Mathématiques et sciences humaines

Mathematics and social sciences

190 | Été 2010

Mathématiques discrètes : théories et usages.

Numéro en hommage à Bruno Leclerc

Classifications en classes recouvrantes ou non, et leurs dissimilarités

Clustering systems with or without overlapping clusters and their dissimilarities

Patrice Bertrand

Édition électronique

URL : <http://journals.openedition.org/msh/11737>

DOI : 10.4000/msh.11737

ISSN : 1950-6821

Éditeur

Centre d'analyse et de mathématique sociales de l'EHESS

Édition imprimée

Date de publication : 10 mars 2010

Pagination : 59-87

ISSN : 0987-6936

Référence électronique

Patrice Bertrand, « Classifications en classes recouvrantes ou non, et leurs dissimilarités », *Mathématiques et sciences humaines* [En ligne], 190 | Été 2010, mis en ligne le 16 octobre 2010, consulté le 23 juillet 2020. URL : <http://journals.openedition.org/msh/11737> ; DOI : <https://doi.org/10.4000/msh.11737>

CLASSIFICATIONS EN CLASSES RECOUVRANTES OU NON, ET LEURS DISSIMILARITÉS

Patrice BERTRAND^{1,2}

RÉSUMÉ – *La bijection de Benzécri-Johnson établit une correspondance biunivoque entre les classifications hiérarchiques et les dissimilarités ultramétriques [Johnson, 1967 ; Benzécri, 1973]. De nouvelles structures de classification en classes recouvrantes, introduites durant la décennie 1980, prolongent cette bijection. Dans ce texte, nous décrivons un cadre général permettant de présenter et de comparer différents prolongements de la bijection de Benzécri-Johnson qui ont été proposés. Nous considérons en particulier les prolongements existant entre certains types généraux de classifications en classes recouvrantes, d’une part, et l’ensemble de toutes les dissimilarités, d’autre part.*

MOTS-CLÉS – Bijection de Benzécri-Johnson, Classification pyramidale, Ensemble maximale-ment relié, Hiérarchie k -faible, Ultramétrie

SUMMARY – *Clustering systems with or without overlapping clusters and their dissimilarities. The bijection of Benzécri and Johnson sets up a one-one correspondence between hierarchical clusterings and ultrametric dissimilarities [Johnson, 1967 ; Benzécri, 1973]. New clustering structures, which were introduced during the 1980’s, include overlapping clusters and extend this bijection. In this text, we are concerned with a general framework that enables us to represent and compare several previously proposed extensions of the bijection of Benzécri and Johnson. We consider, more particularly, the extensions of this bijection that exist between certain general types of clustering systems that include overlapping clusters, on one hand, and the set of all the dissimilarities, on the other hand.*

KEYWORDS – Bijection of Benzécri and Johnson, k -weak hierarchy, Maximal linked subset, Pyramidal clustering, Ultrametric

1. INTRODUCTION

La classification hiérarchique est une technique très populaire en fouille de données. Son usage, fréquent dans de nombreux domaines scientifiques, tient à plusieurs caractéristiques communes aux différentes méthodes de classification hiérarchique : visualisation intuitive et claire de l’ensemble des classes selon un arbre appelé dendrogramme, complexité réduite de la plupart des algorithmes de classification hiérarchique, procédure d’identification automatique du ou des bon(s) nombre(s) de classes. La bijection de Benzécri-Johnson [Johnson, 1967 ; Benzécri, 1973] montre, en outre,

¹Centre de Recherche en Mathématiques de la Décision (CEREMADE), Université Paris-Dauphine, Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, Patrice.Bertrand@ceremade.dauphine.fr.

²Projet AxIS, Inria Rocquencourt, Domaine de Voluceau, B.P. 105, 78153 Le Chesnay.

que chaque classification hiérarchique induit de façon biunivoque une dissimilarité, plus précisément une ultramétrie, sur l'ensemble des objets classés. Il en résulte que l'on peut évaluer la qualité d'une classification hiérarchique en comparant l'ultramétrie qu'elle induit aux données lorsque celles-ci se présentent sous la forme d'une dissimilarité. Cependant, l'utilisation d'une méthode de classification hiérarchique est limitée par la structure même qui la définit : deux classes quelconques d'une classification hiérarchique sont soit disjointes, soit incluses l'une dans l'autre ; en d'autres termes, une hiérarchie ne contient pas de classes recouvrantes, *i.e.* pour toute classe A , il n'existe pas de classe B telle que $A \cap B \notin \{A, B, \emptyset\}$. Or il est relativement fréquent que des données possèdent une structure de classification comportant un ou plusieurs recouvrements. Par conséquent, les classes obtenues par ajustement des données au modèle de classification hiérarchique doivent être évaluées indépendamment de la qualité de l'algorithme utilisé, notamment dans le cas usuel d'absence d'information sur l'existence ou non de recouvrements dans la structure réelle des données. Afin de s'affranchir des limites résultant d'un modèle sans classe recouvrante, plusieurs auteurs ont proposé divers types de *systèmes de classification non floue* qui étendent la classification hiérarchique. Dans ce qui suit, nous désignerons par *système de classification non floue*² toute collection de parties non vides de l'ensemble X des objets à classer, ces parties étant alors appelées *classes* et la partie X étant nécessairement une classe. Un tel système de classification (non floue) est habituellement muni d'une application à valeurs dans $\mathbb{R}^+$ qui indique le degré d'hétérogénéité de chaque classe. Nous ne considérerons que des extensions de la classification hiérarchique qui sont des systèmes de classification non floue que, par la suite, nous désignerons plus simplement sous le terme de *systèmes de classification*.

Parmi les systèmes de classification qui étendent la classification hiérarchique, certains ont donné lieu à des applications à des données réelles, comme les classifications pyramidales introduites par E. Diday [1986] et B. Fichet [1987], ou les hiérarchies faibles proposées par H.-J. Bandelt et A. W. M. Dress [1989]. Pour chacun de ces types de systèmes de classification, l'existence de classes recouvrantes est due à des propriétés particulières de structure. Par exemple, les classes d'une classification pyramidale ont la possibilité de s'intersecter proprement, puisque la propriété qui définit la structure est que toutes les classes soient des intervalles d'un même ordre total. Dans le cas d'une hiérarchie faible, l'intersection de trois classes doit être égale à l'intersection de deux d'entre elles, donc des classes recouvrantes peuvent exister. D'autres extensions, plus générales, ont un intérêt essentiellement théorique, comme les systèmes de classification en bijection avec l'ensemble de toutes les dissimilarités [Batbedat, 1988 ; Bertrand, 2000], et les hiérarchies k -faibles (cf. par exemple, [Bertrand et Janowitz, 2003]).

Dans ce texte, nous proposons une présentation des principales propriétés de la classification hiérarchique et de plusieurs de ses extensions, ainsi que de leurs dissimilarités associées, dans le cadre unifié qui résulte de la généralisation de la

²Une classification est dite *non floue* ou *dure* (*floue*, dans le cas contraire) lorsque chaque élément de X possède, pour toute classe, un degré d'appartenance égal à 0 ou 1 (compris entre 0 et 1, dans le cas contraire). Il en résulte qu'une classification non floue de X est identifiable à un ensemble de parties non vides de X .

bijection de Benzécri-Johnson. Plus formellement, nous présenterons chaque type de système de classification avec le type particulier de dissimilarité qui lui correspond, selon le principe général suivant : à chaque système de classification $\mathcal{C}$ muni d'une application f de $\mathcal{C}$ dans $\mathbb{R}^+$ indiquant le degré d'hétérogénéité de chaque classe appartenant à $\mathcal{C}$, on associe, par la correspondance Φ , l'application $\rho = \Phi(\mathcal{C}, f)$ définie de $X \times X$ sur $\mathbb{R}^+$ par

$$\rho(x, y) = \min\{f(A) : A \in \mathcal{C}, x, y \in A\}.$$

Il est immédiat de vérifier que ρ est une dissimilarité dès que $\bigcup f^{-1}(0) = X$.

La majorité des résultats présentés ci-dessous ont été établis dans la littérature. Afin de rendre le texte autosuffisant, nous indiquerons en détail la plupart des preuves, ou au moins leurs principes. Dans la section suivante, *i.e.* la Section 2, nous rappelons des définitions élémentaires relatives aux structures de classification, et présentons l'application Φ qui induit, par restriction, une bijection entre l'ensemble de toutes les dissimilarités et une classe particulière de systèmes de classification, chaque classe étant munie d'une valeur indiquant son degré d'hétérogénéité. La Section 3 est essentiellement consacrée à l'étude de cette bijection obtenue par restriction de l'application Φ . Dans la Section 4, nous présentons la bijection de Benzécri-Johnson et deux approches pour établir cette bijection. Dans la Section 5, nous nous intéressons aux hiérarchies faibles et aux hiérarchies k -faibles, ainsi qu'aux dissimilarités qui leur sont associées par la correspondance Φ . Dans cette section, nous présentons aussi brièvement la classification pyramidale, qui est une autre extension de la classification hiérarchique.

2. GÉNÉRALITÉS

Nous commençons par rappeler quelques définitions et notations générales en théorie de la classification automatique, comme les notions de système de classification valué, d'indigage et de dissimilarité. Nous présentons ensuite la correspondance Φ qui associe à chaque système de classification valué une dissimilarité. Cette correspondance Φ nous permettra de situer, ultérieurement dans les Sections 4 et 5, la classification hiérarchique et ses diverses extensions dans un cadre unifié.

Nous utiliserons les notations et définitions générales suivantes. Étant donné un ensemble E quelconque, on note, comme c'est l'usage, $|E|$ son cardinal et $\mathcal{P}(E)$ l'ensemble de ses parties. Si k est un entier tel que $0 \leq k \leq |E|$, on note $E^{(k)}$ l'ensemble des parties de E à k éléments. Pour toute partie $A \subseteq E$ et tout $b \in E$, $Ab = A \cup \{b\}$. L'inclusion stricte (resp. large) entre deux ensembles A et B sera notée $A \subset B$ (resp. $A \subseteq B$). Si f est une application définie sur un ensemble E , alors sa restriction à une partie A de E sera notée $f|_A$. Si $\mathcal{F}$ est un ensemble de parties de E , alors $\widehat{\mathcal{F}}$ désignera l'ensemble de toutes les intersections non vides d'éléments de $\mathcal{F}$. L'ensemble $\mathcal{F}$ sera dit *fermé par intersection non vide*, ou simplement *fermé*, si $\mathcal{F} = \widehat{\mathcal{F}}$.

Dans ce qui suit, X désigne l'ensemble des objets à classer. Rappelons qu'un *recouvrement* de X est un ensemble $\mathcal{F}$ de parties non vides de X , deux à deux

incomparables au sens de l'inclusion et telles que $\bigcup \mathcal{F} = X$. Une *partition* de X est un recouvrement dont les éléments sont deux à deux disjoints.

Un *système de classification* sur X est une collection de parties non vides de X qui contient X et dont les éléments sont appelés *classes*. Soit $\mathcal{C}$ un système de classification sur X . Nous dirons que $\mathcal{C}$ est *total* s'il contient tous les singletons de X , *i.e.* si $\{x\}$ est une classe pour tout $x \in X$. Si f est une application de $\mathcal{C}$ dans $\mathbb{R}^+$ telle que $\bigcup f^{-1}(0) = X$, alors nous dirons que $(\mathcal{C}, f)$ est un système de classification *valué* et que f est une *valuation* définie sur $\mathcal{C}$. L'ensemble des systèmes de classification valués sur X sera noté $\mathbb{S}[X]$, ou simplement $\mathbb{S}$ s'il n'y a pas d'ambiguïté sur le choix de X .

Soit $(\mathcal{C}, f)$ un système de classification valué. Nous dirons que f est un *préindice* (resp. *indice strict*) sur $\mathcal{C}$ si f est croissant (resp. strictement croissant) de $(\mathcal{C}, \subseteq)$ dans $(\mathbb{R}^+, \leq)$. Autrement dit, f est un préindice sur $\mathcal{C}$ si pour tous $A, B \in \mathcal{C}$ tels que $A \subseteq B$, on a $f(A) \leq f(B)$. De même, f est un indice strict sur $\mathcal{C}$ si pour tous $A, B \in \mathcal{C}$ tels que $A \subset B$, on a $f(A) < f(B)$. Si f est un préindice (resp. un indice strict) sur $\mathcal{C}$, alors nous dirons que $(\mathcal{C}, f)$ est un système de classification préindiqué (resp. indicé).

DÉFINITION 1. Soit $(\mathcal{C}, f)$ un système de classification valué. Nous notons $\mathcal{M}(\mathcal{C}, f)$ la partie de $\mathcal{C}$ qui est définie par :

$$\mathcal{M}(\mathcal{C}, f) = \{B \in \mathcal{C} : f(B) < f(C) \text{ pour tout } C \in \mathcal{C} \text{ tel que } B \subset C\}.$$

Autrement dit, $\mathcal{M}(\mathcal{C}, f)$ désigne l'ensemble des éléments maximaux (au sens de l'inclusion) de chaque *strate* du système de classification valué $(\mathcal{C}, f)$, une *strate* de $(\mathcal{C}, f)$ désignant toute partie de la forme $f^{-1}(h)$ où h est une valeur prise par f , *i.e.* $h \in f(\mathcal{C})$.

Afin d'établir le Théorème 1, nous introduisons ci-dessous les Lemmes 1, 2 et 3 qui concernent les relations entre deux systèmes de classification sur X , préindiqués, notés $(\mathcal{C}, f)$ et $(\mathcal{B}, g)$, et vérifiant $\mathcal{C} \subseteq \mathcal{B}$ et $g|_{\mathcal{C}} = f$. Ces lemmes seront appliqués, dans la Section 3, dans le cas particulier où $(\mathcal{B}, g) = (\mathcal{P}(X), \text{diam}_\rho)$ et $\rho = \Phi(\mathcal{C}, f)$ (cf. Preuves de la Proposition 5 et du Corollaire 2 en Section 3).

LEMME 1. [Bertrand, 2000]. Soient $(\mathcal{C}, f)$ et $(\mathcal{B}, g)$ deux systèmes de classification sur X , préindiqués et tels que $\mathcal{C} \subseteq \mathcal{B}$. Si, pour tout $B \in \mathcal{B}$, $g(B) = \min\{f(Y) : B \subseteq Y \text{ et } Y \in \mathcal{C}\}$ alors $\mathcal{M}(\mathcal{C}, f) \subseteq \mathcal{M}(\mathcal{B}, g)$.

