

SFC – 2003, Neuchâtel, Suisse

Visualisation de Graphes et de partitions

Ludovic Lebart

*CNRS-ENST, 46 rue Barrault, 75013, Paris, France
lebart@enst.fr*

Pour construire une représentation visuelle d'une partition de n objets décrits par p variables, il existe deux approches courantes :

(A) Construire la partition en s'efforçant d'optimiser un critère, puis, dans un second temps, représenter les classes dans un graphique plan.

(B) Construire simultanément la partition et la représentation, ce qui induit des contraintes sur la partition, mais peut conduire à une meilleure représentation.

Notre but :

combler le vide entre :

- des méthodes de type « boîtes noires »

qui conduisent à des résultats lisibles

(comme les *Kohonen Self Organizing Maps*),

- et les **techniques plus transparentes** de classification (utilisées conjointement avec les techniques factorielles)

dont les « sorties » sont souvent confuses et complexes : **difficiles à lire.**

On insiste ici sur la contribution de l' **analyse de contiguïté** à de telles approches.

La structure de graphe symétrique décrit convenablement certaines structures *a priori* des observations: observations chronologiques, géographique, ...

Mais une telle structure peut être générée par les données multidimensionnelles elles-mêmes, ou par une partie de ces données (relation binaire générale R) :

Graphes des k plus proches voisins (après « symétrisation » !)

$R(i,j)$ signifie que i est un des k plus proches voisins de j (k varie de 1 à n , nombre d'observations).

Graphes engendré par un seuil de distance t (généralement dérivé de quantiles de distances). $R(i,j)$ signifie que $d(i,j) \leq t$.

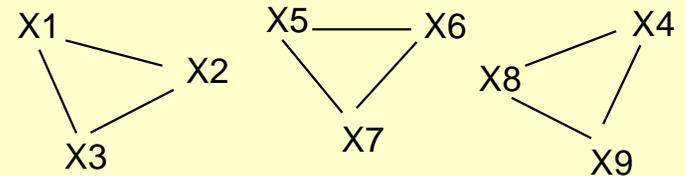
Nous voulons montrer que de tels graphes constituent un (le ?) **chaînon manquant** entre d'une part la **classification**, et d'autre part les **analyses en axes principaux**.

On considère, **n objets** décrits par **p variables**, conduisant à une matrice **Y** , dont les lignes ont une structure de graphe *a priori*.

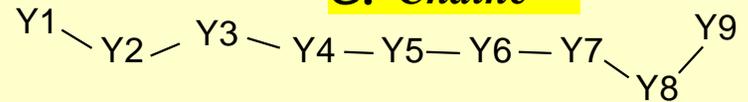
Les **n objets** sont les sommets d'un graphe symétrique **G** dont la matrice **(n, n)** associée est **M** .

(**m_{ii}** = 1 si les sommets i et i' sont joints par une arête, **m_{ii}** = 0 sinon).

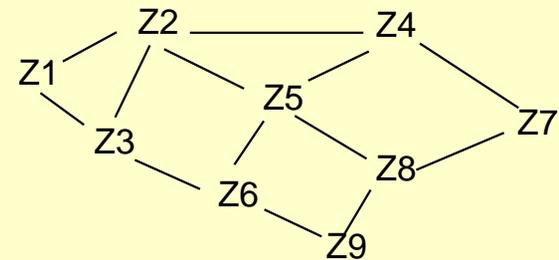
G: Partition: 3 cliques



G: Chaîne



G: Cas Général : relation binaire



1) Dans un premier temps, la structure de graphe peut être considérée comme **externe** (zones géographiques, séries temporelles).

2) Puis on étudie la situation dans laquelle la structure de graphe n'est **pas externe**, mais dérivée du tableau de données lui-même.

(e.g.: le graphe $G(k)$ des **k-plus proches voisins**).

L'idée de déduire des données une **métrique** susceptible de mettre en évidence des **classes** remonte à :

ART, D., GNANADESIKAN, R., KETTENRING, J.R. (1982):

Data Based Metrics for Cluster Analysis, Utilitas Mathematica, 21 A, 75--99.

Les n **objets** (lignes de Y) sont les sommets d'un graphe symétrique G dont la matrice associée (n, n) est M .

$m_{ii'} = 1$ si les sommets i et i' sont joints par une arête, $m_{ii'} = 0$ sinon.

