice that if two points \( x, x' \) belong to the same block of \( \theta_\alpha(t_i) \) then, necessarily, there exists a \( t_i \)-chain, \( x = x_0, x_1, \ldots, x_n = x' \) joining them. In particular, if \( x_j \in B_j \in \theta_\alpha(t_{i-1}), j = 0, \ldots, n \), the corresponding blocks \( \{B_{j-1}, B_j\}, j = 1, n \), satisfy condition \( ii \). This is immediate by construction.

**Example 4.2** Let \( (X, d) \) be the graph from Figure 3 and let \( \alpha = 1 \). Notice that there are eight 1-components, six of them are singletons and two of them, \( N_1, N_2 \), with \( \#(N_1) = \#(N_2) = 4 \). Furthermore, \( x_0 \in N_1 \), \( y_0 \in N_2 \) and \( \dim F_1(N_s) = 3 \) for \( s = 1, 2 \). Then, let us check that applying SL(1) on \( (X, d) \) the clustering \( \{B_1, B_2\} \) is detected.

Let \( \Sigma_D^{SL(1)}(X, d) = \theta_1. \)  \( \theta_1(t) = \{\{x_0\}, \ldots, \{x_6\}, \{y_0\}, \ldots, \{y_6\}\} \) if \( t < 1 \). If \( 1 \leq t < 3 \), \( \theta_1(t) = \{\{x_1\}, \{x_2\}, \{x_3\}, N_1, N_2, \{y_1\}, \{y_2\}, \{y_3\}\} \).

For \( t = 3 \), every pair of clusters in \( N_1, \{x_1\}, \{x_2\}, \{x_3\} \) is merged. Similarly, every pair of clusters \( N_2, \{y_1\}, \{y_2\}, \{y_3\} \) is merged. \( F_3(N_1) \) and \( F_3(N_2) \) have dimension 3 while the unique simplex in \( F_3(X) \) intersecting both \( N_1, N_2 \) has dimension 1. Therefore, condition \( ii \) is not satisfied and the blocks \( N_1, N_2 \) are not merged. Thus, \( \theta_1(3) = \{B_1, B_2\} \).

For \( t = 4 \), \( \dim(F_4(B_1)) = 6 = \dim(F_4(B_2)) \) while the maximal dimension of a simplex intersecting both clusters is 4. Then, by \( ii \), \( B_1, B_2 \) are not merged.

For \( t = 5 \), \( \dim(F_5(B_1)) = 6 = \dim(F_5(B_2)) \). Since the diameter of \( N_1 \cup N_2 \) is 5, these vertices define a simplex \( \Delta \) in \( F_5(X) \) such that \( \dim(\Delta) = 7 \). Clearly, this simplex intersects \( B_1 \) and \( B_2 \). Hence, \( \theta_1(5) = \{X\} \). Therefore, the dendrogram obtained, \( \theta_1 \), is the one on the left in Figure 4.
Modifying the parameter $\alpha$ we can adjust the method to be more or less sensitive to the chaining effect. Increasing $\alpha$ we would need higher dimensions in $F_3(N_s)$ to unchain blocks by condition ii). In particular, if $\alpha \geq 3$, $\theta_\alpha(t) = \{\{x_0\},...\{x_6\},\{y_0\},...,\{y_6\}\}$ if $t < 1$, $\theta_\alpha(t) = \{\{x_1\},\{x_2\},\{x_3\},N_1,N_2,\{y_1\},\{y_2\},\{y_3\}\}$ if $1 \leq t < 3$ and $\theta_\alpha(t) = \{X\}$ if $t \geq 3$ and the resulting dendrogram is the one on the right in Figure 4.

Example 4.3 Consider the set represented in Figure 5. Let us consider three distances, $t_{i1} < t_{i2} < t_{i3}$ which are represented, respectively, by a short segment, a dots line and a thick long segment. Let us assume that the sets $B_1, B_2, B_3$ are $(t_{i3})$-connected and that $\ell^{SL}(B_k,B_{k+1}) = t_{i3}$, $k = 1,2$. Also, we can see that there exist 3-dimensional simplices in $F_{t_{i3}}(X)$ inside $B_1, B_2$ and $B_3$. In Figure 6, we represent the corresponding dendrograms for $SL(1)$ and $SL$. (Notice that, in this case, $SL = SL(\alpha)$ for any $\alpha > 3$.)

It is clear that, $SL HC$ generates a dendrogram where it is impossible to detect the clustering $\{B_1, B_2, B_3\}$ because of the chaining effect. Introducing the parameter $\alpha = 1$, in this example, we obtain a hierarchical clustering which is consistent with the distribution of the sample.

Example 4.4 Let $X$ be the set from Figure 7. Suppose $\varepsilon = t_i$ and let us apply $SL(3)$ first until $t = t_{i-1}$.

In the second square of the figure we can see the 1-dimensional skeleton of the Vietorisi-Rips complex $F_{t_{i-1}}(X)$. As it is shown in the figure, there
Figure 5. For the set of points in the figure appears to be a natural clustering \{B_1, B_2, B_3\}.

Figure 6. Dendrograms \(\theta_1\) and \(\theta_\alpha\) with \(\alpha > 3\) for Example 5.
is no edge joining $B_1$ and $B_2$ yet and there are points inside $B_1, B_2$ producing high dimensional simplices in $F_{t_{i-1}}(X)$. Condition ii) has not applied yet and, therefore, the blocks are just the $t_{i-1}$-components. Let us call $N_1, N_2$ the corresponding blocks in $\theta_3$ defined by the nontrivial $t_{i-1}$ components.

Now $t^{SL}(N_1, N_2) = \varepsilon = t_i$ and there is a single edge joining $N_1$ and $N_2$ in $F_{t_i}(X)$ as we can see in the third square. However, the dimension of $F_{t_i}(N_s), s = 1, 2$ is greater that 3. Then, by condition ii), $N_1$ and $N_2$ are not merged and $\theta_{t_i}(X)$ refines the clustering $\{B_1, B_2\}$. In fact, the clustering from $\theta_3(t_i)$ is given by the connected components in the third square of the figure when the edge of length $\varepsilon$ joining $B_1, B_2$ is eliminated.

Let $t_k = \min\{t \mid B_1, B_2$ are $t$-connected$\}$. In the example, by condition ii), for every $t_i < t_j \leq t_k$ and for any pair of blocks $C_1, C_2 \in \theta_3(t_{j-1})$ with $C_1 \in B_1$ and $C_2 \in B_2$, $C_1, C_2$ are not merged in $\theta(t_j)$. Therefore, $\theta(t_k) = \{B_1, B_2\}$.