Preuve. Soit $M \in \mathcal{M}(\mathcal{C}, f)$ et supposons qu'il existe $B \in \mathcal{B}$ tel que $M \subset B$. Notons que $g(M) = f(M)$ car $M \in \mathcal{C}$. Par définition de g , il existe $C \in \mathcal{C}$ tel que $g(B) = f(C)$ avec $B \subseteq C$. D'où $M \subset B \subseteq C$, et donc $f(M) < f(C)$, car $M \in \mathcal{M}(\mathcal{C}, f)$. Finalement, $g(M) = f(M) < f(C) = g(B)$, ce qui prouve que $M \in \mathcal{M}(\mathcal{B}, g)$. $\square$

Bandelt et Dress [1994] ont établi, sous une forme différente, le lemme suivant. Le lecteur pourra se référer à [Bertrand, 2000] pour la preuve, essentiellement technique, de ce résultat.

LEMME 2. Soit $(\mathcal{C}, f)$ un système de classification préindiqué défini sur X , et soit $(\mathcal{B}, g)$ un système de classification valué défini sur X . Si $\mathcal{C} \subseteq \mathcal{B}$ et $g|_{\mathcal{C}} = f$, alors les conditions suivantes sont équivalentes :

- (i) g est croissante et $\mathcal{M}(\mathcal{B}, g) \subseteq \mathcal{C}$;
- (ii) pour tout $B \in \mathcal{B}$, $g(B) = \min\{f(Y) : B \subseteq Y \text{ et } Y \in \mathcal{C}\}$.

Rappelons qu'une *dissimilarité*, notée δ , est une application de $X \times X$ dans $\mathbb{R}^+$ telle que pour tous $a, b \in X$, $\delta(a, b) = \delta(b, a) \geq \delta(a, a) = 0$. La dissimilarité δ est dite *propre* (ou *définie*) si $\delta(a, b) = 0$ entraîne $a = b$. Si $\delta(a, b) = 0$ implique $\delta(a, c) = \delta(b, c)$ pour tout $c \in X$, alors la dissimilarité δ est dite *semi-propre* (ou *semi-définie*). Par la suite, nous noterons $\mathbb{D}[X]$, ou simplement $\mathbb{D}$, l'ensemble de toutes les dissimilarités définies sur l'ensemble X .

Les systèmes de classification valués sur X ont pour propriété d'être en correspondance naturelle avec les dissimilarités définies sur X . Cette correspondance, notée Φ dans ce texte, consiste à associer à chaque système de classification valué $(\mathcal{C}, f) \in \mathbb{S}$ la dissimilarité $\rho = \Phi(\mathcal{C}, f) \in \mathbb{D}$ qui est définie sur X par, pour tous $x, y \in X$, $\rho(x, y) = \min\{f(A) : A \in \mathcal{C}, x, y \in A\}$.

Il est clair que pour tous $x, y \in X$, on a $\rho(x, y) = \rho(y, x)$ et que $\rho(x, x) = 0$, car $\bigcup f^{-1}(0) = X$. Nous dirons que ρ , qui est donc une dissimilarité définie sur X , est la dissimilarité *induite* par le système de classification valué $(\mathcal{C}, f)$.

Remarque 1. Si f est un préindice sur $\mathcal{C}$ et si $\mathcal{C}$ est stable par intersection non vide, alors la dissimilarité ρ induite par $(\mathcal{C}, f)$ peut se calculer de la façon suivante :

$$\rho(x, y) = f\left(\bigcap\{A \in \mathcal{C} : x, y \in A\}\right).$$

Par la suite, nous notons $\{x, y\}_{\mathcal{C}}^* = \bigcap\{A \in \mathcal{C} : x, y \in A\}$ qui, lorsque $\mathcal{C}$ est stable par intersection non vide, coïncide avec la plus petite classe (au sens de l'inclusion) contenant à la fois x et y . Plus généralement, nous notons $B_{\mathcal{C}}^* = \bigcap\{A \in \mathcal{C} : B \subseteq A\}$ pour toute partie B de X . S'il n'y a pas d'ambiguïté sur le choix de $\mathcal{C}$, nous écrirons B^* au lieu de $B_{\mathcal{C}}^*$.

Soit une dissimilarité $\delta \in \mathbb{D}[X]$. Le *diamètre* d'une partie non vide A de X , noté $\text{diam}_{\delta} A$, ou $\text{diam } A$ s'il n'y a pas de confusion possible sur le choix de δ , est défini par $\text{diam } A = \max\{\delta(a, b) : a, b \in A\}$.

Le résultat suivant est une conséquence directe des définitions précédentes.

LEMME 3. Soit $(\mathcal{C}, f)$ un système de classification fermé, préindiqué et dont la dissimilarité induite est notée ρ . Pour tout $A \subseteq X$, on a $\text{diam}_{\rho} A \leq \text{diam}_{\rho} A^* \leq f(A^*)$.

Après quelques rappels de propriétés vérifiées par les dissimilarités, nous allons montrer que $\Phi : \mathbb{S} \mapsto \mathbb{D}$ est surjective (cf. Corollaire 1). L'étude détaillée de cette correspondance est reportée à la section suivante.

Une partie M de X est dite *maximalement reliée au sens de δ* si, pour toute partie N telle que $M \subset N$, on a $\text{diam}_{\delta} M < \text{diam}_{\delta} N$. Une partie M maximalement

reliée est dite de *niveau* h si $\text{diam}_\delta M = h$. Il est facile de vérifier qu'une partie de X est maximalelement reliée si et seulement si elle est une clique maximale pour au moins un des graphes seuils associés aux relations binaires symétriques $T\delta(h)$ (avec $h \in \mathbb{R}^+$) qui sont définies par $T\delta(h) = \{(a, b) : \delta(a, b) \leq h\}$. L'ensemble des parties maximalelement reliées est noté $M_L(T\delta)$.

Remarque 2. Il est clair que $M_L(T\delta) = \mathcal{M}(\mathcal{P}(X) \setminus \{\emptyset\}, \text{diam}_\delta)$ et que, de plus, $(\widehat{M_L(T\delta)}, \text{diam}_\delta)$ est un système de classification préindiqué, car $\bigcup \text{diam}_\delta^{-1}(0) = X$.

PROPOSITION 1. *Toute dissimilarité δ est induite par $(\widehat{M_L(T\delta)}, \text{diam}_\delta)$.*

Preuve. Soit ρ la dissimilarité induite par $(\widehat{M_L(T\delta)}, \text{diam}_\delta)$, et soient deux éléments quelconques x, y de X . Par définition, il existe une partie M_{xy} maximalelement reliée au sens de δ telle que $x, y \in M_{xy}$ et $\text{diam } M_{xy} = \delta(x, y)$. Donc $\delta(x, y) \geq \rho(x, y)$. De plus, $\text{diam } A \geq \delta(x, y)$ pour toute partie $A \in \widehat{M_L(T\delta)}$ qui contient $x, y \in X$. Donc $\rho(x, y) \geq \delta(x, y)$ et par conséquent $\rho(x, y) = \delta(x, y)$. D'où $\rho = \delta$. $\square$

La Proposition 1 a pour conséquence immédiate le résultat suivant.

COROLLAIRE 1. *L'application Φ est surjective.*

Remarque 3. Soit ρ la dissimilarité induite par un système de classification arbitraire, notée $(\mathcal{C}, f)$. D'après la Proposition 1, on a $\rho = \Phi(\mathcal{C}, f) = \Phi(\widehat{M_L(T\rho)}, \text{diam}_\rho)$. Pour construire une restriction de Φ qui soit injective, cette égalité suggère de considérer l'ensemble, noté $\mathbb{S}_1$, des systèmes de classification valués $(\mathcal{C}, f)$ qui vérifient la condition suivante, notée (1) :

- (1) Il existe $\delta \in \mathbb{D}[X]$ tel que $\mathcal{C} = \widehat{M_L(T\delta)}$ et $f = \text{diam}_\delta$.

Il est immédiat de vérifier que la restriction $\Phi|_{\mathbb{S}_1}$, notée Φ_1 par la suite, est une bijection de $\mathbb{S}_1$ sur l'ensemble $\mathbb{D}$. Nous déterminerons une caractérisation de $\mathbb{S}_1$ (cf. Corollaire 2) qui porte sur les propriétés directement vérifiées par $\mathcal{C}$ et par f pour chaque élément $(\mathcal{C}, f) \in \mathbb{S}_1$, sans recourir au critère (1) qui, fondé sur l'existence d'une dissimilarité δ , s'avère difficile à interpréter et à vérifier.

Les bijections entre des types connus de systèmes de classification valués, d'une part, et les dissimilarités qui leur sont associées, d'autre part, sont des restrictions de l'application Φ , plus précisément des restrictions de $\Phi_1 = \Phi|_{\mathbb{S}_1}$. C'est le cas par exemple de la bijection entre les hiérarchies faibles fermées indicées et les dissimilarités vérifiant la condition des 4 points de Bandelt [Bertrand et Janowitz, 2002]. En outre, d'après la Remarque 3 et la Proposition 1, chaque dissimilarité δ est représentée, au sens de la bijection Φ_1 , par le système de classification $\widehat{M_L(T\delta)}$ qui est préindiqué par la fonction diam_δ . Toutefois, certains auteurs ont représenté chaque dissimilarité δ d'un type particulier par une collection de parties *a priori* distincte du système $\widehat{M_L(T\delta)}$: l'ensemble des boules qui forment une hiérarchie dans le cas d'une ultramétrie [Benzécri, 1973] et celui des 2-boules qui forment une hiérarchie faible fermée dans le cas d'une dissimilarité satisfaisant la condition des 4 points de Bandelt [Diatta et Fichet, 1994, 1998]. Cependant, les collections de parties ainsi

considérées par ces auteurs pour représenter δ coïncident effectivement avec $\widehat{M}_L(T\delta)$ pour les structures de classification spécifiques qu'ils ont envisagées.

On rappelle qu'une *boule fermée de centre $x \in X$ et de rayon $r \geq 0$* , au sens d'une dissimilarité δ , est la partie de X , notée $\mathbf{B}_\delta(x, r)$ ou aussi $\mathbf{B}(x, r)$ s'il n'y a pas d'ambiguïté sur le choix de δ , qui est définie par $\mathbf{B}_\delta(x, r) = \{y \in X : \delta(x, y) \leq r\}$. Par la suite, on note $\text{Boules}(X, \delta)$ l'ensemble des boules fermées au sens d'une dissimilarité δ .

DÉFINITION 2. [Diatta et Fichet, 1994]. *On appelle 2-boule générée par $a, b \in X$, et on note $\mathbf{B}_{ab}^\delta$, ou simplement $\mathbf{B}_{ab}$, la partie définie par*

$$\mathbf{B}_{ab}^\delta = \mathbf{B}_\delta(a, \delta(a, b)) \cap \mathbf{B}_\delta(b, \delta(a, b)).$$

On notera que $\mathbf{B}_{ab}^\delta = \{x \in X : \max\{\delta(a, x), \delta(b, x)\} \leq \delta(a, b)\}$. Prolongeant la notion de 2-boule, on obtient la définition d'une k -boule (avec $k \geq 2$) en remplaçant la paire $\{a, b\} \subseteq X$ par une partie A de X de taille k .

DÉFINITION 3. [Bertrand et Janowitz, 2003]. *Soit A une partie de X de taille $k \geq 2$. On appelle k -boule générée par A , et on note $\mathbf{B}_A^\delta$ ou simplement $\mathbf{B}_A$, la partie définie par*

$$\mathbf{B}_A^\delta = \bigcap_{a \in A} \mathbf{B}_\delta(a, \text{diam}_\delta A) = \{x \in X : \delta(a, x) \leq \text{diam}_\delta A, \text{ pour tout } a \in A\}.$$

Remarque 4. On notera que la Définition 2 qui concerne la 2-boule, notée $\mathbf{B}_{ab}^\delta$, n'exclut pas la possibilité que l'on ait $a = b$. Au contraire, une 2-boule au sens de la Définition 3 exclut la possibilité que $a = b$. Aussi pour éviter les confusions, nous utiliserons la notation $\mathbf{B}_{ab}$ uniquement lorsque la possibilité $a = b$ n'est pas à exclure.

Étant donné une partie A quelconque de X et une dissimilarité δ , nous considérons l'ensemble des éléments de $M_L(T\delta) = \mathcal{M}(\mathcal{P}(E), \text{diam}_\delta)$ qui contiennent A et dont le diamètre est égal à $\text{diam}_\delta(A)$. On notera $\mathcal{M}_\delta(A)$ cet ensemble, voire simplement $\mathcal{M}(A)$ s'il n'y a pas d'ambiguïté sur le choix de δ . Plus formellement, nous avons donc :

$$\mathcal{M}(A) = \{M \in M_L(T\delta) : A \subseteq M, \text{diam } M = \text{diam } A\}.$$

La proposition suivante énonce les relations existant entre les parties maximale-ment reliées appartenant à $\mathcal{M}(A)$, d'une part, et la k -boule générée par A , d'autre part.

PROPOSITION 2. [Bertrand et Janowitz, 2002]. *Pour toute partie A non vide de X et pour toute dissimilarité $\delta \in \mathbb{D}[X]$, on a :*

$$(1) \quad \mathbf{B}_A = \bigcup \mathcal{M}(A),$$

$$(2) \text{ diam } \mathbf{B}_A = \text{diam } A \iff \mathbf{B}_A \in \mathcal{M}(A) \iff |\mathcal{M}(A)| = 1.$$

Preuve. (1). Soient $M \in \mathcal{M}(A)$ et $x \in M$. On a $\delta(a, x) \leq \text{diam } M = \text{diam } A$, pour tout $a \in A$, car $A \subseteq M$. Donc $x \in \mathbf{B}_A$ et $\bigcup \mathcal{M}(A) \subseteq \mathbf{B}_A$. Inversement, si $x \in \mathbf{B}_A$, alors Ax est une partie reliée de niveau $\text{diam } A$. Donc $x \in M$ pour un certain $M \in \mathcal{M}(A)$, et par conséquent, $\mathbf{B}_A \subseteq \bigcup \mathcal{M}(A)$.

(2) se déduit des implications, évidentes d'après (1), qui sont les suivantes :

$$\text{diam } \mathbf{B}_A = \text{diam } A \Rightarrow \mathbf{B}_A \in \mathcal{M}(A) \Rightarrow |\mathcal{M}(A)| = 1 \Rightarrow \text{diam } \mathbf{B}_A = \text{diam } A. \quad \square$$

3. REPRÉSENTATIONS D'UNE DISSIMILARITÉ PAR UN SYSTÈME DE CLASSIFICATION UNIQUE

Nous avons vu à la section précédente (cf. Remarque 3) qu'un choix possible d'un ensemble $\mathbb{S}_1 \subseteq \mathbb{S}$, tel que Φ_1 soit bijective, est l'ensemble $\mathbb{S}_1$ défini par :

$$\mathbb{S}_1 = \{(\mathcal{C}, f) \in \mathbb{S} : \text{il existe } \delta \in \mathbb{D} \text{ tel que } \mathcal{C} = \widehat{M}_L(T\delta) \text{ et } f = \text{diam}_\delta\}.$$

Afin de mettre en évidence les propriétés des éléments $(\mathcal{C}, f) \in \mathbb{S}_1$, nous allons déterminer une caractérisation de l'ensemble $\mathbb{S}_1$ qui ne soit pas fondée sur l'existence d'une dissimilarité pour caractériser chacun de ses éléments.

