Statistique et analyse des données

A. Batbedat

Les dissimilarités médas ou arbas

Statistique et analyse des données, tome 14, n° 3 (1989), p. 1-18
<http://www.numdam.org/item?id=SAD_1989__14_3_1_0>

Statistique et Analyse des Données
1989 - Vol. 14 n°3 - p. 1 - 18

LES DISSIMILARITES MEDAS OU ARBAS

A. BATBEDAT
Mathématiques, Université des Sciences
34000 MONTPELLIER

Résumé

Nous avons déjà construit les bijections réciproques HTS et DIS entre la classe des dissimilarités et la classe des hypergraphes très strictement indicés. Ceci met en valeur deux nouveaux types de dissimilarités, les médas et les arbas, que nous présentons dans cet article.

Mots clés : Ultramétrique, dissimilarité pyra, méda, arba.

Classification STMA : 06120, 06990

Abstract

We have already set out reciprocal bijections between the class of dissimilarities and the class of very strictly indexed hypergraphs . This show off to advantage two new types of dissimilarities, medas and arbas, presented in this article.

Keywords : Ultrametric, dissimilarity pyra, meda, arba.

Manuscrit reçu le 16 mai 1989, révisé le 6 novembre 1989

INTRODUCTION

Les dissimilarités (pyras), médas ou arbas apparaissent naturellement quand on prolonge la bijection de BENZECRI/JOHNSON entre les hiérarchies strictement indicées et les ultras (les ultramétriques).

Ce prolongement a été mis en place dans [7] pout toutes les dissimilarités et selon deux directions différentes appelées respectivement HTS/DIS et HTE/DIS.

Le premier pas (avant [7] et l'article actuel) a été fait au niveau pyramidal pour lequel existent depuis 1984/1985 trois extensions de la bijection de BENZECRI/JOHNSON.

Les trois rappels historiques qui suivent préparent notre sujet sur les médas et les arbas.

*) La bijection de DIDAY 1984 est présentée ainsi dans [13] :
<<... On considère l'ensemble D des indices de dissimilarité d possédant un $M(d, \theta)$ Robinson ; il en résulte que l'ensemble U des ultramétriques est contenu dans D. On sait qu'il existe une bijection entre U et l'ensemble des hiérarchies indicées ; on montre de même qu'il existe une bijection entre D et un ensemble de structures inter-classes appelées pyramides...>>

Dans [13] aussi bien que dans le présent article, un demi-tableau (comme ceux de nos exemples) est dit Robinson lorsqu'il est croissant en lignes et décroissant en colonnes (les deux, au sens large).

**) La bijection de FICHET (1984) est présentée ainsi dans [14] :
<<... On établit alors une bijection entre dissimilarités fortement de Robinson et pseudo-hiérarchies indicées. La pseudo-hiérarchie obtenue est une hiérarchie ssi la dissimilarité est ultramétrique.>>

***) En 1985, l'auteur prolonge la bijection de BENZECRI/JOHNSON aux prépyramides strictement indicées (exposé dans [4]). On précise (c'est utile pour la suite) en utilisant le langage et les résultats de la page 32 de [6] : choisissons un ensemble X avec n éléments, $n > 1$, et considérons d'un côté les dissimilarités sur X (1.7 de [6]) et de l'autre les prépyramides sur X (1.5 de [6]): il est bon de rappeler qu'une prépyramide possède une orientation qui est une chaîne selon laquelle tout palier est connexe). Chaque prépyramide P est accompagnée d'un indice r qui est strict dans le sens habituel : pour des paliers a et b de P, si a est strictement contenu dans b alors $r(a) < r(b)$. En 1.8 de [6] est présentée l'application ḌIS qui à (P,r) associe la dissimilarité α où $\alpha(xy)$ est le plus petit des r(a) pour a dans P et contenant la paire xy (comme dans [6], une partie de X est notée par un mot sur ses éléments). Par définition (1.10 de [6]), les pyras sont les dissimilarités qui peuvent être obtenues de cette façon. Ensuite on montre que pour toute orientation C de P, le C-demi-tableau de DIS(P,r) est Robinson (1.13 de [6]). Réciproquement, à partir d'une dissimilarité β possédant un demi-tableau Robinson, on construit une prépyramide strictement indicée (Q,s), de façon ascendante par seuils selon β : (Q,s) est notée HTS(β) (1.9 de [6]). Alors le résultat est que HTS et DIS sont des bijections réciproques entre les prépyramides strictement indicées et les pyras. Ajoutons que l'on munit chaque pyra des orientations de sa prépyramide : on démontre que ce sont exactement les chaînes qui donnent un demi-tableau Robinson.

