An interval convexity-based framework for multilevel clustering with applications to single-linkage clustering

Patrice Bertrand, Jean Diatta

An interval convexity-based framework for multilevel clustering with applications to single-linkage clustering

Patrice Bertrand[a], Jean Diatta[b]

[a]Ceremade, Université Paris-Dauphine-PSL, Pl. du Maréchal de Lattre de Tassigny, Paris, 75016, France
[b]LIM-EA2525, Université de La Réunion, Parc Technologique Universitaire, 2 rue Joseph Wetzell, Sainte Clotilde, 97490, France

Abstract

Hierarchies of sets, and most multilevel clustering models, have been characterized as convexities induced by interval functions satisfying specific properties, thus giving rise to a unifying framework for characterizing multilevel clustering models. Here, we show that this unifying framework can be relevant to data mining practice. First, we provide a flexible characterization of hierarchies and weak hierarchies as interval convexities. Second, we investigate the Apresjan hierarchy and the Bandelt and Dress weak hierarchy, and characterize them as interval convexities. Third, we propose a method for computing recursively a sequence of path-based dissimilarities which decreases from an arbitrary dissimilarity downto its subdominant ultrametric. We prove that these path-based dissimilarities define two sequences of nested families of interval convexities. One sequence increases from the Apresjan hierarchy to the single-link hierarchy, and the other from a subset of the single-link hierarchy to the Bandelt and Dress weak hierarchy. Applications to the simplification and validation of the single-link hierarchy of an arbitrary dissimilarity are discussed.

Keywords: Apresjan hierarchy, Interval convexity, Single-link hierarchy, Weak hierarchy, Path-based dissimilarities

1. Introduction

Clustering is a popular and valuable data mining technique aimed at revealing a hidden structure within a dataset. The revealed structure is commonly expressed as a collection of homogeneous subsets called clusters. According to Mirkin and Muchnick [23], there are three major approaches to determine a cluster as based on definition, direct algorithm and optimality criterion. The present paper is mainly concerned with some definition-based clusters, namely clusters

Email addresses: patrice.bertrand@ceremade.dauphine.fr (Patrice Bertrand), jean.diatta@univ-reunion.fr (Jean Diatta)

defined as being convex in the sense of an interval function. An interval function I on a dataset S is a symmetric function that maps each ordered pair $(x, y) \in S \times S$ to a subset $I(x, y)$ of S, containing x and y. The subset $I(x, y)$ is called the I-interval of extremities x and y. A subset A of S is said to be convex in the sense of the interval function I, or I-convex, if it contains each I-interval whose extremities belong to A. The collection $\mathscr{C}_I$ of I-convex subsets of S forms a convexity called the interval convexity induced by I. It turns out that the notion of convexity as defined in the most abstract sense (e.g. [14, 27]) coincides with that of multilevel clustering, when the latter is supposed to be closed under arbitrary intersections. A multilevel clustering of S is a collection of nonempty subsets of S, containing S itself and of which at least two members are strictly nested. One of the most known multilevel clustering models is the hierarchical model. A hierarchy is a multilevel clustering for which the intersection of any two clusters is either empty or equal to one of them. In the last decades, several authors have investigated various generalizations of the hierarchical model, such as pyramids [17, 19], paired hierarchies [7, 8] and totally balanced hypergraphs [12, 11]. These models are submodels of the weak hierarchical clustering model. A weak hierarchy is a multilevel clustering for which the intersection of any three clusters is reduced to the intersection of two of them [2, 3, 16]. It is therefore a direct extension of the hierarchical clustering, allowing for some type of overlapping clusters. The weak hierarchical clustering model, and most of its submodels, have been characterized as (types of) interval convexities whose segment functions satisfy specific properties [9]. By generalizing the notion of interval function to more than two arguments, Changat et al. [15] have extended these characterizations to the general case of the k-weak hierarchies for $k \geq 2$, where a k-weak hierarchy is a multilevel clustering for which the intersection of any $(k+1)$ clusters reduces to the intersection of k of them (the case $k = 2$ corresponds to weak-hierarchies). These characterizations, all being based solely on the notion of interval function or its extensions, lead to a unifying framework for multilevel clusterings. The contributions of this paper fall within this line of research, in this general formal framework, together with an attention to specifications allowing to capture clustering models, such as the Apresjan and the single-link hierarchies, which are handled in data mining practice. We propose:

- Flexible characterizations of hierarchies and weak hierarchies as interval convexities;
- Properties of both the Apresjan hierarchy and the Bandelt and Dress weak-hierarchy of an arbitrary dissimilarity, together with their characterization as interval convexities;
- A recursive method for computing (in linear time) each term of a sequence of path-based dissimilarities that decrease from an arbitrary dissimilarity downto its subdominant ultrametric;
- A sequence of a nested hierarchies defined as interval convexities, gradually increasing from the Apresjan hierarchy to the single-link hierarchy;
- A sequence of nested weak hierarchies defined as interval convexities, gradually increasing from a subset of persistent single-linkage clusters (which includes the Apresjan clusters) to the Bandelt and Dress weak-hierarchy.

The rest of the paper is organized as follows. Section 2 presents elementary background notions. Section 3 provides an unifying framework for the construction of interval functions that induce either a hierarchy or a weak hierarchy. Given a function $g : S \times S \mapsto 2^S$ and its two symmetrical versions $J_g = g \cup \overline{g}$ and $M_g = g \cap \overline{g}$ that are defined by $(g \cup \overline{g})(x, y) = g(x, y) \cup \overline{g}(x, y)$

and $(g \cap \overline{g})(x,y) = g(x,y) \cap \overline{g}(x,y)$ for all $x, y \in S$, with $\overline{g}(x,y) = g(y,x)$, we consider the case where J_g and M_g are interval functions on S. We then introduce a condition (H) (resp. (W)) that characterizes the functions g such that J_g (resp. M_g) generates a hierarchical (resp. weakly hierarchical) interval convexity. In the remaining sections, we consider an arbitrary dissimilarity on S, denoted by δ. In Section 4, we show that the two-way Ball-map $g_B^\delta : (x,y) \mapsto B_\delta(x, \delta(x,y)) = \{z \in S \mid \delta(x,z) \leq \delta(x,y)\}$, also denoted simply g_B, satisfies both of the conditions (H) and (W). We also focus on interval convexities $\mathrm{conv}(J_{g_B})$ and $\mathrm{conv}(M_{g_B})$ which coincide respectively with the Apresjan hierarchy of δ and the Bandelt and Dress weak hierarchy of δ. In section 5, we introduce a decreasing sequence $\left(\delta_\ell\right)_{\ell=1,\ldots,n-1}$ of path-based dissimilarities, with $n = |S|$, $\delta_1 = \delta$ and δ_{n-1} is the subdominant ultrametric of δ. Furthermore, we propose a recursive method to compute each dissimilarity value $\delta_\ell(x,y)$, with a complexity bound $O(n)$. In Section 6, we focus on the sequence $\left(\mathrm{conv}(J_{h_\ell})\right)_\ell$ with $h_\ell = g_B^{\delta_\ell}$, and show that this sequence is a nested family of sub-hierarchies of the single-link hierarchy of δ, which gradually increases from the stringent Apresjan hierarchy of δ. In Section 7, we investigate the sequence $\left(\mathrm{conv}(M_{h_\ell})\right)_\ell$, and derive a nested family of sub-weak hierarchies of the Bandelt and Dress weak hierarchy of δ, which gradually increases from a subset of the single-link hierarchy, which contains the Apresjan hierarchy. The final Section 8 discusses possible extensions, in particular by taking advantage of the gradual increase of these nested families to provide guidelines for defining an approach of multilevel clustering simplification.

2. Background

In this text, S denotes the ground (finite) set and $n = |S|$ its size. Let us consider an arbitrary collection $\mathscr{C}$ of subsets of S. It will be said that $\mathscr{C}$ is *hierarchical* if the intersection of any two of its members is either empty or equal to one of the two members. The collection $\mathscr{C}$ is called *weakly hierarchical* if the intersection of any three of its members is equal to the intersection of two of them. The collection $\mathscr{C}$ is said to be a *multilevel clustering* if it contains S and at least one proper subset of S, and if none of its members, called *clusters*, is empty. A *weak hierarchy* [2, 3, 16] (resp. a *hierarchy*) is a multilevel clustering that is weakly hierarchical (resp. hierarchical). By allowing overlapping clusters, the weak hierarchical clustering model extends the hierarchical clustering model. In what follows, multilevel clusterings are viewed through the framework of the so-called abstract convexities introduced since the 1950s [14, 27]. We will use the terminology presented in Van de Vel in [27]. As S is assumed to be finite, a *convexity* on S can be defined as any collection $\mathscr{C} \subseteq 2^S$ which contains both $\varnothing$ and S, and is closed under arbitrary intersections. Clearly, any multilevel clustering closed under nonempty intersections and completed by the empty set is then an abstract convexity. Morover, as shown in [9], up to the completion by the empty set, the most known multilevel clustering models dealt with in the literature are convexities induced by some interval functions. An *interval function* I on S is a symmetric function which maps each ordered pair $(x,y) \in S \times S$ to a subset $I(x,y)$ of S that contains x and y. The subset $I(x,y)$ is called the interval of extremities x and y. A subset A of S is said to be convex according to the interval function I if it contains each interval $I(x,y)$ whose extremities x and y are in A. It is easily checked that each interval function I *induces the convexity* defined as the collection of subsets of S that are convex according to I. Each convexity induced by an interval function I, is called an *interval convexity* and is denoted as $\mathrm{conv}(I)$. Conversely, each convexity $\mathscr{C}$ is associated with an interval function, denoted as $\mathrm{seg}(\mathscr{C})$, and defined by $\mathrm{seg}(\mathscr{C})(x,y) = \bigcap\left\{A \in \mathscr{C} \mid x, y \in A\right\}$ for all $x, y \in S$. This interval function $\mathrm{seg}(\mathscr{C})$

is called the *segment function* associated with convexity $\mathscr{C}$.
Let us recall the axiomatic characterizations of hierarchies and weak hierarchies, as convexities. They are based on the following axioms (H) and (W$_k$) defined as follows. A map $f : S \times S \mapsto 2^S$ satisfies (H) and (W$_k$), respectively, if

(H) for all $x_1, x_2, x_3 \in S$, either $f(x_1, x_2) \subseteq f(x_1, x_3)$ or $f(x_1, x_3) \subseteq f(x_1, x_2)$,

(W$_k$) if there are no $x_1, \ldots, x_{k+1} \in S$ such that for all i, $x_i \notin \bigcup_{x_j \neq x_i \neq x_l} f(x_j, x_l)$,

where k denotes an integer such that $k \geq 2$.

The next two theorems and their respective proofs can be found in [9].

Theorem 1. [9, Theorem 2.5 p. 58] *For all convexity $\mathscr{C}$ on S and all integer $k \geq 3$, we have:*

(i) $\mathrm{seg}(\mathscr{C})$ *satisfies condition* (H) *if and only if $\mathscr{C}$ is hierarchical.*
(ii) $\mathrm{seg}(\mathscr{C})$ *satisfies condition* (W$_2$) *if and only if $\mathscr{C}$ is weakly hierarchical.*
(iii) *If* $\mathrm{seg}(\mathscr{C})$ *satisfies condition* (W$_k$) *then $\mathscr{C}$ is k-weakly hierarchical.*

It is worth noticing that $\mathscr{C} = \mathrm{conv}(\mathrm{seg}(\mathscr{C}))$ whenever $\mathscr{C}$ is a weakly hierarchical convexity (cf. [9], Proposition 3.10, p.61). In general, $\mathrm{seg}(\mathscr{C})$ is not the only interval function which induces $\mathscr{C}$. Indeed, given an interval convexity $\mathscr{C}$, there is a great deal of freedom for choosing an interval function I that induces the convexity $\mathscr{C}$, i.e. such that $\mathscr{C} = \mathrm{conv}(I)$. The following result will be helpful for choosing such an interval function for clustering purposes.

Theorem 2. [9, Theorem 3.5 p.60] *Consider an interval function I on S and an integer $k \geq 2$. Then:*

(i) *If I satisfies* (H) *then* $\mathrm{conv}(I)$ *is hierarchical.*
(ii) *If I satisfies* (W$_k$) *then* $\mathrm{conv}(I)$ *is k-weakly hierarchical.*

3. A convex framework for hierarchies and weak-hierarchies

Let $g : S \times S \longrightarrow 2^S$ be an arbitrary map. The map $\overline{g} : S \times S \longrightarrow 2^S$ is defined by $\overline{g}(x, y) = g(y, x)$ for all $x, y \in S$. In addition, given an arbitrary map $f : S \times S \longrightarrow 2^S$, the maps $g \cup f : S \times S \longrightarrow 2^S$ and $g \cap f : S \times S \longrightarrow 2^S$ are defined by:

$$(g \cup f)(x, y) = g(x, y) \cup f(x, y) \quad \text{and} \quad (g \cap f)(x, y) = g(x, y) \cap f(x, y),$$

for all $x, y \in S$. In this section, we will consider a map $g : S \times S \mapsto 2^S$ satisfying the following axiom (C$_0$):

(C$_0$) For all $x, y \in S$, $\{x, y\} \subseteq g(x, y)$.

We begin by symmetrizing g in two ways. Let $J_g : S \times S \mapsto 2^S$ and $M_g : S \times S \mapsto 2^S$ be the two maps defined by:

$$J_g = g \cup \overline{g} \quad \text{and} \quad M_g = g \cap \overline{g}.$$

It is clear that J_g and M_g are symmetrical and both satisfy (C$_0$), i.e. J_g and M_g are interval functions on S. We will investigate the induced convexities $\mathrm{conv}(J_g)$ and $\mathrm{conv}(M_g)$ from a clustering point of view. We first introduce two axioms, denoted as (W) and (W$'$). Together with (H), these axioms will prove to be decisive in establishing the properties of $\mathrm{conv}(J_g)$ and $\mathrm{conv}(M_g)$. We will say that the map g satisfies, respectively, (W) and (W$'$) if:

(W) for all $x_1, x_2, x_3 \in S$, there exists i, j, k with $\{i, j, k\} = \{1, 2, 3\}$, such that:

$$g(x_i, x_j) \subseteq g(x_i, x_k) \text{ and } g(x_k, x_j) \subseteq g(x_k, x_i);$$

(W′) for all $x_1, x_2, x_3 \in S$, there exists i, j, k with $\{i, j, k\} = \{1, 2, 3\}$, such that:

$$g(x_i, x_j) \subseteq g(x_i, x_k),\ g(x_k, x_j) \subseteq g(x_k, x_i) \text{ and } g(x_j, x_i) \subseteq g(x_j, x_k).$$

Notation 1. *We adopt a delimiter-free notation of subsets defined in extension. For instance, $\{a, b\}$ will be denoted indifferently either as ab or as ba, subset $\{a, b, c\}$ will be denoted as any of the expressions abc, bac, acb, cab, bca, etc. Depending on the context, the notation a may denote either the element a or the singleton $\{a\}$.*

It is easily checked that (W′) ⇒ (H) and (W′) ⇒ (W). As shown by the following two counter-examples, no other implication exists between axioms (H), (W) and (W′).

Counter-example 1. *Let $S = abc$ and consider the map $g_1 : S \times S \mapsto 2^S$ whose values are given in Table 1. Clearly g_1 verifies (C_0), and, in addition, also condition (H) since:*
$g_1(a, b) = ab \subseteq abc = g_1(a, c), g_1(b, c) = bc \subseteq abc = g_1(b, a), g_1(c, a) = ac \subseteq abc = g_1(c, b)$ and $g_1(x, x) = x$ for all $x \in abc$.

g_1	a	b	c
a	a	ab	abc
b	abc	b	bc
c	ac	abc	c

Table 1: (H) ⇏ (W) and (H) ⇏ (W′)

If g_1 satisfies (W), then there exists $x, y, z \in abc$ pairwise distincts such that $g_1(x, z) \subseteq g_1(x, y)$ and $g_1(y, z) \subseteq g_1(y, x)$. Now, we have:

$$g_1(a, c) = abc \nsubseteq ab = g_1(a, b),\ g_1(c, b) = abc \nsubseteq ac = g_1(c, a) \text{ and } g_1(b, a) = abc \nsubseteq bc = g_1(b, c).$$

Thus g_1 does not satisfy (W), which proves that (H) implies neither (W) nor (W′).

Counter-example 2. *Let $S = abc$ and consider the map $g_2 : S \times S \mapsto 2^S$ whose values are indicated in Table 2. Clearly g_2 verifies (C_0), and in addition, g_2 satisfies also (W) since:*

g_2	a	b	c
a	a	abc	ac
b	abc	b	bc
c	ac	bc	c

Table 2: (W) ⇏ (H) and (W) ⇏ (W′)

$$g_2(a, c) = ac \subseteq abc = g_2(a, b), g_2(b, c) = bc \subseteq abc = g_2(b, a) \text{ and } g_2(x, x) = x, \text{for all } x \in S.$$

However g_2 does not verify (H) for $g_2(c, a) = ac$ and $g_2(c, b) = bc$ are not comparable according to set inclusion. Consequently (W) implies neither (H) nor (W′).

Proposition 3, below, together with corollary 4, emphasize the interest of axioms (H) or (W) from a clustering point of view.

Proposition 3. Let g be a map from $S \times S$ into 2^S which satisfies (C_0). Then the following properties hold true.

(i) If g satisfies *(H)*, then the convexity induced by J_g is hierarchical.
(ii) If g satisfies *(W)*, then the convexity induced by M_g is weakly hierarchical.