Remarquons tout d'abord que si $|X| \geq 2$, alors il existe au moins un système de classification valué $(\mathcal{C}, f)$ qui contient deux classes A, B de $\mathcal{C}$ telles que $A \subset B$ et $f(A) = f(B)$. Il est clair que la dissimilarité $\rho = \Phi(\mathcal{C}, f)$ est également induite par le système de classification valué $(\mathcal{C}_A, f|_{\mathcal{C}_A})$, avec $\mathcal{C}_A = \mathcal{C} \setminus \{A\}$. Il en résulte que, pour qu'une restriction de Φ à un sous-ensemble $\mathbb{S}'$ de $\mathbb{S}$ soit injective, il est nécessaire que $\mathbb{S}'$ ne contienne pas deux systèmes de classification valués qui diffèrent entre eux par des opérations d'ajout et/ou de retrait d'une classe contenue dans une autre, les deux classes admettant la même valeur de valuation pour le système de classification les contenant toutes les deux. Un choix possible pour $\mathbb{S}'$ est alors l'ensemble des systèmes de classification dits *faiblement indicés*, ou toute partie de cet ensemble de systèmes de classification dits faiblement indicés.

DÉFINITION 4. Soient $(\mathcal{C}, f)$ et $(\mathcal{B}, g)$ deux systèmes de classification préindicés définis sur le même ensemble X . On suppose de plus que $\mathcal{B}$ est fermé. Nous dirons que $(\mathcal{C}, f)$ est *faiblement indicé par rapport à* $(\mathcal{B}, g)$ si pour tout $A \in \mathcal{C}$ et tout $B \in \mathcal{B}$, on a :

$$(A \subset B \text{ et } f(A) = g(B)) \Rightarrow A = \bigcap \{Y \in \mathcal{B} : A \subset Y\}.$$

Si $(\mathcal{C}, f)$ est faiblement indicé par rapport à $(\mathcal{C}, f)$, alors nous dirons que $(\mathcal{C}, f)$ est *faiblement indicé*. En d'autres termes, un système de classification préindicé $(\mathcal{C}, f)$ est faiblement indicé si et seulement si $\mathcal{C}$ est fermé et, si pour tous $A, B \in \mathcal{C}$, on a :

$$(A \subset B \text{ avec } f(A) = f(B)) \Rightarrow A = \bigcap \{C \in \mathcal{C} : A \subset C\}.$$

Il est clair que si un système de classification fermé est indicé, alors il est aussi faiblement indicé. La proposition suivante donne une caractérisation des systèmes de classification faiblement indicés.

PROPOSITION 3. [Bertrand, 2000]. *Soit $\mathcal{C}$ un système de classification défini sur X , fermé et préindiqué par f . Les deux conditions suivantes sont équivalentes :*

- (i) $(\mathcal{C}, f)$ est faiblement indicé ;
- (ii) $\mathcal{C} = \widehat{\mathcal{M}}(\mathcal{C}, f)$.

Toutefois, la restriction de Φ aux systèmes de classification faiblement indicés n'est en général pas bijective. Plus précisément, la Proposition 4 ci-dessous suggère qu'une partie seulement des relations définissant un élément $(\mathcal{C}, f) \in \mathbb{S}_1$, i.e. $\mathcal{C} = \widehat{M}_L(T\delta)$ et $f = \text{diam}_\delta$, est impliquée par la propriété d'indigage faible. Une seconde démonstration de l'équivalence (i) $\iff$ (ii) de la Proposition 4 est donnée dans la preuve du corollaire 3.2 dans [Bertrand, 2000].

PROPOSITION 4. *Soient $(\mathcal{C}, f)$ un système de classification préindiqué défini sur X et δ une dissimilarité sur X . Les conditions suivantes sont équivalentes :*

- (i) $\mathcal{C} \subseteq \widehat{M}_L(T\delta)$;
- (ii) $(\mathcal{C}, \text{diam}_\delta)$ est faiblement indicé par rapport à $(M_L(T\delta), \text{diam}_\delta)$;
- (iii) $(\mathcal{C}, \text{diam}_\delta)$ est faiblement indicé par rapport à $(\widehat{M}_L(T\delta), \text{diam}_\delta)$.

Preuve. (i) $\Rightarrow$ (ii). Supposons $\mathcal{C} \subseteq \widehat{M}_L(T\delta)$. Soient $A \in \mathcal{C}$ et $B \in M_L(T\delta)$ avec $A \subset B$ et $\text{diam } A = \text{diam } B$. Donc $A \notin M_L(T\delta)$. De $\mathcal{C} \subseteq \widehat{M}_L(T\delta)$, nous déduisons qu'il existe $\{M_i\}_{i \in I} \subseteq M_L(T\delta)$ tel que $A = \bigcap_{i \in I} M_i$. Comme $A \notin M_L(T\delta)$, il en résulte que l'on a $A = \bigcap \{M \in M_L(T\delta) : A \subset M\}$, ce qui prouve (ii).

(ii) $\Rightarrow$ (i). Soit $A \in \mathcal{C}$. Supposons que $A \notin M_L(T\delta)$. Donc il existe $Y \in M_L(T\delta)$ tel que $A \subset Y$ et $\text{diam } A = \text{diam } Y$. D'après (ii), $A = \bigcap \{M \in M_L(T\delta) : A \subset M\}$ et donc $A \in \widehat{M}_L(T\delta)$, ce qui prouve que $\mathcal{C} \subseteq \widehat{M}_L(T\delta)$.

(ii) $\iff$ (iii). Il est facile de vérifier que la condition (ii) équivaut à (iii). $\square$

Afin de trouver une formulation équivalente à (1) pour mettre en évidence les caractéristiques de chaque élément $(\mathcal{C}, f)$ de $\mathbb{S}_1$ (cf. Remarque 3), nous adoptons un point de vue différent qui s'inspire d'une caractérisation des recouvrements qui sont les cliques maximales d'un graphe. Gilmore [1962] a montré qu'un recouvrement $\mathcal{C}$ de X est l'ensemble des cliques maximales d'un graphe si et seulement si, pour tous $C_1, C_2, C_3 \in \mathcal{C}$, la partie $(C_1 \cap C_2) \cup (C_2 \cap C_3) \cup (C_3 \cap C_1)$ est contenue dans un élément de $\mathcal{C}$. Il existe une condition, notée (G_L) , qui étend cette caractérisation de Gilmore au cas de certains systèmes de classification valués (cf. Corollaire 2). La condition (G_L) a d'abord été introduite par Batbedat [1988], et ensuite étendue par Bandelt et Dress [1994] au cas de $n + 1$ parties de $\mathcal{C}$ ($n \geq 2$).

DÉFINITION 5. [Batbedat, 1988]. *Nous dirons qu'un système de classification valué $(\mathcal{C}, f)$ satisfait la condition (G_L) si, pour tous $C_1, C_2, C_3 \in \mathcal{C}$, il existe $B \in \mathcal{C}$ tel que :*

$$\begin{cases} f(B) \leq \max\{f(C_1), f(C_2), f(C_3)\}, \\ B \supseteq (C_1 \cap C_2) \cup (C_2 \cap C_3) \cup (C_3 \cap C_1). \end{cases}$$

De plus, si $(\mathcal{C}, f)$ est un système de classification fermé et préindiqué, alors il est immédiat de vérifier que (G_L) est équivalente à la condition définie par : pour tous $C_1, C_2, C_3 \in \mathcal{C}$, on a :

$$(\text{pour tous } i, j \text{ distincts, } C_i \cap C_j \neq \emptyset) \Rightarrow f\left(\left[\bigcup_{1 \leq i < j \leq 3} (C_i \cap C_j)\right]^*\right) \leq \max_{1 \leq i \leq 3} f(C_i).$$

PROPOSITION 5. [Bertrand, 2000]. *Si $(\mathcal{C}, f)$ est un système de classification fermé et préindiqué dont la dissimilarité induite est notée ρ , alors les conditions suivantes sont équivalentes :*

- (i) $(\mathcal{C}, f)$ satisfait la condition (G_L) ;
- (ii) Pour toute partie non vide A de X , $\text{diam}_\rho A = f(A^*)$;
- (iii) $\widehat{M_L}(T\rho) \subseteq \mathcal{C}$ et $f = (\text{diam}_\rho)_{|\mathcal{C}}$.

Preuve. Dans cette démonstration, nous reprenons, pour l'essentiel mais de façon plus concise, les arguments de la preuve de la proposition 3.5 dans [Bertrand, 2000].
(i) $\Rightarrow$ (ii). Raisonnons par induction sur la taille de la partie A considérée dans la propriété (ii). Cette propriété est clairement vraie si cette taille vaut 1 ou 2. Supposons que (ii) soit vraie pour toute taille de partie de X inférieure ou égale à $m - 1$, où $m \geq 3$. Soit A une partie de taille m . Comme $m \geq 3$, il existe une partition de A de la forme $\{A_1, A_2, A_3\}$. Posons $B_i = A \setminus A_i$ pour tout i . Les parties B_i^* étant deux à deux non disjointes, elles vérifient (G_L) d'après (i). Par hypothèse d'induction, il en résulte :

$$f\left(\left[\bigcup_{1 \leq i < j \leq 3} (B_i^* \cap B_j^*)\right]^*\right) \leq \max_{1 \leq i \leq 3} \text{diam}_\rho B_i^* \leq \text{diam}_\rho A.$$

Comme $A_i \subseteq B_j \cap B_k$ si $\{i, j, k\} = \{1, 2, 3\}$, on en déduit $f(A^*) \leq \text{diam}_\rho A$. D'après le Lemme 3, il en résulte $\text{diam}_\rho A = f(A^*)$, et donc (ii) est vraie pour toute taille de la partie A .

(ii) $\Rightarrow$ (iii). Pour toute partie A de X , on a $\text{diam}_\rho A = \text{diam}_\rho A_\mathcal{C}^*$ d'après (ii) et le Lemme 3. Donc $\mathcal{M}(\mathcal{P}(X), \text{diam}_\rho) \subseteq \mathcal{C}$ d'après le Lemme 2, et par conséquent $M_L(T\delta) \subseteq \mathcal{C}$ d'après la Remarque 2. Finalement, $\widehat{M_L}(T\delta) \subseteq \mathcal{C}$, car $\mathcal{C}$ est fermé. De plus, (ii) entraîne clairement que $f = (\text{diam}_\rho)_{|\mathcal{C}}$, ce qui prouve (iii).

(iii) $\Rightarrow$ (i). Remarquons tout d'abord que, comme $\widehat{M_L}(T\rho) = \mathcal{M}(\mathcal{P}(X), \text{diam}_\rho) \subseteq \mathcal{C}$ d'après (iii), il en résulte, d'après le Lemme 2, que $\text{diam}_\rho A = \text{diam}_\rho A_\mathcal{C}^*$ pour tout $A \subseteq X$. Donc $\text{diam}_\rho A = f(A_\mathcal{C}^*)$ d'après (iii).

Considérons $C_1, C_2, C_3 \in \mathcal{C}$ et notons $A = \bigcup_{1 \leq i < j \leq 3} (C_i \cap C_j)$. Par définition de A , pour tous $x, y \in A$, on a $x, y \in C_i$ pour au moins une valeur de $i \in \{1, 2, 3\}$. Donc

$$\text{diam}_\rho A \leq \max\{\text{diam}_\rho C_i : i \in \{1, 2, 3\}\}.$$

D'après (iii), pour tout i , $\text{diam}_\rho C_i = f(C_i)$. Donc $f(A_\mathcal{C}^*) \leq \max_{1 \leq i \leq 3} f(C_i)$, ce qui, d'après la définition de A , montre que (i) est vérifié. $\square$

COROLLAIRE 2. Soit $(\mathcal{C}, f)$ un système de classification valué défini sur X . Les conditions suivantes sont équivalentes :

- (i) $(\mathcal{C}, f) \in \mathbb{S}_1$;
- (ii) $(\mathcal{C}, f)$ est faiblement indicé, et il existe $\delta \in \mathbb{D}[X]$ tel que $\widehat{M_L}(T\delta) \subseteq \mathcal{C}$ et $f = \text{diam}_\delta$;
- (iii) $(\mathcal{C}, f)$ est faiblement indicé et satisfait la condition (G_L) ;
- (iv) $(\mathcal{C}, f) = (\widehat{M_L}(T\rho), \text{diam}_\rho)$, où ρ est la dissimilarité induite par $(\mathcal{C}, f)$.

Preuve. (i) $\Rightarrow$ (ii). D'après la Proposition 4, la condition (i) du corollaire entraîne que $(\mathcal{C}, \text{diam}_\delta)$ est faiblement indicé par rapport à $(\widehat{M_L}(T\delta), \text{diam}_\delta)$. Mais $f = \text{diam}_\delta$ et $\mathcal{C} = \widehat{M_L}(T\delta)$, donc $(\mathcal{C}, f)$ est faiblement indicé, ce qui prouve (ii).

(ii) $\Rightarrow$ (i). On suppose (ii). D'après la définition de $\mathbb{S}_1$, il suffit de montrer que $\mathcal{C} \subseteq \widehat{M_L}(T\delta)$. D'après la Proposition 5(ii), pour toute partie non vide A de X , on a $\text{diam}_\rho A = f(A^*)$. Il en résulte que $\text{diam}_\rho A = \min\{f(Y) : A \subseteq Y \text{ et } Y \in \mathcal{C}\}$. En appliquant le Lemme 1 avec $\mathcal{B} = \mathcal{P}(X) \setminus \{\emptyset\}$ et $g = \text{diam}_\rho$, nous déduisons que $\mathcal{M}(\mathcal{C}, f) \subseteq \mathcal{M}(\mathcal{P}(X) \setminus \{\emptyset\}, \text{diam}_\rho) = M_L(T\delta)$. Or $(\mathcal{C}, f)$ est faiblement indicé, donc $\mathcal{C} = \widehat{\mathcal{M}}(\mathcal{C}, f)$ d'après la Proposition 3. Donc $\mathcal{C} = \widehat{\mathcal{M}}(\mathcal{C}, f) \subseteq \widehat{M_L}(T\delta)$, ce qui prouve que la condition (i) est vraie.

