BINARY CLUSTERING

JEAN-PIERRE BARTHÉLEMY AND FRANÇOIS BRUCKER

Abstract. In many clustering systems (hierarchies, pyramids and more generally weak hierarchies) clusters are generated by two elements only.

This paper is devoted to such clustering systems (called binary clustering systems). It provides some basic properties, links with (closed) weak hierarchies and some qualitative versions of bijection theorems that occur in Numerical Taxonomy. Moreover, a way to associate a binary clustering system to every clustering system is discussed.

Finally, introducing the notion of weak ultrametrics, a bijection between indexed weak hierarchies and weak ultrametrics is obtained (the standard theorem involves closed weak hierarchies and quasi-ultrametrics).

1. Introduction

One of the aims of classification is to sort a data set $X$ described by a dissimilarity measure $d$ into homogeneous and well-separated clusters (Benzécri [8], Jardine and Sibson [22]). Classically, the clustering systems involved are partitions or hierarchies (dendrograms). However these models do not account for overlapping that is needed in applied fields like biology (hybrids), social networks, or data mining.

Since Jardine and Sibson [22], classification models generalizing classical hierarchies, thus admitting overlapping clusters, have been designed like pyramids [14, 15] (or pseudo-hierarchies [16, 17]), weak-hierarchies [2] or quasi-hierarchies [12]. Bijection theorems [4, 8, 10, 13, 14, 16, 21] make these models equivalent to dissimilarity models (ultrametrics, Robinsonian dissimilarities, quasi-ultrametrics).

Several properties are shared by these models:

- the clusters are generated by two elements (that limit the number of clusters in $O(|X|^2)$),
- maximal cliques of a dissimilarity model are clusters,
- excepted for weak hierarchies, the clustering system is closed under finite nonempty intersections.

The aim of this paper is twofold:

On the one hand, we attempt to make the clustering problem homogeneous. The aim of binary clustering is to get clusters although generated by pairs of elements from a dissimilarity measure $d$. For instance, using the Jardine and Sibson [22] principle, clusters can be viewed as ML-sets (i.e. maximal cliques of the threshold graphs associated with $d$). So they can admit an exponential number of clusters whereas the number of clusters generated by pairs of elements is bounded by $|X|^{|X|-1}$.

On the other hand, the search for a latent structure involved by a data set (which is the main aim of classification in fields like evolution theory where one want to

Key words and phrases. clustering system, Boolean dissimilarity, weak hierarchy.
find the "true" evolutionary tree hidden in the data set) has no real signification in
the general case.

There are indeed a very large number of potential models the data could cor-
respond, and – moreover – searching for the good model is in general an ill-posed
problem because the data generally do not correspond to any model.

Binary clustering offers an alternative to this issue. It constructs what can be
generated by two objects, only, within a complex clustering system.

This paper is organized as follows: after recalling some definitions and results
about clustering (Section 2), Section 3 studies binary clustering systems and states
qualitative versions of known numerical bijection theorems. Section 4 is devoted to
binary realization. The idea is to associate with a pair of objects the intersection
of all clusters containing both of these objects. As a consequence, we establish a
numerical bijection theorem accounting for Bandelt and Dress’s weak-hierarchies
[2].

Section 5 gives results in two directions. First an analysis of the Boolean dissim-
ilarity associated with a (numerical) dissimilarity. Secondly a metric characteriza-
tion of weak hierarchies (which was an open problem since the Bandelt and Dress,
1989, paper [2].

2. Preliminaries

This Section is devoted to some basic notions and some standard results in
classification theory that will be used throughout this paper.

2.1. Class Models. Let $X$ be a finite set with $|X| \geq 3$. A Clustering System (CS)
on $X$ is a set $\mathcal{K}$ of subsets of $X$ such that:

[ CS ] $X \in \mathcal{K}$, $\phi \notin \mathcal{K}$ and for $x \in X$, $\{x\} \in \mathcal{K}$.

The elements of $\mathcal{K}$ are called clusters. The set $X$ and the singletons are the
trivial clusters.

A clustering system $\mathcal{K}$ on $X$ is said to be:

- separated, whenever there exist $u, v \in X$, with $u \neq v$, such that $\{u, v\}$ is
  not included in a nontrivial cluster of $\mathcal{K}$;
- closed whenever $A \in \mathcal{K}$, $B \in \mathcal{K}$ and $A \cap B \neq \phi$ imply $A \cap B \in \mathcal{K}$.

In the following Figures, (Figures 1, 2, 3 and 4), each clustering system in is
represented by its Hasse diagram: edges represent the transitive relation of inclusion
between clusters; clusters of the clustering system are the vertices of the diagram.

Figure 1 shows a nonclosed clustering system (a) and a nonseparated clustering
system (b). Note that clustering systems (c) and (d) are separated and closed.

We denote by $\overline{\mathcal{K}}$ the closure of clustering system $\mathcal{K}$: $\overline{\mathcal{K}}$ is the set of all the
nonempty intersections of the clusters of $\mathcal{K}$. The closure of the clustering system
(a) of Figure 1 is depicted in Figure 2.

If $S \subseteq X$, we denote by $<S>_\mathcal{K}$ the closure of $S$: $<S>_\mathcal{K}= \cap\{A \in \mathcal{K}, S \subseteq A\}$. Thus $\overline{\mathcal{K}} = \{<S>_\mathcal{K}: S \subseteq X\}$.

A hierarchy is a clustering system $\mathcal{K}$ such that: for all $A, B \in \mathcal{K}$, $A \cap B \in
\{A, B, \phi\}$.

A weak hierarchy is a CS so that for all $A, B, C \in \mathcal{K}$, $A \cap B \cap C \in \{A \cap B, B \cap C, A \cap C\}$.

An interval clustering system is a clustering system $\mathcal{K}$ on $X$ such that there
exists a linear order $\theta$ on $X$ such that each cluster of $\mathcal{K}$ is an interval of $\theta$ (i.e. for
A ∈ 𝐾 and \( x, y \in A, x \theta t y \) implies \( t \in A \). The order \( \theta \) is said to be compatible with \( 𝐾 \).

Weak hierarchies were introduced by Batbedat [4, 5] under the name of *medinclus* and by Bandelt and Dress [2] but without the condition [CS] (thus our weak hierarchies are just normalization of Bandelt and Dress’s weak hierarchies). Closed weak hierarchies have been named *quasi-hierarchies* by Diatta and Fichet [12, 13]. Interval clustering systems are the normalization – as a clustering system – of the so-called *interval hypergraphs* (Berge [9]) which are a central topic in hypergraph theory. Closed interval clustering systems have been named *pyramids* by Diday [14, 15] and pseudo-hierarchies by Fichet [16, 17]. In order to avoid the reader being lost in a jungle of terminology, and due to the variety of names that appear in the literature (weak hierarchies have also received some other names in hypergraph theory...), we shall simply speak of *closed weak hierarchies* and *closed interval clustering systems*.