PROOF. (i). Let A and B be two elements of $\operatorname{conv}(J_g)$. Assume that A and B intersect properly. Then, there exist x_1, x_2, x_3 such that:

$$x_1 \in A \setminus B,\ x_2 \in A \cap B \text{ and } x_3 \in B \setminus A.$$

By definition, for all $i, j \in \{1,2,3\}$, $J_g(x_i, x_j) = g(x_i, x_j) \cup g(x_j, x_i)$. Since g verifies (H), we have $g(x_2, x_1) \subseteq g(x_2, x_3)$ or $g(x_2, x_3) \subseteq g(x_2, x_1)$. By convexity of A and B, according to the interval function J_g, we deduce:
If $g(x_2, x_1) \subseteq g(x_2, x_3)$ then $x_1 \in g(x_2, x_3) \subseteq J_g(x_2, x_3) \subseteq B$: contradiction.
If $g(x_2, x_3) \subseteq g(x_2, x_1)$ then $x_3 \in g(x_2, x_1) \subseteq J_g(x_2, x_1) \subseteq A$: contradiction.
Consequently A and B do not intersect properly, and the result is proved.

(ii). By theorem 2, it is sufficient to prove that M_g verifies condition (W_2), or in other words that:

$$\text{No } x_1, x_2, x_3 \in S \text{ exist such that for all } i \in \{1,2,3\},\ x_i \notin \bigcup_{x_j \neq x_i \neq x_k} M_g(x_j, x_k).$$

Within this aim, let us consider three distinct arbitrary elements of S, say x_1, x_2, x_3. Even if it means renumbering the elements x_i, one can assume without loss of generality that:

$$g(x_1, x_2) \subseteq g(x_1, x_3) \text{ and } g(x_3, x_2) \subseteq g(x_3, x_1),$$

since g verifies (W). It results that $x_2 \in g(x_1, x_3) \cap g(x_3, x_1) = M_g(x_1, x_3)$, which proves that M_g verifies (W_2). □

The following corollary results immediately from Proposition 3 and the fact that (W') ⇒ (H) and (W') ⇒ (W).

Corollary 4. If g verifies *(C_0)* and *(W')*, then the convexity induced by J_g (resp. M_g) is hierarchical (resp. weak hierarchical).

Proposition 5. If $g : S \times S \mapsto 2^S$ satisfies axiom *(C_0)*, then the following conditions are equivalent.

(i) g is symmetrical,
(ii) $J_g = M_g$,
(iii) $J_g = M_g = g$.

PROOF. (iii) ⇔ (i) and (iii) ⇒ (ii) are obvious. Moreover (ii) ⇒ (iii) is easy to check. □

Proposition 6. Let $\mathscr{C}$ be a multilevel clustering on S. Then $\mathscr{C}$ is hierarchical if and only if there exists a map $g : S \times S \mapsto 2^S$ satisfying *(C_0)* and *(H)*, such that $\mathscr{C} = \operatorname{conv}(J_g)$.

PROOF. From Proposition 3, we know that if $g : S \times S \mapsto 2^S$ satisfies *(C_0)* and *(H)*, then $\mathscr{C} = \mathrm{conv}(J_g)$ is hierarchical. Assume now that the collection $\mathscr{C}$ is hierarchical. We have $\mathscr{C} = \mathrm{conv}(\mathrm{seg}(\mathscr{C}))$ since $\mathscr{C}$ is weakly-hierarchical (cf. [9], Proposition 3.10, p.61). Now, $\mathrm{seg}(\mathscr{C})$ is clearly a map from $S \times S$ to 2^S which is symmetrical and satifies (C_0). Then, from Proposition 5, it results that $J_{\mathrm{seg}(\mathscr{C})} = \mathrm{seg}(\mathscr{C})$, so that $\mathscr{C} = \mathrm{conv}(J_{\mathrm{seg}(\mathscr{C})})$. In addition, from Theorem 1, we deduce that the map $\mathrm{seg}(\mathscr{C})$ satisfies (H), as required. □

Proposition 7. Let $\mathscr{C}$ be a multilevel clustering on S. Then, $\mathscr{C}$ is weakly hierarchical if and only if it exists a map $g : S \times S \mapsto 2^S$ satisfying *(C_0)* and *(W)*, and such that $\mathscr{C} = \mathrm{conv}(M_g)$.

PROOF. From Proposition 3-(i), it is clear that the existence of a map $g : S \times S \mapsto 2^S$ that satisfies *(C_0)* and *(W)*, together with the relation $\mathscr{C} = \mathrm{conv}(M_g)$, implies that $\mathscr{C}$ is weakly hierarchical. Conversely, assume that $\mathscr{C}$ is weakly hierarchical, and let us consider the segment function $\sigma = \mathrm{seg}(\mathscr{C})$. Then σ is symmetrical, so that $\sigma = M_\sigma$ by Proposition 5. Moreover, $\mathscr{C} = \mathrm{conv}(\sigma)$ since $\mathscr{C}$ is weakly-hierarchical (cf. [9], Proposition 3.10, p.61) and, by definition of $\sigma = \mathrm{seg}(\mathscr{C})$, the map σ satisfies (C_0). Therefore, in order to prove the proposition, we are left to prove that σ satisfies (W). Since $\mathscr{C}$ is weakly hierarchical, its segment function σ satisfies (W_2) by Theorem 1. Consequently, given $x_1, x_2, x_3 \in S$, there exists i, j, k such that $x_j \in \sigma(x_i, x_k)$ with $\{i, j, k\} = \{1, 2, 3\}$. As σ is symmetrical, we have $x_j \in \sigma(x_i, x_k) = \sigma(x_k, x_i)$. By definition of a segment function, it results that

$$\sigma(x_i, x_j) \subseteq \sigma(x_i, x_k) \text{ and } \sigma(x_k, x_j) \subseteq \sigma(x_k, x_i),$$

which proves that σ satisfies (W), as required. □

We end this section with properties of convexities $\mathrm{conv}(J_g)$ and $\mathrm{conv}(M_g)$.

Proposition 8. Let $g : S \times S \mapsto 2^S$ be a map that satisfies axioms *(C_0)* and *(H)*. If C is a subset of S and x_0 is any element of C, then the following statements hold.

(i) The finite collection $g(x_0, C) = \{g(x_0, y) \mid y \in C\}$ is linearly ordered by inclusion. Its greatest element will be denoted as $\max g(x_0, C)$ in the sequel.

(ii) If C is J_g-convex, then $C = g(x_0, y_0)$, for all $y_0 \in C$ such that $g(x_0, y_0) = \max g(x_0, C)$.

PROOF. Let C be a J_g-convex subset and $x_0 \in C$.
(i). Since g satisfies (H), we have $g(x_0, u) \subseteq g(x_0, v)$ or $g(x_0, v) \subseteq g(x_0, u)$ for all $u, v \in C$. Therefore, the finite collection of subsets $g(x_0, C) = \{g(x_0, y) \mid y \in C\}$ is linearly ordered by inclusion, and thus has a (unique) greatest element.
(ii). Let $y_0 \in C$, such that $g(x_0, y_0) = \max g(x_0, C)$. Then, for all element u of C, $u \in g(x_0, y_0)$ since, on the one hand, we have $u \in g(x_0, u)$ for g satisfies (C_0) and, on the other hand, $g(x_0, u) \subseteq g(x_0, y_0)$. Consequently, $C \subseteq g(x_0, y_0)$. Conversely, $g(x_0, y_0) \subseteq J_g(x_0, y_0) \subseteq C$ since C is J_g-convexe. It results that $C = g(x_0, y_0)$. □

Remark 3. *Let $g : S \times S \mapsto 2^S$ satisfy (C_0). Assume first that g satisfies (H). From Proposition 3, the collection $\mathrm{conv}(J_g)$ of J_g-convex subsets of S is hierarchical, and is then closed under intersection. Let C be a J_g-convex subset of S. From Proposition 8, for all $x_0 \in C$, there exists $y_0 \in C$ such that $C = g(x_0, y_0)$, and thus $C = J_g(x_0, y_0)$.*
Assume now that g satisfies (W′). From Corollary 4, the collection $\mathrm{conv}(M_g)$ of M_g-convex

subsets of S is weakly hierarchical. Let C be a M_g-convex subset. A result similar to Proposition 8 would imply that there exist $x_0, y_0 \in C$ such that $C = M_g(x_0, y_0)$. The following counter-example shows that such elements x_0 and y_0 do not exist in the general case. Let $S = abcd$ and g the map from $S \times S$ to 2^S defined by Table 3. It is clear that g satisfies (C_0). Let us prove that g satisfies also (W′), i.e. that for all three distinct elements of S, say x_1, x_2, x_3, there exists some permutation (x_i, x_j, x_k) of (x_1, x_2, x_3) such that:

$$g(x_i, x_j) \subseteq g(x_i, x_k),\ g(x_k, x_j) \subseteq g(x_k, x_i) \text{ and } g(x_j, x_i) \subseteq g(x_j, x_k). \tag{1}$$

g	a	b	c	d
a	a	$abcd$	ac	acd
b	abd	b	$abcd$	bd
c	ac	abc	c	$abcd$
d	$abcd$	bd	bcd	d

Table 3

Consider all distinct triples of elements of S, i.e. abc, abd, acd and bcd. For each of these triples, there exists some permutation of their elements that satisfy (1). For example, for the triple abc, we have :

$$g(c,a) = ac \subseteq g(c,b) = abc, g(b,a) = abd \subseteq g(b,c) = abcd, \text{ and } g(a,c) = ac \subseteq g(a,b) = abcd,$$

and thus the permutation (c, a, b) satifies (1). For other triples, i.e. abd, acd, bcd, one can check similarly that the respective permutations (b, d, a), (a, c, d) and (d, b, c) satisfy (1). It results that g satisfies (W′). Let us then consider subset S which is obviously M_g-convex. It can be observed that there is no $x_0, y_0 \in S$ such that $S = M_g(x_0, y_0)$. Indeed, from Table 3, one can easily check that for all $u, v \in S$, if $g(u, v) = abcd = S$, then $g(v, u) \subset abcd$, which shows that $S \neq M_g(u, v)$.

Notation 2. *We consider the relation defined by $f \preceq g$ if $f(x, y) \subseteq g(x, y)$ for all elements x, y of S. Furthermore, define $f \prec g$ to mean that $f \preceq g$ and there exist $x, y \in S$ such that $f(x, y) \subset g(x, y)$, where $\subset$ denotes the strict subset inclusion order.*

Proposition 9. Let I_1 and I_2 be interval functions on S. Then the following hold:

(i) $I_1 \preceq I_2 \Rightarrow \mathrm{conv}(I_2) \subseteq \mathrm{conv}(I_1)$,

(ii) If for $j \in \{1, 2\}$, $I_j = seg(\mathscr{C}_j)$ with $\mathscr{C}_j$ a weak hierarchical convexity, then $I_1 \preceq I_2 \Leftrightarrow \mathrm{conv}(I_2) \subseteq \mathrm{conv}(I_1)$

PROOF. (i). The proof is elementary (cf. [9] p. 64).
(ii). Considering (i), we are left to prove $\mathrm{conv}(I_2) \subseteq \mathrm{conv}(I_1) \Longrightarrow I_1 \preceq I_2$. Then, assume that $\mathrm{conv}(I_2) \subseteq \mathrm{conv}(I_1)$ and let $x, y \in S$. We aim to prove $I_1(x, y) \subseteq I_2(x, y)$. By hypothesis, the convexity $\mathscr{C}_j$ is weakly hierarchical, so that we have $\mathscr{C}_j = \mathrm{conv}(\mathrm{seg}(\mathscr{C}_j)) = \mathrm{conv}(I_j)$ for all $j \in \{1, 2\}$, and thus $\mathscr{C}_2 \subseteq \mathscr{C}_1$. We obtain

$$\begin{aligned} I_1(x,y) = \mathrm{seg}(\mathscr{C}_1)(x,y) &= \bigcap\{C \in \mathscr{C}_1 \mid x, y \in C\} \\ &\subseteq \bigcap\{C \in \mathscr{C}_2 \mid x, y \in C\} = \mathrm{seg}(\mathscr{C}_2)(x,y) = I_2(x,y), \end{aligned}$$

as required. □

4. Apresjan hierarchies and Bandel and Dress weak-hierarchies as interval convexities

The notion of a cluster is often defined in some broadly and loosely sense, i.e. in a way similar to the following: given a ground set S endowed with some measure of dissimilarity, a cluster is any subset of S having high degrees of cohesion and isolation according to the values taken by this dissimilarity. In this section, we are concerned with two more precise definitions of a cluster. These definitions are based on the notion of a ball according to a dissimilarity δ defined on the set of objects to be clustered. Let us first recall that a *dissimilarity* δ on S is a map from $S \times S$ to $\mathbb{R}^+$, satisfying $\delta(x,y) = \delta(y,x) \geq \delta(x,x) = 0$ for all $x, y \in S$. The diameter of a subset A of S according to δ, denoted as $\mathrm{diam}_\delta(A)$, is defined by $\mathrm{diam}_\delta(A) = \max\{\delta(x,y) \mid x, y \in A\}$. Given $a \in S$ and $\rho \geq 0$, we denote by $B_\delta(a,\rho)$ the closed ball of center a and radius ρ with respect to δ, i.e. $B_\delta(a,\rho) = \{x \in S \mid \delta(x,a) \leq \rho\}$. The collection of all the closed balls with respect to the dissimilarity δ, is called the *ballean* of the space (S,δ) (e.g. [18]) and is denoted hereafter as $Balls(X,\delta)$.
We now consider two interval functions whose respective related convex subsets will turn out to coincide with two distinct cluster types: the Apresjan clusters of δ (cf. Proposition 12) and the Bandelt and Dress clusters of δ (cf. Proposition 17). Let us first introduce some definitions and notations.

Definition 3. *Let $g_{B_\delta} : S \times S \longrightarrow 2^S$ be the map defined for all $x, y \in S$, by:*

$$g_{B_\delta}(x,y) = B_\delta(x, \delta(x,y)) = \{s \in S \mid \delta(x,s) \leq \delta(x,y)\}.$$

Let $\mathbf{D}_\delta = g_{B_\delta} \cup \overline{g_{B_\delta}} = J_{g_{B_\delta}}$ and $\mathbf{B}_\delta = g_{B_\delta} \cap \overline{g_{B_\delta}} = J_{g_{B_\delta}}$. We note that for all $x, y \in S$,

$$\mathbf{D}_\delta(x,y) = \{z \in S \mid \min\{\delta(x,z), \delta(y,z)\} \leq \delta(x,y)\},$$
$$\mathbf{B}_\delta(x,y) = \{z \in S \mid \max\{\delta(x,z), \delta(y,z)\} \leq \delta(x,y)\}.$$

The subset $\mathbf{B}_\delta(x,y)$ is called the 2-ball generated by x and y *according to dissimilarity δ. If there is no ambiguity on the choice of the dissimilarity δ, the maps $g_{B_\delta}, \mathbf{D}_\delta$ and $\mathbf{B}_\delta$ will be denoted respectively as $g_B, \mathbf{D}$ and $\mathbf{B}$.*

It is clear that $x, y \in g_B(x,y)$ for all $x, y \in S$, so that g_B satisfies axiom (C_0). Therefore its associated maps $\mathbf{D} = J_{g_B}$ and $\mathbf{B} = M_{g_B}$ are interval functions defined on S. In addition, as shown by the next proposition, g_B satisfies axiom (W$'$).

Proposition 10. g_B satisfies *(W$'$)*.

PROOF. Let $x_1, x_2, x_3 \in S$. Note that there exist i, j, k such that $\{i,j,k\} = \{1,2,3\}$ and :

$$\delta(x_i, x_j) \leq \delta(x_j, x_k) \leq \delta(x_i, x_k).$$

By definition of g_B, we deduce the following inclusions:

$$g_B(x_i,x_j) \subseteq g_B(x_i,x_k)\ ,\ g_B(x_k,x_j) \subseteq g_B(x_k,x_i) \text{ and } g_B(x_j,x_i) \subseteq g_B(x_j,x_k).$$

In other words, g_B satisfies (W$'$). □

The following result derives from Proposition 10 and Corollary 4.

Corollary 11. The interval functions $\mathbf{D}$ and $\mathbf{B}$ induce convexities that are respectively hierarchical and weakly hierarchical.

4.1. Apresjan hierarchies

Given a dissimilarity δ on S, Apresjan [1] considers the subsets C of S such that for all $x, y \in C$:

$$\delta(x,y) < \min_{z \notin C} \big(\min\{\delta(x,z), \delta(y,z)\} \big). \tag{2}$$

A nonempty subset will be said to be an *Apresjan cluster* of δ, if it satisfies (2). The set of Apresjan clusters of any dissimilarity δ is a hierarchy called the *Apresjan hierarchy* of δ (cf. e.g. [1, 6]). The Apresjan hierarchy is one of the most standard clustering structures, which has provided a common starting point for several constructions of cluster systems (cf. [3, 16, 13]). As pointed out by Bryant and Berry [13], the notion of an Apresjan cluster provides an intuitive and compelling definition of a cluster. However for most dissimilarities, the Apresjan hierarchy is based on a stringent clustering criterion, so that the set of Apresjan clusters of an arbitrary dissimilarity is often a sparse hierarchy which differs from the trivial hierarchy (i.e. the collection of the whole dataset and its singletons) only by a few small non trivial clusters (cf. [13]). It results that in the context of an applied data analysis, the interest per se of the Apresjan hierarchy is often limited.