(ii) $\iff$ (iii) est une conséquence directe de la Proposition 5.

(iv) $\Rightarrow$ (i) par définition de $\mathbb{S}_1$, et (i) $\Rightarrow$ (iv), d'après la Proposition 1. $\square$

Remarque 5. Soit δ une dissimilarité quelconque. Il résulte du Corollaire 2 et de la Proposition 1 que $(\widehat{M_L}(T\delta), \text{diam}_\delta) \in \mathbb{S}_1$, i.e. $(\widehat{M_L}(T\delta), \text{diam}_\delta)$ est faiblement indicé et satisfait la condition (G_L) .

Le théorème suivant, introduit par Batbedat [1988] selon une formulation et une preuve différente, est une conséquence directe de la Remarque 3 et du Corollaire 2.

THÉORÈME 1. [Batbedat, 1988 ; Bertrand, 2000]. L'application Φ_1 est bijective de l'ensemble $\mathbb{S}_1$ des systèmes de classification faiblement indicés qui satisfont la condition (G_L) , sur l'ensemble $\mathbb{D}$ des dissimilarités. Chaque dissimilarité $\delta \in \mathbb{D}$ est associée, par la bijection réciproque Φ_1^{-1} , au système de classification faiblement indicé $(\widehat{M_L}(T\delta), \text{diam}_\delta) \in \mathbb{S}_1$.

Les résultats établis par Batbedat [1988] ne concernent que l'ensemble $\mathbb{D}_0$ des dissimilarités propres, et par conséquent, uniquement l'ensemble des systèmes de classification préindicés $(\mathcal{C}, f)$ qui induisent les dissimilarités propres, c'est-à-dire tels que $f^{-1}(0) = X^{(1)}$. Cependant cette différence avec les hypothèses que nous avons considérées, i.e. les dissimilarités sont quelconques, n'est pas significative. Aussi, dans ce qui suit, par souci d'homogénéité, nous supposons que les dissimilarités considérées sont quelconques lorsque nous présentons les principaux résultats de [Batbedat 1988], qui restent vrais dans ce cadre un peu plus général.

En complément au Théorème 1, Batbedat [1988] montre que les applications $\Phi_1 : \mathbb{S}_1 \mapsto \mathbb{D}$ et $\Phi_1^{-1} : \mathbb{D} \mapsto \mathbb{S}_1$ sont des isomorphismes d'ordre. Ces isomorphismes

sont définis sur les structures ordonnées suivantes. Tout d'abord, l'ensemble $\mathbb{D}$ est muni de l'ordre $\preceq_d$ qui est défini pour tous $\delta_1, \delta_2 \in \mathbb{D}$ par :

$$\delta_1 \preceq_d \delta_2 \text{ si } \delta_1(x, y) \leq \delta_2(x, y), \text{ pour tous } x, y \in X.$$

L'ensemble $\mathbb{S}$ est muni du préordre $\preceq_c$ qui est défini pour tous $(\mathcal{A}, f), (\mathcal{B}, g) \in \mathbb{S}$ par :

$$(\mathcal{A}, f) \preceq_c (\mathcal{B}, g) \text{ si } \forall B \in \mathcal{B} \setminus X^{(1)}, \exists A \in \mathcal{A} \text{ tq } B \subseteq A \text{ et } f(A) \leq g(B).$$

Si $\mathcal{A} \subseteq \mathcal{B}$ avec $g|_{\mathcal{A}} = f$, alors il est immédiat de vérifier que $(\mathcal{B}, g) \preceq_c (\mathcal{A}, f)$. Par ailleurs, la relation $\preceq_c$ n'est pas antisymétrique dans le cas général, comme le montre le contre-exemple suivant.

Contre-exemple. Soit $X = \{a, b, c, d\}$. Considérons les systèmes de classification préinducés $(\mathcal{A}, f)$ et $(\mathcal{B}, g)$ définis par : $\mathcal{A} = \{\{a, b, c\}, \{b, c, d\}, X\} \cup X^{(1)}$ avec pour tout $A \in \mathcal{A}$, $f(A) = |A| - 1$, et $\mathcal{B} = \mathcal{A} \cup \{b, c\}$ avec $g|_{\mathcal{A}} = f$ et $g(\{b, c\}) = 2$ (cf. Figure 1). Les systèmes $(\mathcal{A}, f)$ et $(\mathcal{B}, g)$ sont tels que $f^{-1}(0) = g^{-1}(0) = X^{(1)}$. Nous avons $(\mathcal{B}, g) \preceq_c (\mathcal{A}, f)$, car $\mathcal{A} \subseteq \mathcal{B}$ et $g|_{\mathcal{A}} = f$. Mais on a aussi $(\mathcal{A}, f) \preceq_c (\mathcal{B}, g)$, puisque $\{b, c\}$, qui est le seul élément de $\mathcal{B} \setminus \mathcal{A}$, vérifie $\{b, c\} \subseteq \{a, b, c\} \in \mathcal{A}$ et $f(\{a, b, c\}) = 2 = g(\{b, c\})$. Donc $\preceq_c$ n'est pas antisymétrique.

FIGURE 1. Graphes de couverture des ordres $(\mathcal{A}, \subseteq)$ et $(\mathcal{B}, \subseteq)$. La hauteur $h(C)$ de chaque classe C représente $g(C)$ si $C \in \mathcal{B}$, donc $f(C)$ si $C \in \mathcal{A}$, car $f = g|_{\mathcal{A}}$. Chaque arête $\{C_1, C_2\}$ telle que $h(C_1) < h(C_2)$ vérifie $C_1 \subset C_2$ et est orientée implicitement de C_1 vers C_2 , i.e. du bas vers le haut comme dans un diagramme de Hasse.

Cependant, la restriction de $\preceq_c$ aux systèmes de classification indicés est antisymétrique, et donc est un ordre (cf. [Batbedat, 1988, proposition 2]). Soit $\mathbb{S}_2$ l'ensemble des systèmes de classification indicés qui vérifient (G_L) , et soit $\Phi_2 = \Phi|_{\mathbb{S}_2}$. Avec ces notations, on a le deuxième théorème de bijection suivant.

THÉORÈME 2. [Batbedat, 1988]. *L'application Φ_2 est un isomorphisme d'ordre de l'ensemble $\mathbb{S}_2$ des systèmes de classification indicés qui satisfont la condition (G_L) , muni de l'ordre $\preceq_c$, sur l'ensemble $\mathbb{D}$ des dissimilarités, muni de l'ordre $\preceq_d$. Chaque dissimilarité $\delta \in \mathbb{D}$ est associée, par l'isomorphisme réciproque Φ_2^{-1} , au système de classification indicé $(M_L(T\delta), \text{diam}_\delta) \in \mathbb{S}_2$.*

D'après les Théorèmes 1 et 2, chaque dissimilarité δ peut être représentée, plus précisément induite, de façon unique et indifféremment, par chacun des deux systèmes de classification préinducés $(\widehat{M_L}(T\delta), \text{diam}_\delta)$ et $(M_L(T\delta), \text{diam}_\delta)$. Ces systèmes peuvent admettre un nombre exponentiel de classes, alors que, dans le cas de la représentation de δ par $(\widehat{M_L}(T\delta), \text{diam}_\delta)$, seules les parties $\{x, y\}_{\widehat{M_L}(T\delta)}^*$, avec x, y variant dans X , sont utiles pour déterminer δ . Chaque dissimilarité peut ainsi être représentée par le système de classification $\{\{x, y\}_{\widehat{M_L}(T\delta)}^* : x, y \in X\}$, préinducé par la fonction diamètre, dont la taille est inférieure ou égale à $\frac{1}{2}|X|(|X| - 1)$. Nous concluons cette section avec la Proposition 6 : grâce à ce résultat, il est clair que, pour tout $x, y \in X$, la liste des éléments de chaque partie $\{x, y\}_{\widehat{M_L}(T\delta)}^*$ peut être calculée avec une complexité algorithmique bornée par $\mathcal{O}(|X|^2)$. Par conséquent, il est possible de déterminer le système $\{\{x, y\}_{\widehat{M_L}(T\delta)}^*\}_{x, y \in X}$, qui représente δ , avec un algorithme dont la complexité est bornée par $\mathcal{O}(|X|^4)$.

LEMME 4. Soient $\delta \in \mathbb{D}[X]$ et $x, y \in X$. Si M est une partie maximale reliée, au sens de δ , qui contient x et y , alors $\bigcap_{z \in M} \mathbf{B}(z, \text{diam}_\delta\{x, y, z\}) \subseteq M$.

Preuve. Si $w \notin M$, alors il existe $z \in M$ tel que $\delta(w, z) > \text{diam } M$. Par conséquent $w \notin \mathbf{B}(z, \text{diam } M)$, et donc $w \notin \mathbf{B}(z, \text{diam}\{x, y, z\})$. Il en résulte que w n'appartient pas à $\bigcap_{z \in M} \mathbf{B}(z, \text{diam}\{x, y, z\})$, d'où l'inclusion recherchée. $\square$

PROPOSITION 6. [Brucker et Barthélemy, 2007 ; Barthélemy et Brucker, 2008]. Pour toute dissimilarité δ et pour tous $x, y \in X$, on a :

$$\{x, y\}_{\widehat{M_L}(T\delta)}^* = \bigcap_{z \in X} \mathbf{B}(z, \max\{\delta(x, z), \delta(y, z), \delta(x, y)\}).$$

Preuve. D'après le Lemme 4, si M est une partie maximale reliée, au sens de δ , qui contient x et y , alors $\bigcap_{z \in M} \mathbf{B}(z, \text{diam}\{x, y, z\}) \subseteq M$. Par conséquent, $\bigcap_{z \in X} \mathbf{B}(z, \max\{\delta(x, z), \delta(y, z), \delta(x, y)\}) \subseteq \{x, y\}_{\widehat{M_L}(T\delta)}^*$. Pour l'inclusion inverse, d'après la Proposition 2, pour tout $z \in X$, $\mathbf{B}_{\{x, y, z\}} = \bigcup \mathcal{M}(\{x, y, z\})$. D'où, pour tout $z \in X$, $\{x, y\}_{\widehat{M_L}(T\delta)}^* \subseteq \bigcup \mathcal{M}(\{x, y, z\}) = \mathbf{B}_{\{x, y, z\}} \subseteq \mathbf{B}(z, \text{diam}\{x, y, z\})$. Donc $\{x, y\}_{\widehat{M_L}(T\delta)}^* \subseteq \bigcap_{z \in X} \mathbf{B}(z, \text{diam}\{x, y, z\})$. $\square$

Il est facile de vérifier que la classification hiérarchique et ses extensions les plus connues, *i.e.* les hiérarchies faibles indicées et les représentations pyramidales, sont des sous-ensembles de $\mathbb{S}_1$ ou de $\mathbb{S}_2$. Dans les sections suivantes, nous étudions les restrictions de la bijection Φ_1 à ces sous-ensembles.

4. CLASSIFICATION HIÉRARCHIQUE ET THÉORÈME DE BIJECTION DE BENZÉCRI-JOHNSON

La notion de hiérarchie en classification est d'un usage à la fois extrêmement répandu et très ancien. Il semble que cette notion fut d'abord introduite en botanique, à la

suite d'Aristote et de Théophraste, avant de connaître un plus grand essor avec l'ouvrage *Systema naturae* (1735) du botaniste suédois Carl von Linné, pour la classification des plantes, des animaux et des minéraux. Selon les définitions introduites plus haut, une hiérarchie est un système de classification qui ne contient pas de classes qui s'intersectent proprement, c'est-à-dire tel que deux classes quelconques sont soit disjointes, soit incluses l'une dans l'autre.

DÉFINITION 6. *Hiérarchie. Un système de classification est dit hiérarchique si, pour toute paire de classes $\{A, B\}$, on a $A \cap B \in \{\emptyset, A, B\}$ ou, de façon équivalente, si A et B sont comparables pour l'inclusion dès qu'elles ne sont pas disjointes. On appelle hiérarchie sur X tout système de classification sur X qui est hiérarchique et total. On dit que $(\mathcal{H}, f)$ est une hiérarchie indicée (sur X) si $\mathcal{H}$ est une hiérarchie sur X et f un indice strict sur $\mathcal{H}$ tel que $f(\{x\}) = 0$ pour tout $x \in X$. Par la suite, on notera $\mathbb{H}[X]$, ou plus simplement $\mathbb{H}$, l'ensemble des hiérarchies indicées sur X .*

D'après cette définition, toute hiérarchie est stable par intersection non vide.

Cette section est consacrée au théorème de bijection de Benzécri-Johnson qui établit que la restriction de Φ à l'ensemble des hiérarchies indicées est une bijection sur l'ensemble des ultramétries.³

DÉFINITION 7. *Ultramétrie. On appelle ultramétrie toute dissimilarité δ telle que, pour tous $x, y, z \in X$, on a*

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

L'ensemble des ultramétries sur X sera noté $\mathbb{U}[X]$, ou simplement $\mathbb{U}$.

Il est facile de vérifier qu'une dissimilarité δ est une ultramétrie si et seulement si tout triplet d'éléments de X forme un triangle isocèle avec la longueur de sa base plus petite que ou égale à la longueur des côtés égaux, au sens de δ .

L'existence de la bijection de Benzécri-Johnson (cf. Théorème 3) entraîne que l'on peut mesurer la qualité d'une hiérarchie indicée $(\mathcal{H}, f)$ en comparant la dissimilarité ρ qu'elle induit à la dissimilarité δ qui décrit les données (cette dissimilarité δ pouvant être donnée a priori, ou construite à partir des données, par exemple en calculant une distance euclidienne si les objets à classer sont décrits par plusieurs variables à valeurs réelles). De plus, cette bijection se généralise, via d'autres restrictions de l'application Φ , aux extensions des hiérarchies indicées, comme par exemple les classifications pyramidales ou les hiérarchies faibles (cf. les sections suivantes).

THÉORÈME 3. [Jardine *et al.*, 1967 ; Johnson, 1967 ; Benzécri, 1973]. *La restriction $\Phi|_{\mathbb{H}}$ qui, à chaque hiérarchie indicée, associe sa dissimilarité induite, est une bijection de l'ensemble $\mathbb{H}$ des hiérarchies indicées sur X sur l'ensemble $\mathbb{U}$ des ultramétries sur X .*

³Le lecteur désireux d'approfondir ses connaissances sur les propriétés combinatoires des hiérarchies et des ultramétries pourra se référer à deux articles de Leclerc [1981, 1985] sur ce sujet.