5. **Chaining Through Smaller Blocks: $SL^*(\alpha)$**

The method $SL(\alpha)$ is defined to prevent two adjacent blocks to be chained too soon when the minimal distance between them is small. Thus, condition ii) considers the dimension of the Vietoris-Rips complex restricted to both blocks. However, any cluster $B \in \theta_\alpha(t_{i-1})$ and any isolated point
A Density-Sensitive Hierarchical Clustering Method

\{z\} \in \theta(t_{i-1}) \text{ such that } d(B, z) \leq t_i \text{ are going to be merged in } \theta(t_i). \text{ Therefore, if two clusters are at a certain distance } \varepsilon \text{ from a single point then both blocks will be merged with this point in } \theta(\varepsilon). \text{ Consequently, those clusters will be merged together in } \theta(\varepsilon).

**Example 5.1** Consider the graph represented in Figure 8.

Suppose that the edges in \(N_1, N_2\) have length 1, the edges \(\{x_0, z_0\}\) and \(\{z_0, y_0\}\) have length 2 and the rest have length 3. As we can see in the figure, there are nine 1-components, two of them, \(N_1\) and \(N_2\), have four points and the rest are singletons. The whole space is 3-connected.

Let us fix any \(\alpha \geq 1\). It is trivial to check that, \(\theta(1) = \{x_1, x_2, x_3, N_1, z_0, N_2, y_1, y_2, y_3\}\). Now, for \(t = 2\), since \(\{z_0\}\) is a single point and \(\dim(F_2(\{z_0\})) = 0\), condition ii) is trivially satisfied. Therefore, \(N_1 \sim_{2, \alpha} z_0 \text{ and } z_0 \sim_{2, \alpha} N_1\). Hence, \(\theta(2) = \{x_1, x_2, x_3, \{N_1 \cup \{z_0\} \cup N_2\}, y_1, y_2, y_3\}\).

Similarly, since every block in \(\theta(2)\) except from \(\{N_1 \cup z_0 \cup N_2\}\) is a single point, condition ii) always holds and \(\theta(3) = \{X\}\).

In general, for any pair of clusters \(B_1, B_2 \in \theta(t_{i-1})\) such that \(d(B_1, B_2) \leq t_i\) and \(\#(B_2) \leq \alpha\) the dimension of the Vietoris-Rips complex restricted to \(B_2\) is at most \(\alpha - 1\). Therefore, condition ii) will not apply and the clusters will be merged. Thus, for any chain of clusters \(B_0, ..., B_n\) such that \(\ell_{SL}(B_{i-1}, B_i) = \varepsilon\) for every \(i = 1, ..., n\) and \(\#(B_j) \leq \alpha\) for every \(j = 1, ..., n - 1\), \(B_0, ..., B_n\) are merged together in \(\theta(\varepsilon)\). We call chaining...
through smaller blocks this chaining effect that consists of a tendency to
merge two blocks with a high number of points when they are chained by a
sequence of blocks with (comparatively) few points.

To treat this effect we are going to focus on the “big” blocks. The
selection is done depending on the parameter $\alpha$ (which defines the sensi-
tivity of the whole method to chaining) and on the cardinality of the blocks
involved. We use eqn. (1) below to fix the distinction between big blocks
and small blocks.

Let $X = \{x_1, \ldots, x_n\}$. Let $D = \{t_i : 0 \leq i \leq n\}$ be the ordered set
of distances between points of $X$ with $0 = t_0 < t_1 < \cdots < t_n$.

Let the dendrogram defined by $SL^*(\alpha)$, $\Sigma_D^{SL^*(\alpha)}(X,d) = \theta^\ast_{X,\alpha}$ or
simply $\theta^\ast_{\alpha}$, be as follows:

1) Let $\theta^\ast_{\alpha}(0) := \{\{x_1\}, \ldots, \{x_n\}\}$ and $\theta^\ast_{\alpha}(t) := \theta^\ast_{\alpha}(0)$ \ \forall t < t_1.

Now, given $\theta^\ast_{\alpha}[t_{i-1},t_i) = \theta^\ast_{\alpha}(t_{i-1}) = \{B_1, \ldots, B_m\}$, we define recurs-
vively $\theta^\ast_{\alpha}$ on the interval $[t_i,t_{i+1})$ as follows:

2) Let $G^t_{\alpha}$ be a graph with vertices $V(G^t_{\alpha}) := \{B_1, \ldots, B_m\}$ and edges $E(G^t_{\alpha}) := \{B_j, B_k\}$ such that the following conditions hold:

   i) $\min\{d(x,y) \mid x \in B_j, y \in B_k\} \leq t_i$.
   
   ii) there is a simplex $\Delta \in F_{t_i}(B_j \cup B_k)$ such that $\Delta \cap B_j \neq \emptyset$, $\Delta \cap B_k \neq \emptyset$ and $\alpha \cdot \text{dim}(\Delta) \geq \min\{\text{dim}(F_{t_i}(B_j)), \text{dim}(F_{t_i}(B_k))\}$.

For notational convenience, we may write $B$ to refer both to the block
of $\theta(t_{i-1})$ and to the vertex of $G^t_{\alpha}$.

3) Let us define a relation, $\sim_{t_i,\alpha}$ between the blocks as follows.

Let $cc(G^t_{\alpha})$ be the set of connected components of the graph $G^t_{\alpha}$. Let
$A \in cc(G^t_{\alpha})$ with $A = \{B_{j_1}, \ldots, B_{j_r}\}$.

Let us call big blocks of $A$ those blocks such that

$$\alpha \cdot \#(B_{j_k}) \geq \max_{1 \leq l \leq r}\{\#(B_{j_l})\}. \tag{1}$$

The rest of blocks of $A$ are called small blocks.

Let $H_{\alpha}(A)$ be the subgraph of $A$ whose vertices are the big blocks
and $S_{\alpha}(A)$ be the subgraph of $A$ whose vertices are the small blocks.

Then, $B_{j_k} \sim_{t_i,\alpha} B_{j_{k'}}$ if one of the following conditions holds:

iii) $\exists C \in cc(H_{\alpha}(A))$ such that $B_{j_k}, B_{j_{k'}} \in C$.

iv) $B_{j_k} \in C \in cc(H_{\alpha}(A))$, $B_{j_{k'}} \in C' \in cc(S_{\alpha}(A))$ and there is no
big block in $A\backslash C$ adjacent to any block in $C'$. 
Figure 9. Graph $G_{\alpha}^{t_i}$ with three connected components of big blocks, $C_i$, and six small blocks $B_j$.

Then, $\sim_{t_i,\alpha}$ induces an equivalence relation whose classes are contained in the connected components of $G_{\alpha}^{t_i}$.

4) For every $t \in [t_i, t_{i+1})$, $\theta_{\alpha}^*(t) := \theta_{\alpha}^*(t_{i-1})/\sim_{t_i,\alpha}$.

Notice that if after steps 1) and 2), we merge the clusters corresponding to the connected components of the graph $G_{\alpha}^{t_i}$, we obtain $SL(\alpha)$.