In Figure 1, (c) is a hierarchy, (d) a closed weak hierarchy, (a) a nonclosed weak hierarchy and (b) is not a weak hierarchy. It is easy to check that both hierarchies and interval clustering systems are weak hierarchies and that for \(|X| \leq 3\) a weak

![Figure 1. Examples of clustering systems](image1)

![Figure 2. The closure of the clustering system (a) of Figure 1](image2)
hierarchy is also an interval clustering system. This property no more holds for $|X| > 3$, as shown in Figure 3 (a weak hierarchy on $X$ with $|X| = 4$ which is not an interval clustering system).

![Figure 3](image.png)

**Figure 3.** A weak hierarchy that is not an interval clustering system

Let $K$ be a clustering system. A *pre-index* on $K$ is a real-valued function $f$ defined on $K$ such that:

- $x \in X$ implies $f(\{x\}) = 0$,
- for $A, B \in K$, $A \subseteq B$ implies $f(A) \leq f(B)$.

A pre-index $f$ on $K$ is called a *weak-index* whenever for $A, B \in K$ with $A \subset B$ and $f(A) = f(B)$ we have: $A = \cap\{C | C \in K, A \subset C\}$.

An index is a pre-index such that $A, B \in K$ and $A \subset B$ imply $f(A) < f(B)$.

A pair $(K, f)$ is said to be an indexed clustering system (resp. a pre-indexed clustering system) whenever $f$ is an index (resp. a pre-index) on $K$.

2.2. **Dissimilarity models and bijection theorems.** Dissimilarity models appear as an alternative to class models. In many cases, a clustering method transforms a dissimilarity measure (the data) into an indexed clustering system of a given type. Helpfully, bijection theorems state equivalences between class models and dissimilarity models and make the clustering method homogeneous in terms of dissimilarity fitting. This Section is devoted to the dissimilarity models that will be used in this paper. For a more complete review, see Barthélemy and Brucker [3] for instance.

A *dissimilarity* on $X$ is a real valued function $d$ defined on $X \times X$ such that, for all $x, y \in X$:

- $d(x, y) \geq 0$ and $d(x, x) = 0$,
- $d(x, y) = d(y, x)$.

Let $d$ be a dissimilarity on $X$ and $S \subseteq X$, the *diameter* of $S$ is defined by $\text{diam}_d(S) = \max\{d(x, y) | x, y \in S\}$.

A pre-indexed clustering system $(K, f)$ induces the dissimilarity $\delta_{(K, f)}$ defined by: $\delta_{(K, f)}(x, y) = \min\{f(A) | A \in K, x, y \in A\}$.

An ultrametric is a dissimilarity $d$ such that for all $x, y, z \in X$: $d(x, z) \leq \max\{d(x, y), d(y, z)\}$.

A quasi-ultrametric is a dissimilarity such that for all $x, y, z, t \in X$: $\max\{d(x, z), d(y, z)\} \leq d(x, y)$ implies $d(z, t) \leq \max\{d(x, t), d(t, y), d(x, y)\}$. Clearly an ultrametric is a quasi-ultrametric because for all $x, y, z, t \in X$, $d(z, t) \leq \max\{d(y, t), d(y, z)\}$ and if $\max\{d(x, z), d(y, z)\} \leq d(x, y)$, we have $d(z, t) \leq \max\{d(x, t), d(t, y), d(x, y)\}$.
The dissimilarity $d$ is said to be \textit{proper} whenever $d(x, y) = 0$ implies $x = y$. Up to now, all the dissimilarities considered in this paper (ultrametrics, quasi-ultrametrics and others) will be assumed to be proper.

Let $\mathcal{C}_X$ be the set of all indexed clustering system on $X$ and $\mathcal{D}_X$ be the set of all dissimilarities on $X$. Denote by $\gamma$ the map that assigns to $(\mathcal{K}, f) \in \mathcal{C}_X$ the dissimilarity $\delta(\mathcal{K}, f)$.

\textbf{Proposition 2.1} (Diatta [13]). The restriction of $\gamma$ to the set of all indexed closed weak hierarchies on $X$ induces a bijection to the set of all quasi-ultrametrics on $X$. Moreover $\gamma(\mathcal{K}, f)$ is an ultrametric if and only if $\mathcal{K}$ is a hierarchy.

Proposition 2.1 can be extended to other classes of indexed clustering systems (like indexed and closed interval clustering systems). But since these extensions will not be used in the sequel, we shall not mention them and refer to [3] for the interested reader.

\section*{2.3. A Preparatory Lemma.} The following lemma will be essential for our purpose:

\textbf{Lemma 2.2} (Bandelt and Dress [2]). Let $\mathcal{K}$ be a clustering system. Then the following three assertions are equivalent:

(i) $\mathcal{K}$ is a weak hierarchy,

(ii) $\overline{\mathcal{K}}$ is a weak hierarchy,

(iii) For each $S \subseteq X$ such that $|S| \geq 2$, there exist $u, v \in S$, with $u \neq v$ such that $<S>_{\mathcal{K}} = <\{u, v\}>_{\mathcal{K}}$.

\section*{3. Binary clustering systems}

\subsection*{3.1. Prebinary, binary and strongly binary clustering systems.} A clustering system $\mathcal{K}$ is said to be \textit{prebinary} whenever for $x, y \in X$, the set $\mathcal{K}_{x,y} = \{C \in \mathcal{K} | \{x, y\} \subseteq C\}$ admits one and only one minimal member with respect to the inclusion which is denoted by $\delta_{\mathcal{K}}(x, y)$ (this minimal member is also a minimum one because the cluster collection $\mathcal{K}$ is finite). The justification of this notation will appear in Section 3.4.

Since for all clustering system $\delta_{\mathcal{K}}(x, x) = \{x\}$ (because $\{x\}$ is a cluster), in any case: $\delta_{\mathcal{K}}(x, y) = \cap\{C | C \in \mathcal{K}, x, y \in C\}$. We say that the pair $xy$ \textit{generates} the cluster $\delta_{\mathcal{K}}(x, y)$.

A prebinary clustering system is said to be \textit{binary}, whenever, for each $C \in \mathcal{K}$, there exist $u, v \in X$, such that $C = \delta_{\mathcal{K}}(u, v)$. In Figure 1, only the clustering system (a) is not prebinary ($\mathcal{K}_{2,3}$ admits two minimal clusters $\{1, 2, 3\}$ and $\{2, 3, 4\}$), (b) is prebinary but not binary (the cluster $\{1, 2, 3\}$ is not a $\delta_{\mathcal{K}}(u, v)$ for $u, v \in \{1, 2, 3\}$), (c) and (d) are binary.

A closed clustering system is prebinary (the converse would be false, as shown in Figure 1 (b)). Moreover a binary clustering system is separated and admits at most $\frac{n(n+1)}{2}$ clusters (hence, at most $\frac{n(n-1)}{2}$ non-trivial clusters).

A clustering system $\mathcal{K}$ is said to be \textit{strongly binary} whenever it is prebinary and for each $S \subseteq X$, $S \neq \emptyset$, there exist $u, v \in S$ such that: $S \subseteq \delta_{\mathcal{K}}(u, v)$.