Après les pyras de 1984/85, nous passons dans cet article à deux types plus généraux de dissimilarités : les médas et les arbas. Pour simplifier, nous utiliserons uniquement les bijections réciproques HTS et DIS (nous laissons de côté HTE de [7]). Ces bijections ont d'abord été présentées dans les séminaires [4] et [5], ce deuxième exposé mettant spécialement en valeur les hypergraphes arborés dans un contexte qualifié de "spatial" (au-delà de la visualisation plane du pyramidal). La théorie est maintenant détaillée dans l'article [7].
Le problème essentiel était d'obtenir pour l'image de HTS, une bonne caractérisation des hypindicés (K, t) (K est un hypergraphe et t un indice sur K, simplement croissant dans le cas général).
Nous allons rappeler la condition retenue (appelée "très stricte"), mais il nous faut introduire la médiane ensembliste pour trois parties a, b, c, de X :
$\text{Méd}(a, b, c) = (a \cap b) \cup (b \cap c) \cup (c \cap a)$,
ce qui est aussi bien $\text{Méd}(a, b, c) = (a \cup b) \cap (b \cup c) \cap (c \cup a)$.

Condition très stricte : L'indice t est strict et pour tous paliers a, b, c, dans k, non singletons, il existe un palier e de K qui contient Méd(a, b, c) et vérifie : $t(e) \leq \text{Max}(t(a), t(b), t(c))$.

Alors nous avons démontré (théorème 2 en page 55 de [7]):
HTS et DIS sont des bijections réciproques entre la classe des hypindicés très stricts et la classe de toutes les dissimilarités.

Ainsi se pose le double problème général :
Problème (H) : Pour une classe remarquable de dissimilarités, étudier son image par HTS.
Problème (D) : Pour une classe remarquable d'hypindicés (d'hypergraphes et/ou d'indices très stricts) étudier son image par DIS.

Parmi les hypergraphes, les arborés ont une place de choix et se situent comme une généralisation directe des prépyramides : les arborés donnent les arbas. De façon imprévue, la théorie impose une autre généralisation directe des prépyramides : nous avons appelé médinclus ces hypergraphes ; les médinclus donnent les médas.

1. LES HYPERGRAPHES MEDINCLUS ET LES DISSIMILARITES MEDAS

1.1. Les Médinclus

On dit qu'un triplet (a, b, c) de parties de X est médinclus si Méd(a, b, c) (introduction) est contenue dans a ou b ou c.

On dit qu'un hypergraphe K est médinclus lorsque tout triplet de paliers de K est médinclus.

Propriété 1

| Un hypergraphe K est médinclus ssi tout indice strict est très strict.

Preuve :

*) Si K est médinclus et t un incice sur K, pour des paliers a, b, c, de K, Méd(a, b, c) est contenue par exemple dans a : alors trivialement $t(a) \leq Max(t(a), t(b), t(c))$.