By showing that the set of $\mathbf{D}$-convex subsets coincide with the collection of Apresjan clusters, the next proposition gives the semantics of $\mathbf{D}$-convex subsets.

Proposition 12. For all dissimilarity δ on S and all nonempty subset C of S, the following statements are equivalent.

(i) C is an Apresjan cluster of δ.

(ii) C is $\mathbf{D}_\delta$-convex.

(iii) $\delta(x,y) < \delta(x,z)$ for all $(x,y,z) \in C \times C \times (S \setminus C)$.

Proof. $(i) \Leftrightarrow (ii)$. Let C be a nonempty subset of S and δ a dissimilarity defined on S. First, assume that C is an Apresjan cluster of δ, and consider two elements x, y of C. By definition of an Apresjan cluster, if $z \in S$ fulfills one of the inequalities $\delta(x,z) \leq \delta(x,y)$ or $\delta(y,z) \leq \delta(x,y)$, then z belongs necessarily to C. Therefore C is $\mathbf{D}$-convex. Conversely, assume that C is $\mathbf{D}$-convex. Consider $x, y \in C$ and $z \notin C$. Neither $\delta(x,z) \leq \delta(x,y)$ nor $\delta(y,z) \leq \delta(x,y)$ hold, otherwise z would belong to C since C is $\mathbf{D}$-convex. Therefore, for all $x, y \in C$ and all $z \notin C$, we have $\min\{\delta(x,z), \delta(y,z)\} > \delta(x,y)$, which proves that C is an Apresjan cluster of δ.
$(i) \Leftrightarrow (iii)$. Condition (iii) is simply a reformulation of (i). □

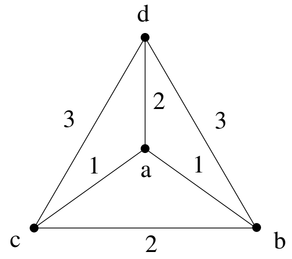
From proposition 12, it results that the collection of Apresjan clusters coincides with the interval convexity $\mathrm{conv}(\mathbf{D}) = \mathrm{conv}(J_{g_B})$, which, by Corollary 11, is a hierarchy. In the sequel, the Apresjan hierarchy of a dissimilarity δ will be denoted by $\mathcal{H}_\mathcal{A}(\delta)$.

Proposition 13. Let $A \subseteq S$ and $\mathscr{F}_\delta(A)$ be the collection of subsets of diameter equal to the diameter of A, according to a dissimilarity δ. If A is an Apresjan cluster of δ, then A is a maximal subset of $\mathscr{F}_\delta(A)$ (w.r.t the set-inclusion order).

Proof. Let $A, B \subseteq S$, and suppose that $A \subset B$. Let a_1 and a_2 be two elements of A such that $\delta(a_1, a_2) = \mathrm{diam}_\delta A$ and $b \in B \setminus A$. Since A is an Apresjan cluster of δ, we have :

$$\delta(a_1, a_2) < \min\{\delta(a_1, b), \delta(a_2, b)\},$$

which implies that $\mathrm{diam}_\delta A < \mathrm{diam}_\delta B$, as required. □



(a) Dissimilarity δ

	a	b	c	d
a	a	abc	abc	S
b	$-$	b	abc	S
c	$-$	$-$	c	S
d	$-$	$-$	$-$	d

(b) Interval function $\mathbf{D}_\delta$

Figure 1: (a) The weight of each edge $\{u, v\} \subset S = abcd$ indicates the value $\delta(u, v)$ taken by a dissimilarity δ defined on S. (b) The values of $\mathbf{D}_\delta$ are derived from δ. It is easily checked that abc is the only non trivial Apresjan cluster of δ, so that $\mathcal{H}_{\mathcal{A}}(\delta) = \{a, b, c, d, abc, S\}$. Subset ab is a maximal subset of diameter 1, but $ab \notin \mathcal{H}_{\mathcal{A}}(\delta)$.

The converse of Proposition 13, however, is not true as shown by the example in Figure 1.

Notation 4. *Recall that the distance between any two vertices x and y in a graph is defined as the length (i.e., the number of edges) of a shortest path joining x and y. Given a vertex subset A in a graph G, $\mathrm{Diam}_G(A)$ will denote its diameter in the graph G, i.e. the maximal distance between two vertices of A. Given $A \subseteq S$, we also denote by $\mu_\delta(A)$, or simply $\mu(A)$ if the choice of δ is clear, the threshold value defined by:*

$$\mu(A) = \mu_\delta(A) = \min\{\alpha \geq 0 \mid A \textit{ is connected in the graph } \Gamma_\delta(\alpha)\},$$

where $\Gamma_\delta(\alpha)$, or simply $\Gamma(\alpha)$ if there is no ambiguity on the choice of δ, denotes the (lower) threshold-based graph of dissimilarity δ at level α. The graph $\Gamma(\alpha)$, also called the Vietoris-Rips graph (see e.g. [10]), is defined as the (undirected) graph whose vertex set is S, and for which $xy \subseteq S$ is an edge iff $\delta(x, y) \leq \alpha$. For all $\alpha \geq \mathrm{diam}_\delta(S)$, the threshold-based graph $\Gamma_\delta(\alpha)$ coincides with the complete graph K_S defined on S. Instead of $\Gamma_\delta\big(\mu_\delta(A)\big)$, we will write $\Gamma_\delta(A)$, or simply $\Gamma(A)$ if there is no ambiguity on the choice of δ.

The following proposition characterizes the $\mathbf{D}$-convex subsets as a type of connected components of $\Gamma(A) = \Gamma_\delta(A)$.

Proposition 14. Let δ be a dissimilarity on S. A nonempty subset $A \subseteq S$ is $\mathbf{D}$-convex if and only if the following statements hold.

- (i) A is a connected component of $\Gamma(A)$.
- (ii) If A is not a clique in $\Gamma(A)$, then $\delta(a, b) < \delta(a, x)$ for all triple $(a, b, x) \in C \times C \times (S \smallsetminus C)$ such that $\delta(a, b) > \mu(A)$.

Proof. Assume first that A is $\mathbf{D}$-convex. Suppose that A is not a connected component of $\Gamma(A)$. By definition of $\mu(A)$, the subset A is connected in the graph $\Gamma(A)$, so that A must be not maximally connected in $\Gamma(A)$, or equivalently there must exist $x \notin A$ and $a \in A$ such that $\delta(x, a) \leq \mu(A)$. Now, suppose that $\delta(a, b) < \mu(A)$ for all $b \in A \smallsetminus a$. Then, we deduce that A is a connected subset of the graph $\Gamma(\alpha)$ for $\alpha = \max\limits_{b \in A \smallsetminus a} \delta(a, b)$. Since $\alpha < \mu(A)$, this is contradictory. Therefore, there exists $c \in A \smallsetminus a$ such that $\delta(a, c) \geq \mu(A)$. Thus $\delta(x, a) \leq \delta(a, c)$, and consequently $x \in \mathbf{D}(a, c)$. We deduce that $x \in A$ since A is $\mathbf{D}$-convex, which is contradictory. It results that A is a connected component of $\Gamma(A)$, which proves property (i). Property (ii) holds also by Proposition 12-(iii).

Conversely, assume that properties (i) and (ii) both hold. Then, let a and b be two elements of A which, by (i), is a connected component of $\Gamma(A)$.
Let us first examine the case where A is a clique. In this case, $\delta(a,b) \leq \mu(A)$. Consider an element x of $\mathbf{D}(a,b) = B(a,\delta(a,b)) \cup B(b,\delta(b,a))$ and assume, without loss of generality, that $x \in B(a,\delta(a,b))$. Then, $\delta(x,a) \leq \delta(a,b) \leq \mu(A)$. Thus x and a belong to the same connected component of $\Gamma(A)$ which must be equal to A. Consequently, $x \in A$, which proves that A is $\mathbf{D}$-convex.
Consider now the case where A is not a clique, and let $(a,b,x) \in A \times A \times (S \setminus A)$.
Assume that $\delta(a,b) \leq \mu(A)$. If $\delta(a,x) \leq \delta(a,b)$, then $\delta(a,x) \leq \mu(A)$, which implies that $x \in A$ since, by (i), A is the connected component of $\Gamma(A)$ that contains a and b. This contradicts the fact that $(a,b,x) \in A \times A \times (S \setminus A)$. Therefore $\delta(a,b) < \delta(a,x)$.
Assume that $\delta(a,b) > \mu(A)$. Applying (ii), we deduce that $\delta(a,b) < \delta(a,x)$.
We conclude that if A is not a clique, then $\delta(a,b) < \delta(a,x)$ for all $(a,b,x) \in A \times A \times (S \setminus A)$. This implies that A is $\mathbf{D}$-convex by Proposition 12-(iii). Finally, it results that if (i) and (ii) are valid, then A is $\mathbf{D}$-convex. □

Propositions 15 and 16 below refer to the notion of an indexed multilevel clustering. An *indexed multilevel clustering* is a pair $(\mathscr{C}, f)$ where $\mathscr{C}$ is a multilevel clustering, and f a map from $\mathscr{C}$ into $\mathbb{R}^+$ such that $f(X) = 0$ when X is minimal in $\mathscr{C}$, and $f(A) < f(B)$ when $A \subset B$. The map $\rho : S \times S \mapsto \mathbb{R}^+$ defined, for all $x, y \in S$, by:

$$\rho(x,y) = \min\{f(A) \mid x, y \in A \in \mathscr{C}\}, \tag{3}$$

is a dissimilarity called the *dissimilarity induced by* $(\mathscr{C}, f)$, or the *cophenetic dissimilarity of* $(\mathscr{C}, f)$. In the sequel, we denote by Φ the map that associates each indexed multilevel clustering $(\mathscr{C}, f)$ with its induced dissimilarity ρ. It is known that the dissimilarity induced by an indexed hierarchy is ultrametric. A dissimilarity δ is an *ultrametric* on S if

$$\delta(x,y) \leq \max\{\delta(x,z), \delta(z,y)\}, \tag{4}$$

for all $x, y, z \in S$. Inequality (4) is sometimes called the *strong triangle inequality*.

Proposition 15. If δ is an ultrametric on S, then $\Phi(\mathit{Balls}(S,\delta), \mathrm{diam}_\delta) = \delta$.

PROOF. Assume that δ is an ultrametric on S and denote by $\mathscr{B}_\delta$ the set of its 2-balls. By Theorem 2 in [16], the indexed multilevel clustering $(\mathscr{B}_\delta, \mathrm{diam}_\delta)$ induces δ. Then, it is sufficient to prove that $\mathscr{B}_\delta = \mathit{Balls}(S,\delta)$. Now, in an ultrametric space, every point of a ball is a center of this ball (cf. for example [18]). Consequently, we deduce that $B(a,\delta(a,b)) = B(b,\delta(a,b))$, and thus $\mathbf{B}(a,b) = B(a,\delta(a,b)) = B(b,\delta(a,b))$. We deduce that $\mathscr{B}_\delta = \mathit{Balls}(S,\delta)$ and finally it results that the indexed multilevel clustering $(\mathit{Balls}(S,\delta), \mathrm{diam}_\delta)$ induces δ, which completes the proof. □

Proposition 16. Let δ be a dissimilarity on S, and ρ the cophenetic dissimilarity of the indexed Apresjan hierarchy $(\mathcal{H}_\mathcal{A}(\delta), \mathrm{diam}_\delta)$. If δ is an ultrametric, then $\rho = \delta$.

PROOF. Let δ be an ultrametric. On the one hand, we have $\Phi(\mathcal{H}_\mathcal{A}(\delta), \mathrm{diam}_\delta) = \rho$ by hypothesis and, on the other hand, we have $\Phi(\mathit{Balls}(X,\delta), \mathrm{diam}_\delta) = \delta$ from Proposition 15. Therefore, it is sufficient to prove that $\mathcal{H}_{\mathcal{A}}(\delta) = Balls(X,\delta)$. We first prove that $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq Balls(X,\delta)$. Let x_0 be an element of an arbitrary cluster of Apresjan, say C. From Proposition 8, there exists $y_0 \in E_C(x_0)$ such that $C = g_B(x_0,y_0)$ since $C \in \mathcal{H}_{\mathcal{A}}(\delta) = \mathrm{conv}(J_{g_B})$ and g_B satisfies (W′) and thus (H) also. Therefore $C = B(x_0,\delta(x_0,y_0))$, which proves that $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq Balls(X,\delta)$. Conversely, let us prove that $Balls(X,\delta) \subseteq \mathcal{H}_{\mathcal{A}}(\delta)$. Consider an arbitrary closed ball $B(a,\lambda) = B_\delta(a,\lambda)$, with $a \in S$ and λ a nonnegative real number. We are then left to prove that $B(a,\lambda)$ is an Apresjan cluster of δ. Consider $x,y \in B(a,\lambda)$ and $z \notin B(a,\lambda)$. Since δ is an ultrametric, elements x and y are centers of $B(a,\lambda)$. Therefore $\delta(x,y) \le \lambda < \delta(x,z)$ and $\lambda < \delta(y,z)$, since $z \notin B(a,\lambda)$. Thus we deduce that $\delta(x,y) < \min\limits_{u\notin C}\{\delta(x,u),\delta(y,u)\}$, i.e. $B(a,\lambda)$ is an Apresjan cluster of δ, as required. □

4.2. Bandelt and Dress weak hierarchies

Bandelt and Dress [3] and Bandelt [2], use a criterion weaker than (2), to define the notion of weak cluster. A *weak cluster* is a nonempty subset C of S such that:

$$\text{For all } x,y \in C, \quad \delta(x,y) < \min_{z\notin C}\left(\max\{\delta(x,z),\delta(y,z)\}\right). \tag{5}$$

In the sequel, a nonempty subset is called a *Bandelt and Dress cluster of* δ, if it satisfies (5). By showing that the $\mathbf{B}$-convex subsets coincide with the clusters of Bandelt and Dress, the next proposition 17 determines the semantics of $\mathbf{B}$-convex subsets. The proof of this property is left to the reader, for it is straightforward and very similar to the proof of Proposition 12.

Proposition 17. A nonempty subset of S is a Bandelt and Dress cluster of a dissimilarity δ if and only if it is $\mathbf{B}$-convex.

From this proposition, it results that the collection of Bandelt and Dress clusters of δ is a weak hierarchy [3, 2, 16]. In the sequel, this weak hierarchy is called the *Bandelt and Dress weak hierarchy of* δ, and denoted by $\mathcal{W}_{\mathcal{BD}}(\delta)$.

Proposition 18. For all dissimilarity δ on S, we have

- (i) $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq \mathcal{W}_{\mathcal{BD}}(\delta)$,
- (ii) If $C \in \mathcal{W}_{\mathcal{BD}}(\delta)$, then $C = \mathbf{B}(a,b)$ for all $a,b \in C$ such that $\mathrm{diam}_\delta(C) = \delta(a,b)$.

PROOF. (i). Since $\mathbf{B}(x,y) \subseteq \mathbf{D}(x,y)$ for all $x,y \in S$, we have $\mathbf{B} \le \mathbf{D}$. Then $\mathrm{conv}(\mathbf{D}) \subseteq \mathrm{conv}(\mathbf{B})$ by Proposition 9 (i). Thus, by Propositions 12 and 17, the Apresjan hierarchy is included in the Bandelt and Dress weak hierarchy.
(ii). Let a,b such that $\mathrm{diam}_\delta(C) = \delta(a,b)$. Therefore, $\max\{\delta(x,a),\delta(x,b)\} \le \delta(a,b)$ for all $x \in C$, which implies that $C \subseteq \mathbf{B}(a,b)$. Conversely, let $x \in \mathbf{B}(a,b)$. Since C is $\mathbf{B}$-convex and $a,b \in C$, we deduce that $x \in C$, and consequently $\mathbf{B}(a,b) \subseteq C$, as required. □

Remark 4. From Proposition 18, it results that the number of clusters in the Bandelt and Dress weak hierarchy of δ is bounded by $n(n-1)/2$. This bound is tight in the sense that there exists dissimilarities δ such that $|\mathcal{W}_{\mathcal{BD}}(\delta)| = n(n-1)/2$, as is the case for the euclidean distance on $S = \{1,\dots,n\}$.

	b	c	d	e
a	7	4	4	8
b	–	4	4	8
c	–	–	6	2
d	–	–	–	2

(a) values of δ

	b	c	d	e
a	$abcd$	ac	ad	S
b	–	bc	bd	S
c	–	–	S	ce
d	–	–	–	ed

(b) Function $\mathbf{B}_\delta$

Figure 2: (a) Values of a dissimilarity δ defined on $S = abcde$. (b) Values of $\mathbf{B}_\delta$ which are derived from those of δ. It is easily checked that $\mathcal{H}_\mathcal{A}(\delta) = \mathcal{T}(S)$ where $\mathcal{T}(S) = \{a, b, c, d, e, S\}$ is the trivial hierarchy on S. Moreover, $\mathcal{W}_{\mathcal{BD}}(\delta) = \mathrm{conv}(\mathbf{B}_\delta) = \mathcal{T}(S) \cup \{ac, ad, bc, bd, ce, de\}$. Furthermore, $\mathbf{B}(a, b) = abcd$ is such that $\mathrm{diam}_\delta(\mathbf{B}(a, b)) = 7 = \delta(a, b)$ but $\mathbf{B}(a, b)$ is not $\mathbf{B}$-convex since $\mathbf{B}(c, d)$ contains $e \notin \mathbf{B}(a, b)$.

Figure 2 shows that the inclusion of the Apresjan hierarchy into the Bandelt and Dress weak hierarchy may be strict. This figure shows also that the converse of Proposition 18-(ii) does not hold.