In brief, a clustering system is binary if and only if each cluster is generated by two elements. It is strongly binary if and only if a smallest cluster containing a subset $S$ of $X$ is generated by two elements of $S$. Obviously, a strongly binary clustering system is binary.
Proposition 3.1. Let $\mathcal{K}$ be a clustering system. The following two assertions are equivalent:

(i) $\mathcal{K}$ is strongly binary,

(ii) $\mathcal{K}$ is a closed weak hierarchy.

Proof. From Lemma 2.2, it suffices to show that if $\mathcal{K}$ is strongly binary, then $\mathcal{K}$ is closed. Let $A, B \in \mathcal{K}$ such that $|A \cap B| \geq 2$. We know from strongly binarity that there exist $u, v \in A \cap B$ such that $\delta_{\mathcal{K}}(u, v)$ is minimum among the clusters containing $A \cap B$. Because $\mathcal{K}$ is prebinary and $u, v \in A$ (resp. $u, v \in B$), we have $\delta_{\mathcal{K}}(u, v) \subseteq A$ (resp. $\delta_{\mathcal{K}}(u, v) \subseteq B$). Hence $A \cap B \subseteq \delta_{\mathcal{K}}(u, v) \subseteq A \cap B$ and $\mathcal{K}$ is closed. \qed

3.2. Clusters as maximal cliques of graphs. Define a nested family of graphs on $X$ as a sequence $\mathcal{G} = (G_0, \ldots, G_p)$, $G_i = (X, E_i)$, with $X$ a vertex set and with $E_0 = \emptyset; E_i \subseteq E_{i+1}$ for $0 \leq i < p$ and $G_p$ is the complete graph on $X$. The integer $p$ is called the length of $\mathcal{G}$.

With the nested family of graphs $\mathcal{G}$ is associated the clustering system $\mathcal{K}[\mathcal{G}]$ whose clusters are all the maximal cliques of the graphs $G_i$ ($0 \leq i \leq p$). Bertrand [10] has shown that if $\mathcal{K}[\mathcal{G}]$ is closed, then it is a closed weak hierarchy. As a consequence of Proposition 3.1, Proposition 3.2 provides a binary counter-part of this observation.

Proposition 3.2. Let $\mathcal{G} = (G_0, \ldots, G_p)$ be a nested family of graphs. Then the following two assertions are equivalent:

(i) $\mathcal{K}[\mathcal{G}]$ is a binary clustering system

(ii) $\mathcal{K}[\mathcal{G}]$ is a closed weak hierarchy

Proof. We know, from Proposition 3.1 that (ii) implies (i). To prove that (i) implies (ii), we just have to show that the binariness of $\mathcal{K}[\mathcal{G}]$ implies its strong binariness. For $x, y \in X$, denote by $i(x, y)$ the smallest integer $i$ between 0 and $p$ such that $xy$ is an edge of $G_i$. We will first prove that the binariness of $\mathcal{K}[\mathcal{G}]$ implies that there is only one maximal clique containing $xy$ in $G_{i(x, y)}$. Suppose that there exist two distinct maximal cliques $C_1$ and $C_2$ containing $xy$ in $G_{i(x, y)}$. Thus, $\delta_{\mathcal{K}[\mathcal{G}]}(x, y) \subseteq C_1 \cap C_2 \subseteq C_1$. Since $\mathcal{K}[\mathcal{G}]$ is binary, $\delta_{\mathcal{K}[\mathcal{G}]}(x, y)$ is a maximal clique of some $G_k$, $k \geq i(x, y)$ but $\delta_{\mathcal{K}[\mathcal{G}]}(x, y)$ is strictly included in $C_1$, hence a contradiction.

Consider $S \subseteq X$ and choose $u, v \in S$ such that $i(u, v)$ is maximum. Then $S$ is a clique of $G_{i(u, v)}$ and can then be completed into a unique maximal clique containing $uv$ in $G_{i(u, v)}$. Moreover, this maximal clique is $\delta_{\mathcal{K}[\mathcal{G}]}(u, v)$ because of the binariness of $\mathcal{K}[\mathcal{G}]$. We then have $S \subseteq \delta_{\mathcal{K}[\mathcal{G}]}(u, v)$; $\mathcal{K}[\mathcal{G}]$ is a strongly binary clustering system. \qed

3.3. Boolean dissimilarities. Boolean metrics were introduced by Melter [23] in order to take into account for set-valuated distances. Then they were studied by Harary et al. [18] and others. The definition given by Melter is essentially a mimic of the definition of a (numerical) distance function (definiteness, symmetry and triangle inequality). A Boolean metric on a finite set is a function $\beta$ from $X \times X$ to $2^X$ such that:

- for $x, y \in X$, $\beta(x, y) = \emptyset$ if and only if $x = y$,
- for $x, y \in X$, $\beta(x, y) = \beta(y, x)$,
- for $x, y, z \in X$, $\beta(x, y) \subseteq \beta(x, z) \cup \beta(y, z)$. 

We will adapt and generalize this notion into Boolean dissimilarities with the modifications that allow us to capture the sets $\delta_K(x, y)$. Essentially we do not need the triangle inequality and we require that $x, y \in \beta(x, y)$. This forces us to change the definiteness property into $\beta(x, x) = \{x\}$.

A Boolean dissimilarity on $X$ is a function $\delta$ from $X \times X$ to $2^X$ such that:

- **BD$_1$**: for each $x \in X$, $\delta(x, x) = \{x\}$,
- **BD$_2$**: for each $x, y \in X$, $\delta(x, y) = \delta(y, x)$,
- **BD$_3$**: for each $x, y \in X$, with $x \neq y$, $\{x, y\} \subseteq \delta(x, y)$.

A Boolean dissimilarity $\delta$ is said to be separated whenever:

- **BD$_4$**: there exist $u, v \in X$ such that $\delta(u, v) = X$.

$\delta$ is said to be convex whenever:

- **BD$_5$**: for each $x, y \in X$ and $z, t \in \delta(x, y)$, $\delta(z, t) \subseteq \delta(x, y)$.

An example of a convex Boolean dissimilarity is the interval function of a graph (Mulder [24]): $\delta(x, y)$ is the set of vertices lying on a shortest path between $x$ and $y$. This Boolean dissimilarity is not separated except for some graphs like hypercubes.

If $K$ is a prebinary clustering system then $\delta_K$ is a convex Boolean dissimilarity. In this case $\delta_K$ is separated if and only if the clustering system $K$ is separated.

A **Boolean quasi-ultrametric** on $X$ is a convex and separated Boolean dissimilarity such that, for all $x, y, z \in X$,

$$\delta(x, y) \cap \delta(y, z) \cap \delta(x, z) \cap \{x, y, z\} \neq \phi.$$  

A **Boolean ultrametric** is a convex and separated Boolean dissimilarity such that, for all $x, y, z \in X$,

$$\delta(x, y) \cap \delta(y, z) \in \{\delta(x, y), \delta(y, z)\}.$$  

Clearly, a Boolean ultrametric is a Boolean quasi-ultrametric.