**) Réciproquement, si K possède un triplet (a, b, c) non médinclus on prend un indice s strict sur K (il en existe, voir l'exemple 3), on note v la valeur $(1 + Max(s(a), s(b), s(c)))$, puis on sollicite le nouvel indice r qui coïncide avec s sauf sur les paliers d qui contiennent Méd(a, b, c) pour lesquels on pose $r(d) = s(d)+v$: r est strict mais n'est pas très strict.

Propriété 2

| Toute prépyramide est médincluse.

Preuve :

Soit une prépyramide et $a = [x,y]$, $b = [z,t]$, $c = [u,v]$, des paliers de P présentés en intervalles d'une orientation C de P. Si c ne contient pas l'intersection de a et b alors $z < y$ puis par exemple (sinon, opposer C) $y < u$: par conséquent l'intersection de a et c est vide, donc le triplet (a, b, c) est médinclus.

Exemple 1

Pour X = xyzu, voici un médinclus non prépyramide : outre X et les singletons, les paliers sont les paires xy, yz, zu et ux.

Exemple 2

Pour X = xyzu, voici un hypergraphe non médinclus : outre X et les singletons, les paliers sont xuy, yuz et zux.

Propriété 3

| Pour un hypergraphe K sont équivalentes i, ii, iii, iv :
| i) K est médinclus
| ii) Aucun triplet (a, b, c) de paliers de K ne possède un vrai triangle constitué par trois
| singletons x, y, z, avec x dans $((b \cap c)-a)$, y dans $((c \cap a)-b)$, et z dans $((a \cap b)-c)$
| iii) Pour toute famille F' de trois paliers de K, l'intersection peut être obtenue par
| l'intersection de deux paliers de F'
| iv) Pour toute famille F" de paliers de K, l'intersection peut être obtenue par
| l'intersection deux paliers de F".

Preuve :

Méd(a, b, c) $\subseteq$ a ssi $(b \cap c) \subseteq$ a ssi $((b \cap c)-a)$ est vide.

La propriété ii apparait dans les travaux de RYSER [24, 25, 26] et ANSTEE [1] pour des (0,1) -matrices (le titre de [1] est "Properties of (0,1)-matrices with no triangles"). La propriété iv relie les médinclus au contexte de GOLUMBIC/JAMISON [18] sur l'intersection de familles de paliers.

1.2. Les médas

Les médas sont les dissimilarités associées par HTS/DIS aux médinclus strictement indicés.

Corollaire (de la propriété 2)

| Les pyras sont des médas.

En particulier, pour n = 3 toute dissimilarité est méda.

Exemple 3
Voici la méda non pyra associée à l'exemple 1 pour l'indice cardinal (du nombre de singletons) :

	y	z	u
x	2	4	2
y		2	4
z			2

2. LES HYPERGRAPHES ARBORES ET LES DISSIMILARITES ARBAS

2.1. **Les arborés**

On trouvera dans le livre récent de BARTHELEMY/GUENOCHE [2], des propriétés et des références sur les arbres. Nous citons aussi BERGE [9]. Un arbre est un graphe symétrique, connexe et sans cycle : pour toute paire de sommets x, y, il existe un unique chemin entre x et y. Relativement à un arbre A sur X, une partie a de X est connexe si pour tous x, y, de a, le A-chemin de xy est contenu dans a. Ceci prolonge aux arbres la notion d'intervalle d'une chaîne.

Maintenant on étend la notion de prépyramide.
Pour un hypergraphe K, un arbre A est une assise lorsque tout palier de K est A-connexe. Le même lien est présenté dans l'autre sens en analyse de similitude sous la forme de rigidité de K par rapport à l'arbre A. Les assises généralisent les orientations et ouvrent la voie aux dendrogrammes spatiaux introduits dans [5] pour la classification arborée. Ensuite on dit qu'un hypergraphe K est arboré lorsqu'il possède une assise. Sur les hypergraphes arborés on peut voir [17], [19], [23] et [27].