Preuves. Johnson [1967] et Benzécri [1973] ont donné deux preuves de ce théorème, qui commencent de façon identique par établir que $\Phi_{\mathbb{H}}$ est à valeurs dans $\mathbb{U}$. Montrons cette propriété. Soit $(\mathcal{H}, f)$ une hiérarchie indicée sur X . Pour tous $u, v \in X$, notons $A_{uv} = \{u, v\}_{(\mathcal{H}, f)}^*$. D'après ce qui précède, pour tous $x, y \in X$, on a $\rho(x, y) = f(A_{xy})$ avec $\rho = \Phi(\mathcal{H}, f)$. Soient $x, y, z \in X$. Les classes A_{xy} , A_{xz} et A_{yz} étant deux à deux d'intersection non vide, elles sont deux à deux comparables au sens de l'inclusion. Quitte à permuter x, y, z , nous pouvons donc supposer que $A_{xy} \subseteq A_{yz} \subseteq A_{xz}$, d'où $x \in A_{yz} \subseteq A_{xz}$, et par suite $A_{yz} = A_{xz}$. Par conséquent, $\rho(x, y) \leq \rho(y, z) = \rho(x, z)$, puisque f est croissant (strictement). Il en résulte que, pour tous $u, v, w \in X$, on a bien $\rho(u, v) \leq \max\{\rho(v, w), \rho(u, w)\}$. Par ailleurs, f est strictement croissant et $f(\{x\}) = 0$ pour tout $x \in X$, donc ρ est une dissimilarité propre et, par conséquent, une ultramétrie.

Pour montrer que $\Phi_{\mathbb{H}}$ est surjective, Johnson [1967] considère une ultramétrie δ et construit de façon algorithmique une hiérarchie indicée qui induit δ . L'algorithme commence avec la partition triviale $\mathcal{P}_0$ égale à l'ensemble $X^{(1)}$ des singletons de X et affectée d'un niveau $\alpha_0 = 0$. À l'étape $i \geq 0$, on dispose d'un partitionnement $\mathcal{P}_{i-1}$ et d'une matrice de dissimilarité inter-classes. Notons α_i la plus petite valeur de cette matrice. L'algorithme consiste alors à réunir deux classes qui sont comparables selon la valeur de dissimilarité α_i . On obtient ainsi une nouvelle partition $\mathcal{P}_i$ et il est prouvé que la nouvelle matrice de distances inter-classes n'engendre aucune perte d'information pour reconstituer δ . Il en résulte que l'ensemble des classes ainsi construites, et munies des niveaux α_i , forme une hiérarchie indicée dont la dissimilarité induite est égale à δ . Donc l'application $\Phi_{\mathbb{H}} : \mathbb{H} \mapsto \mathbb{U}$ est surjective. La preuve de Johnson ne contient pas de mention explicite de l'injectivité de $\Phi_{\mathbb{H}}$.

Une seconde preuve, plus formelle, de Benzécri [1973] est fondée sur les propriétés des boules fermées au sens d'une ultramétrie. Dans le cas où δ est une ultramétrie, le rayon d'une boule fermée B coïncide avec son diamètre et, de plus, tout point de B peut être choisi pour centre de B . Par ailleurs, cette preuve est fondée sur la proposition suivante :

PROPOSITION 7. [Benzécri, 1973]. *Soit $(\mathcal{H}, f)$ une hiérarchie indicée sur X , et soit ρ l'ultramétrie induite par $(\mathcal{H}, f)$. On a $\mathcal{H} = \text{Boules}(X, \rho)$ et $f = \text{diam}_\rho$.*

Il résulte directement de la Proposition 7 que $\Phi_{\mathbb{H}}$ est injective. Notons ray l'application qui, à chaque boule fermée $B \in \text{Boules}(X, \delta)$, associe son rayon $\text{ray}(B)$. Pour montrer que $\Phi_{\mathbb{H}}$ est surjective, on remarque tout d'abord qu'il est facile de vérifier que si δ est une ultramétrie, alors $(\text{Boules}(X, \delta), \text{ray})$ est une hiérarchie indicée. Il suffit donc d'établir que $\Phi(\text{Boules}(X, \delta), \text{ray}) = \delta$. D'après la Proposition 7, on a $\text{Boules}(X, \delta) = \text{Boules}(X, \rho)$ où ρ est l'ultramétrie induite par $(\text{Boules}(X, \delta), \text{ray})$. D'où, pour tous $x, x' \in X$, nous avons $\rho(x, x') = \text{ray}(B(x, x'))$, où $B(x, x')$ désigne la plus petite boule, au sens de δ , qui contient x et x' . Soit $\mathbf{B}_\delta(y, \lambda)$ une boule (fermée) contenant x et x' . Supposons qu'il existe un élément z dans $\mathbf{B}_\delta(x, \delta(x, x')) \setminus \mathbf{B}_\delta(y, \lambda)$. Donc $\delta(z, y) > \lambda \geq \delta(x, y)$. De plus, δ étant ultramétrie, le rayon de $\mathbf{B}_\delta(y, \lambda)$ coïncide avec son diamètre, par conséquent $\delta(x, z) \leq \delta(x, x') \leq \lambda$. Donc $\delta(z, y) > \max\{\delta(x, y), \delta(x, z)\}$: contradiction car δ est ultramétrie. D'où $\mathbf{B}_\delta(x, \delta(x, x')) \subseteq \mathbf{B}_\delta(y, \lambda)$, et donc on a

$B(x, x') = \mathbf{B}_\delta(x, \delta(x, x'))$. Il en résulte que $\rho(x, x') = \delta(x, x')$, ce qui prouve $\rho = \delta$, c'est-à-dire $\Phi(\text{Boules}(X, \delta), \text{ray}) = \delta$ et $\Phi_{|\mathbb{H}}$ est surjective. $\square$

Remarque 6. Une autre démonstration du théorème de bijection de Benzécri-Johnson consiste à utiliser les résultats généraux sur l'application Φ obtenus dans les sections précédentes. Plus précisément, tout d'abord, on établit comme précédemment que $\Phi_{|\mathbb{H}}$ est à valeurs dans $\mathbb{U}$. Puis, on remarque que si δ est une ultramétrie alors $\mathbf{B}_{ab} = \mathbf{B}(a, \delta(a, b))$ et $\text{diam } \mathbf{B}_{ab} = \delta(a, b)$ pour tous $a, b \in X$. D'après la Proposition 2, il en résulte que $\mathbf{B}_{ab}$ est l'unique partie maximale reliée, contenant a et b , et de niveau $\delta(a, b)$. Donc $M_L(T\delta) = \text{Boules}(X, \delta) = \widehat{M_L}(T\delta)$, d'où l'on déduit que $\Phi_{|\mathbb{H}}$ est surjective sur $\mathbb{U}$, d'après la Proposition 1. Donc $\Phi_{|\mathbb{H}}$ est bijective sur $\mathbb{U}$, puisque $\Phi_{|\mathbb{S}_1}$ est injective et qu'il est clair que $\mathbb{H} \subseteq \mathbb{S}_1$. $\square$

5. QUELQUES EXTENSIONS DE LA CLASSIFICATION HIÉRARCHIQUE

Dans cette section, nous nous intéressons aux hiérarchies faibles, aux hiérarchies k -faibles qui les généralisent et aux classifications pyramidales, ainsi qu'aux dissimilarités induites par ces systèmes de classification.

Pour tout entier $l > 0$, la notation $[l]$ désignera l'ensemble des l premiers entiers non nuls : $[l] = \{1, \dots, l\}$.

Soit $\mathcal{F}$ une famille de parties d'un ensemble quelconque E et soit $k \in [|E|]$. Nous dirons que $\mathcal{F}$ est *fermée par intersection de cardinalité $\geq k$* si pour tout $\mathcal{G} \subseteq \mathcal{F}$ tel que $|\bigcap \mathcal{G}| \geq k$, on a $\bigcap \mathcal{G} \in \mathcal{F}$.

5.1. HIÉRARCHIES FAIBLES ET HIÉRARCHIES k -FAIBLES

Les hiérarchies faibles ont été introduites par Bandelt et Dress [1989] et par Batbedat [1989]. Plusieurs publications se sont ensuite intéressées à leurs propriétés, *e.g.* [Diatta et Fichet, 1998 ; Bandelt, 1992 ; Diatta, 1996].

DÉFINITION 8. Hiérarchie faible. *Une hiérarchie faible sur X est un système de classification $\mathcal{C}$ sur X tel que, pour tous $A, B, C \in \mathcal{C}$, on a $A \cap B \cap C \in \{A \cap B, A \cap C, B \cap C\}$.*

La propriété caractéristique d'une hiérarchie est que, pour toute paire de classes $\{A, B\}$, on a $A \cap B \in \{A, B, \emptyset\}$. Il est donc clair que les hiérarchies faibles constituent une extension de la structure hiérarchique. Bandelt [1992] et Diatta et Fichet [1994] ont établi une bijection entre les hiérarchies faibles fermées et indicées et les dissimilarités satisfaisant la condition des 4 points de Bandelt [1992].

DÉFINITION 9. Condition des 4 points de Bandelt. [Bandelt, 1992]. *On dira qu'une dissimilarité δ satisfait la condition des 4 points de Bandelt si, pour tous $a, b, c_1, c_2 \in X$, on a :*

$$\max\{\delta(a, c_1), \delta(a, c_2)\} \leq \delta(c_1, c_2) \Rightarrow \delta(a, b) \leq \max\{\delta(b, c_1), \delta(b, c_2), \delta(c_1, c_2)\}.$$

Étant donné $A, C \subseteq X$, notons :

$$\widetilde{\delta}(A, C) = \max\{\delta(a, c) : a \in A, c \in C\}.$$

Il est alors immédiat de vérifier que la condition des 4 points de Bandelt est vérifiée si et seulement si, pour tous $a, b \in X$ et tout $C \in X^{(1)} \cup X^{(2)}$, on a :

$$\tilde{\delta}(\{a\}, C) \leq \text{diam } C \Rightarrow \text{pour tout } b \in X, \delta(a, b) \leq \text{diam}(Cb).$$

Si on impose, dans cette condition, à C d'être de taille arbitraire $k \geq 1$, au lieu d'être de taille 1 ou 2, alors on obtient une condition, plus générale que la condition des 4 points de Bandelt, qui fait intervenir $k + 2$ points.

DÉFINITION 10. Condition des $(2, k)$ -points. [Bertrand et Janowitz, 2003]. *Étant donné un entier $k \geq 1$, nous dirons qu'une dissimilarité δ satisfait la condition des $(2, k)$ -points si pour tous $a, b \in X$ et tout $C \in X^{(k)}$, on a :*

$$\tilde{\delta}(\{a\}, C) \leq \text{diam } C \Rightarrow \text{pour tout } b \in X, \delta(a, b) \leq \text{diam}(Cb).$$

Remarque 7. La condition des $(2, 1)$ -points est équivalente à :

$$\forall a, b, c \in X, \quad \delta(a, c) = 0 \Rightarrow \delta(a, b) \leq \delta(c, b).$$

Cette condition équivaut à $(\delta(a, c) = 0 \Rightarrow \delta(a, b) = \delta(c, b))$, car a et c sont interchangeables. Donc, la condition des $(2, 1)$ -points est vraie si et seulement si δ est semi-propre. Par conséquent, une dissimilarité vérifie la condition des 4 points de Bandelt si et seulement si elle est semi-propre et vérifie la condition des $(2, 2)$ -points.

Le Lemme suivant est une reformulation immédiate de la condition des $(2, k)$ -points.

LEMME 5. *Pour tout entier $k \geq 1$, une dissimilarité δ satisfait la condition des $(2, k)$ -points si et seulement si, pour tout $C \in X^{(k)}$ et tout $x \in X$, on a $\mathbf{B}_C \subseteq \mathbf{B}_{Cx}$.*

Il est naturel de généraliser la structure de hiérarchie faible, de la façon suivante.

DÉFINITION 11. Hiérarchie k -faible. [Bandelt et Dress, 1994 ; Diatta, 1997]. *Soit un entier $k \geq 2$ et soient $k + 1$ parties non vides de X , notées $A_1, \dots, A_{k+1}$. Nous considérons la condition (kF) définie par :*

$$\bigcap_{i \in [k+1]} A_i \in \left\{ \bigcap_{i \in [k+1] \setminus \{j\}} A_i : 1 \leq j \leq k+1 \right\}. \quad (kF)$$

Une hiérarchie k -faible sur X est un système de classification sur X tel que tout $(k + 1)$ -uplet de classes satisfait (kF) .

Remarque 8. On établit facilement que la condition (kF) est équivalente à chacune des conditions suivantes :

$$\text{Il existe } \alpha \in [k+1] \text{ tq } \bigcup_{i \in [k+1]} \left(\bigcap_{i \in [k+1] \setminus \{j\}} A_i \right) \subseteq A_\alpha; \quad (kF')$$

$$\text{Il n'existe pas } x_1, \dots, x_{k+1} \in X \text{ tq } x_i \in A_j \iff i \neq j. \quad (kF'')$$

Introduisons le Lemme 6 qui est facile à établir par applications successives de la propriété (kF) .

LEMME 6. *Si $\mathcal{C}$ est une hiérarchie k -faible, alors $\widehat{\mathcal{C}}$ est aussi une hiérarchie k -faible.*

La condition suivante, notée (I_k) , est un type particulier d'indigage faible. Appliqué aux hiérarchies k -faibles, ce type d'indigage permet d'induire les dissimilarités satisfaisant la condition des $(2, k)$ -points (cf. Théorème 5).

DÉFINITION 12. Représentations hiérarchiques k -faibles. [Bertrand et Janowitz, 2003]. *Étant donné un système préindiqué $(\mathcal{C}, f)$, notons (I_k) la condition suivante :*

$$\text{Pour tous } A, B \in \mathcal{C} \text{ tel que } |A| \geq k, \quad A \subset B \Rightarrow f(A) < f(B) \quad (I_k)$$

Nous dirons qu'un système de classification valué $(\mathcal{C}, f) \in \mathbb{S}_1$ est une représentation hiérarchique k -faible si $\mathcal{C}$ est une hiérarchie k -faible et si $(\mathcal{C}, f)$ satisfait (I_k) . Nous noterons $\mathbb{H}_k$ l'ensemble des représentations hiérarchiques k -faibles.

Bandelt et Dress [1994] et Diatta [1997] ont établi une bijection entre les hiérarchies k -faibles, lorsqu'elles sont fermées et indicées, et un type particulier de dissimilarité k -voies. Dans ce qui suit, nous établissons une bijection, par restriction de l'application Φ , entre les représentations hiérarchiques k -faibles et les dissimilarités 2-voies satisfaisant la condition des $(2, k)$ -points. Afin d'établir cette bijection, nous déterminons tout d'abord les propriétés de ces dissimilarités. Nous considérerons deux types particuliers de parties de X : les parties k -fermées et les parties à k -diamètre.