A direct consequence of the next proposition is that the Apresjan hierarchy coincides with the Bandelt and Dress weak hierarchy, when they are constructed from the same ultrametric dissimilarity.

Proposition 19. Let δ be a dissimilarity on S. Then the following statements are equivalent.

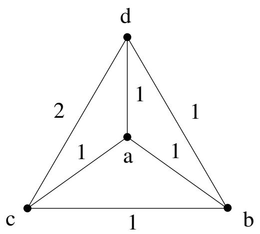
(i) δ is ultrametic,

(ii) $B_\delta(x, \delta(x, y)) = B_\delta(y, \delta(y, x))$,

(iii) $\mathbf{D}_\delta = \mathbf{B}_\delta$.

PROOF. (i) $\Longleftrightarrow$ (ii) holds true by [16, Proposition 3]. Moreover (ii) $\Longleftrightarrow$ (iii) results directly from the equivalence $A \cup B = A \cap B \Longleftrightarrow A = B$. $\square$

If δ is ultrametric then $\mathcal{H}_\mathcal{A}(\delta) = \mathcal{W}_{\mathcal{BD}}(\delta)$, since in this case $\mathbf{D}_\delta = \mathbf{B}_\delta$ by Proposition 19. However, the converse does not hold in general, as shown by Figure 3.



(a) values of δ

	a	b	c	d
a	a	S	S	S
b	–	b	S	S
c	–	–	c	S
d	–	–	–	d

(b) Function $\mathbf{D}_\delta$

	a	b	c	d
a	a	S	abc	abd
b	–	b	abc	abd
c	–	–	c	S
d	–	–	–	d

(c) Function $\mathbf{B}_\delta$

Figure 3: (a) Values of a dissimilarity δ defined on $S = abcd$. (b) and (c) Values of functions $\mathbf{D}_\delta$ and $\mathbf{B}_\delta$, respectively. It is easily checked that $\mathcal{H}_\mathcal{A}(\delta)$ and $\mathcal{W}_{\mathcal{BD}}(\delta)$ both coincide with the trivial hierarchy $\mathcal{T}(S) = \{a, b, c, d, S\}$. However, δ is clearly not ultrametric since, for example, we have $\delta(c, d) = 2 > 1 = \delta(a, c) = \delta(a, d) = 1$.

5. A sequence of path-based dissimilarities decreasing from any dissimilarity to its subdominant ultrametric

Let us consider an arbitrary dissimilarity δ on S, and its associated complete undirected graph K_S defined by weighting each edge uv (with $u \neq v$) by $\delta(u,v)$. Let P be any simple (i.e. without loops) path of graph K_S, and denote by $\mathrm{val}_\delta(P)$ the *maximum δ-jump within* P, that is the *maximum edge-weight of* P. More precisely, if $P = u_0 u_1 \dots u_m$, we denote

$$\mathrm{val}_\delta P = \begin{cases} \max\limits_{1 \le i \le m} \delta(u_{i-1}, u_i), \text{ if } u_0 \neq u_m, \\ 0, \text{ otherwise, i.e. if } P \text{ is reduced to one vertex } u_0 = u_m. \end{cases}$$

Let $\mathcal{P}_{x-y}$ be the set of simple paths of K_S that join vertex x to vertex y. It is well known that the map $(x,y) \in S \times S \mapsto \min\limits_{P \in \mathcal{P}_{x-y}} \mathrm{val}_\delta(P)$ coincides with the subdominant ultrametric of δ: see for example [5, 4]. In order to improve the path-based clustering method which was proposed by Fischer et al. [21, 20], Xu *et al.* and Yu *et al.* have recently introduced a family of path-based dissimilarities that contains the subdominant ultrametric of δ: see [28, 30, 29]. Let us now recall the definition of these path-based dissimilarities hereafter denoted by δ_ℓ with $\ell \ge 1$.

Definition 5. *Let ℓ be any integer greater than or equal to 1. We denote by $\mathcal{P}^{(\ell)}$ the set of simple paths of length at most ℓ in K_S. For all elements x, y of S, let $\mathcal{P}_x^{(\ell)}$ be the subset of $\mathcal{P}^{(\ell)}$ whose first vertex is x, and $\mathcal{P}_{x-y}^{(\ell)}$ be the subset of all $x - y$* paths *in $\mathcal{P}^{(\ell)}$, i.e. the set of paths of length at most ℓ in graph K_S, and whose first and last vertices are respectively x and y. Note that $\mathcal{P}_{x-y}^{(n-1)} = \mathcal{P}_{x-y}^{(n)} = \dots = \mathcal{P}_{x-y}^{(m)} = \mathcal{P}_{x-y}$ for all $m \ge n$. We will denote by δ_ℓ the map that associates each $(x,y) \in S \times S$ to the* smallest maximum δ-jump *within the paths in $\mathcal{P}_{x-y}^{(\ell)}$. Formally, for all $x, y \in S$,*

$$\delta_\ell(x,y) = \min_{P \in \mathcal{P}_{x-y}^{(\ell)}} \mathrm{val}_\delta(P).$$

Remark 5. It may be observed that δ_ℓ is well defined on $S \times S$, since $\mathcal{P}_{x-y}^{(\ell)} \neq \varnothing$ for all $x, y \in S$ and all $\ell \ge 1$. Notice also that $\delta_1 = \delta$ and that δ_ℓ is clearly a dissimilarity on S for all $\ell \ge 1$. Furthermore, Xu *et al.* and Yu *et al.* employ the intuitive term *"the transitive distance with order* ℓ" to designate δ_ℓ (cf. [28, 30, 29]), although δ_ℓ does not necessarily satisfy the metric inequality.

The rest of this section is devoted to the properties of the so-called transitive distances with order $\ell \ge 1$. In what follows, we will write simply $\mathrm{val}\, P$ instead of $\mathrm{val}_\delta P$ if the choice of δ is clear from the context.

Proposition 20. Let δ be a dissimilarity on S, and ℓ_1, ℓ_2 be strictly positive integers. The following assertions hold.

(i) If $\ell_1 \le \ell_2$, then $\delta_{\ell_1} \ge \delta_{\ell_2}$, where $\ge$ designates the point-wise order defined on dissimilarities.

(ii) If $\ell_1, \ell_2 \ge n-1$, then $\delta_{\ell_1} = \delta_{\ell_2} = \delta_{n-1}$.

(iii) δ_{n-1} is the subdominant ultrametric of δ

PROOF. (i). Let $\ell_1 \leq \ell_2$ and $x, y \in S$. Then $\mathcal{P}_{x-y}^{(l_1)} \subseteq \mathcal{P}_{x-y}^{(l_2)}$, and thus $\delta_{\ell_1}(x,y) \geq \delta_{\ell_2}(x,y)$.

(ii). When $\ell \geq n-1$, then $\mathcal{P}^{(\ell)}$ is clearly the set of all simple paths in K_S. Thus $\delta_{\ell_1} = \delta_{\ell_2} = \delta_{n-1}$.

(iii). It is easily checked that δ_{n-1} is an ultrametric. By (i), δ_{n-1} is lower than $\delta_1 = \delta$. Indeed, it is well-known that δ_{n-1} is the subdominant ultrametric of δ: see e.g. [5, 4]. □

Definition 6. *Let x and y be two arbitrary elements of S, and Q a simple path of length m and from x to y. The edge of Q of which one extremity is y, is called the extremal edge of Q before y. Let $Q = u_0u_1\ldots u_m$ with $u_0 = x$ and $u_m = y$, and let $z \neq u_i$ for all $i \in [1,m]$. Then, we define the paths $Q \smallsetminus y, zQ$ and Qz as follows:*

$$Q \smallsetminus y = u_0u_1\ldots u_{m-1}, \quad zQ = zu_0u_1\ldots u_m \ \text{ and } \ Qz = u_0u_1\ldots u_m z.$$

*Moreover, we will say that a path Q from x to y, is ℓ-*optimal *if $Q \in \arg\min_{P \in \mathcal{P}_{x-y}^{(\ell)}} \operatorname{val} P$.*

The following proposition is concerned with the complexity of computing the dissimilarity δ_ℓ (with $\ell \in [2, n-1]$). By induction on ℓ, property (i) enables to compute the values of $\delta_\ell(x,y)$ in a linear time w.r.t the size of S, for each $x, y \in S$ and $\ell \in [2, n-1]$. Properties (ii) and (iii) provide conditions for further reduction of this complexity.

Proposition 21. Let $x, y \in S$ and an integer $\ell \in [2, n-1]$. The following properties hold.

(i) $\delta_\ell(x,y) = \min_{z \in S} \max\{\delta_{\ell-1}(x,z), \delta(z,y)\}$.

(ii) If $\delta_{\ell-1}(x,y) = \delta_\ell(x,y)$ for all $y \in S$, then:

$$\text{for all } y \in S, \qquad \delta_{\ell-1}(x,y) = \delta_\ell(x,y) = \delta_{\ell+1}(x,y) = \ldots = \delta_{n-1}(x,y).$$

(iii) If $\delta_{\ell-1} = \delta_\ell$, then $\delta_{\ell-1} = \delta_\ell = \delta_{\ell+1} = \ldots = \delta_{n-1}$.

PROOF. Let $x, y \in S$ and consider an integer $\ell \in [2, n-1]$.
(i). We aim to compute $\delta_\ell(x,y)$ from the dissimilarity $\delta_{\ell-1}$. Recall that, by definition:

$$\delta_\ell(x,y) = \min_{P \in \mathcal{P}_{x-y}^{(\ell)}} \operatorname{val} P = \min_{P \in \mathcal{P}_{x-y}^{(\ell)}} \max_{uv \in P} \delta(u,v).$$

Let us first examine the case where there exists some ℓ-optimal path from x to y, whose length is less than or equal to $\ell - 1$. Denote by $C_\ell(x,y)$ such a path. Then, $C_\ell(x,y) \in \mathcal{P}_{x-y}^{(\ell-1)} \subseteq \mathcal{P}_{x-y}^{(\ell)}$ and, since $\delta_\ell \leq \delta_{\ell-1}$ by Proposition 20 (ii), it results that

$$\delta_\ell(x,y) = \operatorname{val}(C_\ell(x,y)) \geq \min_{P \in \mathcal{P}_{x-y}^{(\ell-1)}} \operatorname{val}(P) = \delta_{\ell-1}(x,y) \ \text{ and } \ \delta_\ell(x,y) \leq \delta_{\ell-1}(x,y).$$

This implies that

$$\delta_\ell(x,y) = \delta_{\ell-1}(x,y) = \max\{\delta_{\ell-1}(x,y), \delta(y,y)\}. \tag{6}$$

Now, suppose that there exists $z_0 \in S$ that satisfies the following inequality (7) defined by

$$\max\{\delta_{\ell-1}(x,z_0), \delta(z_0,y)\} \lneqq \max\{\delta_{\ell-1}(x,y), \delta(y,y)\} = \delta_\ell(x,y). \tag{7}$$

Denote by $C_{\ell-1}(x,z_0)$ an $(\ell-1)$-optimal path from x to z_0. If y is a vertex of $C_{\ell-1}(x,z_0)$, then let P be the sub-path of $C_{\ell-1}(x,z_0)$ from x to y, and let $P = C_{\ell-1}(x,z_0)y$ otherwise. In either cases, we have $P \in \mathcal{P}^{(\ell)}_{x-y}$. Moreover, we deduce, from (7), that P satisfies $\operatorname{val} P < \delta_\ell(x,y)$. This contradicts the definition of $\delta_\ell(x,y)$. Therefore, no element $z_0 \in S$ satisfies (7), and consequently, for all $z \in S$, we have

$$\max\{\delta_{\ell-1}(x,z),\delta(z,y)\} \geq \delta_\ell(x,y) = \max\{\delta_{\ell-1}(x,y),\delta(y,y)\}.$$

It results that $\delta_\ell(x,y) = \min_{z\in S}\max\{\delta_{\ell-1}(x,z),\delta(z,y)\}$, as required.

We are now left to examine the case where each ℓ-optimal path from x to y is of length equal to ℓ. Let $C_\ell(x,y)$ be an arbitrary such ℓ-optimal path, and z_1 the vertex linked to y in the extremal edge of $C_\ell(x,y)$ before y. Note that $z_1 \neq x$ since $\ell \geq 2$. Let $Q = C_\ell(x,y) \smallsetminus y$. Clearly, we have $Q \in \mathcal{P}^{(\ell-1)}_{x-z_1}$, and moreover

$$\operatorname{val} C_\ell(x,y) = \max(\operatorname{val} Q, \delta(z_1,y)) \geq \max\{\delta_{\ell-1}(x,z_1),\delta(z_1,y)\}.$$

Suppose now that this inequality is strict, i.e.

$$\operatorname{val} C_\ell(x,y) > \max\{\delta_{\ell-1}(x,z_1),\delta(z_1,y)\}. \tag{8}$$

Let $C_{\ell-1}(x,z_1)$ be an $(\ell-1)$-optimal path from x to z_1, and let Q' be the path defined as follows

$$Q' = \begin{cases} \text{the sub-path of } C_{\ell-1}(x,z_1) \text{ from } x \text{ to } y, & \text{if } y \text{ is a vertex of } C_{\ell-1}(x,z_1),\\ C_{\ell-1}(x,z_1)y, & \text{otherwise.}\end{cases}$$

From (8), it results that

$$\operatorname{val} Q' \leq \max(\delta_{\ell-1}(x,z_1),\delta(z_1,y)) < \operatorname{val} C_\ell(x,y) = \delta_\ell(x,y),$$

which is contradictory, for $Q' \in \mathcal{P}^{(\ell)}_{x-y}$. It results:

$$\operatorname{val} C_\ell(x,y) = \max\{\delta_{\ell-1}(x,z_1),\delta(z_1,y)\}. \tag{9}$$

Let z be an arbitrary element of S. First observe that $\delta_\ell(x,y) \leq \max\{\delta_{\ell-1}(x,y),\delta(y,y)\}$ and $\delta_\ell(x,y) \leq \max\{\delta_{\ell-1}(x,x),\delta(x,y)\}$, so that

$$\text{For all } \ z \in \{x,y\}, \quad \delta_\ell(x,y) \leq \max\{\delta_{\ell-1}(x,z),\delta(z,y)\}. \tag{10}$$

Now, suppose that $z \notin \{x,y\}$ and consider an $(\ell-1)$-optimal path from x to z, denoted as $C_{\ell-1}(x,z)$.
If $y \notin C_{\ell-1}(x,z)$, then $R_1 = C_{\ell-1}(x,z)y$ is a (simple) path from x to y, and thus

$$\delta_\ell(x,y) \leq \operatorname{val} R_1 = \max\{\operatorname{val} C_{\ell-1}(x,z),\delta(z,y)\} = \max\{\delta_{\ell-1}(x,z),\delta(z,y)\}. \tag{11}$$

If $y \in C_{\ell-1}(x,z)$, then denote as R_2 the sub-path of $C_{\ell-1}(x,z)$ from x to y. We have

$$\delta_\ell(x,y) \leq \delta_{\ell-1}(x,y) \leq \operatorname{val} R_2 \leq \operatorname{val} C_{\ell-1}(x,z) \leq \max\{\delta_{\ell-1}(x,z),\delta(z,y)\}. \tag{12}$$

From inequalities (10), (11) and (12), it follows that

$$\delta_\ell(x,y) \leq \max\{\delta_{\ell-1}(x,z),\delta(z,y)\},$$

for all $z \in S$. Moreover, using (9), we conclude that $\delta_\ell(x,y) = \min_{z\in S} \max\{\delta_{\ell-1}(x,z), \delta(z,y)\}$, as required.

(ii). We assume that $\delta_\ell(x,y) = \delta_{\ell-1}(x,y)$ for all $y \in S$. We aim to prove that $\delta_{m-1}(x,y) = \delta_m(x,y)$ for all $y \in S$ and all $m \in [\ell, n-1]$. We proceed by induction on $m \in [\ell, n-1]$. By hypothesis, the property is satisfied for $m = \ell$. Assume now that $\delta_{m-1}(x,y) = \delta_m(x,y)$ for all $y \in S$ and some $m \in [\ell, n-1]$. Using property (i), we have :

$$\begin{aligned}
\text{for all } y \in S, \quad \delta_{m+1}(x,y) &= \min_{z\in S} \max\{\delta_m(x,z), \delta(z,y)\} \\
&= \min_{z\in S} \max\{\delta_{m-1}(x,z), \delta(z,y)\}, \text{ by induction hypothesis} \\
&= \delta_m(x,y), \text{ by property } (i),
\end{aligned}$$

which proves (ii).

(iii) is a direct consequence of (ii). □

Notation 7. *Given an arbitrary dissimilarity δ defined on S, we will denote by $r(\delta)$ the minimum value of ℓ such that $\delta_\ell = \delta_{\ell+1}$. Formally $r(\delta) = \min\{\ell \in [1, n-1] \mid \delta_\ell = \delta_{\ell+1}\}$.*

From Proposition 20-(iv), we derive Corollary 22 which relates the ultrametricity of δ to $r(\delta)$.

Corollary 22. A dissimilarity δ is ultrametric iff $r(\delta) = 1$.

Given any dissimilarity δ, we define the function g_ℓ^δ from δ_ℓ as follows.