Dans [3], [9] et [17], on dit qu'un hypergraphe K a la propriété de Helly si pour toute famille & de paliers de K qui ont deux à deux une intersection non vide, l'intersection des éléments de & n'est pas vide. Lorsqu'un hypergraphe possède une assise A, l'unicité du A-chemin entre deux singletons impose la propriété de Helly (énoncé dans [9] page 383 et repris dans [17] page 224). Mais il existe des hypergraphes non arborés qui ont la propriété de Helly, d'où les propriétés complémentaires proposées dans [17] pour retrouver les arborés.

Dans ce contexte, voici un cas particulier intéressant :
Disons ici (pour la propriété qui va suivre) qu'un hypergraphe a pleinement la propriété de Helly lorsque l'intersection de tous ses paliers non singletons n'est pas vide.

Propriété 4 (facile)

| i) Un hypergraphe K a pleinement la propriété de Helly ssi il possède une assise étoile
| (donc il est arboré).
| ii) Alors une étoile A est une assise de K ssi le centre de A est dans l'intersection de
| tous les
| paliers non singletons.

Propriété 5

| Toute prépyramide est arborée.

Une prépyramide possède des assises puisqu'elle est arborée, mais elle est prépyramide parce qu'elle possède en particulier (l'un n'exclut pas l'autre) des assises linéaires (des assises chemins) : chacune donne deux orientations opposées.

Exemple 4

L'hypergraphe de l'exemple 2, qui est non médinclus et non prépyramide,
est arboré étoile de centre u.

Exemple 5

L'hypergraphe de l'exemple 1 est médinclus, non arboré
(étendant la notions d'assise, on pourrait dire qu'il est assis sur un cyle).

En résumé, les médinclus et les arborés constituent deux prolongements distincts des prépyramides.

2.2. Les arbas

Les arbas sont les dissimilarités associées par HTS/DIS aux arborés très strictement incicés. Compte-tenu de ce qui précède, la précision " très " est indispensable.

Corollaire (de la propriété 5)

| Les pyras sont des arbas.

En particulier, pour n = 3 toute dissimilarité est arba.

Exemple 6

La méda de l'exemple 3 (ici, n = 4) n'est pas arba (son hypergraphe n'est pas arboré).

Exemple 7

Considérons la dissimilarité α :

	y	z	u	v
x	2	2	1	1
y		2	1	1
z			3	2
u				1

Voici son image (K,s) par HTS (entre parenthèses la valeur de l'indice) : outre les singletons, xuv(1), yuv(1), xyuv(2), xyzv(2) et X(3). On vérifie que K possède l'assise étoile de centre v et que le triplet (xuv, yuv, xyzv) n'est pas médinclus : par conséquent α est une arba non méda.

Ainsi se présente la sous-classe des médarbas, dissimilarités qui sont à la fois arbas et médas (toute pyra est médarba).

Exemple 8

Sur X = xyzu, considérons la dissimilarités β suivante :

	y	z	u
x	2	2	1
y		2	1
z			1

Voici l'image de β par HTS : outre les singletons, xu(1), yu(1), zu(1) et X(2). Ainsi β est une médarba non pyra.

Propriété 6

| Pour n = 4, toute arba est méda (est médarba).
| Pour n = 4, il existe des médas non arbas (exemple 6).
| Pour n = 5, il existe des arbas non médas (exemple 7).

Preuve de la première assertion :
Nous voulons tout d'abord un arboré K non médinclus : K possède donc une assise étoile (disons de centre u) sur X = xyzu. De plus il existe dans K un triplet (a, b, c) non médinclus : ce ne peut être que xuy, yuz, zux.
Ensuite il faut pour K un indice très strict : c'est impossible.

3. LES TABLEAUX SPATIAUX

3.1. Le cas général

Pour un arbre A sur X, on note chem(xy) le A-chemin entre x et y puis CHEM(A) la famille des A chemins non singletons.