### 3.4. Bijection theorems

We shall state below a qualitative version of Proposition 2.1. By **qualitative**, we mean that (numerical) dissimilarities are replaced by Boolean dissimilarities.

Let $\Phi$ be the map from the set $\mathcal{K}_X$ of all prebinary clustering systems on $X$ to the set of all convex Boolean dissimilarities on $X$: $\Phi$ associates $\delta_K$ with $K$. Note that $\Phi$ is onto, but not one-to-one. Figure 4 provides a counter-example: $K \neq K'$ but $\delta_K = \delta_{K'}$.

![Figure 4. $K \neq K'$ but $\delta_K = \delta_{K'}$](image)
Let us denote by $\mathcal{K}_X^b$ the set of all binary clustering systems on $X$.

**Proposition 3.3.** The restriction of $\Phi$ to the set $\mathcal{K}_X^b$ defines a bijection from $\mathcal{K}_X^b$ to the set of all convex and separated Boolean dissimilarities on $X$. Moreover:

(i) $\Phi(\mathcal{K}) = \delta_\mathcal{K}$ is a Boolean quasi-ultrametric if and only if $\mathcal{K}$ is a closed weak hierarchy.

(ii) $\delta_\mathcal{K}$ is a Boolean ultrametric if and only if $\mathcal{K}$ is a hierarchy.

**Proof.** We have already seen that, if $\mathcal{K}$ is binary, then $\delta_\mathcal{K}$ is a convex and separated Boolean dissimilarity. Conversely, let $\delta$ be such a Boolean dissimilarity. Set $\mathcal{K}[\delta] = \{\delta(x, y) | x, y \in X\}$. Remark that $\delta(x, y)$ is the (only) minimum member of $\{A[x, y] | A \in \mathcal{K}[\delta]\}$: if $C \in \{A[x, y] | A \in \mathcal{K}[\delta]\}$, $C = \delta(u, v)$ for some $u, v \in C$ and, from $BD_3 \delta(x, y) \subseteq \delta(u, v)$. Hence $\delta(\mathcal{K})$ is binary and $\delta_{\mathcal{K}[\delta]} = \delta$. Hence the result.

Assume now that $\mathcal{K}$ is a weak hierarchy. Then, for $x, y, z \in X$, $\delta_\mathcal{K}(x, y) \cap \delta_\mathcal{K}(y, z) \cap \delta_\mathcal{K}(x, z) \in \{\delta_\mathcal{K}(x, y) \cap \delta_\mathcal{K}(y, z), \delta_\mathcal{K}(x, y) \cap \delta_\mathcal{K}(x, z), \delta_\mathcal{K}(y, z) \cap \delta_\mathcal{K}(x, z)\}$. When the weak hierarchy $\mathcal{K}$ is closed, then it is binary (Proposition 3.1) and $\delta_\mathcal{K}$ is separated. To prove the converse, it suffices to show that, if $\delta$ is a Boolean quasi-ultrametric, then $\mathcal{K}[\delta]$ is strongly binary. Consider $S \subseteq X$, with $|S| \geq 2$ and choose $u, v \in S$ such that $\delta(u, v)$ is maximal for the inclusion. If $S \subseteq \delta(u, v)$, then $\delta(u, v)$ is the minimum cluster of $\mathcal{K}[\delta]$ containing $S$ (because if $S \subseteq C$, with $C$ a cluster, then $u, v \in C$ and, by convexity, $\delta(u, v) \subseteq C$). Otherwise there exist $x \in S$ such that $x \notin \delta(u, v)$. Set $I = \delta(x, u) \cap \delta(x, v) \cap \delta(u, v) \cap \{x, u, v\}$. We know that $x \notin I$. Thus $u \in I$ or $v \in I$. If $u \in I$, then $\delta(u, v) \subseteq \delta(x, v)$. In case of equality, $x \in \delta(u, v)$ which leads to a contradiction. So, $\delta(u, v) \subseteq \delta(x, v)$, but this inclusion contradicts the maximality of $\delta(u, v)$. The argument remains the same for $v \in I$, hence the result.

Let us now examine (ii). If $\mathcal{K}$ is a hierarchy, then $\delta_\mathcal{K}$ is obviously a Boolean ultrametric. Conversely, let $\delta$ be a Boolean ultrametric. Let $A, B \in \mathcal{K}[\delta]$ such that $A \cap B \neq \emptyset$ and $A \not\subseteq B$. Consider $x \in A \cap B$ and $u \in A, u \notin B$. Then for each $y \in B$, we have $\delta(x, y) \cap \delta(x, u) = \delta(x, y)$ (otherwise $u \in \delta(x, y) \subseteq B$). Hence $B \subseteq A$. □

Now we shall extend the notion of closed interval clustering systems to Boolean dissimilarities. We say that a linear order $\theta$ is compatible with the Boolean dissimilarity $\delta$ whenever: $x \theta y \theta z$ implies $\delta(x, y) \cup \delta(y, z) \subseteq \delta(x, z)$. In the sequel $\theta$ will just be called an order. It is easy to check that if $\theta$ is compatible with the Boolean dissimilarity $\delta$, then $\delta$ is convex and separated. The following Proposition asserts that, in the bijection of Proposition 3.3, the closed interval clustering system corresponds to the Boolean dissimilarity admitting a compatible order.

**Proposition 3.4.** Let $\mathcal{K}$ be a clustering system. Then $\mathcal{K}$ is a closed interval clustering system if and only if $\delta_\mathcal{K}$ admits a compatible order. In this case the orders compatible with $\mathcal{K}$ are exactly the orders compatible with $\delta_\mathcal{K}$.

**Proof.** Assume that $\mathcal{K}$ is a closed interval clustering system and consider a compatible order $\theta$. Let $x, y, z \in X$ be such that $x \theta y \theta z$. Then $\delta_\mathcal{K}(x, z)$ is an interval of $\theta$ containing $x$ and $z$. Hence $y \in \delta_\mathcal{K}(x, z)$ and by convexity: $\delta_\mathcal{K}(x, y) \subseteq \delta_\mathcal{K}(x, z)$ and $\delta_\mathcal{K}(y, z) \subseteq \delta_\mathcal{K}(x, z)$.

Conversely assume that $\theta$ is compatible with $\delta_\mathcal{K}$. Then, for $x, y, z \in X$, we can assume that $x \theta y \theta z$. So, $y \in \delta_\mathcal{K}(x, y) \cap \delta_\mathcal{K}(y, z) \cap \delta_\mathcal{K}(x, z)$. Thus $\delta_\mathcal{K}$ is a Boolean quasi-ultrametric and $\mathcal{K}[\delta]$ is closed. Now, for $u, v \in X$, with $u \theta v$, denote by $u^*$ the smallest element of $\delta_\mathcal{K}(u, v)$ and by $v^*$ its largest one. From the definition of
compatibility, we get that $t \in [u^*, v^*]$ implies $t \in \delta_K(u, v)$. Hence $\delta_K(u, v) = [u^*, v^*]$ and $K$ is an interval clustering system. 