By iii), if two big blocks, $B, B'$, are joined by an edge in $G_{\alpha}^{t_i}$, then $B \sim_{\alpha,t_i} B'$. Thus, the connected components of big blocks are merged.

By iv), a connected component of small blocks $C'$ is merged with a component of big blocks $C$ if $C$ is the unique component of big blocks adjacent to $C'$. Otherwise, the blocks of $C'$ stay as separated blocks in $\theta_{\alpha}^*(t_i)$. This can be seen also as follows. By iv), if a small block is connected by chains of small blocks to two different components of big blocks we will consider it as a block apart in $\theta(t_i)$. See Example 5.2.

**Example 5.2** Suppose $A \in cc(G_{\alpha}^{t_i})$ is as represented in Figure 9: $H_{\alpha}(A)$ has three connected components, $C_i$, $i = 1, 3$ and $A \setminus H_{\alpha}(A)$ consists of six small blocks $B_j$, $j = 1, 6$. The components $C_j$ are merged by iii). The edges in the figure represent the resulting edges from $G_{\alpha}^{t_i}$ after identifying the components $C_i$ by iii).

Now, the component of small blocks formed by $B_1, B_2$ is only adjacent to $C_1$. Therefore, by iv), $C_1 \cup B_1 \cup B_2$ is contained in some block of $\theta(t_i)$. The same happens with $B_5$ which is a component in $S_{\alpha}(A)$ which is only adjacent to the component $C_2$. Thus, $C_2 \cup B_5$ is contained in some block of $\theta(t_i)$. However, the component of small blocks given by $B_3, B_4$ is adjacent
to two different components of big blocks, $C_1$ and $C_2$. Therefore, $B_3, B_4$ are independent blocks in $\theta_\alpha(t_i)$. The same happens with $B_6$. Thus, $\theta_\alpha(t_i) = \{\{C_1 \cup B_1 \cup B_2\}, \{C_2 \cup B_5\}, \{C_3\}, B_3, B_4, B_6\}$.

**Remark 3.** At step $iii)$, if $H_\alpha(A)$ is connected, then $B_{i_1} \cup \cdots \cup B_{i_r}$ defines a block of $\theta_\alpha(t_i)$.

**Remark 4.** Notice that Remark 2 still applies. In fact, if two points $x, x'$ belong to the same block of $\theta_\alpha^*(t_i)$ then, necessarily, there exists a $t_i$-chain, $x = x_0, x_1, ..., x_n = x'$ joining them so that if $x_j \in B_j \in \theta_\alpha^*(t_{i-1})$ for every $0 \leq j \leq n$, the corresponding edges $\{B_{j-1}, B_j\}, 1 \leq j \leq n$, satisfy condition $ii)$.

**Example 5.3** Let $(X, d)$ be the graph from Figure 8 and let $\alpha = 1$. Then, let us check that applying SL(1) on $(X, d)$ the dendrogram generated is the one from Figure 10. Clearly, $\theta_1(t) = \{\{x_0\}, ..., \{x_6\}, \{y_0\}, ..., \{y_6\}\}$ if $t < 1$. If $1 \leq t < 2$, $\theta_1(t) = \{\{x_1\}, \{x_2\}, \{x_3\}, N_1, \{z_0\}, N_2, \{y_1\}, \{y_2\}, \{y_3\}\}$. There are nine 1-components, seven of them are singletons and two of them, $N_1, N_2$, with $\#(N_1) = \#(N_2) = 4$. Furthermore, $x_0 \in N_1, y_0 \in N_2$ and $\dim F_1(N_s) = 3$ for $s = 1, 2$.

For $t = 2$, conditions $i)$ and $ii)$ induce edges in $G_2^1$ between $N_1$ and $\{z_0\}$ and between $\{z_0\}$ and $N_2$. Then, $G_2^1$ has one component, $A$, which is not a single point: $A = \{N_1, \{z_0\}, N_2\}$. $\{z_0\}$ is a single point and $\#N_1 = \#N_2 = 4$. Then, there are two big blocks in $A$, $N_1$ and $N_2$, and one small block, $\{z_0\}$. Since the small block is connected to both big blocks, by condition $iv)$, these blocks are not merged. Thus, for every $1 \leq t < 3$, $\theta_1(t) = \{\{x_1\}, \{x_2\}, \{x_3\}, N_1, \{z_0\}, N_2, \{y_1\}, \{y_2\}, \{y_3\}\}$.
For $t = 3$, there are edges in $G_1^3$ between $N_1$ and $x_i$ for every $1 \leq i \leq 3$ and between $N_2$ and $y_i$ for every $1 \leq i \leq 3$. Thus, $G_1^3$ is connected. Now, there are two big blocks, $N_1$, $N_2$ and 7 small blocks. By conditions $iii)$ and $iv)$, it is readily seen that $\theta_1(3) = \{B_1, \{z_0\}, B_2\}$.

$\theta_1(t) = \{B_1, \{z_0\}, B_2\}$ for every $t < 6$ since the minimal distance $t$ such that $B_1, B_2$ are connected by an edge in $G_1^t$ is $t = 6$. Finally, $\theta_1(t) = \{X\}$ for every $t \geq 6$.

6. Unchaining Properties of $SL(\alpha)$

In this section we try to give some theoretical background to the treatment of the chaining effect.

**Definition 6.1** Let $X$ be a finite metric space. We say that two $b$-connected subsets of $X$, $B_1, B_2$, are $(a,b)$-**chained** if they satisfy that

i) $\min\{ t \mid B_1 \text{ is } t\text{-connected} \} = b$,

ii) there exist $x_0 \in B_1$ and $y_0 \in B_2$ such that $d(x_0, y_0) = a \leq b$.

If the parameters $a, b$ are not relevant, we say simply that $B_1, B_2$ are **chained**.

Notice that not every case of chained subsets is going to induce an “undesired” chaining effect. See Figure 11. The idea here is that this definition of chained subsets includes all the situations where we are considering an undesired chaining effect.

Inside the cases of chained subsets let us consider some specific type which are the subsets chained by a single edge. We define this as the prototypical case of chained subsets on which the algorithm should be tested to check its unchaining properties.

**Definition 6.2** Let $X$ be a finite metric space. We say that two $b$-connected subsets of $X$, $B_1, B_2$, are $(a,b)$-**chained by a single edge** if it holds that

i) $\min\{ t \mid B_1 \text{ is } t\text{-connected} \} = b$,

ii) there exist $x_0 \in B_1$ and $y_0 \in B_2$ such that $d(x_0, y_0) = a \leq b$

iii) $\forall (x_0, y_0) \neq (x, y) \in B_1 \times B_2, d(x, y) > b$.

If the parameters $a, b$ are not relevant, we say simply that $B_1, B_2$ are **chained by a single edge**.