Nous introduisons une généralisation spatiale de la notion de demi-tableau (ceci joue un grand rôle dans la classification arborée et va être utilisé ici) : on appelle tableau spatial sur A, une application, disons S, de CHEM(A) vers les réels strictement positifs. La valeur de S sur chem(xy) est notée S(xy).

Pour A fixé, les tableaux spatiaux correspondent bijectivement aux dissimilarités sur X (dans la correspondance naturelle pour chaque xy).

Dans le cas particulier où A est lui-même un chemin, on peut aussi parler de tableau plan. Alors chacune des deux chaînes de A donne un demi-tableau associé.

3.2. Le cas de Robinson

On dit qu'un tableau spatial est Robinson lorsqu'il est croissant pour l'inclusion des chemins. Un tableau plan est Robinson ssi ses demi-tableaux sont Robinson au sens de l'introduction.

4. LES ASSISES D'UNE ARBA

4.1. **Mise en place**

On définit les assises d'une arba α comme étant celles de l'hypergraphe de HTS(α).

Ceci prolonge la mise en place des orientations d'une pyra (introduction). Maintenant, pour une pyra deux orientations opposées correspondent à une assise chemin.

Propriété 7 (conséquence directe des définitions)

| Une dissimilarité est arba ssi elle possède une assise.

4.2. **Caractérisation spatiale**

Théorème 1

| Soit α une dissimilarité sur X et A un arbre sur X :

| A est une assise de α ssi le A-tableau spatial S de α est Robinson.

Preuve :

Soit (K,s) = HTS(α).

*) Si A est une assise de α, alors A est une assise de K.

Considérons chem(xy) puis z, u, sur ce chemin et notons v la valeur S(xy) $= \alpha$(xy) : il existe un palier a de K qui contient xy avec s(a) = v (voir le retour par DIS). Mais a contient chem(xy) puisque A est assise de K, donc a contient z et u : ceci donne α (zu) $\leq$ v et donc S(zu) $\leq$ S(xy).

**) On suppose que S est Robinson.

On prend un palier a dans K avec s(a) = v, puis x, y, dans a, on note b = chem(xy) et on étudie le cas défavorable où il existe z dans (b-a). Pour u dans (a $\cap$ b), S(uz) $\leq$ S(xy) $\leq$ v. Ensuite, pour u dans (a-b), z est sur chem(xu) (sinon , échanger x et y) donc S(ux) $\leq$ S(uz) $\leq$ v : il en résulte que a n'est pas maximale au seuil v, ce qui est contradictoire avec HTS.

4.3. **Les arbres minimums**

Corollaire (du théorème 1)

| Pour une arba toute assise est un arbre minimum.

Preuve :
Soit α une arba d'assise A : on utilise le A-Robinson S de α pour construire un arbre minimum de α, si possible A lui-même. Dans une étape à valeur v effective, nous avons une paire xy à valeur v, cette paire ne bouclant pas un cycle avec le graphe déjà construit. Donc il existe une arête zu de A dans chem(xy), zu ne bouclant pas un cycle à cette étape. S étant Robinson, $S(zu) \leq S(xy) = v$, donc on peut prendre l'arête zu.

Ce corollaire étend le résultat semblable pour les chaînes en pyramidal, énoncé dans [12].

Montrons que la réciproque n'est pas vraie (une arba peut posséder un arbre minimum qui n'est pas une assise). On considère la dissimilarité α

du demi-tableau T que voici :

	y	z	t
x	1	2	3
y		1	1
z			1

T est Robinson donc α est pyra et la chaîne $C = (x < y < z < t)$ est une orientation. Il est clair que les seules orientations de α sont C et son opposée. Par contre x-y-t-z est un arbre minimum.

5. LOCALISATION/GLOBALISATION

De façon concrète, on peut dire qu'un arbre est localement dirigé par ses chemins. C'est pourquoi il est intéressant d'étudier les restrictions (des phénomènes considérés) sur les chemins et il est normal (efficace) d'accorder l'intérêt principal aux chemins maximaux (ce que nous faisons dans le théorème 2 ci-après) : On parle de localisation.