DÉFINITION 13. *Soient un entier $k \geq 1$ et $C \subseteq X$ de taille supérieure ou égale à k . Nous dirons que C est k -fermée si, pour toute partie $A \in C^{(k)}$, $\mathbf{B}_A \subseteq C$. Nous dirons que C est à k -diamètre si, pour toute partie $A \in C^{(k)}$ telle que $\text{diam } C = \text{diam } A$, on a $\mathbf{B}_A = C$. On vérifie aisément que cette condition est vraie si et seulement si, pour toute partie $A \in C^{(k)}$ telle que $\text{diam } C = \text{diam } A$, on a $\mathbf{B}_A \subseteq C$.*

Remarque 9. Considérons un entier $k \geq 2$. En utilisant la Définition 13, ainsi que la Proposition 2(2), il est facile de montrer les inclusions suivantes sont satisfaites pour toute dissimilarité δ :

$$\{\text{parties } k\text{-fermées}\} \subseteq \{\text{parties à } k\text{-diamètre}\} \subseteq \{k\text{-boules}\} \cap M_L(T\delta). \quad (I_c)$$

Par la suite, nous considérons des types particuliers de dissimilarités pour lesquels une ou deux des inclusions ci-dessus sont remplacées par des égalités.

Diatta et Fichet [1994, 1998] ont caractérisé les dissimilarités satisfaisant la condition des 4 points de Bandelt à l'aide de la condition d'inclusion et de la condition du diamètre.

DÉFINITION 14. Conditions d'inclusion et du diamètre. [Diatta et Fichet, 1994]. *Soit δ une dissimilarité sur X . On dit que δ satisfait la condition d'inclusion si pour tous $a, b \in X$, on a $\mathbf{B}_{cd} \subseteq \mathbf{B}_{ab}$ pour tous $c, d \in \mathbf{B}_{ab}$. On dit que δ satisfait la condition du diamètre si $\text{diam } \mathbf{B}_{ab} = \delta(a, b)$ pour tous $a, b \in X$.*

PROPOSITION 8. [Diatta et Fichet, 1998]. *Une dissimilarité satisfait la condition des 4 points de Bandelt si et seulement si elle satisfait les conditions d'inclusion et du diamètre.*

Nous introduisons, à présent, les conditions de k -inclusion et du k -diamètre qui généralisent, respectivement, les conditions d'inclusion et du diamètre.

DÉFINITION 15. Condition de k -inclusion. [Bertrand et Janowitz, 2003] *Soit un entier $k \geq 1$ et soit $\delta \in \mathbb{D}[X]$. Nous dirons que δ satisfait la condition de k -inclusion si chaque k -boule est k -fermée, i.e. si, pour tout $A_2 \in X^{(k)}$ et pour tout $A_1 \in \mathbf{B}_{A_2}^{(k)}$, on a $\mathbf{B}_{A_1} \subseteq \mathbf{B}_{A_2}$.*

DÉFINITION 16. Condition du k -diamètre. [Bertrand et Janowitz, 2003]. *Soit un entier $k \geq 1$ et soit $\delta \in \mathbb{D}[X]$. Nous dirons que δ satisfait la condition du k -diamètre si, pour tout $A \in X^{(k)}$, $\text{diam } \mathbf{B}_A = \text{diam } A$.*

Remarque 10. La condition de 1-inclusion est équivalente à la condition du 1-diamètre qui peut s'écrire sous la forme :

$$\forall a \in X, \quad \delta(x, a) = \delta(y, a) = 0 \Rightarrow \delta(x, y) = 0.$$

Donc si δ est semi-propre, autrement dit si la condition des $(2, 1)$ -points est vérifiée, alors les conditions de 1-inclusion et du 1-diamètre sont vérifiées. Comme le montre le lemme suivant, cette propriété se généralise à la condition des $(2, k)$ -points pour $k \geq 1$.

LEMME 7. *Pour tout entier $k \geq 1$, la condition des $(2, k)$ -points implique les conditions de k -inclusion et du k -diamètre.*

Preuve. D'après la Remarque 10, il suffit d'établir la propriété pour $k \geq 2$. Soient $k \geq 2$ et δ une dissimilarité vérifiant la condition des $(2, k)$ -points. Montrons que δ vérifie la condition du k -diamètre. Soit $A \in X^{(k)}$ et soient $x, y \in \mathbf{B}_A$. Donc $\text{diam } Ax = \text{diam } A$. Comme $y \in \mathbf{B}_A$ et $\mathbf{B}_A \subseteq \mathbf{B}_{Ax}$, d'après le Lemme 5, on a $\delta(x, y) \leq \text{diam } Ax = \text{diam } A$. Donc on a bien $\text{diam } \mathbf{B}_A = \text{diam } A$.

Montrons que δ vérifie aussi la condition de k -inclusion. Supposons que l'on ait $A_1, A_2 \in X^{(k)}$ et $A_1 \subseteq A_2$. D'après la condition du k -diamètre, $\text{diam } A_1 \leq \text{diam } A_2$. De plus, pour tout $(a_1, a_2) \in A_1 \times A_2$, on a $\delta(a_1, a_2) \leq \text{diam } A_2$. D'où, pour tout $a_2 \in A_2$, $\text{diam } A_1 a_2 \leq \text{diam } A_2$. Soit $x \in \mathbf{B}_{A_1}$. D'après le Lemme 5, $x \in \mathbf{B}_{A_1} \subseteq \mathbf{B}_{A_1 a_2}$, pour tout $a_2 \in A_2$. D'où $\delta(x, a_2) \leq \text{diam } A_1 a_2 \leq \text{diam } A_2$, pour tout $a_2 \in A_2$. Par conséquent $x \in \mathbf{B}_{A_2}$, ce qui prouve la condition de k -inclusion. $\square$

La Proposition 9 (resp. 10) établit plusieurs caractérisations différentes de la condition de k -inclusion (resp. condition du k -diamètre). Les caractérisations définies par les Propositions 9(ii) et 10(v) expriment le fait qu'une ou les deux inclusions figurant dans (I_c) sont remplacées par une égalité.

PROPOSITION 9. *Soit δ une dissimilarité définie sur X . Pour tout entier $k \geq 2$, les conditions suivantes sont équivalentes :*

- (i) δ satisfait la condition de k -inclusion ;

- (ii) *Les trois ensembles suivants sont identiques : l'ensemble des k -boules, celui des parties k -fermées et celui des parties à k -diamètre ;*
- (iii) *Pour toute partie A de $X^{(k)}$, $\mathbf{B}_A$ est la plus petite k -boule contenant A .*

Preuve. Compte tenu de la Remarque 9, il est immédiat de vérifier que les propriétés (ii) et (iii) sont toutes les deux équivalentes à l'assertion que chaque k -boule est k -fermée. $\square$

PROPOSITION 10. [Bertrand et Janowitz, 2003]. *Soit δ une dissimilarité définie sur X . Pour tout entier $k \geq 2$, les conditions suivantes sont équivalentes :*

- (i) *δ satisfait la condition du k -diamètre ;*
- (ii) *Pour tout $A \in X^{(k)}$, il existe une unique partie maximale reliée contenant A et de niveau $\text{diam } A$;*
- (iii) *Pour tout $A \in X^{(k)}$, $\mathbf{B}_A$ est l'unique partie maximale reliée contenant A et de niveau $\text{diam } A$;*
- (iv) *Chaque partie maximale reliée, de taille au moins égale à k , est une partie à k -diamètre ;*
- (v) *Les trois ensembles suivants sont identiques : l'ensemble des parties maximale reliées et de taille au moins égale à k , celui des k -boules et celui des parties à k -diamètre.*

Preuve. D'après la Proposition 2(2), les conditions (ii) et (iii) sont des formulations équivalentes de (i).

(ii) $\Rightarrow$ (iv). Soient $M \in M_L(T\delta)$ et $A \in M^{(k)}$ tel que $\text{diam } A = \text{diam } M$. D'après (ii) et la Proposition 2(2), $\mathbf{B}_A = M$, ce qui prouve (iv).

(iv) $\Rightarrow$ (i). Soit $A \in X^{(k)}$ et soit $M \in \mathcal{M}(A)$, donc $\text{diam } M = \text{diam } A$. D'après (iv), M est une partie à k -diamètre, d'où $M = \mathbf{B}_A$. Par conséquent $\text{diam } \mathbf{B}_A = \text{diam } M$ et donc $\text{diam } \mathbf{B}_A = \text{diam } A$, ce qui prouve (i). D'où (i) $\iff$ (ii) $\iff$ (iii) $\iff$ (iv).

(iv) $\Rightarrow$ (v). D'après (iv), chaque partie maximale reliée, et de taille supérieure ou égale à k , est une partie à k -diamètre, donc aussi une k -boule, par définition. Inversement, si $A \in X^{(k)}$, alors d'après (iii), $\mathbf{B}_A$ est une partie maximale reliée de taille au moins k , donc d'après (iv), $\mathbf{B}_A$ est une partie à k -diamètre. Or une partie à k -diamètre est, par définition, une k -boule, donc, d'après (iii), est aussi une partie maximale reliée de taille au moins égale à k . D'où l'égalité entre les trois ensembles. Enfin (v) $\Rightarrow$ (iv) est évident. $\square$

PROPOSITION 11. [Bertrand et Janowitz, 2003]. *Soit δ une dissimilarité définie sur X . Pour tout entier $k \geq 2$, les conditions suivantes sont équivalentes :*

- (1) *δ satisfait la condition des $(2, k)$ -points ;*
- (2) *Chaque $(k + 1)$ -boule est k -fermée.*

Si $k \geq 3$, alors (1) et (2) sont aussi équivalentes à (3).

(3) δ satisfait la condition de k -inclusion.

Preuve. Supposons tout d'abord que $k \geq 2$. D'après le Lemme 5, il est clair que l'on a (2) $\Rightarrow$ (1). Montrons (1) $\Rightarrow$ (2). Étant donné $A \in X^{(k+1)}$ et $Y \in \mathbf{B}_A^{(k)}$, il s'agit de montrer que $\mathbf{B}_Y \subseteq \mathbf{B}_A$. On note $A = \{a_1, \dots, a_k, a_{k+1}\}$ avec $\delta(a_k, a_{k+1}) = \text{diam } A$. Notons également $A_1 = A \setminus \{a_1\}$. Il résulte de ces définitions que $\text{diam } A = \text{diam } A_1$ et $\mathbf{B}_A \subseteq \mathbf{B}_{A_1}$. Du Lemme 5, on déduit que $\mathbf{B}_{A_1} \subseteq \mathbf{B}_{A_1 a_1} = \mathbf{B}_A$. D'où $\mathbf{B}_A = \mathbf{B}_{A_1}$. Or, d'après le Lemme 7, δ satisfait la condition de k -inclusion, donc $\mathbf{B}_{A_1}$ est k -fermée. D'où $\mathbf{B}_Y \subseteq \mathbf{B}_{A_1} = \mathbf{B}_A$, puisque $Y \in \mathbf{B}_A^{(k)} = \mathbf{B}_{A_1}^{(k)}$.

Supposons maintenant $k \geq 3$. L'implication (1) $\Rightarrow$ (3) résulte du Lemme 7.

Montrons (3) $\Rightarrow$ (2). Soit $A \in X^{(k+1)}$. Il s'agit de montrer que $\mathbf{B}_A$ est k -fermée. Considérons a_1 et a_2 distincts dans A tels que $\delta(a_1, a_2) = \text{diam } A$, ce qui est possible car $k \geq 2$. D'où :

$$\mathbf{B}_A = \bigcap_{a \in A} \mathbf{B}(a, \delta(a_1, a_2)) = \bigcap_{a \in A} \left\{ u \in X : \delta(u, a) \leq \delta(a_1, a_2) \right\}.$$

Notons $\mu(A) = \delta(a_1, a_2) = \text{diam } A$. Pour tout recouvrement $\{Y_i\}_{i \in I}$ de A , on a :

$$\mathbf{B}_A = \bigcap_{i \in I} \bigcap_{y \in Y_i} \left\{ u \in X : \delta(u, y) \leq \mu(A) \right\} = \bigcap_{i \in I} \left\{ u \in X : \forall y \in Y_i, \delta(u, y) \leq \mu(A) \right\}.$$

Or $\{Y \in A^{(k)} : \{a_1, a_2\} \subset Y\}$ est un recouvrement de A , car $k \geq 3$, d'où :

$$\mathbf{B}_A = \bigcap_{\{a_1, a_2\} \subset Y \in A^{(k)}} \left\{ u \in X : \delta(u, y) \leq \mu(A), \text{ pour tout } y \in Y \right\}.$$

Par conséquent, $\mathbf{B}_A = \bigcap_{\{a_1, a_2\} \subset Y \in A^{(k)}} \mathbf{B}_Y$. Donc, d'après (3), $\mathbf{B}_A$ est k -fermée en tant qu'intersection de parties k -fermées. $\square$

PROPOSITION 12. *Si $k \geq 3$, alors la condition de k -inclusion implique la condition du k -diamètre.*

Preuve. Supposons qu'une dissimilarité δ vérifie la condition de k -inclusion, avec $k \geq 2$, et soit $A \in X^{(k)}$. Il s'agit de montrer que $\text{diam } \mathbf{B}_A = \text{diam } A$. Si $u, v \in A$, alors $\delta(u, v) \leq \text{diam } A$. Soient $u \in \mathbf{B}_A \setminus A$ et $v \in \mathbf{B}_A$. Comme $k \geq 3$, d'après la Proposition 11, on a $\mathbf{B}_A \subseteq \mathbf{B}_{Au}$ car $|Au| = k + 1$. Donc $v \in \mathbf{B}_{Au}$, et par conséquent $\delta(u, v) \leq \text{diam } Au = \text{diam } A$ car $u \in \mathbf{B}_A$. Donc on a bien $\text{diam } \mathbf{B}_A = \text{diam } A$. $\square$

Réciproquement, pour $k \geq 3$, la condition du k -diamètre n'implique pas la condition de k -inclusion comme le montre le contre-exemple suivant.