Suppose that $X = B_1 \cup B_2$ and $B_1, B_2$ are $(a,b)$-chained by a single edge. Notice that given the Vietoris-Rips complex $F_b(X)$ and the edge $e := \{x_0, y_0\} \in F_b(X)$, then $F_b(X) \setminus \{e\}$ has exactly two connected components:
Figure 11. $B_1$ and $B_2$ are $(a, b)$-chained subsets although we consider that there is no undesired chaining effect if these blocks are merged.

$B_1$ and $B_2$. Notice that this partition is not trivially detected by the minimal spanning trees approach since the minimal distance between the clusters is not greater than the maximal distance of the minimal spanning tree of $B_1$. See Theorem 3 in Zahn (1971).

**Example 6.3** The subsets $B_1, B_2$ from Figure 2 are $(\varepsilon, t_i)$-chained by a single edge, $\{x_0, y_0\}$.

$B_1, B_2$ are $t_i$-connected. There is a pair of points $x_0 \in B_1, y_0 \in B_2$ with $d(x_0, y_0) = \varepsilon < t_i$ and for any pair of points $x \in B_1, y \in B_2$, if $(x, y) \neq (x_0, y_0)$, then $d(x, y) > t_i$.

**Example 6.4** Consider the graph represented in Figure 3 where the edges in $N_1, N_2$ have length 1 and the rest have length 3. The distance between the vertices is measured as the length of the minimal path joining them. The whole space is 3-connected with $d(x_1, x_2) = d(y_1, y_2) = 3 > 1$.

Thus, $B_1$ and $B_2$ are $(3, 3)$-chained by a single edge.

**Definition 6.5** Let $\mathcal{X}$ be a HC method and $\mathcal{X}_{\mathcal{D}}(X) = \theta$. We say that $\mathcal{X}$ is strongly chaining if for any set $X$, any pair of $(a, b)$-chained subsets $B_1, B_2$ of $X$ with $x_0 \in B_1, y_0 \in B_2$ and $d(x_0, y_0) = a \leq b$ and any $t > 0$, if $B_1$ is contained in some block $B$ of $\theta(t)$, then $y_0 \in B$.

It is immediate to check that $SL HC$ is strongly chaining. In general, we can prove the following:

**Theorem 6.6** Let $\mathcal{X}$ be a hierarchical clustering method. If for every metric space $X$ and every $x, y, z, t \in X$, $u_{SL}(x, y) \leq u_{SL}(z, t)$ implies that $u(x, y) \leq u(z, t)$, then $\mathcal{X}$ is strongly chaining. In particular, SL HC is strongly chaining.
Proof. First, let us see that $\mathcal{S}$ is strongly chaining. Consider two $(a, b)$-chained subsets $B_1, B_2$. By hypothesis, there exist $x_0 \in B_1, y_0 \in B_2$ such that $u_{SL}(x_0, y_0) = a$. Also, there exist $x_1, x_2 \in B_1$ with $u_{SL}(x_1, x_2) = b \geq a$. Thus, $u(x_1, x_2) \geq u(x_0, y_0)$.

If $B_1$ is contained in some block $B$ of $\theta(t)$, then $t \geq u(x_1, x_2) \geq u(x_0, y_0)$ and $y_0 \in B$.

$\square$

$AL$ and $CL$ $HC$ are not strongly chaining:

Example 6.7 Consider the graph from Figure 3. Suppose that, in addition, we include edges of length 3 from $x_1, x_2, x_3$ to every vertex in $N_1$ and from $y_1, y_2, y_3$ to every vertex in $N_2$. Also, suppose that $d(x_0, y_0) = 2.5$.

Thus, every pair of points in $N_1$ (resp. $N_2$) are at distance 1, $d(x_1, x_j) = 3$ (resp. $d(y_1, y_j) = 3$) for every $i \neq j$, $0 \leq i, j \leq 3$, $d(x_i, x') = 3$ for every $x' \in N_1$ and every $1 \leq i \leq 3$, $d(y_i, y_j) = 3$ for every $i \neq j$, $0 \leq i, j \leq 3$, $d(y_i, y') = 3$ for every $y' \in N_2$ and every $1 \leq j \leq 3$.

d($x_0, y_0) = 2.5$ and $d(x, y) > 3$ for every $(x_0, y_0) \neq (x, y) \in B_1 \times B_2$.

Then, $B_1$ and $B_2$ are $(2.5, 3)$-chained subsets. However, $\theta_{AL}(1) = \theta_{CL}(1) = \{\{x_1\}, \{x_2\}, \{x_3\}, N_1, N_2, \{y_1\}, \{y_2\}, \{y_3\}\}$ and $\theta_{AL}(3) = \theta_{CL}(3) = \{B_1, B_2\}$.

Now, we define weakly unchaining for methods that are able to detect at least certain blocks chained by a single edge. We prove that $SL(\alpha)$ is weakly unchaining and check that $CL$ and $AL$ $HC$ are not weakly unchaining although, as seen in Example 6.7, they are not strongly chaining either.

Definition 6.8 Let $\mathcal{S}$ be a HC method and $\mathcal{S}_D(X) = \theta$. We say that $\mathcal{S}$ is weakly unchaining for the parameter $\alpha$ if the following implication holds:

Let $X$ be a finite metric space such that $X = B_1 \cup B_2$, with $B_1, B_2$ a pair of subsets $(a, b)$-chained by a single edge $\{x_0, y_0\}$. Suppose there exist $N_1 \in B_1, N_2 \in B_2$ such that

- $N_s$ is contained in some block $B_s$ of $\theta(t)$ for $s = 1, 2$ with $t < a$,
- $\dim F_a(N_s) > \alpha$, $s = 1, 2$,
- $x_0 \in N_1, y_0 \in N_2$,
- $\sup_{x, x' \in B_1} \{d(x, x')\} \leq b$ and $\sup_{y, y' \in B_2} \{d(y, y')\} \leq b$.

Then, there exists $t' > 0$ such that $\theta(t') = \{B_1, B_2\}$.

We say that $\mathcal{S}$ is weakly unchaining if it is weakly unchaining for some parameter $\alpha$. 
Remark 5. Notice that in the definition above we consider two chained subsets with further conditions. Therefore, if a HC method is strongly chaining, in particular, it is not weakly unchaining.

Theorem 6.9 Let $X$ be a finite metric space such that $X = B_1 \cup B_2$, with $B_1, B_2$ a pair subsets of $(t_j, t_i)$-chained by a single edge $\{x_0, y_0\}$. Suppose there exist $N_1 \in B_1, N_2 \in B_2$ such that

- $N_s$ is contained in some block $B_s^{j-1}$ of $\theta(t_{j-1})$, $s = 1, 2$,
- $\dim F_{t_i}(N_s) > \alpha$, $s = 1, 2$,
- $x_0 \in N_1, y_0 \in N_2$.