Note that a non-closed interval clustering system is not necessarily binary (Figure 1 (a)).

3.5. Colonius-Schulze and Bandelt characterizations revisited. First recall that a ternary relation on $X$ is a subset of $X \times X \times X$. We shall write $T(x, y, z)$ instead of $(x, y, z) \in T$ and denote by $\overline{T}$ the complement of $T$ ($\overline{T}$ is the relation defined by $(x, y, z) \in \overline{T}$ if and only if $(x, y, z) \notin T$).

Colonius and Schulze [11] have provided a characterization of hierarchies in terms of ternary relations. Bandelt [1] extended this result to closed weak-hierarchies. Both characterizations are based on the so-called separation relation $S_K$: $S_K(x, y, z)$ means that there is a cluster containing $x$ and $y$ but not $z$. Up to a minor change (separation has to be replaced by aggregation) such characterizations appear as immediate consequences of Proposition 3.3.

Let $K$ be a prebinary clustering system on $X$. The aggregation relation $T_K$ induced by $K$ is the ternary relation on $X$ defined by $T_K(x, y, z)$ if and only if $z \in \delta_K(x, y)$ (in the binary case, the separation relation corresponds to $S(x, y, z)$ if and only if $z \notin \delta_K(x, y)$).

The conditions defining various kinds of Boolean dissimilarities can be easily translated in terms of ternary relations (since a function from $X \times X$ to $2^X$ “is” exactly a subset of $X \times X \times X$). These conditions can be stated as follows:

- $T_1$: $T(x, x, y)$ if and only if $x = y$,
- $T_2$: $T(x, y, z)$ implies $T(y, x, z)$,
- $T_3$: $T(x, y, x)$,
- $T_4$: there exist $u, v \in X$ such that for each $t \in X$, $T(u, v, t)$,
- $T_5$: for all $x, y, z, t, u \in X$, $T(x, y, z)$ and $T(x, y, t)$ and $T(z, t, u)$ imply $T(x, y, u)$,
- $T_6$: for each $x, y, z \in X$, $T(x, y, z)$ or $T(x, z, y)$ or $T(y, z, x)$,
- $T_7$: if there is $t \in X$ such that $T(x, y, t)$ and $\overline{T}(y, z, t)$, then for each $u \in X$, $T(y, z, u)$ imply $T(x, y, u)$.

Conditions $T_i$ for $1 \leq i \leq 5$ are just the conditions $BD_i$. $T_6$ corresponds to the quasi-ultrametricity condition and $T_7$ to the ultrametric condition. For a ternary relation $T$ on $X$, we define the Boolean dissimilarity $\delta_T$ by $z \in \delta_T(x, y)$ if and only if $T(x, y, z)$. The set $K^T = \{\delta_T(x, y) | x, y \in X\}$ is a clustering system when $T$ satisfies $T_1$ and $T_5$. We shall denote by $\psi$ the map which associates $K^T$ to the ternary relation $T$. Rephrasing Proposition 3.3, we get:

\textbf{Proposition 3.5.} The restriction of $\psi$ to the set of all ternary relations on $X$ fulfilling $T_1$ to $T_5$ induces a bijection to the set of all binary clustering systems. Moreover $\psi$ satisfies $T_6$ if and only if $\psi(T) = K^T$ is a closed weak hierarchy and $T$ satisfies $T_7$ if and only if $K^T$ is a hierarchy.

Proposition 3.4 could also be restated in terms of ternary relations. We say that $T$ is compatible with the order $\theta$ whenever $x \theta y \theta z$ implies $T(x, z, y)$. Obviously, $\psi$ induces a bijection from the set of ternary relations on $X$ satisfying $T_i$, $1 \leq i \leq 5$, and admitting a compatible order, to the set of all closed interval clustering systems.
4. Binary realizations

4.1. Basic concepts. The notions discussed in Section 3 extend to any clustering system (prebinary or not) via the concept of binary realization. The binary realization of the clustering system $\mathcal{K}$ is the Boolean dissimilarity $\delta_{\mathcal{K}}$ defined by:

$$\delta_{\mathcal{K}}(x, y) = \cap\{C \in \mathcal{K}, x, y \in C\}.$$  

The following observations are straightforward:

(i) $\delta_{\mathcal{K}}$ is a convex Boolean dissimilarity.

(ii) In general $\delta_{\mathcal{K}}$ is not separated, $\delta_{\mathcal{K}}$ is separated if and only if $\mathcal{K}$ is itself separated.

(iii) Set $\mathcal{K} = \{\delta_{\mathcal{K}}(x, y)|x, y \in X\}$. In general, $X \notin \mathcal{K}$, thus $\mathcal{K}$ is not a clustering system. However, when $\mathcal{K}$ is separated, then $\mathcal{K}$ is a binary clustering system and $\delta_{\mathcal{K}} = \delta_{\hat{\mathcal{K}}}$.

(iv) $\mathcal{K}$ is a prebinary clustering system if and only if $\hat{\mathcal{K}} \subseteq \mathcal{K}$.

Proposition 4.1 below extends the results of Propositions 3.1 and 3.4, as well as Lemma 2.2, for (i).

**Proposition 4.1.** Let $\mathcal{K}$ be a clustering system. Then:

(i) $\mathcal{K}$ is a weak hierarchy if and only if $\hat{\mathcal{K}}$ is a closed weak hierarchy,

(ii) $\mathcal{K}$ is an interval clustering system if and only if $\hat{\mathcal{K}}$ is a closed interval clustering system. Moreover, the orders compatible with $\mathcal{K}$ are exactly the orders compatible with $\hat{\mathcal{K}}$,

(iii) $\mathcal{K}$ is a hierarchy if and only if $\hat{\mathcal{K}}$ is a hierarchy (and in this case $\mathcal{K} = \hat{\mathcal{K}}$).

**Proof.** To prove that if $\mathcal{K}$ is a weak hierarchy, then $\hat{\mathcal{K}}$ is a closed weak hierarchy, it suffices to show that $\hat{\mathcal{K}} = \overline{\mathcal{K}}$ (Lemma 2.2). Clearly $\hat{\mathcal{K}} \subseteq \overline{\mathcal{K}}$. Consider $S \in \overline{\mathcal{K}}$. From Lemma 2.2, there exist $u, v \in S$ such that $<S>_{\mathcal{K}} = <\{u, v\}>_{\mathcal{K}} = \delta_{\mathcal{K}}(u, v)$, thus $S = \delta_{\mathcal{K}}(u, v)$ and $\overline{\mathcal{K}} \subseteq \hat{\mathcal{K}}$.