Ainsi, dans son article page 27 de [19], Vergès s'intéresse aux chaînes maximales et dit <<on ne repère que des propriétés locales>>. Précisons qu'une des propriétés locales considérée comme importante est la "régularité", ce qui signifie que l'on rencontre localement un Robinson. La situation où la régularité est partout valable a été envisagée par DEGENNE et VERGES dans le paragraphe 2.5.2, page 498 de [11] : on verra le lien avec ce qui suit.

Théorème 2

| On se donne une dissimilarité α et un arbre A (arbitrairement) :
| α est une arba d'assise A ssi pour chaque chemin maximal de A,
| le tableau plan de la restriction de α est Robinson.

Preuve :

Soit S le A-tableau spatial de α. On note M un chemin maximal de A puis P le tableau plan sur M, restriction de S.

*) Il est clair que si S est Robinson, alors P est Robinson.

**) On suppose que pour tout M le tableau P est Robinson.

Il faut montrer que S est Robinson. On considère chem(zu) contenu dans chem(xy) puis un M qui contient xy.

Ceci nous conduit à la notion de famille globalisante.

D'une façon générale nous avons un arbre A auquel on associe la famille (Mj) de ses chemins maximaux ; puis on considère une famille de dissimilarités μj, chacune sur l'ensemble de Mj. Il faut préciser les conditions pour que cette famille représente en localisation, une dissimilarité α sur l'ensemble X (des sommets de A). Tout cela est développé dans la classification arborée.

Compte-tenu du théorème 2, une arba dans cette situation sera représentée par une famille de pyras : c'est pourquoi nous disons que la classification arborée est localement pyramidale.

Ce contexte, dans un sens large, est actuellement très étudié par de nombreux chercheurs : les raisons sont nombreuses et variées : ici localisation (avec un retour "parfait"), ailleurs réduction de la taille des données ou mise en place de séries parallèles pour gros ordinateurs.

6. DES TESTS DE RECONNAISSANCE

6.1. Pour les dissimilarités

On se donne une dissimilarité δ.

*) On veut savoir si δ est méda.

Nous disposons uniquement de la définition :
prendre HTS(δ) et voir si son hypergraphe est médinclus.

**) On veut savoir si δ est arba et trouver une assise.
-) Lorsque δ a peu d'arbres minimums, on en cherche un qui soit une assise. Bien entendu, il faut parcourir les arbres minimums jusqu'à un résultat positif ou épuisement des possibilités.
-) Nous ne possédons pas un test direct sur la dissimilarité.
C'est pourquoi nous avons retenu le test arboré de Leclerc qui s'applique à l'hypergraphe de HTS(δ) : dans un premier temps il précise le caractère arboré éventuel, puis il donne les assises (voir [23]).

6.2. **Pour les hypergraphes**

On se donne un hypergraphe K.
*) On veut savoir si K est une prépyramide.
Nous avons présenté dans [6] les algorithmes de Booth/Lueker et de Dubost/Oubina (le premier par les PQ-arbres et le second par les formules).
**) On veut savoir si K est médinclus : on applique la définition.
***) On veut savoir si K est arboré : on applique par exemple le test arboré de Leclerc.

6.3. **Pour les hypergraphes avec DIS**

On se donne un hypergraphe K et on veut le reconnaître grâce à la dissimilarité α = DIS(K,ic), où ic est l'indice strict du nombre de singletons.
On fait un test médinclus.
*) Lorsque ce test est positif, le lien entre α et (K, ic) est celui de la théorie dans [7] : les propriétés de α et de (K, ic) sont en harmonie.
Exemple : Si un test pyra sur α est positif, alors K est une prépyramide. De plus, souvent ces tests donnent les orientations de α : ce sont celles de K.
**) A contrario, voici une situation avec le test médinclus négatif l'hypergraphe K est sur X = xyzu, avec pour paliers, outre les singletons et X : xy, yz, xyz, yzu et uxy.
Pour l'indice ic, sa dissimilarité α est :

	y	z	u
x	2	3	3
y		2	3
z			3

Puisque ce demi-tableau est Robinson, α est pyra.