Then, $\theta_{\alpha}(t_i)$ refines $\{B_1, B_2\}$. If, in addition, $\sup_{x,x' \in B_i}\{d(x, x')\} \leq t_i$ and $\sup_{y,y' \in B_2}\{d(y, y')\} \leq t_i$, then $\theta_{\alpha}(t_i) = \{B_1, B_2\}$.

Proof. Let us recall that, by definition, $t_{j-1} < t_j \leq t_i$.

For the first part it suffices to check that for every pair $(x, y) \in B_1 \times B_2$, $\{x, y\}$ is not contained in any block of $\theta(t_i)$, this is, $u_{\alpha}(x, y) \geq t_i$.

Let $(x, y) \in B_1 \times B_2$. First, notice that for any $t < t_j$, there is no $t$-chain joining $x$ to $y$. Thus, $u_{\alpha}(x, y) \geq t_j$. Let us check that $u_{\alpha}(x, y) > t_{j+k}$, $k = 0, i - j$.

For $k = 0$, since $x_0 \in N_1 \subset B_1^{j-1}$ and $y_0 \in N_2 \subset B_2^{j-1}$, condition ii) implies that there is no edge in $G_{\alpha}^j$ between $B_1^{j-1}$ and $B_2^{j-1}$. Since $d(x_1, y_1) > t_i$ for every $(x_0, y_0) \neq (x_1, y_1) \in B_1 \times B_2$ there is no $t_i$-chain joining $x$ to $y$ which does not contain the edge $\{x_0, y_0\}$. In particular, there is no $t_j$-chain joining $x$ to $y$ which does not contain the edge $\{x_0, y_0\}$. Therefore, by Remark 2, it follows that $u_{\alpha}(x, y) > t_j$.

The same argument works for every $0 < k \leq i - j$. Thus, $u_{\alpha}(x, y) > t_i$ and $\theta_{\alpha}(t_i)$ refines $\{B_1, B_2\}$.

Suppose, in addition, that $\sup_{x,x' \in B_i}\{d(x, x')\} \leq t_i$ and $\sup_{y,y' \in B_2}\{d(y, y')\} \leq t_i$. We already proved that $\theta_{\alpha}(t_i)$ refines $\{B_1, B_2\}$. Clearly, since $\sup_{x,x' \in B_i}\{d(x, x')\} \leq t_i$ (respectively, for $B_2$), all the blocks contained in $B_1$ (resp. $B_2$) are joined by an edge in $G_{\alpha}^{t_i}$. Therefore, $B_1$ (resp. $B_2$) is a block of $\theta_{\alpha}(t_i)$.

\[ \square \]

Corollary 6.10 $SL(\alpha)$ is weakly unchaining for the parameter $\alpha$.

See Example 4.2.

Corollary 6.11 $SL^*(\alpha)$ is weakly unchaining for the parameter $\alpha$.

Remark 6. AL and CL HC are not weakly unchaining.

Consider the graph in Figure 3. To check that CL HC is not weakly unchaining suppose that we add some edges between $N_1$ and $N_2$ so that $\forall (x_0, y_0) \neq (x, y) \in N_1 \times N_2, d(x, y) = 4$. 
Notice that this graph satisfies the conditions in the definition of weakly unchaining. Then, it suffices to check that $\theta_{CL}(t)$ is never $\{B_1, B_2\}$.

It is immediate to check that $\ell_{CL}(N_1, N_2) = 4$. Then, it is readily seen that $\theta_{CL}(t) = \{x_1, x_2, x_3, N_1, N_2, y_1, y_2, y_3\}$ for every $1 \leq t < 4$ and $\theta_{CL}(4) = \{X\}$.

To check that $AL HC$ is not weakly unchaining suppose that in Figure 3, we made $d(x_0, y_0) = 3 - \frac{3}{4}$. Then, let us see that $\theta_{AL}(t)$ is never $\{B_1, B_2\}$.

First, notice that this graph satisfies the conditions in the definition of weakly unchaining. Also, it is immediate to check that $\ell_{AL}(N_1, N_2) = \frac{3}{4} + 3 = \ell_{AL}(x_i, N_1) = \ell_{AL}(y_j, N_2), i, j = 1, 3$. Thus, it is readily seen that $\theta_{AL}(t) = \{x_1, x_2, x_3, N_1, N_2, y_1, y_2, y_3\}$ for every $1 \leq t < 3 + \frac{3}{4}$ and $\theta_{AL}(3 + \frac{3}{4}) = \{X\}$.

DBSCAN is a density-based algorithm for clustering. See Ester et al. (1996). Although this is not a hierarchical clustering method it is worth comparing its results with $SL(\alpha)$. DBSCAN requires two parameters: some distance $\varepsilon > 0$ and a minimal number of points $minPts$. A point is a core point of a cluster if there are at least $minPts$ in its $\varepsilon$-neighborhood. Then, the density-reachable points from a core point define a cluster. Let us recall here the formal definition.

A $\varepsilon$-neighbourhood of a point $p$, $N_\varepsilon(p) = \{q \in X \mid d(p, q) \leq \varepsilon\}$.

A minimum number of points, $minPts$ is defined so that if $N_\varepsilon(p)$ has at least $minPts$, then $p$ is a core point.

A point $p$ is directly density-reachable from a point $q$ with respect to $\varepsilon$, $minPts$ if

1) $p \in N_\varepsilon(q)$ and
2) $|N_\varepsilon(q)| \geq minPts$

A point $p$ is density-reachable from a point $q$ with respect to $\varepsilon$, $minPts$ if there is a chain of points $q = p_1, ..., p_n = p$ such that $p_{i+1}$ is directly density-reachable from $p_i$.

A point is density-connected to a point $q$ with respect to $\varepsilon$, $minPts$ if there is a point $o$ such that both, $p$ and $q$ are density-reachable from $o$ with respect to $\varepsilon$, $minPts$.

A cluster is defined to be a set of density connected points which is maximal with respect to density-reachability. The points which do not belong to any cluster are considered noise.

One of the advantages of DBSCAN is that it is capable of detecting isolated points and eliminating them as noise. Also, the chaining effect through a chain of points is reduced. In fact, the type of unchaining DBSCAN does is more related to the chaining through smaller blocks studied.
below and it is not so effective to detect the chaining effect produced between two blocks when the minimal distance between them is small.

In general, suppose \( X = B_1 \cup B_2 \) with \( B_1, B_2 \) two clusters \((a,b)\)-chained by a single edge \( \{x_0, y_0\} \). Let us assume that \( \varepsilon = a \) and that \( x_0, y_0 \) are core points. Then, \( x_0, y_0 \) are density connected and they belong to the same cluster in the output of DBSCAN. Therefore, the clustering \( \{B_1, B_2\} \) is not detected by DBSCAN. See Example 6.12. However, if \( x_0, y_0 \) belong to simplices with dimension at least \( \alpha \) in \( F_t(B_1), F_t(B_2) \) respectively for some \( t < a \), then \( SL(\alpha) \) detects this clustering. In fact: \( \theta_\alpha(b) = \{B_1, B_2\} \).