Definition 8. *For $\ell \in \{1, \dots, n-1\}$, the map $g_\ell^\delta : S \times S \mapsto 2^S$ is defined by:*

$$g_\ell^\delta(x,y) = \bigcup \{P \in \mathcal{P}_x^{(\ell)} \mid \mathrm{val}_\delta(P) \le \delta_\ell(x,y)\}.$$

The map g_ℓ^δ trivially satisfies axiom (C_0), i.e. $x, y \in g_\ell^\delta(x,y)$. Thus, we can consider its two associated interval functions $J_{g_\ell^\delta}$ and $M_{g_\ell^\delta}$ according to the general convex framework introduced in section 3. In the sequel, when the choice of the dissimilarity δ is clear from the context, we will write simply g_ℓ, J_ℓ and M_ℓ instead of g_ℓ^δ, $J_{g_\ell^\delta}$ and $M_{g_\ell^\delta}$, respectively.

Proposition 23. Let δ be an arbitrary dissimilarity and $\ell \in \{1, \dots, n-1\}$. The following properties hold.

(i) For all $x, y \in S$, $g_\ell(x,y) = B_{\delta_\ell}(x, \delta_\ell(x,y))$; in other words $g_\ell = g_{B_{\delta_\ell}}$.

(ii) $g_1 = g_B$, $J_1 = \mathbf{D}_\delta$ and $M_1 = \mathbf{B}_\delta$. More generally, $J_\ell = \mathbf{D}_{\delta_\ell}$ and $M_\ell = \mathbf{B}_{\delta_\ell}$.

(iii) A nonempty subset is J_ℓ-convex iff it is an Apresjan cluster of δ_ℓ.

(iv) A nonempty subset is M_ℓ-convex iff it is a Bandelt and Dress cluster of δ_ℓ.

PROOF. (i). Let δ be an arbitrary dissimilarity defined on S, and x, y be any two elements of S. By definition, the condition $z \in g_\ell(x,y)$ is equivalent to

$$z \in \bigcup \{P \in \mathcal{P}_x^{(\ell)} \mid \mathrm{val}_\delta(P) \leq \delta_\ell(x,y)\}.$$

This condition can be rewritten as:

$$\exists\, P \in \mathcal{P}_{x-z}^{(\ell)} \text{ s.t. } \mathrm{val}(P) \leq \delta_\ell(x,y).$$

Now, this formulation amounts to assert that $\min_{P \in \mathcal{P}_{x-z}^{(\ell)}} [\mathrm{val}(P)] \leq \delta_\ell(x,y)$, or equivalently that we have $\delta_\ell(x,z) \leq \delta_\ell(x,y)$. But this in turn is equivalent to $z \in B_{\delta_\ell}(x, \delta_\ell(x,y))$.
Therefore, for all $x, y \in S$, we have $g_\ell(x,y) = B_{\delta_\ell}(x, \delta_\ell(x,y))$. By definition 3, this amounts to assert that $g_\ell = g_{B_{\delta_\ell}}$.

(ii) This is a direct consequence of property (i) and previous definitions.

(iii) Derives from $J_\ell = J_{g_{B_{\delta_\ell}}} = \mathbf{D}_{\delta_\ell}$, which holds by (ii), and from Proposition 12.

(iv) Derives from $M_\ell = M_{g_{B_{\delta_\ell}}} = \mathbf{B}_{\delta_\ell}$, which holds by (ii), and from Proposition 17. □

Note that (iii) and (iv) of Proposition 23 amount to assert that, for each $\ell \in [1, n-1]$, the interval functions J_ℓ and M_ℓ induce two convexities that are, respectively, the Apresjan hierarchy $\mathcal{H}_\mathcal{A}(\delta_\ell)$ of δ_ℓ and the Bandelt and Dress weak-hierarchy of δ_ℓ. In next sections 6 and 7, we will investigate the two sequences defined by interval convexities $\mathrm{conv}(J_\ell)$ and $\mathrm{conv}(M_\ell)$, when ℓ varies.

6. From the Apresjan hierarchy to the single-link hierarchy

In this section, we investigate the sequence $(\mathrm{conv}(J_\ell))_{1 \leq \ell \leq n-1}$. According to Proposition 23-(iii), it is the sequence of Apresjan hierarchies of δ_ℓ when ℓ varies.

Proposition 24. Let $\ell \in \{1, \ldots, n-1\}$ and $A \subseteq S$. If A is J_ℓ-convex then A is a connected component of the graph $\Gamma(A)$. The converse is true when $\mathrm{Diam}_{\Gamma(A)}(A) \leq \ell$.

PROOF. Let A be a J_ℓ-convex subset of S. By Proposition 23 (iii), the subset A is an Apresjan cluster of δ_ℓ. Therefore, using Proposition 14, we deduce that A is a connected component of $\Gamma_{\delta_\ell}(A)$. Now, it is clear that, for each $\alpha \geq 0$, a subset of S is connected in $\Gamma_{\delta_\ell}(\alpha)$ if and only if it is connected in $\Gamma_\delta(\alpha)$. Thus, we have $\mu_{\delta_\ell}(A) = \mu_\delta(A)$ and A is a connected component of $\Gamma_\delta(A) = \Gamma(A)$.

Assume now that A is a connected component of the graph $\Gamma(A)$ and that $\mathrm{Diam}_{\Gamma(A)}(A) \leq \ell$. Let a and b be two elements of A. Since A is connex and its diameter in graph $\Gamma(A) = \Gamma(\mu(A))$ is not greater than ℓ, it results that there exists at least one path in $\Gamma(A)$ that joins a and b and whose length is not greater than ℓ. In other words, $\delta_\ell(a,b) \leq \mu(A)$. Now, consider an element x of $J_\ell(a,b)$. By proposition 23-(ii), we have $J_\ell(a,b) = B_{\delta_\ell}(a, \delta_\ell(a,b)) \cup B_{\delta_\ell}(b, \delta_\ell(b,a))$. Assume w.l.o.g. that $x \in B_{\delta_\ell}(a, \delta_\ell(a,b))$. Then, $\delta_\ell(x,a) \leq \delta_\ell(a,b) \leq \mu(A)$, which proves that there exists some path between x and a in graph $\Gamma(A) = \Gamma(\mu(A))$. Thus x and a belong to the same connected component of $\Gamma(A)$ which must be equal to A. Therefore, $x \in A$, which proves that A contains $J_\ell(a,b)$, and thus is J_ℓ-convex. □

In the case where $1 \leq \ell < \mathrm{Diam}_{\Gamma(A)}(A)$, the converse of Proposition 24 is not always true. In other words, if $1 \leq \ell < \mathrm{Diam}_{\Gamma(A)}(A)$ and A is a connected component of the graph $\Gamma(A)$, then A may or may not be J_ℓ-convex, as shown by Examples 6 and 7 below.

Example 6. Consider the dissimilarity δ defined in Figure 1, and let $A = abc$. In this case, $\mu(A) = 1$, A is a connected component of $\Gamma(A)$, and its diameter in $\Gamma(A)$ is 2. Now, since $\mathcal{H}_{\mathcal{A}}(\delta) = \{a, b, c, d, abc, S\}$, A is $\mathbf{D}$-convex or, equivalently, J_ℓ-convex for $\ell = 1$. So, in this case, the connected component A is J_ℓ-convex with $\ell < 2 = \mathrm{Diam}_{\Gamma(A)}(A)$.

Example 7. Consider the dissimilarity δ defined in Figure 4 and let $A = abc$. The interval function $\mathbf{D}_\delta$ is given in Table 4.

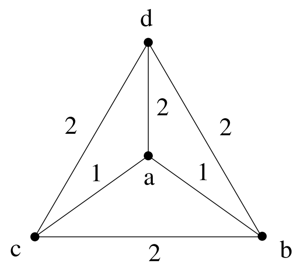


Figure 4: dissimilarity δ

	a	b	c	d
a	a	abc	abc	S
b	–	b	S	S
c	–	–	c	S
d	–	–	–	d

Table 4: interval function $\mathbf{D}_\delta$

It is easily checked that $\mu(A) = 1$, A is a connected component of $\Gamma(A)$ and $\mathrm{Diam}_{\Gamma(A)}(A) = 2$. However, for $\ell = 1$, we have $1 = \ell < \mathrm{Diam}_{\Gamma(A)}(A) = 2$, but the connected component A is not J_ℓ-convex, since $J_1(b, c) = \mathbf{D}_\delta(b, c) = S \notin A$.

Proposition 25. Let δ be a dissimilarity on S and $\ell \in [1, n-1]$. The following statements hold.

(i) g_ℓ satisfies axiom (W$'$).

(ii) g_{n-1} is symmetrical.

Proof. (i). Let $x_1, x_2, x_3 \in S$ and $\ell \in [1, n-1]$. Then there exist i, j, k with $\{i, j, k\} = \{1, 2, 3\}$ such that $\delta_\ell(x_i, x_j) \leq \delta_\ell(x_j, x_k) \leq \delta_\ell(x_i, x_k)$. We deduce the following inequalities.

$$\delta_\ell(x_i, x_j) \leq \delta_\ell(x_i, x_k),\ \delta_\ell(x_k, x_j) \leq \delta_\ell(x_k, x_i) \text{ and } \delta_\ell(x_j, x_i) \leq \delta_\ell(x_j, x_k).$$

As a consequence, we obtain:

$$\begin{aligned}
\{P \in \mathcal{P}^{(\ell)}_{x_i} \mid \mathrm{val}(P) \leq \delta_\ell(x_i, x_j)\} &\subseteq \{P \in \mathcal{P}^{(\ell)}_{x_i} \mid \mathrm{val}(P) \leq \delta_\ell(x_i, x_k)\},\\
\{P \in \mathcal{P}^{(\ell)}_{x_k} \mid \mathrm{val}(P) \leq \delta_\ell(x_k, x_j)\} &\subseteq \{P \in \mathcal{P}^{(\ell)}_{x_k} \mid \mathrm{val}(P) \leq \delta_\ell(x_k, x_i)\},\\
\{P \in \mathcal{P}^{(\ell)}_{x_j} \mid \mathrm{val}(P) \leq \delta_\ell(x_j, x_i)\} &\subseteq \{P \in \mathcal{P}^{(\ell)}_{x_j} \mid \mathrm{val}(P) \leq \delta_\ell(x_j, x_k)\}.
\end{aligned}$$

Therefore $g_\ell(x_i, x_j) \subseteq g_\ell(x_i, x_k)$, $g_\ell(x_k, x_j) \subseteq g_\ell(x_k, x_i)$ and $g_\ell(x_j, x_i) \subseteq g_\ell(x_j, x_k)$, which proves (i).

(ii). As previously noticed, δ_{n-1} is an ultrametric (more precisely, the subdominant ultrametric of δ), thus $\mathbf{D}_{\delta_{n-1}} = \mathbf{B}_{\delta_{n-1}}$ by Proposition 19. Therefore $M_{g_{n-1}} = \mathbf{B}_{\delta_{n-1}} = \mathbf{D}_{\delta_{n-1}} = J_{g_{n-1}}$ by Proposition 23 (i), so that g_{n-1} is symmetrical by Proposition 5. □

By Proposition 9-(i), $\mathrm{conv}(J_{\ell_1}) \subseteq \mathrm{conv}(J_{\ell_2})$ whenever $J_{\ell_2} \preceq J_{\ell_1}$. Thus, it is natural to investigate whether $\ell_1 < \ell_2 \Rightarrow J_{\ell_2} \preceq J_{\ell_1}$, which would provide a direct proof of Theorem 26 (ii). Unfortunately, this implication is generally not true as shown by the counterexample in Figure 5.

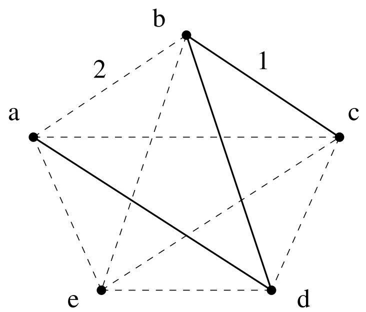


Figure 5: A dissimilarity δ defined as follows: each full line (resp. dashed line) indicates that the value of δ is equal to 1 (resp. 2). We have: $J_1^\delta(b,c) = bcd \subset J_2^\delta(b,c) = abcd$ and $J_1^\delta(c,d) = S \supset J_2^\delta(c,d) = abcd$. Therefore δ is an example of a dissimilarity such that interval functions J_1^δ and J_2^δ are uncomparable according to the order $\preceq$.

While Theorem 26 (ii) is not a direct consequence of previous results, Theorem 26 (i) is an almost immediate consequence of Proposition 24.

Theorem 26. *For all dissimilarity δ on S, the following statements hold.*

(i) *The convexity induced by J_{n-1} is the single-link hierarchy of δ.*
(ii) *For $\ell_1, \ell_2 \in \{1, \dots, n-1\}$, if $\ell_1 \leq \ell_2$ then $\mathrm{conv}(J_{\ell_1}) \subseteq \mathrm{conv}(J_{\ell_2})$.*

Proof. (i). Since the diameter of any vertex subset A in graph $\Gamma(A)$ is less than or equal to $n-1$, this property is a direct consequence of Proposition 24, and the well-known characterization of the single-link hierarchy as the collection of all connected components of all threshold-based graphs $\Gamma[\alpha]$ for $\alpha \geq 0$ (see e.g. [5, 4]).

(ii). Let $\ell \in \{1, \dots, n-1\}$ and A a J_ℓ-convex nonempty subset of S. To prove (ii), it is sufficient to prove that A is $J_{\ell+1}$-convex. According to (iii) in Proposition 12 and characterization (iii) in Proposition 23, we are left to prove that given any $(x,y,z) \in A \times A \times (S \setminus A)$, we have $\delta_{\ell+1}(x,y) < \delta_{\ell+1}(x,z)$. Then, let $(x,y,z) \in A \times A \times (S \setminus A)$. By definition of $\delta_{\ell+1}$, we have:

$$\delta_{\ell+1}(x,z) = \min_{P \in \mathcal{P}_{x-z}^{(\ell+1)}} [\mathrm{val}(P)].$$

Let $Q^\star$ denote an optimal path in $\mathcal{P}_{x-z}^{(\ell+1)}$, i.e. such that $\delta_{\ell+1}(x,z) = \mathrm{val}(Q^\star)$. Let w denote the vertex of $Q^\star$ which is adjacent to the extremity z of $Q^\star$. Suppose that $w = x$. Then, $Q^\star = xz$, so that by Proposition 20-(ii) and the hypothesis that A is J_ℓ-convex, we deduce

$$\delta_{\ell+1}(x,z) = \mathrm{val}(Q^\star) = \delta(x,z) = \delta_1(x,z) \geq \delta_\ell(x,z) > \delta_\ell(x,y) \geq \delta_{\ell+1}(x,y),$$

as required. Thus we are left to consider the case $w \neq x$ in the rest of the proof.

Let us then denote $Q^\star = R^\star z$ where $R^\star \in \mathcal{P}_{x-w}^{(\ell)}$ is the (non empty) path defined as $Q^\star$ deprived from the edge wz. By definition, we have $\mathrm{val}(Q^\star) = \max\{\mathrm{val}(R^\star), \delta(w,z)\}$. Moreover, since

$Q^\star$ minimizes the valuation val among the paths in $\mathcal{P}^{(\ell+1)}_{x-z}$, without loss of generality we may assume that $R^\star$ minimizes the valuation val among the paths in $\mathcal{P}^{(\ell)}_{x-w}$. Thus $\mathrm{val}(R^\star) = \delta_\ell(x, w)$.

Case 1. Assume that $\delta_\ell(x, w) = \mathrm{val}(R^\star) \geq \delta(w, z)$. If $w \in A$, then $(x, w, z) \in A \times A \times (S \smallsetminus A)$, and since A is J_ℓ-convex, we deduce that $\delta_\ell(w, z) > \delta_\ell(w, x)$. By Proposition 20-($ii$), we have $\delta(w, z) = \delta_1(w, z) \geq \delta_\ell(w, z)$. Therefore,

$$\delta(w, z) \geq \delta_\ell(w, z) > \delta_\ell(w, x) = val(R^\star) \geq \delta(w, z),$$

which is contradictory. Therefore $w \notin A$, so that $(y, x, w) \in A \times A \times (S \smallsetminus A)$. Then, using again the fact that A is J_ℓ-convex and Proposition 23-(iii), we obtain:

$$\delta_{\ell+1}(x, y) \leq \delta_\ell(x, y) < \delta_\ell(x, w) = val(R^\star) = val(Q^\star) = \delta_{\ell+1}(x, z).$$

Case 2. Assume now that $\delta_\ell(x, w) = val(R^\star) < \delta(w, z)$. Therefore

$$\delta_{\ell+1}(x, z) = \delta(w, z) = \mathrm{val}(Q^\star). \tag{13}$$

We will distinguish two subcases depending on whether $w \in A$ or $w \notin A$.