Mais K n'est pas médinclus car Méd (xyz, yzu, uxy) = X.
Donc K n'est pas une prépyramide.
Ajoutons que K possède l'assise étoile de centre y : par conséquent le test médinclus ne peut pas être remplacé par un test arboré.

7. COMPLEMENTS

-) Dans la ligne tracée par le problème (D) de l'introduction, nous venons de présenter les deux types de dissimilarités médas et arbas. En fait est apparu aussi le type intermédiaire des médarbas : elles sont strictement plus générales que les pyras ; il pourrait être intéressant de leur consacrer une étude approfondie.

-) Un des référés a posé la question du prolongement de la propriété "forte" au sens de [14] en pyramidal. Dans le cas général, ceci concerne le problème (D) de la façon suivante : une dissimilarité α sera forte dans le sens précédent si l'hypergraphe de HTS(α) est stable par intersection non vide (puisque nous sommes dans les compléments, ajoutons que ceci a lieu ssi à partir de α, HTS et HTE de [7] donnent la même image).
D'où les cas particuliers des médarbas, médas ou arbas, fortes. Dans ce contexte, nous annonçons en octobre 1989 la mise au point d'une méthode directe de représentation par une méda forte : nous allons la tester en liaison avec des spécialistes de cluster analysis.

-) Dans cet article, le problème (D) a été limité aux hypindicés très stricts afin de profiter du retour par HTS. Il y a cependant une exception en 6.3.** : (K, ic), avec K non prépyramide, donne une pyra par DIS (ic n'est pas très strict). Dans le cadre d'un élargissement du champ du problème (D), nous avons obtenu les deux résultats suivants : pour un hypindicé (K,t), si K est médinclus (resp. arboré) alors DIS(K,t) est méda (resp. arba).

-) Selon la ligne opposée tracée par le problème (H), chaque classe remarquable de dissimilarités ouvre la voie à une recherche vers les hypindicés associés : c'est pourquoi nous citons quelques références récentes sur les dissimilarités (choix non exclusifs) : [15] et [22] en L1, [16] en signes de présence absence, [20] en puissances d'une distance, [2] et [10] en distances quadrangulaires (qui vérifient la condition des quatre points. Ajoutons que le langage varie selon les auteurs : distances "additives" ou encore "arborées").

-) La classification additive (développée dans [2]) représente une dissimilarité δ par un arbre additif. On peut dire aussi que δ est représentée par une distance quadrangulaire (la représentante obtenue dépend bien entendu de la méthode choisie).

-) La classification arborée représente δ par un dendrogramme spatial assis sur un arbre, ou de façon équivalente par une arba : un livre est en cours d'évaluation, pour les méthodes de l'auteur et les programmes Pascal de J.P. Bordat.
Nous montrerons que la dissimilarité représentative dans l'article de DEGENNE|VERGES [11] (obtenue après une suite de transformations de filtrants), cette dissimilarité correspond à une arba.
Nous préciserons que toute distance quadrangulaire est une arba (liaison fondamentale entre la classification additive et la classification arborée).

BIBLIOGRAPHIE

[1] ANSTEE, *Properties of (0,1)-matrices with no triangles*. J. Comb. th., A29, p. 186-198, 1980.

[2] BARTHELEMY/GUENOCHE, *Les arbres et les représentations des proximités*. Masson, Paris, 1988.

[3] BANDELT/PRISNER, *Clique graphs and Helly graphs*. A paraître dans J. Comb. Th. série B.