Example 6.12 Let us analyze the case of Example 6.4. Let \( \varepsilon = 3 \) and \( \text{minPts} = 4 \). Then, notice that \( x_0, y_0 \) are core points. Therefore, since \( d(x_0, y_0) = 3 \), they are density-connected. Thus, DBSCAN does not detect the clustering \( \{B_1, B_2\} \).

Minimal spanning trees, see Zahn (1971), also offer a useful and efficient approach to detect touching clusters. However, it may be noticed that choosing the spanning trees one looses more information about the data set than in our method. For example, given an \( n \)-simplex, the minimal spanning tree can be any tree with \( n + 1 \) vertices. Hence, examples can be easily defined where this method creates undesired “necks”.

7. Unchaining Properties of \( SL^*(\alpha) \)

Definition 7.1 Let \( X \) be a finite metric space, \( B_0, ..., B_k \) be \( b \)-connected subsets of \( X \) and \( a \leq b \). We say that \( B_0 \) and \( B_k \), are \((a,b)\)-chained through \( \alpha \)-smaller blocks if the following conditions hold

i) \( \min\{t \mid B_0 \text{ is } t\text{-connected} \} = b \),

ii) there exists an \( a \)-chain \( x_0, ..., x_k \) with \( x_s \in B_s \) for every \( 0 \leq s \leq k \)

iii) \( \forall (x,y) \in B_0 \times B_k, d(x,y) > b \).

iv) \( \alpha \cdot \#(B_s) < \min\{\#(B_1), \#(B_k)\} \) for every \( 1 \leq s \leq k - 1 \).

If the parameters \( a, b, \alpha \) are not relevant, we simply say that \( B_1, B_2 \) are chained through smaller blocks.

Definition 7.2 Let \( \Xi \) be a strongly chaining HC method and \( \Xi_D(X) = \emptyset \). We say that \( \Xi \) is completely chaining if for any set \( X \), any pair of components \( B_0, B_k \) of \( X \) chained through smaller blocks and any \( t > 0 \), if \( B_0 \) is contained in some block \( B \) of \( \theta(t) \), then \( \{x_0, ..., x_k\} \subseteq B \).

Theorem 7.3 Let \( \Xi \) be a hierarchical clustering method. If for every metric space \( X \) and every \( x, y, z, t \in X \), \( u_{SL}(x,y) \leq u_{SL}(z,t) \) implies that \( u(x,y) \leq u(z,t) \), then \( \Xi \) is completely chaining. In particular, \( SL \) HC is completely chaining.
Proof. By Theorem 6.6, we already know that \( \mathcal{T} \) is strongly chaining.

Let \( B_0, B_k \) two \( b \)-connected subsets \((a, b)\)-chained through smaller blocks. Let \( x_0, ..., x_k \) be the corresponding chain. Then, \( u_{SL}(x_r, x_s) \leq a \) for every \( 1 \leq r, s \leq k \) and there exist \( x, x' \in B_0 \) such that \( u_{SL}(x, x') = b \geq a \). Thus, \( u(x, x') \geq u(x_r, x_s) \) for every \( 1 \leq r, s \leq k \).

Now, suppose \( t > 0 \) such that \( B_0 \) is contained in some block \( B \) of \( \theta(t) \). Then, \( t \geq u(x, x') \geq u(x_r, x_s) \) for every \( 1 \leq r, s \leq k \) and \( \{x_0, ..., x_k\} \in B \).

\[ \mathcal{T} \]

**Definition 7.4** \( \mathcal{T} \) is **\( \alpha \)-bridge-unchaining** if it is weakly unchaining for the parameter \( \alpha \) and the following implication holds:

Let \( X \) be a finite metric space, \( \mathcal{T}_D(X) = \theta \) and let

\[ \theta(t) = \{B_1, B_2, \{z_0\}, ..., \{z_k\}, \{x_1\}, ..., \{x_n\}, \{y_1\}, ..., \{y_m\}\} \]

for some \( t < t_i \) with \( z_j, x_r, y_s \) single points for every \( j, r, s \). Suppose that

a) \( d(z_{j-1}, z_j) = t_i \) for every \( 1 \leq j \leq k \),

b) \( d(z_j, z_{j+1}) > t_i \) for every \( |j_1 - j_2| > 1 \),

c) \( \ell_{SL}(x_r, B_1) \leq t_i \) for every \( r \)

d) \( \ell_{SL}(y_s, B_2) \leq t_i \) for every \( s \)

e) \( \ell_{SL}(z_0, B_1) = t_i \) and \( \ell_{SL}(z_k, B_2) = t_i \)

f) \( \min_{1 \leq j \leq k} \{\ell_{SL}(z_j, B_1)\} > t_i \) and \( \min_{0 \leq j \leq k-1} \{\ell_{SL}(z_j, B_2)\} > t_i \)

g) \( \min_{r,j,s} \{d(x_r, z_j), d(z_j, y_s), \ell_{SL}(x_r, B_2), \ell_{SL}(y_s, B_1), \ell_{SL}(B_1, B_2)\} > t_i \)

h) \( \alpha < \min\{#(B_1), #(B_2)\} \), \( \alpha \cdot #(B_1) > #(B_2) \) and \( \alpha \cdot #(B_2) > #(B_1) \).

Then, there exists \( t > 0 \) such that

\[ \theta(t) = \{\{B_1 \cup x_1 \cup \cdots \cup x_n\}, z_0, ..., z_k, \{B_2 \cup y_1 \cup \cdots \cup y_m\}\} \]

\( \mathcal{T} \) is **bridge-unchaining** if it is \( \alpha \)-bridge-unchaining for some parameter \( \alpha \).

**Remark 7.** Notice that in the conditions above, if \( \min\{t \mid B_1 \text{ is } t\text{-connected}\} = t_i \), then \( B_1 \) and \( B_2 \) are \((t_i, t_i)\)-chained through the \( \alpha \)-smaller blocks \( z_0, ..., z_k \).