Conversely, assume that $\mathcal{K}$ is not a weak hierarchy. Then, there exist $A, B, C \in \mathcal{K}$ such that $A \cap B \cap C \notin \{A \cap B, B \cap C, A \cap C\}$. Thus, there exist $x, y, z \in X$ such that: $x \in A \cap B$ and $x \notin C$; $y \in B \cap C$ and $y \notin A$; $z \in A \cap C$ and $z \notin B$. Hence $x \notin \delta_{\mathcal{K}}(y, z)$, $y \notin \delta_{\mathcal{K}}(x, z)$ and $z \notin \delta_{\mathcal{K}}(x, y)$, and from Proposition 3.3, $\hat{\mathcal{K}}$ is not a closed weak hierarchy.

Now look at (ii). We know that if $\mathcal{K}$ or $\hat{\mathcal{K}}$ admits a compatible order $\theta$ then it is a weak hierarchy. Hence from (i), $\hat{\mathcal{K}} = \overline{\mathcal{K}}$. Since the orders compatible with $\overline{\mathcal{K}}$ are exactly those compatible with $\mathcal{K}$, $\mathcal{K}$ is an interval clustering system if and only if $\hat{\mathcal{K}}$ is a closed interval clustering system and they share the same compatible orders.

Finally, examine (iii). Clearly, if $\mathcal{K}$ is a hierarchy, then $\hat{\mathcal{K}}$ is a hierarchy and since $\mathcal{K}$ is closed $\mathcal{K} = \overline{\mathcal{K}} = \hat{\mathcal{K}}$. Conversely, if $\hat{\mathcal{K}}$ is a hierarchy, $\mathcal{K}$ is a quasi-hierarchy and $\overline{\mathcal{K}} = \hat{\mathcal{K}}$. Since $\overline{\mathcal{K}}$ is a hierarchy, $\mathcal{K}$ is also a hierarchy, which concludes the proof. □

Sine a clustering system $\mathcal{K}$ is binary whenever $\mathcal{K} = \hat{\mathcal{K}}$, as a consequence of Proposition 4.1, a weak-hierarchy is closed if and only if it is binary.

4.2. Clusters as maximal cliques of graphs. Section 3.2 was devoted to the binary clustering system whose clusters are maximal cliques of graphs from a given nested family. Here we shall study the binary realizations of the (non-binary) clustering system whose clusters are maximal cliques of graphs.

We can associate with any graph $G = (X, E)$ the nested family of graphs $\mathcal{G}[G]$ defined by:
• $G = (G_0, G_1, G_2)$ with $G = G_1$ if $G$ is not discrete (the discrete graph on $X$ is the graph having no edges) nor complete, $G_0$ being the discrete graph on $X$ and $G_2$ the complete graph on $X$;
• $G = (G_0, G_1)$ with $G = G_0$ if $G$ is discrete and $G = G_1$ if $G$ is complete.

$G[G]$ will simply be denoted as $G$ in the following, and we shall also write $K[G]$ instead of $K[G[G]]$.

We know that $K[G]$ is separated. Moreover, $K[G]$ is binary if and only if two maximal cliques of $G$ have at most one vertex in common. This condition is equivalent to excluding from $G$ the configuration of Figure 5 as an induced subgraph [3].

**Figure 5.** Forbidden induced subgraph for $G$ if $K[G]$ is binary

Since $K[G]$ is separated, $\overrightarrow{K[G]}$ is a binary clustering system. We shall simply denote by $\delta_G(x, y)$ the clusters $\delta_{K[G]}(x, y) = \delta_{\overrightarrow{K[G]}}(x, y)$. Thus we have:

- $\delta_G(x, x) = \{x\}$,
- $\delta_G(x, y) = X$ if $x \neq y$ and $xy \notin E$,
- $\delta_G(x, y)$ is the intersection of all maximal cliques of $G$ containing $x$ and $y$ whenever $xy \in E$.

For $u \in X$, we denote by $B_G(u) = \{v|uv \in E\} \cup \{u\}$ the ball centered at $u$ of $G$. It corresponds to the neighborhood of $u$ in $G$ plus the element $u$.

**Proposition 4.2.** Let $G = (X, E)$ be a graph. Then for all $x, y \in X$ such that $xy \in E$, $\delta_G(x, y)$ is equal to the intersection of all the closed neighborhoods containing $x$ and $y$. That is:

$$\delta_G(x, y) = \cap\{B_G(u)|u \in X; x, y \in B_G(u)\}.$$

**Proof.** If $G$ is the complete graph on $X$, the property clearly holds. We assume then that $G$ is not the complete graph. Thus for all $x$ and $y$ such that $xy \in E$, $\delta_G(x, y) \subsetneq X$ (because $\delta_G(x, y)$ is a clique of $G$ if $xy \in E$).

Let $x$ and $y$ be two elements of $X$ such that $xy \in E$ and set $\delta_G(x, y) = \cap\{B_G(u)|u \in X; x, y \in B_G(u)\}$. We first prove that $\delta_G^0(x, y) \subseteq \delta_G(x, y)$. Let $z \notin \delta_G(x, y)$. There exist a maximal clique $C$ of $G$ such that $x, y \in C$ and $z \notin C$. Thus we get a $t \in C$ such that $zt \notin E$. Since $x, y \in C$, we have $x, y \in B_G(t)$ but $z \notin B_G(t)$, thus $z \notin \delta_G^0(x, y)$ and $\delta_G^0(x, y) \subseteq \delta_G(x, y)$.

Conversely, let $z \notin \delta_G^0(x, y)$. Thus there exists $u \in X$ such that $x, y \in B_G(u)$ and $z \notin B_G(u)$. Since $xy \in E$, there exists a maximal clique of $G$ containing $x$, $y$ and $u$ but not $z$ (since $z \notin B_G(u)$, $zu \notin E$) thus $z \notin \delta_G(x, y)$ and $\delta_G(x, y) \subseteq \delta_G^0(x, y)$. □

As a consequence of Proposition 4.2, $\delta_G(x, y)$ – hence $\overrightarrow{K[G]}$ – can be constructed in polynomial time (even if the generation of all the maximal cliques of a graph
Since the computation of their intersections makes the task much more difficult. The number of operations needed to compute a neighborhood is in \( O(|X|) \), and the number of operations needed to compute the intersection of two clusters is also in \( O(|X|) \). Thus \( \delta_G(x,y) \) is the intersection of at most \( |X| \) neighborhoods, we need \( O(|X|^2) \) operations to compute it.