[4] BATBEDAT, *Isomorphismes d'ordonnés entre graphes et surgraphes*. Comptes rendus du séminaire argentin en graphes 1985, Editeur Université des Sciences, La plata Argentine, p. 30-32, 1985.

[5] BATBEDAT, *Deux prolongements de la bijection de Benzecri/Johnson pour toutes les dissimilarités*. Séminaire 1987 des mathématiques discrètes et sciences sociales, centre d'analyse et mathématiques sociales, Paris (un résumé est disponible chez l'auteur), 16 mars 1987.

[6] BATBEDAT, *L'algorithme PROXEL pour les dissimilarités*. Math. Inf. Sci. hum., n° 102, p. 31-38, 1988.

[7] BATBEDAT, *les isomorphismes HTS et HTE*. METRON, Vol. XLVI, n°1-4, p. 47-59, 31-XII-1988.

[8] BENZECRI, *L'analyse de données*. 1. *La taxinomie*. Dunod-Paris, 1973.

[9] BERGE, *Graphes et hypergraphes*. Dunod-Paris, 1970.

[10] BROSSIER, *Problèmes de représentation des données par des arbres*. Thèse Université de Rennes-2, 1986.

[11] DEGENNE/VERGES, *Introduction à l'analyse de similitude*. Revue française de sociologie, XIV, p. 471-512, 1973.

[12] DIDAY, *Croisements, ordres et ultramétriques*. Math. Sci. hum., 21ème année, n°83, p. 31-54, 1983.

[13] DIDAY, *Une représentation visuelle des classes empiétantes, les pyramides*. Journées de la Grande-Motte en Statistiques, Editeur INRA-Biométrie Montpellier, 1984.

[14] FICHET, *Sur une extension de la notion de hiérarchie et son équivalence avec certaines matrices de Robinson*. Journées de La Grande-Motte en Statistiques, Editeur INRA-Biométrie Montpellier, 1984.

[15] FICHET, *The role played by L1 in data analysis. Statistical data analysis based on the L1-norm and related methods*, Y. Dodge editor, North Holland, p. 185-193, 1987.

[16] FICHET/LE CALVE, *Structure géométrique des principaux indices de dissimilarité sur signe de présence absence*. Statistique et Analyse des Données, 9, n° 3, p. 11-14, 1984.

[17] FLAMENT, *hypergraphes arborés*. Discrete Mathematics, 21, p. 223-227, 1978.

[18] GOLUMBIC/JAMISON, *The intersection graphs of paths in a tree*. Comb. Th., B38, p. 8-22, 1985.

[19] Informatique et sciences humaines, n°67, Analyse de similitude, 1985.

[20] JOLY/LE CALVE, *Etude des puissances d'une distance*. Statistiques et analyse des données, 11, n° 3, p. 30-50, 1986.

[21] JOHNSON, *Hierarchical clustering schemes*. Psychometrika, 22, p. 241-254, 1967.

[22] LE CALVE, *L1-embeddings of a data structure (I,D). Statistical data analysis based on the L1-norm and related methods*, Y. Dodge editor, North Holland, p. 195-202, 1987.

[23] LECLERC, *Arbres minimaux communs et compatibilité de données de types variés.* Math. Sc. Hum., 98, p. 41-67, 1987.

[24] RYSER, *Combinatorial configurations.* Siam J. Appl. Math., 17, p. 593-602, 1969.

[25] RYSER, *A fundamental matrix equation for finite sets.* Proc. Amer. Math. Soc., 34, p. 332-336, 1972.

[26] RYSER, *Intersection properties of finite sets.* J. Comb. Th., 14, p. 79-92, 1973.

[27] TARJAN/YAKANNAKIS, *Simple linear time algorithms to test chordiality of graphs, test aciclycity of hypergraphs, and selectively reduce acyclics hypergraphs.* Siam J. of Computing, 13, p. 566-579, 1984.