Subcase 1: $w \notin A$. From (13), and since $\delta_\ell(x, w) > \delta_\ell(x, y)$ for A is J_ℓ-convex, we deduce:

$$\delta_{\ell+1}(x, z) = \delta(w, z) > val(R^\star) = \delta_\ell(x, w) > \delta_\ell(x, y) \geq \delta_{\ell+1}(x, y).$$

Subcase 2: $w \in A$. Let v be the vertex of $R^\star$ which is adjacent to the extremity x of $R^\star$. Note that $v \neq z$ since otherwise $w = x$, but v may eventually be equal to w.
Let us first prove that $\delta_{\ell+1}(x, z) = \delta_\ell(v, z)$. Recall that $\mathrm{val}(Q^\star) = \delta_{\ell+1}(x, z) = \delta(w, z)$. Then, the path $Q^\star$ deprived from the edge xv, is a path in $\mathcal{P}^{(\ell)}_{v-z}$ whose edge valuations are less than or equal to $\delta(w, z)$. Therefore:

$$\delta_\ell(v, z) \leq \delta(w, z) = \delta_{\ell+1}(x, z).$$

Moreover, xv is an edge of $R^\star$ and, as assumed, $\mathrm{val}(R^\star) < \delta(w, z)$, which implies that $\delta(x, v) < \delta(w, z)$. Suppose that $\delta_\ell(v, z) < \delta(w, z)$. Then, by definition, there would exist some path $T \in \mathcal{P}^{(\ell)}_{v-z}$ such that $\mathrm{val}(T) < \delta(w, z)$. Let us distinguish two alternatives depending on whether x is or is not a vertex in T. If $x \in T$, then denote by U the subpath of T that connects x and z, and let m denote the length of U. Since $U \subseteq T$, we have $m < \ell$, and consequently

$$U \in \mathcal{P}^{(m)}_{x-z} \subseteq \mathcal{P}^{(\ell+1)}_{x-z} \text{ and } \mathrm{val}(U) \leq \mathrm{val}(T) < \delta(w, z) = \delta_{\ell+1}(x, z),$$

which is contradictory. If $x \notin T$, then the path $xT \in \mathcal{P}^{(\ell+1)}_{x-z}$ would satisfy

$$\mathrm{val}(xT) = \max\{\delta(x, v), \mathrm{val}(T)\} < \delta(w, z) = \delta_{\ell+1}(x, z),$$

which is again contradictory. Therefore,

$$\delta_\ell(v, z) = \delta(w, z) = \delta_{\ell+1}(x, z).$$

Once more, let us distinguish two alternatives according to whether $v \in A$ or not. Using the J_ℓ-convexity of A, we deduce the following:

- if $v \in A$, then $\delta_{\ell+1}(x,z) = \delta(w,z) = \delta_\ell(v,z) > \max\{\delta(x,v), \delta_\ell(v,y)\} \geq \delta_{\ell+1}(x,y)$;
- if $v \notin A$, then $\delta_{\ell+1}(x,z) = \delta(w,z) > \mathrm{val}(R^\star) \geq \delta(x,v) \geq \delta_\ell(x,v) > \delta_\ell(x,y) \geq \delta_{\ell+1}(x,y)$.

We conclude that $\delta_{\ell+1}(x,z) > \delta_{\ell+1}(x,y)$, which proves that A is $J_{\ell+1}$-convex, as required. □

Notation 9. *In what follows, $\mathcal{H}_{\mathrm{SL}}(\delta)$ denotes the single-link hierarchy of the dissimilarity δ, and $\mathcal{H}_\ell(\delta)$ the Apresjan hierarchy of δ_ℓ. Therefore $\mathcal{H}_\ell(\delta) = \mathrm{conv}(J_\ell)$ by Proposition 23-(iii). It results that $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_1(\delta)$ and $\mathcal{H}_{\mathrm{SL}}(\delta) = \mathcal{H}_{n-1}(\delta) = \mathrm{conv}(J_{n-1})$. For an arbitrary cluster $C \in \mathcal{H}_{\mathrm{SL}}(\delta)$, we adopt the following notations:*

$$\lambda(C,\delta) = \min\{\ell \,|\, C \in \mathcal{H}_\ell(\delta)\} \quad \textit{and} \quad \overline{\lambda}(\delta) = \max_{X \in \mathcal{H}_{\mathrm{SL}}(\delta)} \lambda(X,\delta).$$

Moreover, $\lambda(C,\delta)$ will be called the chaining-length *of cluster C. If there is no ambiguity on the choice of δ, $\lambda(C,\delta)$ and $\overline{\lambda}(\delta)$ will be simply denoted respectively by $\lambda(C)$ and $\overline{\lambda}$.*

Definition 10. Assume that $\mathscr{C}$ is a multilevel clustering of a type, say T, and $k \geq 2$ is an integer. Then, we will say that sequence $\mathcal{F} = (\mathscr{C}_1, \ldots, \mathscr{C}_j, \ldots, \mathscr{C}_k)$ is a *filtration* of $\mathscr{C}$ if

(i) $\mathscr{C}_1 \subseteq \ldots \subseteq \mathscr{C}_j \subseteq \ldots \subseteq \mathscr{C}_k = \mathscr{C}$,

(ii) For each $1 \leq j \leq k$, the multilevel clustering $\mathscr{C}_j$ is of the given type T.

The proof of the following proposition is straightforward.

Proposition 27. The size of every cluster of $\mathcal{H}_{\mathrm{SL}}(\delta)$ is strictly greater than its chaining–length.

Next corollary results from Theorem 26 together with Notations 9.

Corollary 28. If C is a cluster of the single link hierarchy of δ, then

(i) $1 \leq \lambda(C) \leq \overline{\lambda} \leq r(\delta) \leq n-1$.

(ii) The family $\big(\mathcal{H}_\ell(\delta)\big)_{1 \leq \ell < n}$ is a filtration of the single-link hierarchy of δ. More precisely,

$$\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_1(\delta) \subseteq \ldots \subseteq \mathcal{H}_{\overline{\lambda}-1}(\delta) \subset \mathcal{H}_{\overline{\lambda}}(\delta) = \ldots = \mathcal{H}_{n-1}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta). \tag{14}$$

(iii) $C \in \mathcal{H}_\ell(\delta)$ iff $\lambda(C) \leq \ell$.

Corollary 28-(ii) shows that $\big(\mathcal{H}_\ell(\delta)\big)_{\ell=1,\ldots,\overline{\lambda}(\delta)}$, whose first term is the Apresjan hierarchy of δ, is an increasing sequence of Apresjan hierarchies, with respect to the set inclusion order. The hierarchies $\mathcal{H}_\ell(\delta)$ are then built according to a criterion which is less and less stringent, as ℓ increases. The question of whether this criterion is less and less demanding, as ℓ increases, from the clustering point of view, will be discussed in section 8.

7. From the Apresjan hierarchy to the Bandelt and Dress weak-hierarchy, through a persistent set of single-linkage clusters

According to Proposition 23, each $\mathrm{conv}(M_\ell)$ with $\ell \in \{1, \dots, n-1\}$, coincides with the Bandelt and Dress weak-hierarchy of δ_ℓ. In particular, $\mathrm{conv}(M_1)$ is the Bandelt and Dress weak-hierarchy of $\delta_1 = \delta$. In this section, we are mainly concerned with the study of the interval convexities $\mathrm{conv}(M_\ell)$ for an arbitrary value of $\ell \in [1, n]$.

Given a dissimilarity δ, we first compare the Bandelt and Dress weak hierarchy $\mathcal{W}_{\mathcal{BD}}(\delta) = \mathrm{conv}(M_1)$, the convexity $\mathrm{conv}(M_{n-1})$ induced by the interval function M_{n-1}, the Apresjan hierarchy $\mathcal{H}_{\mathcal{A}}(\delta)$ and the single-link hierarchy $\mathcal{H}_{\mathrm{SL}}(\delta)$.

Proposition 29. For each dissimilarity δ on S, we have:

(i) $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq \mathcal{H}_{\mathrm{SL}}(\delta) \cap \mathcal{W}_{\mathcal{BD}}(\delta)$,
(ii) If δ is ultrametric then $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta) = \mathcal{W}_{\mathcal{BD}}(\delta)$.
(iii) $\mathrm{conv}(M_{n-1}) = \mathcal{H}_{n-1}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta)$.

PROOF. (i). From Proposition 18-(i), we have $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq \mathcal{W}_{\mathcal{BD}}(\delta)$. Moreover, from Theorem 26, $\mathcal{H}_{\mathcal{A}}(\delta) = conv(J_1) \subseteq conv(J_{n-1}) = \mathcal{H}_{\mathrm{SL}}(\delta)$. These two inclusions prove (i).

(ii). Let δ be an ultrametric. From Proposition 19, it results that $\mathrm{conv}(\mathbf{D}_\delta) = \mathrm{conv}(\mathbf{B}_\delta)$. In other words $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{W}_{\mathcal{BD}}(\delta)$, so it is sufficient to prove $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta)$. By Proposition 16, the indexed hierarchy $(\mathcal{H}_{\mathcal{A}}(\delta), \mathrm{diam}_\delta)$ induces the ultrametric δ. Moreover, it is well known that if δ is an ultrametric, then we have:

$$\Phi\big((\mathcal{H}_{\mathrm{SL}}(\delta), f^{\delta}_{\mathrm{SL}})\big) = \delta,$$

where f^{δ}_{SL} denotes the index function of the indexed hierarchy generated by the single-link hierarchical clustering applied to the dissimilarity δ. Therefore, we deduce that

$$\Phi\big((\mathcal{H}_{\mathcal{A}}(\delta), \mathrm{diam}_\delta)\big) = \Phi\big((\mathcal{H}_{\mathrm{SL}}(\delta), f^{\delta}_{\mathrm{SL}})\big).$$

As Φ is bijective, we get $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta)$, as required.

(iii). First recall that δ_{n-1} is ultrametric. Then, applying (ii) with $\delta = \delta_{n-1}$, we deduce that

$$\mathcal{H}_{n-1}(\delta) = \mathcal{H}_{\mathcal{A}}(\delta_{n-1}) = \mathcal{H}_{\mathrm{SL}}(\delta_{n-1}) = \mathcal{W}_{\mathcal{BD}}(\delta_{n-1}) = \mathrm{conv}(M_{n-1}).$$

Now, since δ_{n-1} is the ultrametric subdominant of δ, it is well known that $\mathcal{H}_{\mathrm{SL}}(\delta_{n-1}) = \mathcal{H}_{\mathrm{SL}}(\delta)$, as required. □

In the example of Figure 3, it is easily checked that $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{W}_{\mathcal{BD}}(\delta) = \mathcal{T}(S)$, which shows that the converse of Proposition 29 (ii) does not hold in general.

Remark 8. *Consider the dissimilarity δ given in Figure 2 (a). It is easily checked that the single-link hierarchical clustering applied to δ, generates only one non trivial cluster, in other words $\mathcal{H}_{\mathrm{SL}}(\delta) = \mathcal{T}(S) \cup cde$. Since $\mathcal{W}_{\mathcal{BD}}(\delta) = \mathcal{T}(S) \cup \{ac, ad, bc, bd, ce, de\}$, we deduce that $\mathcal{H}_{SL}(\delta)$ and $\mathcal{W}_{\mathcal{BD}}(\delta)$ are not comparable according to the set-inclusion order. Now,*

$\mathcal{H}_{SL}(\delta) = \mathrm{conv}(M_{n-1})$ and $\mathcal{W}_{\mathcal{BD}}(\delta) = \mathrm{conv}(M_1)$, therefore this example shows that the sequence $\big(\mathrm{conv}(M_\ell)\big)_{\ell\in[1,n-1]}$ is not monotone for an arbitrary dissimilarity δ, and consequently is not a filtration of $\mathcal{W}_{\mathcal{BD}}(\delta) = \mathrm{conv}(M_1)$. Otherwise, filtrations of $\mathcal{W}_{\mathcal{BD}}(\delta)$ can be easily derived from $\big(\mathrm{conv}(M_\ell)\big)_\ell$. Two of them are the sequences $\big(\mathcal{V}_\ell(\delta)\big)_\ell$ and $\big(\mathcal{W}_\ell(\delta)\big)_\ell$ which are defined simply by:

$$\mathcal{V}_\ell(\delta) = \bigcap_{k\leq n-\ell} \mathrm{conv}(M_k) \quad \text{and} \quad \mathcal{W}_\ell(\delta) = \mathrm{conv}(M_1) \cap \bigcup_{k\leq \ell} \mathrm{conv}(M_{n-k}),$$

for all $\ell \in [1, n-1]$. Note that $\mathcal{V}_\ell(\delta) \subseteq \mathcal{V}_{\ell+1}(\delta) \subseteq \mathrm{conv}(M_1) = \mathcal{W}_{\mathcal{BD}}(\delta)$, which shows that $\big(\mathcal{V}_\ell(\delta)\big)_\ell$ is a filtration of $\mathcal{W}_{\mathcal{BD}}(\delta)$. Similarly, it is easily checked that $\big(\mathcal{W}_\ell(\delta)\big)_\ell$ is a filtration of $\mathcal{W}_{\mathcal{BD}}(\delta)$ (see also Corollary 30). Moreover:

$$\bigcap_{k\leq n-\ell} \mathrm{conv}(M_k) \subseteq \mathrm{conv}(M_1) \cap \mathrm{conv}(M_{n-\ell}) \subseteq \mathrm{conv}(M_1) \cap \bigcup_{k\leq \ell} \mathrm{conv}(M_{n-k}),$$

which proves that $\mathcal{V}_\ell(\delta) \subseteq \mathcal{W}_\ell(\delta)$. Note that an element of $\mathcal{V}_\ell(\delta)$ is necessarily M_k-convex for each $k \leq n-\ell$, whereas an element of $\mathcal{W}_\ell(\delta)$ has only to be M_k-convex for at least one value $k \geq n-\ell$. Therefore, the elements of $\mathcal{V}_\ell(\delta)$ satisfy a clustering criterion which is significantly more stringent than the criterion satisfied by the elements of $\mathcal{W}_\ell(\delta)$. It results that, for each ℓ, the weak hierarchy $\mathcal{V}_\ell(\delta)$ is significantly more sparse than the weak hierarchy $\mathcal{W}_\ell(\delta)$. For this reason, we will now focus on filtration $\big(\mathcal{W}_\ell(\delta)\big)_\ell$ of $\mathcal{W}_{\mathcal{BD}}(\delta)$.

Notation 11. *Given any $C \in \mathcal{W}_{\mathcal{BD}}(\delta) = \mathrm{conv}(M_1) = \mathcal{W}_{n-1}(\delta)$, let*

$$\mu(C,\delta) = \min\{\ell \mid C \in \mathcal{W}_\ell(\delta)\} \quad \text{and} \quad \overline{\mu}(\delta) = \max_{C\in\mathcal{W}_{\mathcal{BD}}(\delta)} \mu(C,\delta).$$

If there is no ambiguity on the choice of δ, $\mu(C,\delta)$ and $\overline{\mu}(\delta)$ will be simply denoted respectively by $\mu(C)$ and $\overline{\mu}$.

Corollary 30. Let δ be an arbitrary dissimilarity on S and $k, \ell \in \{1, \dots, n-1\}$.

- (i) For all $C \in \mathcal{W}_{BD}(\delta)$, we have $1 \leq \mu(C) \leq \overline{\mu} \leq n-1$.
- (ii) If $k \leq \ell$, then $\mathcal{H}_k(\delta) \subseteq \mathrm{conv}(M_\ell)$,
- (iii) $\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_1(\delta) \subseteq \mathcal{W}_\ell(\delta)$,
- (iv) The family $\mathcal{F}_M = \big(\mathcal{W}_\ell(\delta)\big)_{1\leq\ell<n}$ is a filtration of $\mathcal{W}_{\mathcal{BD}}(\delta)$.
  More precisely, for all $\ell \in \{1, \dots, \overline{\mu}-1\}$, we have:

$$\mathcal{W}_1(\delta) \subseteq \ldots \subseteq \mathcal{W}_{\overline{\mu}}(\delta) = \ldots = \mathcal{W}_\ell(\delta) = \ldots \subseteq \mathcal{W}_{n-1}(\delta) = \mathcal{W}_{BD}(\delta). \tag{15}$$

PROOF. (i) results directly from the definitions of the compared quantities.

(ii). Let $k, \ell \in \{1, \dots, n-1\}$ with $k \leq \ell$. By definition of J_ℓ and M_ℓ, we have $M_\ell \leq J_\ell$, and thus $\mathcal{H}_\ell = \mathrm{conv}(J_\ell) \subseteq \mathrm{conv}(M_\ell)$. By Theorem 26, $\mathcal{H}_k = \mathrm{conv}(J_k) \subseteq \mathrm{conv}(J_\ell) = \mathcal{H}_\ell$. It results that (ii) is satisfied.

(iii). is a direct consequence of (ii).

(iv). This filtration results immediately from the definition of $\mathcal{W}_\ell(\delta)$ and $\overline{\mu}$. □

Remark 9. *We have $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq \mathcal{W}_1(\delta)$ by Corollary 30-(iii), and from Proposition 29-(iii),* $\mathrm{conv}(M_{n-1}) = \mathcal{H}_{n-1}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta)$ *so that $\mathcal{W}_1(\delta) \subseteq \mathrm{conv}(M_{n-1}) = \mathcal{H}_{\mathrm{SL}}(\delta) = \mathcal{H}_{\overline{\lambda}}(\delta)$. It results that filtration $\big(\mathcal{H}_\ell(\delta)\big)_{1 \le \ell < n}$ introduced in corollary 28, can be compared to filtration $\big(\mathcal{W}_\ell(\delta)\big)_{1 \le \ell < n}$ as follows:*

$$\mathcal{H}_{\mathcal{A}}(\delta) = \mathcal{H}_1(\delta) \subseteq \ldots \subseteq \mathcal{H}_{\overline{\lambda}}(\delta) = \ldots = \mathcal{H}_{n-1}(\delta) = \mathcal{H}_{\mathrm{SL}}(\delta) \tag{14}$$

$$\rotatebox{-45}{$\subseteq$} \qquad \rotatebox{45}{$\supseteq$}$$

$$\mathcal{W}_1(\delta) \subseteq \ldots \subseteq \mathcal{W}_{\overline{\mu}}(\delta) = \ldots = \mathcal{W}_{n-1}(\delta) = \mathcal{W}_{BD}(\delta) \tag{15}$$

This diagram points out that $\mathcal{W}_1(\delta)$ is a hierarchy that is intermediary w.r.t. set inclusion order, between the Apresjan hierarchy of δ and the single-link hierarchy of δ.