Contre-exemple. Soit $X = \{a, b, c, d, e\}$, et soit δ la dissimilarité définie sur X par : $\delta(a, b) = 3$, $\delta(a, c) = 2$ et $\delta(u, v) = 1$ pour tout $\{u, v\}$ distinct de $\{a, b\}$ et de $\{a, c\}$ (cf. Figure 2). Comme $\{b, c, d, e\} = \mathbf{B}_{bde} \not\subseteq \mathbf{B}_{abe} = \{a, b, d, e\}$, il en résulte que δ ne vérifie pas la condition de 3-inclusion. Vérifions que δ satisfait la condition du 3-diamètre. Pour toute partie A ne contenant ni $\{a, b\}$ ni $\{a, c\}$, on vérifie facilement que $\text{diam } A = 1 = \text{diam } \mathbf{B}_A$. Si A contient $\{a, b\}$ mais pas c , alors $A = \{a, b, d\}$ ou $A = \{a, b, e\}$, et donc $\text{diam } A = 2 = \text{diam } \mathbf{B}_A$ car $\mathbf{B}_A = \{a, b, d, e\}$. Si A contient $\{b, c\}$, alors $\text{diam } A = 3 = \text{diam } \mathbf{B}_A$ car $\mathbf{B}_A = X$ puisque $\text{diam } X = 3$. Il en résulte que δ satisfait la condition du 3-diamètre.

FIGURE 2. Graphe valué représentant δ : la valuation des arêtes $\{a, b\}$ et $\{a, c\}$ est explicite, celle des autres arêtes vaut implicitement 1. Pour tous $u, v \in X$, $\delta(u, v)$ est égal à la valuation de l'arête $\{u, v\}$.

LEMME 8. *Soit un entier $k \geq 2$. Si $A_1, \dots, A_{k+1}$ sont des parties de X non vides et k -fermées, alors le $(k+1)$ -uplet $(A_1, \dots, A_{k+1})$ satisfait la condition (kF) .*

Preuve. Soient $A_1, \dots, A_{k+1}$ des parties de X non vides et k -fermées. Supposons que la condition (kF) ne soit pas satisfaite par $A_1, \dots, A_{k+1}$. D'après la Remarque 8, il existe $x_1, \dots, x_{k+1}$ tels que pour tout $i \in [k+1]$, $x_i \in \left(\bigcap_{j \neq i} A_j\right) \setminus A_i$.

Notons $X_i = \{x_1, \dots, x_{k+1}\} \setminus \{x_i\}$ pour tout $i \in [k+1]$. Pour chaque valeur de $i \in [k+1]$, on a $\mathbf{B}_{X_i} \subseteq A_i$ puisque A_i est k -fermée. D'où :

$$x_i \in \left(\bigcap_{j \neq i} \mathbf{B}_{X_j}\right) \setminus \mathbf{B}_{X_i}. \quad (\star)$$

Soient x_u et x_v tels que $\delta(x_u, x_v) = \text{diam}\{x_1, \dots, x_{k+1}\}$. Comme $k \geq 2$, nous pouvons choisir $\alpha \in [k+1] \setminus \{u, v\}$. Donc $x_u, x_v \in X_\alpha$, et par conséquent :

$$\text{diam } X_\alpha = \delta(x_u, x_v) = \text{diam}\{x_1, \dots, x_{k+1}\}.$$

D'où $x_\alpha \in \mathbf{B}_{X_\alpha}$, ce qui contredit $(\star)$. □

Le théorème suivant généralise plusieurs résultats proposés dans [Bertrand, 2000] (*i.e.* la proposition 5 et le théorème 5) qui ont été établis pour les dissimilarités satisfaisant la condition des 4 points de Bandelt.

THÉORÈME 4. *Pour tout entier $k \geq 2$ et toute dissimilarité δ définie sur X , les conditions suivantes sont équivalentes :*

- (i) δ satisfait la condition des $(2, k)$ -points ;
- (ii) $M_L(T\delta)$ est fermé par intersections de cardinalité $\geq k$;
- (iii) $(\widehat{M_L}(T\delta), \text{diam}_\delta)$ satisfait (I_k) ;

- (iv) *Pour tout $A \in X^{(k)}$, l'ensemble des éléments de $M_L(T\delta)$ qui contiennent A admet un plus petit élément.*

Preuve. (i) $\Rightarrow$ (ii). Supposons (i). D'après le Lemme 7 et les Propositions 9 et 10, chaque élément de $M_L(T\delta)$ ayant une taille supérieure ou égale à k est une partie k -fermée. Donc toute intersection d'éléments de $M_L(T\delta)$, qui est de taille $\geq k$, est k -fermée. Or, d'après la Proposition 9, toute partie k -fermée est une partie à k -diamètre, qui d'après la Proposition 10, appartient à $M_L(T\delta)$. Donc (ii) est vérifié.

(ii) $\Rightarrow$ (iii). D'après (ii), les éléments de $\widehat{M_L}(T\delta)$ qui sont de taille au moins égale à k appartiennent à $M_L(T\delta)$, donc $(\widehat{M_L}(T\delta), \text{diam}_\delta)$ satisfait (I_k) .

(iii) $\Rightarrow$ (iv). Soit $A \in X^{(k)}$, et soit M un élément de l'ensemble $\mathcal{M}(A)$ qui n'est pas vide. Supposons qu'il existe un autre élément, noté N , de $M_L(T\delta)$ qui contienne A . Si $M \cap N \subset M$, alors $\text{diam}(M \cap N) < \text{diam} M = \text{diam} A$, d'après la condition (I_k) , ce qui est contradictoire. D'où $M \cap N = M$, et $M \subseteq N$. Par conséquent, l'ensemble des éléments de $M_L(T\delta)$ qui contiennent A admet un plus petit élément.

(iv) $\Rightarrow$ (i). Soit $A \in X^{(k)}$. D'après (iv), on a nécessairement $|\mathcal{M}(A)| = 1$. De la Proposition 2, on en déduit que $\mathbf{B}_A$ est le plus petit élément (au sens de l'inclusion) de $M_L(T\delta)$ contenant A . Soit $x \in X$, et soit $M \in \mathcal{M}(Ax)$. Il en résulte que $\mathbf{B}_A \subseteq M$. D'où $\mathbf{B}_A \subseteq \bigcup \mathcal{M}(Ax) = \mathbf{B}_{Ax}$, ce qui prouve, d'après le Lemme 5, que δ satisfait la condition des $(2, k)$ -points. $\square$

THÉORÈME 5. *Pour tout entier $k \geq 2$, l'application $\Phi_{\mathbb{H}_k}$ est une bijection de l'ensemble des représentations hiérarchiques k -faibles sur l'ensemble des dissimilarités satisfaisant la condition des $(2, k)$ -points.*

Preuve. Soit $\mathbb{U}_k$ l'ensemble des dissimilarités vérifiant la condition des $(2, k)$ -points, et soit $(\mathcal{C}, f) \in \mathbb{H}_k$. D'après le Corollaire 2(iv), on a $(\mathcal{C}, f) = (\widehat{M_L}(T\rho), \text{diam}_\rho)$ où ρ est la dissimilarité induite par $(\mathcal{C}, f)$. Donc $(\widehat{M_L}(T\rho), \text{diam}_\rho)$ satisfait (I_k) , ce qui prouve que ρ vérifie la condition des $(2, k)$ -points, d'après le Théorème 4. Donc $\Phi_{\mathbb{H}_k}$ est à valeurs dans $\mathbb{U}_k$.

Montrons que $\Phi_{\mathbb{H}_k}$ est surjective sur $\mathbb{U}_k$. Soit $\delta \in \mathbb{U}_k$. Remarquons tout d'abord que $(\widehat{M_L}(T\delta), \text{diam}_\delta) \in \mathbb{S}_1$, d'après la Remarque 5, et que $(\widehat{M_L}(T\delta), \text{diam}_\delta)$ satisfait (I_k) , d'après le Théorème 4. Soit $M_1, \dots, M_{k+1}$ un $(k+1)$ -uplet d'éléments de $M_L(T\delta)$. Montrons que ce $(k+1)$ -uplet vérifie (kF) . Pour cela, d'après (kF'') , on peut supposer que chaque taille $|M_i|$ est supérieure ou égale à k . En appliquant le Lemme 7 et les Propositions 9 et 10, il résulte que les parties M_i sont k -fermées, et donc que le $(k+1)$ -uplet $M_1, \dots, M_{k+1}$ satisfait (kF) , d'après le Lemme 8. Donc $M_L(T\delta)$ est une hiérarchie k -faible. D'après le Lemme 6, il en résulte que $\widehat{M_L}(T\delta)$ est aussi une hiérarchie k -faible. D'où $(\widehat{M_L}(T\delta), \text{diam}_\delta) \in \mathbb{H}_k$. Or $\Phi(\widehat{M_L}(T\delta), \text{diam}_\delta) = \delta$, d'après la Proposition 1, donc $\Phi_{\mathbb{H}_k}$ est surjective sur $\mathbb{U}_k$.

Comme Φ est injective, nous en concluons que $\Phi_{\mathbb{H}_k}$ est bijective sur $\mathbb{U}_k$. $\square$

5.2. CLASSIFICATION PYRAMIDALE

La classification pyramidale, introduite par Diday [1984, 1986], et par Fichet [1987] sous le nom de pseudo-hiérarchie, constitue également une extension naturelle des

hiérarchies. Remarquons que les représentations hiérarchiques k -faibles peuvent être caractérisées uniquement par des conditions définies à partir de Φ (cf. Théorème 4(iii)). Par contre, un système de classification pyramidale $\mathcal{C}$ est défini par l'existence d'un graphe externe, plus précisément, par l'existence d'une chaîne dont les sommets sont les éléments de X et telle que chacune des classes soit un intervalle de cette chaîne. Cette spécificité de la classification pyramidale permet une visualisation commode de l'ensemble des classes d'une classification pyramidale, grâce au diagramme de Hasse de la relation d'inclusion entre classes, qui, dans ce cas, est planaire. En contrepartie, la validité d'une classification pyramidale repose sur l'hypothèse qu'une sériation des objets à classer est compatible avec la dissimilarité initialement définie sur l'ensemble des objets. Par souci de concision, nous ne présenterons, dans ce qui suit, que les deux principales bijections associant des systèmes de classification pyramidale à leurs dissimilarités induites.

DÉFINITION 17. [Diday, 1984 ; Diday, 1986 ; Fichet, 1987]. *Nous dirons que $\mathcal{C}$ est un système de classification pyramidale si $\mathcal{C}$ est un système de classification fermé et s'il existe un ordre total (également dit linéaire) sur X , noté $\preceq$, tel que chaque élément de $\mathcal{C}$ est un intervalle de $\preceq$. Un tel ordre sera dit compatible avec $\mathcal{C}$.*

On peut remarquer que tout système de classification pyramidale est une hiérarchie faible fermée, puisque l'intersection de trois intervalles d'un ordre total peut se réduire à l'intersection de deux d'entre eux.

DÉFINITION 18. [Diday, 1984, 1986]. *Un système de classification pyramidale, défini sur X et faiblement indicé, est appelé représentation pyramidale sur X . On notera $\mathbb{P}[X]$, ou simplement $\mathbb{P}$, l'ensemble des représentations pyramidales définies sur X . On notera également $\mathbb{P}_i$ l'ensemble des systèmes de classification pyramidale, définis sur X , qui sont indicés.*

Comme tout système de classification pyramidale est fermé, il est clair que $\mathbb{P}_i \subseteq \mathbb{P}$. Considérons tout d'abord l'ensemble $\mathbb{P}$ des représentations pyramidales qui induisent des dissimilarités particulières appelées *robinsoniennes*.

Remarque 11. Il est facile de vérifier que si $\mathcal{I}$ est un ensemble d'intervalles d'un ordre total, fermé par intersection non vide, et si $J = \bigcap \mathcal{I} \neq \emptyset$, alors $\mathcal{I} \setminus \{J\}$ admet exactement deux éléments minimaux (au sens de l'inclusion), notés I_1 et I_2 , tels que $J = I_1 \cap I_2$. La condition d'indication faible introduite dans la Définition 4 peut alors être réécrite de la façon suivante : $A \subset B$ et $f(A) = f(B) \Rightarrow (\{C \in \mathcal{C} : A \subset C\})$ admet exactement deux éléments minimaux et leur intersection est égale à A .

DÉFINITION 19. [Robinson, 1951 ; Hubert, 1974]. *Une dissimilarité δ définie sur X est dite robinsonienne s'il existe un ordre total $\preceq$ sur X tel que :*

$$a \preceq b \preceq c \Rightarrow \delta(a, c) \geq \max\{\delta(a, b), \delta(b, c)\}$$

On dit alors que l'ordre total $\preceq$ est compatible avec δ .

Dans la proposition suivante, les caractérisations (2) et (3) ont été introduites par Diatta et Fichet [1998].

PROPOSITION 13. [Bertrand et Janowitz, 2002]. *Soient δ une dissimilarité définie sur X et $\preceq$ un ordre total sur X . Les conditions suivantes sont équivalentes :*

- (i) *L'ordre $\preceq$ est compatible avec δ ;*
- (ii) *Chaque boule de δ est un intervalle de l'ordre $\preceq$;*
- (iii) *Chaque 2-boule $\mathbf{B}_{ab}$ est un intervalle de l'ordre $\preceq$;*
- (iv) *Chaque partie maximale reliée, au sens de δ , est un intervalle de l'ordre $\preceq$;*
- (v) *Les parties maximale reliées, au sens de δ , coïncident avec les 2-boules de la forme $\mathbf{B}_{ab} = [a, b]$.*

Preuve. Les preuves des implications (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i) sont des vérifications.

(i) $\Rightarrow$ (iv). Supposons qu'un ordre $\preceq$ soit compatible avec δ . Soit $M \in M_L(T\delta)$ et soit a (resp. b) le minimum (resp. maximum) de M selon $\preceq$. Donc $M \subseteq [a, b]$, d'où $\text{diam } M \leq \text{diam}[a, b]$. Mais, d'après (i), $\text{diam}[a, b] = \delta(a, b)$. Par ailleurs, $\text{diam } M \geq \delta(a, b)$ car $a, b \in M$. Il en résulte que $\text{diam } M = \delta(a, b) = \text{diam}[a, b]$ avec $M \subseteq [a, b]$. D'où $[a, b] \subseteq M$ d'après (i), et donc $M = [a, b]$, ce qui prouve (iv).

(iv) $\Rightarrow$ (v). Soit $M \in M_L(T\delta)$. D'après (iv), il existe $a, b \in X$ tels que $M = [a, b]$ au sens de l'ordre $\preceq$. Soit $N \in \mathcal{M}(\{a, b\})$. D'après (iv), $M = [a, b] \subseteq N$. Donc $M = N$ car M et N ont même diamètre. D'où $|\mathcal{M}(\{a, b\})| = 1$ et par conséquent $M = \mathbf{B}_{ab}$, d'après la Proposition 2. Donc M est bien de la forme $\mathbf{B}_{ab} = [a, b]$. Inversement, soient a, b deux éléments de X tels que $\mathbf{B}_{ab} = [a, b]$. Soit $M \in \mathcal{M}(\{a, b\})$. D'après (iv), M contient $[a, b]$. Or $M \subseteq \mathbf{B}_{ab}$ d'après la Proposition 2(1). Donc $M = \mathbf{B}_{ab}$, ce qui prouve (v).