Finally, the construction of \( \widehat{\mathcal{K}[G]} \) can be performed in \( O(|X|^4) \) operations because there are \( |X|^2 \) pairs of elements of \( X \). For instance, the non-trivial \( \delta_G(x,y) \)'s associated with the graph depicted in Figure 6 are:

- \( \delta_G(x,y) = \delta_G(x,z) = \{x, y, z\} \)
- \( \delta_G(y,z) = \{y, z\} \)
- \( \delta_G(y,t) = \delta_G(z,t) = \{y, z, t\} \)
- \( \delta_G(t,u) = \{t, u\} \)
- \( \delta_G(t,v) = \{t, v\} \)

![Figure 6. An example graph \( G \)](image)

Proposition 4.2 extends to any nested family of graphs \( \mathcal{G} \). Set \( \delta_\mathcal{G}(x,y) = \delta_{\widehat{\mathcal{K}[G]}}(x,y) \). Then:

**Proposition 4.3.** Let \( \mathcal{G} = (G_0, \ldots, G_\rho) \) be a nested family of graphs. Then for all \( x, y \in X \):

\[
\delta_\mathcal{G}(x,y) = \bigcap_{1 \leq i \leq \rho} \delta_{G_i}(x,y).
\]

Moreover, if we denote by \( i(x,y) \) the smallest integer \( i \) between 0 and \( \rho \) such that \( xy \) is an edge of \( G_i \) (see Proposition 3.2), and \( i(x,y,z) = \max\{i(x,y), i(x,z), i(y,z)\} \), we have for all \( x, y \in X \):

\[
\delta_\mathcal{G}(x,y) = \bigcap_{z \in E (G_{i(x,y,z)})} B_{G_{i(x,y,z)}}(z).
\]

**Proof.** Since \( \delta_\mathcal{G}(x,y) \) is the intersection of all the maximal cliques of \( \widehat{\mathcal{K}[G]} \) containing \( x \) and \( y \), and \( \delta_{G_i}(x,y) \) the intersection of all the maximal cliques of \( G_i \) if \( xy \in E_i \) and \( X \) otherwise, the first equality is clear.

To prove the second equality, set \( \delta_\mathcal{G}^*(x,y) = \bigcap_{z \in X} B_{G_{i(x,y,z)}}(z) \). We will first show that \( \delta_\mathcal{G}^*(x,y) \subseteq \delta_\mathcal{G}(x,y) \). Let \( z \notin \delta_\mathcal{G}(x,y) \). Due to the first equality, there exists \( i \) such that \( z \notin \delta_{G_i}(x,y) \), thus \( i(x,y,x) = i(x,y,y) < i(x,y,z) \) and \( z \notin B_{G_{i(x,y,z)}}(y) \cap B_{G_{i(x,y,z)}}(x) \) if \( z \notin \delta_{G_i}(x,y) \). Hence \( \delta_\mathcal{G}^*(x,y) \subseteq \delta_\mathcal{G}(x,y) \).

Conversely, let \( z \notin \delta_\mathcal{G}^*(x,y) \). Then there exists \( u \in X \) such that \( z \notin B_{G_{i(x,y,u)}}(u) \): either \( xu \notin E_{i(x,y,u)} \), or \( yu \notin E_{i(x,y,u)} \). Thus, \( z \notin \delta_{G_{i(x,y,u)}}(x,y) \) and \( z \notin \delta_\mathcal{G}(x,y) \). Hence \( \delta_\mathcal{G}(x,y) \subseteq \delta_\mathcal{G}^*(x,y) \). \( \square \)
It results from Proposition 4.3 that if $\mathcal{G}$ is a nested family of graphs, then its binary realization $K[\mathcal{G}]$ can be computed in $O(|X|^4)$ operations.

4.3. **Interpretation in terms of Boolean dissimilarities.** As noted above, if $K$ is a clustering system then $\delta_K$ is a convex Boolean dissimilarity. When $K$ is a weak hierarchy, then $\delta_K$ is separated. Proposition 4.4 below is just a rephrasing of Proposition 4.1.

**Proposition 4.4.** Let $K$ be a clustering system, then:

(i) $K$ is a weak hierarchy if and only if $\delta_K$ is a Boolean quasi-ultrametric,

(ii) $K$ is an interval clustering system if and only if $\delta_K$ admits a compatible order,

(iii) $K$ is a hierarchy if and only if $\delta_K$ is a Boolean ultrametric.

Note that (excepted for (iii)), Proposition 4.4 is not a bijection theorem (several weak hierarchies admit the same closure).

5. **Boolean dissimilarities and (numerical) dissimilarities**

5.1. **Boolean dissimilarities associated with a dissimilarity.** The definitions given in Section 2 extend to Boolean dissimilarities. In particular, a convex Boolean dissimilarity $\delta$ is said to be pre-indexed (resp. indexed) by $f$ whenever $f$ is a function from $K[\delta]$ to $\mathbb{R}$ such that: $u, v \in \delta(x,y)$ imply $f(\delta(u,v)) \leq f(\delta(x,y))$ (resp. $u, v \in \delta(x,y)$ and $\delta(u,v) \neq \delta(x,y)$ implies $f(\delta(u,v)) < f(\delta(x,y))$).

The dissimilarity $d_{(\delta,f)}(x,y)$ defined by $d_{(\delta,f)}(x,y) = f(\delta(x,y))$ is associated with each convex pre-indexed Boolean dissimilarity $(\delta,f)$. Note that from convexity this definition is consistent with those (the dissimilarity associated with a pre-indexed clustering system) given in Section 2.2:

$$d_{(\delta,f)}(x,y) = \min\{f(\delta(u,v)) | x, y \in \delta(u,v)\}.$$ 

Conversely, if we denote by $0 < d_1 < \cdots < d_p$ all the different values taken by a proper dissimilarity $d$, we get the nested family of graphs $\mathcal{G}_d = (G_0, G_1, \ldots, G_p)$, with $G_i = (X, E_i)$ and $xy \in E_i$ if and only if $d(x,y) \leq d_i$ (the $G_i$’s are called the threshold graphs of $d$). For simplicity, we denote by $K[d]$ instead of $K[\mathcal{G}_d]$ the clustering system associated with $\mathcal{G}_d$ (i.e. the set of all the maximal cliques of the threshold graphs of $d$). Thus, we associate with $d$ a convex and separated Boolean dissimilarity $\delta[d]$ defined by $\delta[d] = \delta_{\mathcal{G}_d}$.

It follows from Proposition 3.2 that $d$ is a quasi-ultrametric if and only if $K[d] = \delta[d]$. Finally, if we denote by $B_d(x, \rho) = \{y | y \in X, d(x,y) \leq \rho\}$, according to Proposition 4.2, we get an alternative definition of $\delta[d]$:

**Proposition 4.5.** For every proper dissimilarity $d$ on $X$:

$$\delta[d](x,y) = \bigcap_{1 \leq i \leq p} \delta_{G_i}(x,y) = \bigcap_{z \in X} B_d(z, \max\{d(x,z), d(y,z), d(x,y)\}).$$

**Proof.** It follows directly from Proposition 4.2. \qed

5.2. **A “bijection” theorem.** We now establish a bijection theorem between a special class of pre-indexed Boolean dissimilarities on $X$ and all the proper dissimilarities on $X$. With that respect, convex and separated Boolean dissimilarities appear as a general (and large) qualitative counter-part of data described by dissimilarity measures.
Let \( \nu \) be the function from the set \( \mathcal{D}_X \) of all proper dissimilarities on \( X \) to the set \( \mathcal{B}_X \) of all pre-indexed Boolean dissimilarities defined by: \( \nu(d) = (\delta[d], \text{diam}_d) \).