8. Discussion

The main contributions of this paper can be summarized as follows. In the line of research carried out by [9, 15], we first proposed new flexible characterizations of hierarchical and weak-hierarchical clustering models as interval convexities. In detail, we considered an arbitrary map $g : S \times S \longrightarrow 2^S$ such that $\{x, y\} \subseteq g(x, y)$ for all $x, y \in S$, and proved that $J_g = g \cup \overline{g}$ (resp. $M_g = g \cap \overline{g}$) is an interval function that induces a hierarchical (resp. weakly hierarchical) clustering, if and only if g satisfies a specific property (Propositions 6 and 7). In particular, if g denotes the two-way Ball-map of an arbitrary dissimilarity δ, it results that J_g (resp. M_g) induces the Apresjan hierarchy (resp. the Bandelt and Dress weak hierarchy) of δ. Then, we focussed on the path-based dissimilarities δ_ℓ, also known as transitive distances with order ℓ [28, 30, 29], such that, for all $x, y \in S$, the value $\delta_\ell(x, y)$ is the smallest maximum δ-jump along all $x-y$ paths of length at most ℓ. First, we determined a recursive method for computing the sequence $(\delta_\ell)_{1 \le \ell \le n-1}$ which turns out to be strictly decreasing for $1 \le \ell \le r(\delta)$ downto to the subdominant ultrametric of δ, and then stationary for $r(\delta) < \ell \le n-1$ (Proposition 21 and notation 7). Finally, defining g_ℓ as the two-way Ball-map of δ_ℓ, we proved that the sequences whose general terms are $\mathcal{H}_\ell(\delta) = \mathrm{conv}(J_{g_\ell})$ and $\mathcal{W}_\ell(\delta) = \mathrm{conv}(M_{g_1}) \cap \Big[\bigcup_{k \le \ell} \mathrm{conv}(M_{g_{n-k}})\Big]$, with $1 \le \ell < n-1$, are respectively a filtration of the single-link hierarchy of δ and a filtration of the Bandelt and Dress weak-hierarchy of δ (Corollaries 28 and 30).
These theoretical results may be viewed as one step toward introducing interval convexity-based tools for achieving and interpreting multilevel clustering based on dissimilarities. We then conclude the paper by pointing out some final remarks and discussing potential extensions of this approach, in particular for data mining practice involving the single-linkage clustering.

Remark 10 (Distortion of δ_ℓ w.r.t. the subdominant ultrametric of δ). If δ is an arbitrary dissimilarity then, by Propositions 20 and 21, we have:

$$\delta = \delta_1 \succ \ldots \succ \delta_\ell \succ \ldots \succ \delta_{r(\delta)} = \ldots = \delta_{n-1}. \tag{16}$$

As a consequence, if δ is not ultrametric, then for all $\ell \in [1, r(\delta)-1]$, the subdominant ultrametric of δ_ℓ is $\delta_{r(\delta)} = \delta_{n-1}$, which shows incidentally that δ_ℓ is not ultrametric. Let $\ell \in [1, r(\delta)-1]$. Denoting by $\| \ \|$ the L_1-norm, and using the fact that $\delta_\ell \succ \delta_{\ell+1} \ge \delta_{n-1} \ge 0$, we have

$$\|\delta_\ell - \delta_{\ell+1}\| = \|\delta_\ell\| - \|\delta_{\ell+1}\| \text{ and } \|\delta_\ell\| > 0.$$

Denoting $\Delta(\delta_\ell) = \dfrac{\|\delta_\ell - \delta_{n-1}\|}{\|\delta_\ell\|}$, which is a measure of the relative distortion between δ_ℓ and its subdominant ultrametric δ_{n-1}, it results that $\Delta(\delta_\ell) = 1 - \dfrac{\|\delta_{n-1}\|}{\|\delta_\ell\|}$. Note that $\Delta(\delta_\ell)$ is strictly decreasing as ℓ increases from 1 to $r(\delta)$. Figure 6 highlights the decrease of $\Delta(\delta_\ell)$ when ℓ increases, in the case of the euclidean distance δ computed for the Fisher's Iris dataset.

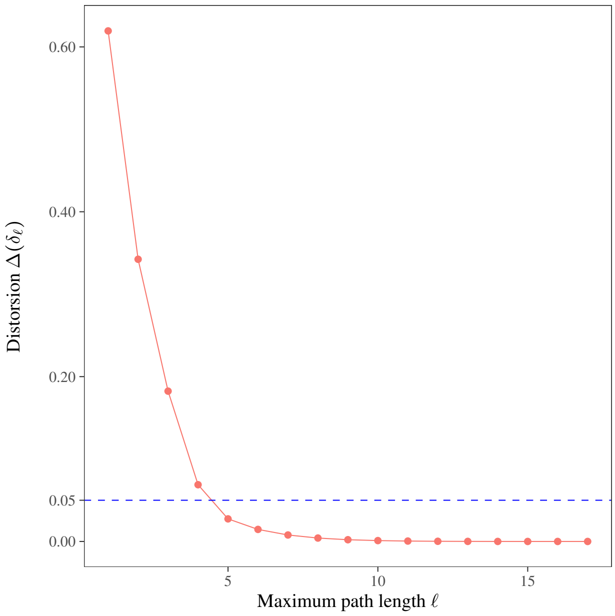


Figure 6: The curve of $\Delta(\delta_\ell)$ as a function of ℓ, with the euclidean distance δ computed on the Iris dataset.

The index $\Delta : \delta \mapsto \Delta(\delta)$ was first introduced by Rammal *et al.* (cf. [26]) in order to measure the degree of ultrametricity of any dissimilarity δ. Other ultrametricity indices have been previously proposed [22, 24, 25, 10]. However, these indices estimate the degree of ultrametricity of a dissimilarity without measuring its distortion w.r.t. its subdominant ultrametric, for example by measuring the proportion of approximately ultrametric triples (cf. [24, 25]).

Remark 11 (Single link dendrogram simplification). Consider the dendrogram $\mathcal{D}_{\mathrm{SL}}(\delta)$ generated by the single-linkage clustering applied to an arbitrary dissimilarity δ, and let f_{SL} be the level function of this dendrogram. In addition, let $\mathcal{D}_\ell(\delta)$ be the dendrogram which represents the hierarchy $\mathcal{H}_\ell(\delta)$ indexed by the restriction of f_{SL} to $\mathcal{H}_\ell(\delta)$. Notice that $\mathcal{D}_{\mathrm{SL}}(\delta) = \mathcal{D}_{n-1}(\delta) = \mathcal{D}_{r(\delta)}(\delta)$. By (14), we have: $\mathcal{H}_{\mathcal{A}}(\delta) \subseteq \ldots \subseteq \mathcal{H}_\ell(\delta) \subseteq \ldots \subseteq \mathcal{H}_{\overline{\lambda}-1}(\delta) \subset \mathcal{H}_{\overline{\lambda}}(\delta) = \ldots = \mathcal{H}_{\mathrm{SL}}(\delta)$. This filtration and the example of the curve of $\Delta(\delta_\ell)$ in Figure 6, suggest that $\mathcal{D}_\ell(\delta)$ is a faithful simplified subdendrogram of $\mathcal{D}_{\mathrm{SL}}(\delta)$ whenever ℓ is large enough to ensure that the discrepancy between $\mathcal{D}_\ell(\delta)$ and $\mathcal{D}_{\mathrm{SL}}(\delta)$ be small enough. Given the relation $\mathcal{H}_\ell(\delta) = \mathrm{conv}(J_{g_\ell}) = \mathcal{H}_{\mathcal{A}}(\delta_\ell)$, we can associate uniquely each dissimilarity δ_ℓ with the dendrogram $\mathcal{D}_\ell(\delta)$ representing $\mathcal{H}_\ell(\delta)$. As it is easy to show examples of dissimilarities δ and path lengths ℓ such that $\mathcal{H}_{\mathcal{A}}(\delta_\ell) = \mathcal{H}_\ell(\delta) = \mathcal{H}_{\ell+1}(\delta) = \mathcal{H}_{\mathcal{A}}(\delta_{\ell+1})$ and $\delta_\ell > \delta_{\ell+1}$, it results that this map $\delta_\ell \mapsto \mathcal{D}_\ell(\delta)$, with $1 \leq \ell \leq r(\delta)$, is clearly non injective. Consequently, the distortion $\Delta(\delta_\ell)$ between δ_ℓ and δ_{n-1} is not appropriate to measure the discrepancy between $\mathcal{D}_\ell(\delta)$ and $\mathcal{D}_{\mathrm{SL}}(\delta)$. An alternative is to consider the bijection that maps each dendrogram to its cophenetic ultrametric. Denoting by ρ_ℓ the cophenetic ultrametric induced by the dendrogram $\mathcal{D}_\ell(\delta)$, the distortion $\Delta(\rho_\ell)$ is then appropriate to estimate the discrepancy between dendrograms $\mathcal{D}_\ell(\delta)$ and $\mathcal{D}_{\mathrm{SL}}(\delta)$. Note that, if $\ell \in [1, n-2]$ then $\mathcal{H}_\ell(\delta) \subseteq \mathcal{H}_{\ell+1}(\delta)$, so that $\Delta(\rho_\ell)$ decreases when ℓ increases. This leads us to propose $\mathcal{D}_\ell(\delta)$ as a consistent simplification of the dendrogram $\mathcal{D}_{\mathrm{SL}}(\delta)$, whenever $\ell \geq \ell_0$ with $\Delta(\rho_{\ell_0}) \leq \alpha$ and α is a user-chosen threshold, e.g. $\alpha = 0.05$. Based on filtration $\left(\mathcal{W}_\ell(\delta)\right)_\ell$ of $\mathcal{W}_{\mathcal{BD}}(\delta)$ and the bijection due to [16], a similar approach applies in order to determine a simplified sub-diagram of the Hasse diagram of $\mathcal{W}_{\mathcal{BD}}(\delta)$ indexed by the diameter function. More precisely, if τ_ℓ is the cophenetic quasi-ultrametric induced by the weak-hierarchy $\mathcal{W}_\ell(\delta)$ indexed by the function diam_δ, then the distortion $\Delta(\tau_\ell)$ can be used to estimate the discrepancy between the Hasse diagrams of $\mathcal{W}_\ell(\delta)$ and of $\mathcal{W}_{\mathcal{BD}}(\delta)$ both indexed by the function diam_δ. Then $\mathcal{W}_\ell(\delta)$ is a simplified Hasse sub-diagram of $\mathcal{W}_{\mathcal{BD}}(\delta)$ provided that $\Delta(\tau_\ell) \leq \alpha$ where α is a user-chosen threshold.

Example 12. The Single-Link dendrogram simplification method (cf. Remark 11), was applied to dendrogram $\mathcal{D}_{\mathrm{SL}}(\delta)$ where δ is the Euclidean distance computed on the Iris dataset.

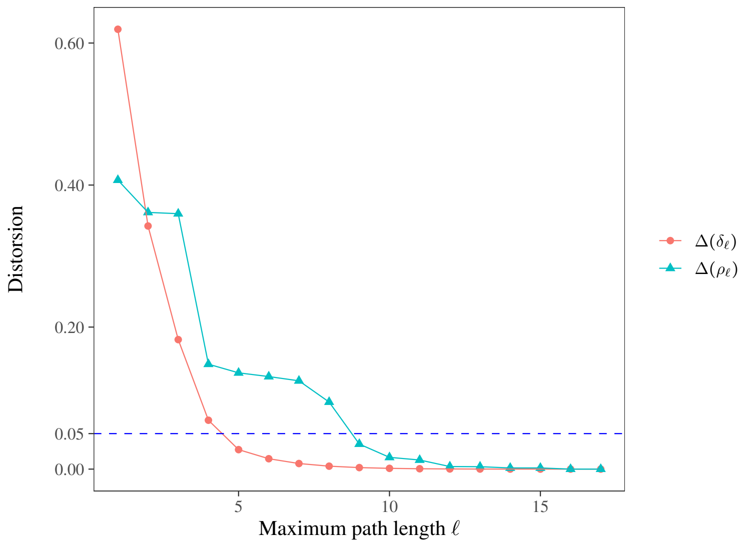
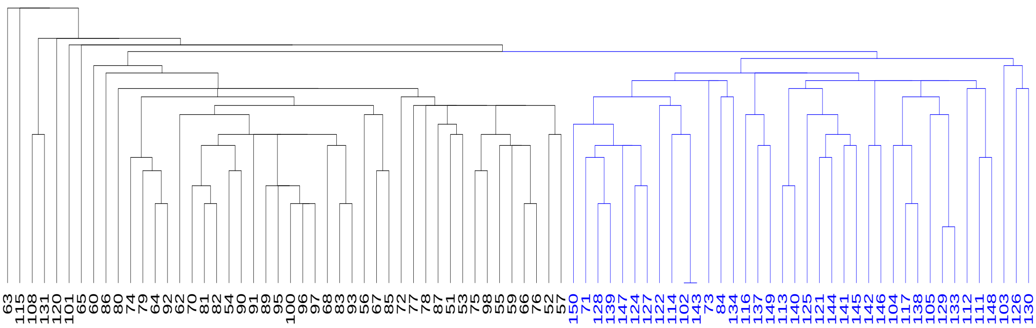


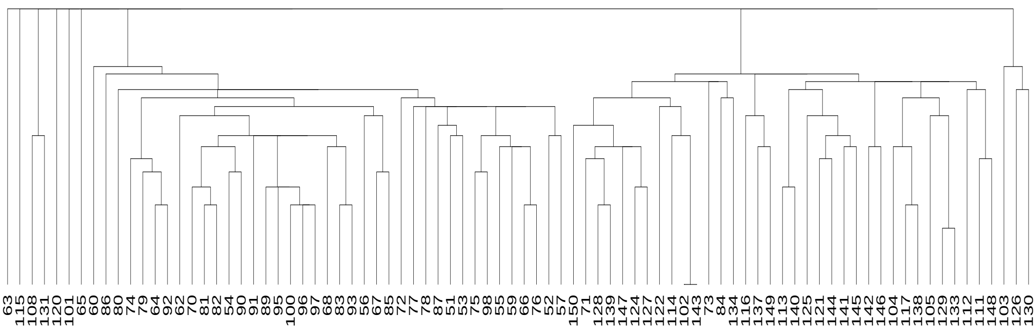
Figure 7: Distortions $\Delta(\rho_\ell)$ and $\Delta(\delta_\ell)$ as functions of ℓ, for the euclidean distance δ on the Iris dataset.

To shorten our notation, we write $\mathcal{D}_{\mathrm{SL}}$ (resp. $\mathcal{D}_\ell$) instead of notations $\mathcal{D}_{\mathrm{SL}}(\delta)$ (resp. $\mathcal{D}_\ell(\delta)$). Figure 7 displays the curves of $\Delta(\rho_\ell)$ and $\Delta(\delta_\ell)$ when ℓ varies. We observe that $\Delta(\rho_{10}) < 0.05$, which indicates that $\mathcal{D}_{10}$ is a faithful simplification of dendrogram $\mathcal{D}_{\mathrm{SL}}$. In order to display clearly the differences between $\mathcal{D}_{\mathrm{SL}}$ and its simplified dendrogram $\mathcal{D}_{10}$, we restrict our comparison between a (non trivial) subdendrogram of $\mathcal{D}_{\mathrm{SL}}$ (cf. Figure 8 (a)) and its corresponding part in $\mathcal{D}_{10}$ (cf. Figure 8 (b)). Figure 8 (a) represents the restriction of $\mathcal{D}_{\mathrm{SL}}$ to one of its cluster, say C, which belongs also to $\mathcal{D}_{10}$. Cluster C consists of all the elements displayed on Figure 8 (a), and is indeed a subset of size 84 containing only Virginica and Versicolor iris samples. Several of the clusters of the subdendrogram of Figure 8 (a) do not belong to $\mathcal{D}_{10}$. For this dataset, these missing clusters are nested, and the smallest of them, say D, is displayed in blue color in Figure

8 (a). Figure 8 (b) represents the restriction of $\mathcal{D}_{10}$ to cluster C which appears, in this case, to be the smallest cluster of $\mathcal{H}_{10}$ that contains subset D.



(a)



(b)

Figure 8: (a) Subdendrogram $\mathcal{D}_{SL}$ restricted to cluster C. The cluster highlighted in blue colour has a chaining-length greater than 10. (b) Subdendrogram $\mathcal{D}_{10}$ restricted to cluster C.

Our approach was applied on a few simulated and real datasets that are either structured into a partition or uniformly distributed. The obtained results lead us to the following comments:

(i) In each case of our experimentations, we observed that the shape of curve $\Delta(\rho_\ell)$ was similar to the curve represented by Figure 7. Notice that this shape of curve implies that clusters with the longest chaining-length are those clusters which contribute significantly the least to the distortion reduction measured by $\Delta(\rho_\ell)$.

(ii) For the Iris dataset, we observe that $\lambda(C) = 10 < \lambda(D) = 11$ whereas $D \subset C$, so that $f_{\mathrm{SL}}(D) < f_{\mathrm{SL}}(C)$. Therefore the order defined by λ is not compatible with the order defined by f_{SL}. For a cluster $X \subset S$, this implies that the information provided by the chaining-length $\lambda(X)$ is different from that provided by $f_{\mathrm{SL}}(X)$ which measures the degree of connectivity of X. Recall that $f_{\mathrm{SL}}(X)$ is the largest weight of an edge connecting two elements of X in a minimum spanning tree weighted with the values taken by δ.