(v) $\Rightarrow$ (iii). D'après la Proposition 2(1), chaque 2-boule $\mathbf{B}_{ab}$ est l'union de parties maximale reliées au sens de δ . D'après (v), chacune de ces parties est un intervalle de $\preceq$ et contient a et b . Donc $\mathbf{B}_{ab}$ est un intervalle de $\preceq$, et (iii) est vérifiée. $\square$

THÉORÈME 6. [Diday, 1986 ; Durand et Fichet, 1988]. *L'application $\Phi_{\mathbb{P}}$ est une bijection de l'ensemble des représentations pyramidales sur l'ensemble des dissimilarités robinsoniennes.*

Preuve. Montrons tout d'abord que $\Phi_{\mathbb{P}}$ prend ses valeurs dans l'ensemble des dissimilarités robinsoniennes. Soit $(\mathcal{C}, f)$ une représentation pyramidale, et soit $\preceq$ un ordre compatible avec $(\mathcal{C}, f)$. D'après le Corollaire 2, $(\mathcal{C}, f) = (\widehat{M_L(T\rho)}, \text{diam}_\rho)$, où $\rho = \Phi(\mathcal{C}, f)$. Donc chaque élément de $M_L(T\rho)$ est un intervalle de $\preceq$, et par conséquent $\rho = \Phi(\mathcal{C}, f)$ est bien une dissimilarité robinsonienne, d'après la Proposition 13(iv). Comme Φ est injective, il reste donc à montrer que $\Phi_{\mathbb{P}}$ est surjective sur l'ensemble des dissimilarités robinsoniennes. Soit δ une dissimilarité robinsonienne, et soit $\preceq$ un ordre compatible avec δ . D'après la Proposition 13(iv), chaque élément de $M_L(T\delta)$ est un intervalle de $\preceq$. L'intersection non vide de deux intervalles étant un intervalle, $\widehat{M_L(T\delta)}$ est donc un système de classification pyramidale. Par ailleurs, $(\widehat{M_L(T\delta)}, \text{diam}_\delta) \in \mathbb{S}_1$ d'après la Remarque 5, donc $(\widehat{M_L(T\delta)}, \text{diam}_\delta)$

est une représentation pyramidale. Comme $\Phi(\widehat{M}_L(T\delta), \text{diam}_\delta) = \delta$ d'après le Théorème 1, on en conclut que $\Phi_{\mathbb{P}}$ est bien surjective, donc bijective. $\square$

Les dissimilarités *fortement robinsonniennes*, introduites par Fichet [1987], sont induites par les systèmes de classification pyramidale indicés.

DÉFINITION 20. [Fichet, 1987 ; Durand et Fichet, 1988]. Soit $\preceq$ un ordre total sur X qui est compatible avec δ . On dit que l'ordre $\preceq$ est *fortement compatible* avec δ si pour tous $a, b, c \in X$ tels que $a \preceq b \preceq c$, les conditions (α) et (β) suivantes sont vérifiées :

$$(\alpha) \quad \delta(a, c) = \delta(b, c) \Rightarrow \text{pour tout } x \succeq c, \delta(a, x) = \delta(b, x) ;$$

$$(\beta) \quad \delta(a, b) = \delta(a, c) \Rightarrow \text{pour tout } x \preceq a, \delta(x, b) = \delta(x, c).$$

Une dissimilarité δ est dite *fortement robinsonnienne* si elle admet au moins un ordre *fortement compatible*.

Remarque 12. Désignons par l'expression « triangle ultramétrique de sommet w » tout triangle uvw vérifiant $\delta(u, v) \leq \delta(u, w) = \delta(v, w)$ (cf. Définition 7 et son commentaire), et supposons les hypothèses de la Définition 20 vérifiées : $a, b, c \in X$ satisfont $a \preceq b \preceq c$, l'ordre $\preceq$ étant total et compatible avec δ . La condition (α) est alors équivalente à chacune des deux conditions suivantes :

(α') si abc est un triangle ultramétrique de sommet c alors, pour tout $x \succeq c$, le triangle abx est ultramétrique et de sommet x ;

(α'') si $\delta(a, c) = \delta(b, c)$ alors, pour tout $x \succeq c$ et tout $y \in [a, b]$, on a $\delta(x, y) = \delta(x, a)$.

Il en résulte que (α) s'interprète à la fois comme une propriété de transport de l'inégalité ultramétrique (condition (α')) et comme une propriété d'homogénéité partielle de l'intervalle $[a, b]$ (condition (α'')). Il est clair que (β) peut s'interpréter de la même façon que (α) .

Il est facile d'établir que si δ est fortement robinsonnienne, alors δ est semi-propre.

THÉORÈME 7. [Fichet, 1987 ; Durand et Fichet, 1988]. L'application $\Phi_{\mathbb{P}_i}$ est une bijection de l'ensemble des systèmes de classification pyramidale indicés sur l'ensemble des dissimilarités fortement robinsonniennes.

Preuve. Montrons tout d'abord que $\Phi_{\mathbb{P}_i}$ prend ses valeurs dans l'ensemble des dissimilarités fortement robinsonniennes. Soit $\rho = \Phi(\mathcal{C}, f)$ avec $(\mathcal{C}, f) \in \mathbb{P}_i$. Comme $\mathbb{P}_i \subseteq \mathbb{P}$, la dissimilarité ρ est robinsonnienne d'après le Théorème 7. Soit $\preceq$ un ordre total sur X qui est compatible avec ρ , et soient $a, b, c \in X$ tels que $a \preceq b \preceq c$. Par raison de symétrie, il suffit de montrer (α) pour établir que ρ est fortement robinsonnienne. Nous supposons donc que $\rho(a, c) = \rho(b, c)$ et considérons $x \in X$ tel que $c \preceq x$. Comme $\mathcal{C}$ est fermé, nous avons $\rho(a, x) = f(A_{ax})$ et $\rho(b, x) = f(A_{bx})$, où $A_{uv} = \bigcap_{u, v \in B \in \mathcal{C}} B$ est un intervalle de l'ordre $\preceq$, pour tous $u, v \in X$. Il est clair que $A_{bx} \subseteq A_{ax}$ et que $A_{bc} = A_{ac}$ car $A_{bc} \subseteq A_{ac}$ avec $\rho(b, c) = \rho(a, c)$. Comme $A_{bx} \cap A_{ac}$ contient c , on a $f(A_{bx} \cap A_{ac}) = f(A_{ac}) = f(A_{bc})$. Or f est un indice strict, donc

$A_{bx} \cap A_{ac} = A_{ac}$, ce qui prouve que $a \in A_{bx}$. Donc $A_{bx} = A_{ax}$, et par conséquent $\rho(a, x) = \rho(b, x)$, ce qui établit (α) . Donc ρ est fortement robinsonienne.

Comme Φ est injective, il reste donc à montrer que $\Phi_{|\mathbb{P}_i}$ est surjective sur l'ensemble des dissimilarités fortement robinsoniennes. Soit δ une dissimilarité fortement robinsonienne. D'après les Théorèmes 1 et 7, $(\widehat{M}_L(T\delta), \text{diam}_\delta) = \Phi^{-1}(\delta)$ est une représentation pyramidale. Montrons que δ vérifie la condition des $(2, 2)$ -points. D'après le Lemme 5, il suffit de montrer que pour tout $a \neq b$, et tout $x \in X$, on a $\mathbf{B}_{ab} \subseteq \mathbf{B}_{abx}$. Soit $\preceq$ un ordre fortement compatible avec δ . Sans perte de généralité, on suppose $a \prec b$. Le cas $x \in [a, b]$ étant évident, on suppose que l'on a $x \notin [a, b]$, et plus précisément $a \prec b \prec x$, par raison de symétrie. Soit $y \in \mathbf{B}_{ab}$. Si $y \in [a, x]$ ou si $x \prec y$, alors il est facile de vérifier que $y \in \mathbf{B}_{abx}$ car $\preceq$ est (fortement) compatible avec δ . Considérons alors le seul cas restant, c'est-à-dire $y \prec a$. Comme $y \in \mathbf{B}_{ab}$, on a $\delta(y, b) \leq \delta(a, b)$. Mais $\delta(y, b) \geq \delta(a, b)$ puisque $y \prec a \prec b$. Donc $\delta(y, b) = \delta(a, b)$. D'où $\delta(y, x) = \delta(a, x)$ car $\preceq$ est fortement compatible avec δ . Comme $\max\{\delta(y, a), \delta(y, b)\} \leq \delta(a, b)$, il en résulte que $y \in \mathbf{B}_{abx}$. Nous concluons que $\mathbf{B}_{ab} \subseteq \mathbf{B}_{abx}$, et que δ vérifie la condition des $(2, 2)$ -points. Donc d'après le théorème 4, $(\widehat{M}_L(T\delta), \text{diam}_\delta)$ satisfait (I_2) . Par ailleurs, les parties de taille 1 sont de diamètre 0, et il est clair que $\text{diam}_\delta^{-1}(0)$ est une partition de X , car δ est semi-propre (cf. aussi lemme 4.7 dans [Bertrand, 2000]). Il en résulte directement que $(\widehat{M}_L(T\delta), \text{diam}_\delta)$ est indicé. Donc $\Phi_{|\mathbb{P}_i}$ est surjective, et par conséquent, bijective. $\square$

Remarque 13. Par définition, $\mathbb{P}_i$ est l'intersection de $\mathbb{P}$ avec l'ensemble des hiérarchies faibles, fermées et indicées. Les bijections établies ci-dessus permettent d'en conclure qu'une dissimilarité est fortement robinsonienne si et seulement si elle est robinsonienne et si elle vérifie la condition des 4 points de Bandelt.

Remerciements. L'auteur remercie les deux lecteurs anonymes pour leurs commentaires constructifs sur une première version du texte.

BIBLIOGRAPHIE

- BANDELT H.-J., DRESS A. W. M. (1989), "Weak hierarchies associated with similarity measures - an additive clustering technique", *Bull. Math. Biol.* 51, p. 133-166.
- BANDELT H.-J., DRESS A. W. M. (1994), "An order theoretic framework for overlapping clustering", *Discrete Math.* 136, p. 21-37.
- BANDELT H.-J. (1992), *Four-point characterization of the dissimilarity functions obtained from indexed closed weak hierarchies*, Mathematisches Seminar, Hamburg, Universität Hamburg (Germany).
- BARTHÉLEMY J.-P., BRUCKER F. (2008), "Binary clustering", *Journal of Discrete Applied Mathematics* 156, p.1237-1250.
- BATBEDAT A. (1988), « Les isomorphismes HTS et HTE (après la bijection de Benzécri / Johnson) », *Metron* 46, p. 47-59.
- BATBEDAT A. (1989), « Les dissimilarités Médas et Arbas », *Stat. Anal. Données* 14, p. 1-18.
- BENZÉCRI J.-P. (1973), *L'Analyse des Données*, Paris, Dunod.
- BERTRAND P. (2000), "Set Systems and Dissimilarities", *Europ. J. Combinat.* 21, p. 727-743.
- BERTRAND P., JANOWITZ M.F. (2002), "Pyramids and Weak Hierarchies in The Ordinal Model for Clustering", *Discrete Appl. Math.* 122, p. 55-81.
- BERTRAND P., JANOWITZ M.F. (2003), "The k -weak hierarchical representations: an extension of the indexed closed weak hierarchies", *Discrete Appl. Math.* 127, p. 199-220.
- BRUCKER F., BARTHÉLEMY J.-P. (2007), *Éléments de classification*, Londres, Hermes publishing.
- CRITCHLEY F., VAN CUTSEM B. (1992), *An order-theoretic unification and generalisation of certain fundamental bijections in mathematical classification*, Grenoble, Univ. Joseph Fourier, LMC-Imag Rapports de recherche, RR 874-M and RR 875-M.
- DIATTA J. (1996), *Une extension de la classification hiérarchique : les quasi-hiérarchies*, Thèse de Doctorat de l'Université de Provence, Marseille.
- DIATTA J. (1997), « Dissimilarités multivoies et généralisations d'hypergraphes sans triangles », *Math. Sci. hum.* 138, p. 57-73.
- DIATTA J., FICHET B. (1994), "From Apresjan hierarchies and Bandelt-Dress weak hierarchies to quasi-hierarchies", E.Diday *et al.* (eds), *New Approaches in Classification and Data Analysis*, Berlin, Springer, p. 111-118.
- DIATTA J., FICHET B. (1998), "Quasi-ultrametrics and their 2-Ball Hypergraphs", *Discrete Math.* 192, p. 87-102.
- DIDAY E. (1984), *Une représentation visuelle des classes empiétantes : les pyramides*, Rapport de recherche I.N.R.I.A. no. 291, Rocquencourt.
- DIDAY E. (1986), "Orders and Overlapping clusters in pyramids", J. De Leeuw *et al.* (eds), *Multidimensional Data Analysis*, Leiden, DSWO Press, p. 201-234.
- DURAND C., FICHET B. (1988), "One-to-one correspondences in pyramidal representation: a unified approach", H.H. Bock (ed.), *Classification and Related Methods of Data Analysis*, Amsterdam, North-Holland, p. 85-90.
- FICHET B. (1987), "Data Analysis: geometric and algebraic structures", Y. A. Prohorov, V. V. Sazonov (eds), *Proceedings of the First World Congress of the Bernoulli Society*, Tachkent, 1986, V.N.U. Science Press vol. 2, p. 123-132.

- GILMORE P.C. (1962), *Families of sets with faithful graph representation*, Technical report, Yorktown Heights, Thomas J. Watson Research Center.
- HUBERT L. (1974), "Some applications of graph theory and related non-metric techniques to problems of approximate seriation: the case of symmetric proximity measures", *British J. of Math. and Stat. Psychology* 27(2), p. 133-153.
- JARDINE J.-P., JARDINE N., SIBSON R. (1967), "The Structure and Construction of Taxonomic Hierarchies", *Mathematical Biosciences* 1, p. 173-179.
- JOHNSON S.-C. (1967), "Hierarchical clustering schemes", *Psychometrika* 32(3), p. 241-254.
- LECLERC B. (1981), "Description combinatoire des ultramétries", *Mathématiques et Sciences humaines* 73, p. 5-37.
- LECLERC B. (1985), « Les hiérarchies de parties et leur demi-treillis », *Mathématiques et Sciences humaines* 89, p. 5-34.
- ROBINSON W.S. (1951), "A method for chronological ordering of archeological deposits", *American Antiquity* 16(4), p. 293-301.