**Proposition 5.2.** \( \nu \) is one-to-one.

*Proof.* First observe that for \( x, y \in X \), \( d(x, y) = \text{diam}_d(\delta[d](x, y)) \) (because \( \delta[d](x, y) \) is the intersection of all the clusters of \( d \) containing \( x \) and \( y \)). Consider two dissimilarities \( d \) and \( d' \) such that \( (\delta[d], \text{diam}_d) = (\delta[d'], \text{diam}_{d'}) \). Then from the preceding observation, \( d(x, y) = \text{diam}_d(\delta[d](x, y)) = \text{diam}_{d'}(\delta[d'](x, y)) = d'(x, y) \). Hence the result. \( \Box \)

Proposition 5.2 does not give the codomain of \( \nu \) (the set of pre-indexed clustering systems \((\mathcal{K}, f)\) such that it exists a dissimilarity \( d \) for which \( \nu(d) = (\mathcal{K}, f) \)), hence it does not deserve to be called a “bijection theorem”. However to check if a pre-indexed Boolean dissimilarity \((\delta, f)\) corresponds to a proper dissimilarity can be performed in polynomial time. It suffices to set \( d(x, y) = f(\delta(x, y)) \) and to use Proposition 5.1 \( (O(|X|^4) \text{ time is needed to check if all the } \delta(x, y) \text{ can be recovered from the corresponding intersection of balls}).

5.3. **Weak ultrametrics.** We shall now use Propositions 5.1 and 3.3 (i) to characterize a dissimilarity model associated with weak hierarchies. We say that a proper dissimilarity \( d \) is a weak-ultrametric whenever for each \( x, y, z \in X \), at least one of the following three conditions always holds:

- for each \( t \in X \): \( d(x, t) \leq \max\{d(y, z), d(y, t), d(z, t)\} \)
- for each \( t \in X \): \( d(y, t) \leq \max\{d(x, z), d(x, t), d(z, t)\} \)
- for each \( t \in X \): \( d(z, t) \leq \max\{d(x, y), d(x, t), d(y, t)\} \).

It is immediate to check that a quasi-ultrametric is a weak ultrametric.

**Proposition 5.3.** Let \( d \) be a proper dissimilarity. The following three assertions are equivalent:

- (i) \( \mathcal{K}[d] \) is a weak hierarchy
- (ii) \( \hat{\delta}[d] \) is a Boolean quasi-ultrametric
- (iii) \( d \) is a weak ultrametric.

*Proof.* If \( \mathcal{K}[d] \) is a weak hierarchy, then \( \hat{\mathcal{K}}[d] \) is a closed weak hierarchy. But \( \delta[d] = \hat{\delta}[d]_{\hat{\mathcal{K}}[d]} \). Hence (i) implies (ii). If \( \hat{\delta}[d] \) is a Boolean quasi-ultrametric, then for \( x, y, z \in X \), \( \delta[d](x, y) \cap \delta[d](y, z) \cap \delta[d](x, z) \cap \{x, y, z\} \neq \phi \) (Proposition 3.3) and we can assume, up to a symmetry, that \( x \in \delta[d](y, z) \). From Proposition 5.1, we deduce that: for each \( t \in X \) \( d(x, t) \leq \max\{d(y, z), d(y, t), d(z, t)\} \). The other two inequalities are obtained from \( y \in \delta[d](x, z) \) and \( z \in \delta[d](x, y) \). Hence (ii) implies (iii).

If (iii) holds, then from Proposition 5.1, for each \( x, y, z \in X \), \( x \in \delta[d](y, z) \), or \( y \in \delta[d](x, z) \), or \( z \in \delta[d](x, y) \). Hence \( \delta[d] \) is a Boolean quasi-ultrametric. Assume that \( \mathcal{K}[d] \) is not a weak hierarchy. Then there exist \( A, B, C \in \mathcal{K}[d] \) such that \( A \cap B \cap C \notin \{A \cap B, A \cap C, B \cap C\} \). Thus there exist \( x, y, z \in X \) with \( x \in A \cap B \) and \( x \notin C \), \( y \in A \cap C \) and \( y \notin B \), \( z \in B \cap C \) and \( z \notin A \). We have then that \( x \notin \delta[d](y, z) \), \( y \notin \delta[d](x, z) \) and \( z \notin \delta[d](x, y) \) and \( \delta[d] \) is not a Boolean quasi-ultrametric. \( \Box \)

**Proposition 5.4.** Let \( G \) be a graph, then \( \mathcal{K}[G] \) is a weak hierarchy if and only if the configuration of Figure 7 is forbidden as an induced subgraph of \( G \).
**Proof.** Consider the distance $d$ defined for $x \neq y$ by $d(x, y) = 1$ if $xy$ is an edge of $G$ and $d(x, y) = 2$ otherwise. From Proposition 5.3, $K[G]$ is a weak hierarchy if and only if $d$ is a weak-ultrametric. Suppose that it exist $x, y, z, t_1, t_2, t_3 \in X$ such that Figure 7 is a subgraph of $G$. Thus:

- $d(x, t_1) > \max\{d(y, z), d(y, t_1), d(z, t_1)\}$
- $d(y, t_2) > \max\{d(x, z), d(x, t_2), d(z, t_2)\}$
- $d(z, t_3) > \max\{d(x, y), d(x, t_3), d(y, t_3)\}$

By definition, $d$ cannot be a weak ultrametric.

Conversely, suppose that $d$ is not a weak ultrametric. It exist then $x, y, z, t_1, t_2, t_3 \in X$ such that the three above equations are satisfied. Since $d(u, v) \in \{1, 2\}$ for all $u \neq v \in X$, $x, y, z, t_1, t_2$ and $t_3$ are distinct and:

- $d(x, t_1) = d(y, t_2) = d(z, t_3) = 2$,
- $d(y, z) = d(y, t_1) = d(z, t_1) = d(x, z) = d(x, t_2) = d(z, t_2) = d(x, y) = d(x, t_3) = d(y, t_3) = 1$

Figure 7 is then a subgraph of $G$. $\square$

**Figure 7.** Forbidden subgraph for $G$ if $K[G]$ is a weak hierarchy

6. Conclusion

This paper has explored, in different ways, clustering systems whose clusters are generated by two elements (excepted for the singletons). Binary clustering systems has their own interest as generalization of classical clustering systems like hierarchies, quasi-hierarchies, weak hierarchies, ... They are also naturally associated to any dissimilarity.

Binary clustering leads to the notion of Boolean dissimilarities (close to the boolean distances as introduced by Melter, 1964 [23]. Dissimilarities taking their values in other sets than the real numbers have also been considered by Benkaraache [6, 7] and Janowitz [19, 20].

**References**


GET - ENST BRETAGNE, POINTE DU DIABLE, CS 83818, 29238 BREST CEDEX 3, FRANCE
E-mail address: jp.barthelemy@enst-bretagne.fr
E-mail address: francois.brucker@enst-bretagne.fr