**Theorem 7.5** Let \( X \) be a finite metric space and let

\[ \theta_\alpha(t_{j-1}) = \{B_0, B_1, ..., B_{k-1}, B_k, B_1', ..., B_n', B_1'', ..., B_m''\} \]

with \( t_j \leq t_i < 2t_j \). Suppose that
\[\ell^{SL}(B_{\ell-1}, B_{\ell}) = t_j \text{ for every } 1 \leq \ell \leq k,\]
\[\ell^{SL}(B_{1 \ell}, B_{\ell_2}) > t_i \text{ for every } |\ell_1 - \ell_2| > 1,\]
\[\ell^{SL}(B'_r, B_0) \leq t_i \text{ for every } r\]
\[\ell^{SL}(B''_s, B_k) \leq t_i \text{ for every } s\]
\[\ell^{SL}(B'_r, B_{\ell}) > t_i \text{ for every } r \text{ and every } 1 \leq \ell \leq k\]
\[\ell^{SL}(B'_r, B_{\ell}) > t_i \text{ for every } s \text{ and every } 0 \leq \ell \leq k - 1\]
\[\alpha \max_{1 \leq \ell \leq k-1} \{\#(B_{\ell})\} < \max\{\#(B_0), \#(B_k)\}, \alpha \cdot \#(B_0) > \#(B_k) \text{ and } \alpha \cdot \#(B_k) > \#(B_0)\]
\[\alpha > \dim(F_{t_i}(B_{\ell})) \text{ for every } 1 \leq \ell \leq k - 1, \alpha > \dim(F_{t_i}(B'_r)), \alpha > \dim(F_{t_i}(B''_s)) \text{ for every } r, s.\]

Then,
\[\theta^*_\alpha(t_i) = \{B_0 \cup B'_1 \cup \cdots \cup B'_n, B_1, \ldots, B_{k-1}, \{B_k \cup B''_1 \cup \cdots \cup B''_m\}\}.\]

**Proof.** Let
\[\theta^*_\alpha(t_{j-1}) = \{B_0, B_1, \ldots, B_{k-1}, B_k, B'_1, \ldots, B'_r, B''_1, \ldots, B''_m\}\]
satisfying the conditions above. For \(t = t_j\) let us apply conditions \(i)\) and \(ii)\) of \(SL(\alpha)\). Since, \(\alpha > \dim(F_{t_i}(B_{\ell})) > \dim(F_{t_j}(B_{\ell}))\) for every \(1 \leq \ell \leq k - 1\), we obtain edges \(B_{\ell-1}, B_{\ell}\) for every \(1 \leq \ell \leq k\). Since \(\alpha > \dim(F_{t_i}(B'_r))\) and \(\alpha > \dim(F_{t_i}(B''_s))\) for every \(r, s\), we also obtain edges \(B'_r, B_0\) and \(B''_s, B_k\) for every \(r, s\) such that the distance is less or equal than \(t_j\). Thus, the blocks \(B_\ell, 0 \leq \ell \leq k\) are in the same connected component of \(G^{t_j}_\alpha\). Since \(\alpha \max_{1 \leq \ell \leq k-1} \{\#(B_{\ell})\} < \max\{\#(B_0), \#(B_k)\}\), by \(iiv)\), \(B_{\ell}\) is an independent block in \(\theta^*_\alpha(t_j)\) for every \(1 \leq \ell \leq k - 1\). Also, by \(iiv)\), the blocks \(B'_s\) joined by an edge to \(B_0\) (resp. the blocks \(B''_r\) joined by an edge to \(B_k\)) are merged with \(B_0\) (resp. \(B_k\)).

The same argument holds for every \(t_j < t \leq t_i\). Thus,
\[\theta^*_\alpha(t_i) = \{B_0 \cup B'_1 \cup \cdots \cup B'_n, B_1, \ldots, B_{k-1}, \{B_k \cup B''_1 \cup \cdots \cup B''_m\}\}.\]

\[\square\]

Renaming the corresponding blocks, it is immediate to obtain the following:

**Corollary 7.6** \(SL^*(\alpha)\) is \(\alpha\)-bridge unchaining.

See Example 5.3.

**Remark 8.** \(CL\) and \(AL\) are not bridge-unchaining. As we mentioned in the introduction, \(CL\) and \(AL\) show a tendency to merge isolated points before joining them to a pre-existing cluster. This tendency can be seen applying the algorithms in the following cases.
Let $\theta_{CL}(t_{i-1}) = \{B_1, x_1, \ldots, x_n, z_0, \ldots, z_k, B_2, y_1, \ldots, y_n\}$ and suppose the conditions from Definition 7.4 hold with equalities on conditions $c)$ and $d)$. See Figure 12. Assuming that $\ell_{CL}(z_0, B_1) > t_i$ and $\ell_{CL}(z_k, B_2) > t_i$, then in $\theta_{CL}(t_i)$, $\{z_0 \cup \cdots \cup z_k\}$ is a cluster.

Similarly, let $\theta_{AL}(t_{i-1}) = \{B_1, x_1, \ldots, x_n, z_0, \ldots, z_k, B_2, y_1, \ldots, y_n\}$ and suppose the conditions from Definition 7.4 hold with equalities on conditions $c)$ and $d)$. Assuming that $\ell_{AL}(z_0, B_1) > t_i$ and $\ell_{AL}(z_k, B_2) > t_i$, then in $\theta_{AL}(t_i)$, $\{z_0 \cup \cdots \cup z_k\}$ is a cluster.

In either case, the point $z_0$ (resp. $z_k$) would appear to be far away from $B_1$ (resp. $B_2$) while $z_0, z_k$ appear to be very close to each other.

This type of chaining through points or through smaller blocks may be also avoided by DBSCAN. In the conditions from Definition 7.4, if $\varepsilon < d(B_1, B_2)$ and the points $z_0, \ldots, z_k$ are not core points (with $k > 0$), DBSCAN would not merge $B_1$ and $B_2$ either. However, it would not necessarily return the clustering $\{\{B_1 \cup x_1 \cup \cdots \cup x_n\}, z_0, \ldots, z_k, \{B_2 \cup y_1 \cup \cdots \cup y_n\}\}$. The result depends on the density distribution of the points in $B_1$ and $B_2$ and the parameters $\varepsilon, \text{minPts}$ involved. Let us assume that every point in $B_1$, $B_2$ is a core point and $\varepsilon \geq t_i$. In this case, the output of DBSCAN will be two clusters, $B, B'$ with $\{B_1 \cup x_1 \cup \cdots \cup x_n \cup z_0\} \subset B$, $\{B_2 \cup y_1 \cup \cdots \cup y_n \cup z_k\} \subset B'$ and some single points (noise) from the sequence $z_1, \ldots, z_{k-1}$.

8. Conclusions

Herein, we treat a particular type of chaining effect which is characteristic from single linkage. This effect is reduced if the algorithm shows some sensitivity to the density distribution of the data set. Our aim was to define an algorithm such that it encodes information about the density distribution with a very simple input. $SL(\alpha)$ is able to detect clusters affected by this kind of chaining effect. This may be useful, for example, in data analytics and image segmentation.

We also provide some theoretical background to the study of the chaining effect. Thus, a hierarchical clustering method is strongly chaining if every pair of chained clusters is automatically merged into one cluster. This is the case of single linkage. On the contrary, a hierarchical clustering method is weakly unchaining if at least it detects some type of chained clusters. We prove that $SL(\alpha)$ is weakly unchaining while complete linkage and average linkage are not. Also, compared with DBSCAN, our method seems to have a more natural and powerful treatment of this problem.