(iii) As noticed by [13], Apresjan clusters of an arbitrary dissimilarity satisfy a stringent clustering criterion and, in general, are small-sized, which is a major limiting factor to their usefulness. However, the Apresjan hierarchy of an ultrametric dissimilarity coincides with its single link hierarchy (cf. Proposition 29), so that it may contain clusters of arbitrary sizes. By an argument of continuity, the Apresjan hierarchy of a dissimilarity δ whose distortion $\Delta(\delta)$ is small, may also contain clusters of arbitrary size. If the distortion $\Delta(\delta)$ is not small, our experimentations have confirmed that Apresjan clusters of such datasets are, in general and as expected, small sized. Moreover, these experimentations indicate that the curves of functions $\ell \mapsto \Delta(\delta_\ell)$ have a similar shape to the one in Figure 7. This shows that the distortion $\Delta(\delta_\ell)$ decreases drastically when ℓ increases in a short interval whose minimal value is 2, so that the ultrametric degree of δ_ℓ, when this degree is defined in an approximate sense, increases significantly as ℓ increases from value 2, which finally implies that the Apresjan hierarchy of δ_ℓ may contain clusters of arbitrary size if $\ell \geq 2$. Note that, since $\mathcal{H}_{\mathcal{A}}(\delta_\ell) \subseteq \mathcal{H}_{\mathcal{A}}(\delta_{\ell+1})$, Apresjan clusters of δ_ℓ satisfy a clustering criterion which is less and less demanding as ℓ increases. More specifically, each Apresjan cluster of δ_ℓ, say A, satisfies by definition the following property (P) related to cluster isolation: For all $x \notin A$ and all $y, z \in A$,

$$\exists\, P_{yz} \in \mathcal{P}^{(\ell)} \text{ s.t. for all } uv \in P_{yz}, \text{we have } \ \delta(u,v) < \delta(x,y),$$

or equivalently, $P_{yz} \in \mathcal{P}^{(\ell)}$ exists such that $\mathrm{val}(P_{yz}) < \delta(x,y)$. Moreover, it can be proved that if $\mathrm{val}(P_{yz}) = \delta_\ell(y,z)$ then the path P_{yz} is included in A. Therefore, property (P) indicates that A is isolated insofar as the dissimilarity between two elements of cluster A can be defined as the valuation of an internal path of A whose length is at most ℓ. In addition, (P) also indicates that cluster A is all the more compact that ℓ is small in comparison with $|A|$. To conclude, on the basis of property (P) completed with the claim that $P_{yz} \subseteq A$ is true if $\mathrm{val}(P_{yz}) = \delta_\ell(y,z)$, and according to the fact that the single-linkage criterion is strengthened when ℓ is small, we can consider as plausibly valid each Apresjan cluster A of δ_ℓ such that ℓ is small and $|A| \gg \ell$, or equivalently each cluster $A \in \mathcal{H}_{SL}(\delta)$ such that its chaining-length $\lambda(A)$ is small and $\lambda(A) \ll |A|$. This assertion outlines guidelines for identifying single-linkage clusters whose size is large enough, and which are both compact and isolated in the sense of a more or less strengthening of the single-linkage criterion. By way of example, there are forty-seven non trivial Apresjan clusters for the Iris flower dataset: forty-one are of size 2, five of size 3 and one of size 4, which is consistent with the assertion that Apresjan clusters of an arbitrary dissimilarity are generally small-sized. Regarding clusters of interest to data analysts, it is well known that applying a standard partitionning method to the Iris dataset leads usually to identify clearly two clusters: the cluster of Setosa iris, say A, and the cluster that gathers Virginica and Versicolor irises, say B. Computations show that A and B are single-linkage clusters such that:

$$\lambda(A) = 2 \ \text{ with } \ |A| = 50 \ \text{ and } \ \lambda(B) = 4 \ \text{ with } \ |B| = 100.$$

With respect to the guidelines roughly defined above for identifying relevant single-linkage clusters, these numerical results and the graph presented in Figure 9 are consistent with the common knowledge that clusters A and B are the most pertinent single-link clusters of the Iris dataset. However, additional theoretical and empirical insights are clearly required to make these guidelines both more precise and reliable.

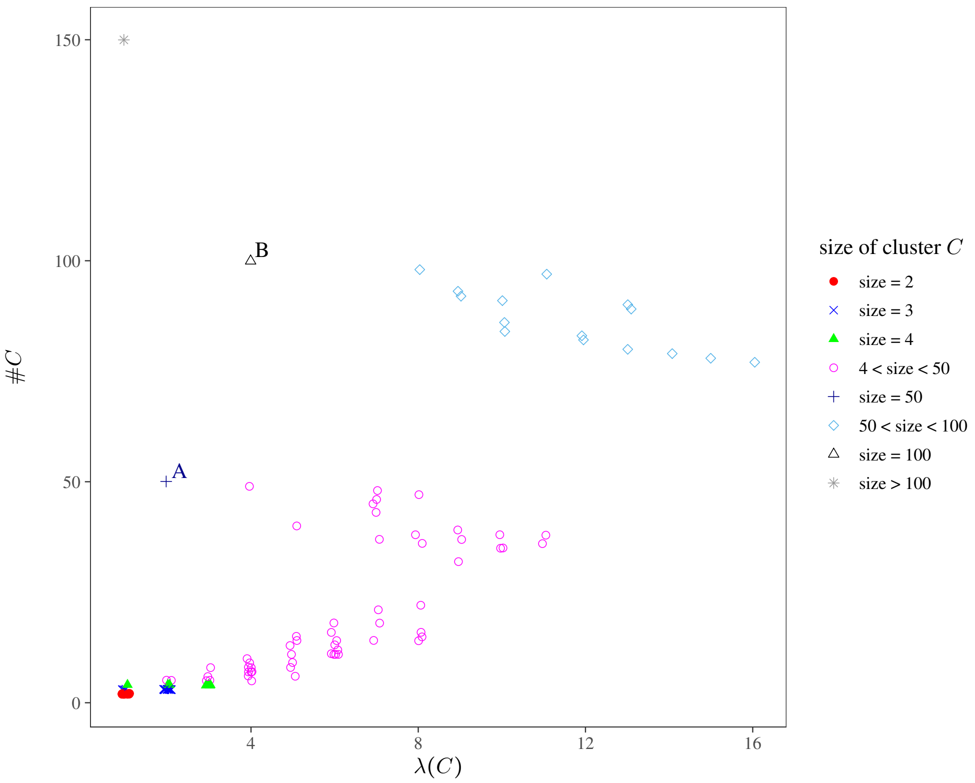


Figure 9: A biplot that displays the size of clusters versus their chaining-length, when the set of clusters is the single-link hierarchy built on the Iris dataset using the euclidean distance.

Remark 13. It may be observed that, for $\ell \in [1, r(\delta) - 1]$, we have

$$\Delta(\delta_\ell) - \Delta(\delta_{\ell+1}) = \left(1 - \frac{\|\delta_{n-1}\|}{\|\delta_\ell\|}\right) - \left(1 - \frac{\|\delta_{n-1}\|}{\|\delta_{\ell+1}\|}\right) = \frac{\|\delta_{n-1}\|}{\|\delta_{\ell+1}\|\|\delta_\ell\|}\left(\|\delta_\ell\| - \|\delta_{\ell+1}\|\right),$$

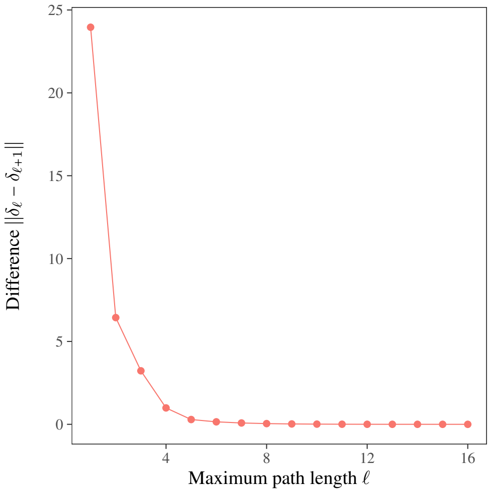
so that $\Delta(\delta_\ell) - \Delta(\delta_{\ell+1}) > 0$ for all $\ell \in [1, r(\delta) - 1]$ since $\delta_\ell > \delta_{\ell+1}$. Figure 7 suggests that the successive differences $\Delta(\delta_\ell) - \Delta(\delta_{\ell+1})$, are decreasing when ℓ increases, whereas the sequence of successive differences of $\Delta(\rho_\ell)$ is not decreasing. Consider now the simpler statement that the successive differences of $\|\delta_\ell\|$ are decreasing when ℓ increases, i.e. that $\left(\|\delta_\ell - \delta_{\ell+1}\|\right)_{\ell=1,\dots,r(\delta)-1}$ is a decreasing sequence. Based on experimentations on simulated dissimilarities, our intuition leads us to view this assertion as a conjecture whose self-contained formulation is as follows.

Conjecture: *Let δ be an arbitrary dissimilarity on S and, for all $\ell \in \mathbb{N}^\star$, denote by δ_ℓ the dissimilarity on S defined recursively by:*

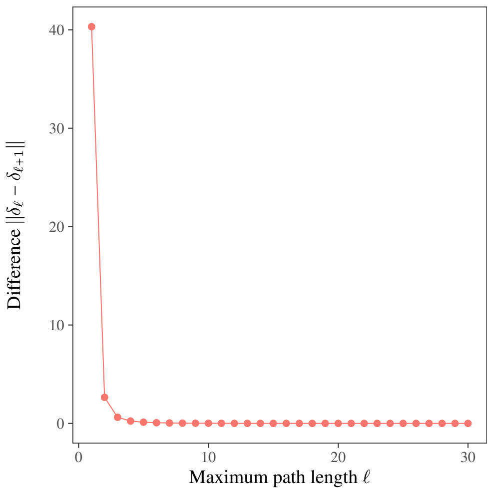
$$\textit{for all } (x,y) \in S^2, \quad \delta_\ell(x,y) = \begin{cases} \delta(x,y), \textit{ if } \ell = 1, \\ \min_{z \in S} \max\{\delta_{\ell-1}(x,z), \delta(z,y)\}, \textit{ otherwise.} \end{cases}$$

For all $\ell \in \mathbb{N}^\star$, we have $\|\delta_\ell - \delta_{\ell+1}\| \geq \|\delta_{\ell+1} - \delta_{\ell+2}\|$, where $\|\ \|$ denotes the L_1-norm.

Recall that $\delta_\ell = \delta_{\ell+1}$ if $\ell \geq r(\delta)$, so that this inequality is obvious for $\ell \geq r(\delta) - 1$, and therefore the conjecture holds true when $r(\delta) \in \{1, 2\}$. Figures 10 (a) and (b) point out that the conjecture is true for two datasets such that $r(\delta) \notin \{1, 2\}$. Note that, in each of these Figures 10 (a) and (b), the decreasing curve of $\|\delta_\ell - \delta_{\ell+1}\|$ as a function of ℓ has an exponential shape.



(a)



(b)

Figure 10: Decrease of $\|\delta_\ell - \delta_{\ell+1}\|$ when ℓ increases, for two datasets: (a) the famous Iris flower dataset, and (b) a simulated dataset S of 300 samples x_i where the values $\delta(x_i, x_{i'})$ are i.i.d. according to the uniform law on $[0, 1]$.

This conjecture suggests a general property of dissimilarities, whose formulation is rather simple, but whose study is not straightforward for $r(\delta) \geq 3$. As a consequence, the conjecture seems beyond the scope of this paper.

While the idea of an unifying framework based on abstract convexity in order to characterize several multilevel clustering models has been discussed before (cf. [9, 15]), the main contribution of this paper lies in the finding of interval convexity-based results relevant to the practice of data mining, and more specifically concerning the single-link hierarchical clustering, Apresjan hierarchies, and Bandelt and Dress weak hierarchies. For future research, we point out the following open issues:

1 – Recall that the sequence $\left(\delta_\ell\right)_{1\leq\ell\leq n-1}$ satisfies the property that $\mathrm{conv}(J_{g_{\ell_1}}) \subseteq \mathrm{conv}(J_{g_{\ell_2}})$ if $\ell_1 \leq \ell_2$. Is it possible de characterize the sequences of arbitrary path-based dissimilarities which satisfy this property?

2 – Another direction for future research would be to extend our approach based on interval convexity, to hierarchical clustering schemes such as the complete-link clustering and the unweighted average linkage clustering.

References

1. J. Apresjan. An algorithm for constructing clusters from a distance matrix. *Mashinnyi perevod: prikladnaja lingvistika*, 9:3–18, 1966.
2. H. J. Bandelt. Four point characterization of the dissimilarity functions obtained from indexed closed weak hierarchies. Technical report, Mathematiche Seminar der Universität, Hamburg, 1992.

3. H.-J. Bandelt and A.W.M. Dress. Weak hierarchies associated with similarity measures: an additive clustering technique. *Bull. Math. Biology*, 51:133–166, 1989.
4. J. M. Bayod and J. Martinez-Maurica. Subdominant ultrametrics. *Proceedings of the American Mathematical Society*, 108(3), 1990.
5. J.-P. Benzécri. *L'Analyse des Données*. Dunod, Paris, 1973.
6. J.P. Benzécri. Description mathématique des classifications. *Revue de Statistique Appliquée*, 20(3):23—56, 1972.
7. P. Bertrand. Set systems for which each set properly intersects at most one other set - Application to cluster analysis. *Discrete Applied Mathematics*, 156(8):1220–1236, 2008.
8. P. Bertrand and F. Brucker. On lower-maximal paired-ultrametrics. In P. Brito, P. Bertrand, G. Cucumel, and F. De Carvalho, editors, *Selected Contributions in Data Analysis and Classification*, pages 455–464. Springer-Verlag, 2007.
9. P. Bertrand and J. Diatta. Multilevel clustering models and interval convexities. *Discrete Applied Mathematics*, 222:54–66, 2017.
10. P. E. Bradley. Finding ultrametricity in data using topology. *Journal of Classification*, 34(1):76–84, 2017.
11. F. Brucker and A. Gély. Crown-free lattices and their related graphs. *Order*, 28(3):443–454, 2011.
12. F. Brucker, P. Préa, and C. Châtel. Totally balanced dissimilarities. *Journal of Classification*, 2019 (accessed March 30, 2019). https://doi.org/10.1007/s00357-019-09320-w.
13. D. Bryant and V. Berry. A family of clustering and tree construction methods. *Advances in Applied Mathematics*, 27(4):705–732, 2001.
14. J. Calder. Some elementary properties of interval convexities. *J. Lond. Math. Soc.*, s2-3(3):422—428, 1971.
15. M. Changat, P.-G. Narasimha-Shenoi, and P.-F. Stadler. Axiomatic characterization of transit functions of weak hierarchies. *The Art of Discrete and Applied Mathematics*, 2:#P1.01, 2019.
16. J. Diatta and B. Fichet. Quasi-ultrametrics and their 2-ball hypergraphs. *Discrete Math.*, 192:87–102, 1998.
17. E. Diday. Orders and overlapping clusters in pyramids. In Jan De Leeuw et al., editor, *Multidimensional Data Analysis*, pages 201–234. DSWO Press, 1986.
18. O. Dovgoshey. Finite ultrametric balls. *p-Adic Numbers, Ultrametric Analysis and Applications*, 11(3):177–191, 2019.
19. C. Durand and B. Fichet. One-to-one correspondence in pyramidal representation: a unified approach. In H. H. Bock, editor, *Classification and Related Methods of Data Analysis*, pages 85–90. North-Holland, 1988.
20. Bernd Fischer and Joachim Buhmann. Path-based clustering for grouping of smooth curves and texture segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25:513– 518, 05 2003.
21. Bernd Fischer, Thomas Zöller, and Joachim Buhmann. Path based pairwise data clustering with application to texture segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition, Third International Workshop, EMMCVPR 2001*, volume 2134, pages 235–250, 09 2001.
22. I. Lerman. *Classification et Analyse Ordinale des Données*. Dunod, Paris, 1981.
23. B. Mirkin and I. Muchnik. Combinatorial optimization in clustering. In D.-Z. Du and P. Pardalos, editors, *Handbook of Combinatorial Optimization*, volume 2, pages 261–329. Kluwer Academic Publishers, 1998.
24. F. Murtagh. On ultrametricity, data coding, and computation. *Journal of Classification*, 21(2):167–184, 2004.
25. F. Murtagh. Identifying the ultrametricity of time series. *The European Physical Journal B*, 43(45):573–579, 2005.
26. R. Rammal, J.-C. Anglès d'Auriac, and B. Douçot. On the degree of ultrametricity. *Journal de Physique Lettres*, 46(20):945–952, 1985.
27. M. Van de Vel. *Theory of Convex Structures*. Elsevier, Amsterdam, North-Holland, 1993.
28. Chunjing Xu, Jianzhuang Liu, and Xiaoou Tang. Clustering with transitive distance and k-means duality. *ArXiv*, abs/0711.3594, 2007.
29. Zhiding Yu, Weiyang Liu, Wenbo Liu, Yingzhen Yang, Ming Li, and B. V. K. Vijaya Kumar. On order-constrained transitive distance clustering. In *AAAI*, 2016.
30. Zhiding Yu, Chunjing Xu, Deyu Meng, Zhuo Hui, Fanyi Xiao, Wenbo Liu, and Jianzhuang Liu. Transitive distance clustering with k-means duality. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 987–994, 06 2014.