One weakness of $SL(\alpha)$ is that it fails to detect when two blocks are chained by a single point or a small block. $SL^*(\alpha)$ deals with that weakness. We prove that $SL^*(\alpha)$ is $\alpha$-bridge-unchaining showing that $SL^*(\alpha)$ is capable of detecting this kind of chaining.
Figure 12. If $\theta_\alpha(t_{i-1})$ satisfies the conditions from Definition 7.4 with equalities on conditions $c)$ and $d)$, then $G_{\alpha}^{t_i}$ is the graph above and $\theta_\alpha(t_i)$ is as indicated.

Figure 13. If $t_j = t_i$ and $\theta_\alpha(t_{i-1})$ satisfies the conditions from Proposition 7.5 with equalities on conditions $c)$ and $d)$, then $G_{\alpha}^{t_i}$ is the graph above and $\theta_\alpha(t_i)$ is as indicated.

We focused here on the theoretical problem of chaining and providing a new approach. Future work should consider the computational problem involved and study the efficiency of $SL(\alpha)$ compared with other methods.

Martínez-Pérez (2016) studies further properties of $SL(\alpha)$ and $SL^*(\alpha)$ and compares the results with the properties satisfied by some classical linkage-based hierarchical clustering methods.

One of the main advantages of $SL$ is that it is stable in the Gromov-Hausdorff sense which is a really interesting property for a clustering algor-
-ithm. If the algorithm is too sensitive to small perturbation of the data, the output may be easily meaningless. Unfortunately, our method does not share with SL the good stability properties. Modifying the algorithm to deal with the chaining effect we lost that advantage. The problem of stability of linkage-based clustering methods and the difficulties to define linkage-based algorithms, other than SL, stable in the Gromov-Hausdorff sense is studied in Martínez-Pérez (2015).

References

ACKERMAN, M., and BEN-DAVID, S. (2008), “Measures of Clustering Quality: A Working Set of Axioms for Clustering”, Neural Information Processing Systems Conference (NIPS).

ACKERMAN, M., BEN-DAVID, S., and LOKER, D. (2010), “Towards Property-Based Algorithms Among Hierarchical Clustering Methods”, Advances in Neural Information Processing Systems, 23, 10–18.

ACKERMAN, M., and BEN-DAVID, S. (2011), “Discerning Linkage-Based Algorithms Among Hierarchical Clustering Methods”, International Joint Conference on Artificial Intelligence (IJCAI).

ACKERMAN, M., BROWN, D., and LOKER, D. (2012), “Effects of Rooting via Outgroups on Ingroup Topology in Phylogeny”, IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS).

BURIAGO, D., BURIAGO, Y., and IVANOV, S. (2001), A Course in Metric Geometry, Graduate Studies in Mathematics, 33, Providence, RI: AMS.

CARLSSON, G., and MÉMOLI, F. (2008), “Persistent Clustering and a Theorem of J. Kleinberg”, arXiv:0808.2241.

CARLSSON, G., and MÉMOLI, F. (2010a), “Classifying Clustering Schemes”, arXiv:1011.5270v2.

CARLSSON, G., and MÉMOLI, F. (2010b), “Characterization, Stability and Convergence of Hierarchical Clustering Methods”, Journal of Machine Learning Research, 11, 1425–1470.

CARLSSON, G., and MÉMOLI, F. (2010c), “Multiparameter Hierarchical Clustering Methods”, in Classification as a Tool for Research, ed. C.W.H. Locarek-Junge, (Proceedings of the 11th Conference of the International Federation of Classification Societies, IFCS-09), Heidelberg: Springer, pp. 63–70.

ESTER, M., KIREGEL, H., SANDER, J., and XU, X. (1996), “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”, in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI Press, pp. 226–231.

GROMOV, M. (2007), Metric Structures for Riemannian and Non-Riemannian Spaces, Modern Birkhäuser Classics, Boston MA: Birkhäuser Boston Inc.

HUGHES, B. (2004), “Trees and Ultrametric Spaces: A Categorical Equivalence”, Advances in Mathematics, 189, 148–191.

JARDINE, N., and SIBSON, R. (1971), Mathematical Taxonomy, New York: Wiley.

JOHNSON, S.C. (1967), “Hierarchical Clustering Schemes”, Psychometrika, 32, 241–254.
KLEINBERG, J.M. (2002), “An Impossibility Theorem for Clustering”, in Advances in Neural Information Processing Systems, 15 (NIPS 2002), eds. S. Becker, S. Thrun, and K. Obermayer, Cambridge MA: MIT Press, pp. 446–453.

LANCE, G.N., and WILLIAMS, W.T. (1967), “A General Theory of Classificatory Sorting Strategies 1. Hierarchical Systems”, Computer Journal, 9 (4), 373–380.

MARTÍNEZ-PÉREZ, A., and MORÓN, M A. (2009), “Uniformly Continuous Maps Between Ends of \( \mathbb{R} \)-Trees”, Mathematische Zeitschrift, 263(3), 583–606.

MARTÍNEZ-PÉREZ, A. (2015), “Gromov-Hausdorff Stability of Linkage-Based Hierarchical Clustering Methods”, Advances in Mathematics, 279, 234–262.

MARTÍNEZ-PÉREZ, A. (2016), “On the Properties of \( \alpha \)-Unchaining Single Linkage Hierarchical Clustering”, Journal of Classification, 33(1), 118–140.

NARENDRA, P.M., and GOLDBERG, M. (1980), “Image Segmentation with Directed Trees”, IEEE Transactions PAMI, 2, 185–191.

ROHLF, F.J. (1982), “Single-Link Clustering Algorithms”, in Handbook of Statistics, 2, ed. P.R. Krishnaiah, Amsterdam: North Holland, pp. 267–284.

SNEATH, P.H.A., and SOKAL, R.R. (1973), Numerical Taxonomy. The Principles and Practice of Numerical Classification, San Francisco: W.H. Freeman.

WILLIAMS, W.T., LAMBERT, J.M., and LANCE, G.N. (1966), “Multivariate Methods in Plant Ecology: V,” The Journal of Ecology, 54(2), 427–445.

WISHART, D. (1969), “Mode Analysis: A Generalization of Nearest Neighbor Which Reduces Chaining Effects”, in Numerical Taxonomy, ed. A.J. Cole, New York: Academic Press, pp. 282–311.

ZADEH, R.B., and BEN-DAVID, S. (2009), “A Uniqueness Theorem for Clustering”, Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI), arXiv: 1205.2600 [cs.LG].

ZAHN, C.T. (1971), “Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters”, IEEE Transactions on Computers, C-20, 68–86.