$m = \sum m_{ii'}$ (nombre d'arêtes du graphe G)

Variance locale

$$v^c(y) = (1/2m) \sum m_{ii'} (y_i - y_{i'})^2$$

Variance globale

$$v(y) = (1/2n(n-1)) \sum (y_i - y_{i'})^2$$

(La variance empirique est un cas particulier de la variance locale lorsque le graphe est complet, i.e.: $m_{ii'} = 1$ pour tout i et i')

Le coefficient de contiguïté (Geary, 1954; après Moran et Von Neumann)

$$c(y) = v^*(y) / v(y)$$

« Correction » de la définition de la variance locale, nouveau coefficient de contiguïté

Nouvelle variance locale

$$m_i^* = (1/n_i) \sum_k m_{ik} y_k$$

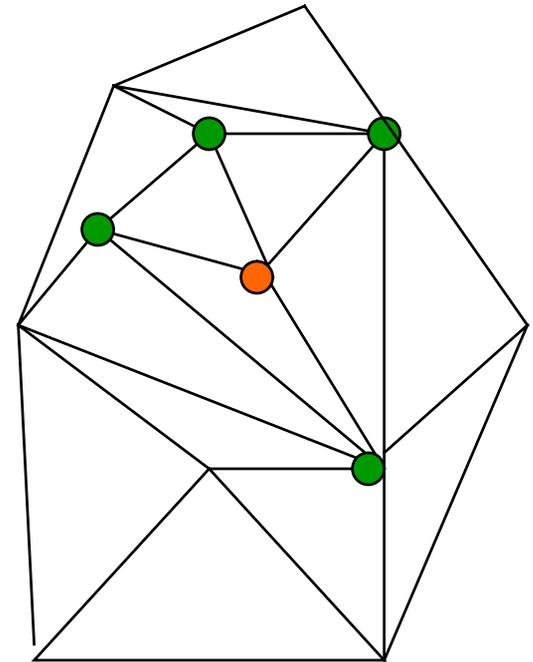
$$v^*(y) = (1/n) \sum (y_i - m_i^*)^2$$

Nouveau coefficient de contiguïté

$$c(y) = v^*(y) / v(y)$$

Avec, comme d'habitude :

$$v(y) = 1/n \sum_{i=1}^n (y_i - m)^2$$



La matrice diagonale \mathbf{N} (matrice des degrés) est telle que :

$$n_i = \sum_k m_{ik}$$

$c(y)$ s'écrit, en notations matricielles (\mathbf{U} = matrice associée au graphe complet):

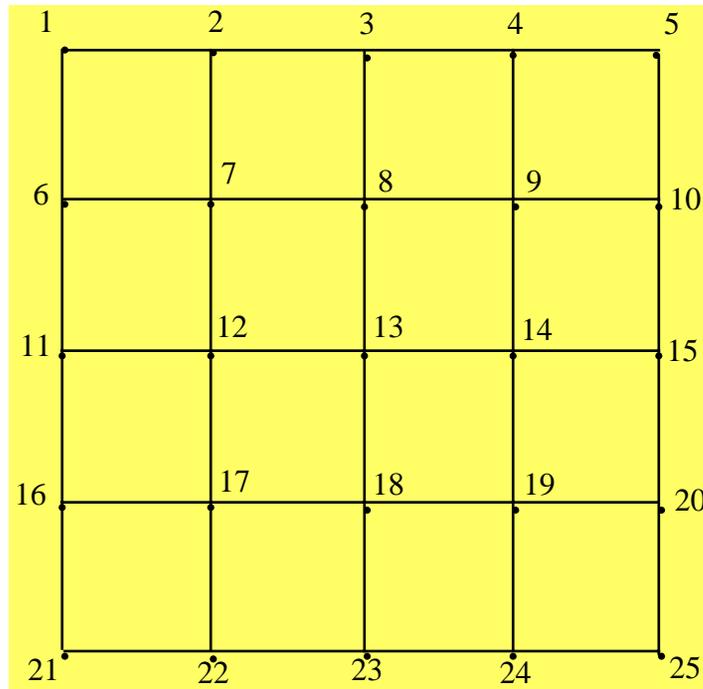
$$c(y) = \mathbf{y}' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \mathbf{y} / \mathbf{y}' (\mathbf{I} - (1/n)\mathbf{U}) \mathbf{y}$$

La (p, p) matrice de covariance locale \mathbf{V}^* est définie comme :

$$\mathbf{V}^* = (1/n) \mathbf{Y}' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \mathbf{Y}$$

Cette matrice définit un puissant outil de mesure de corrélation partielles, si le tableau \mathbf{Y} est n tableau de variables instrumentales.

Exemple d'un graphe \mathbf{G} ($n = 25$) associé à un damier



... et sa matrice associée $\mathbf{M} \rightarrow$

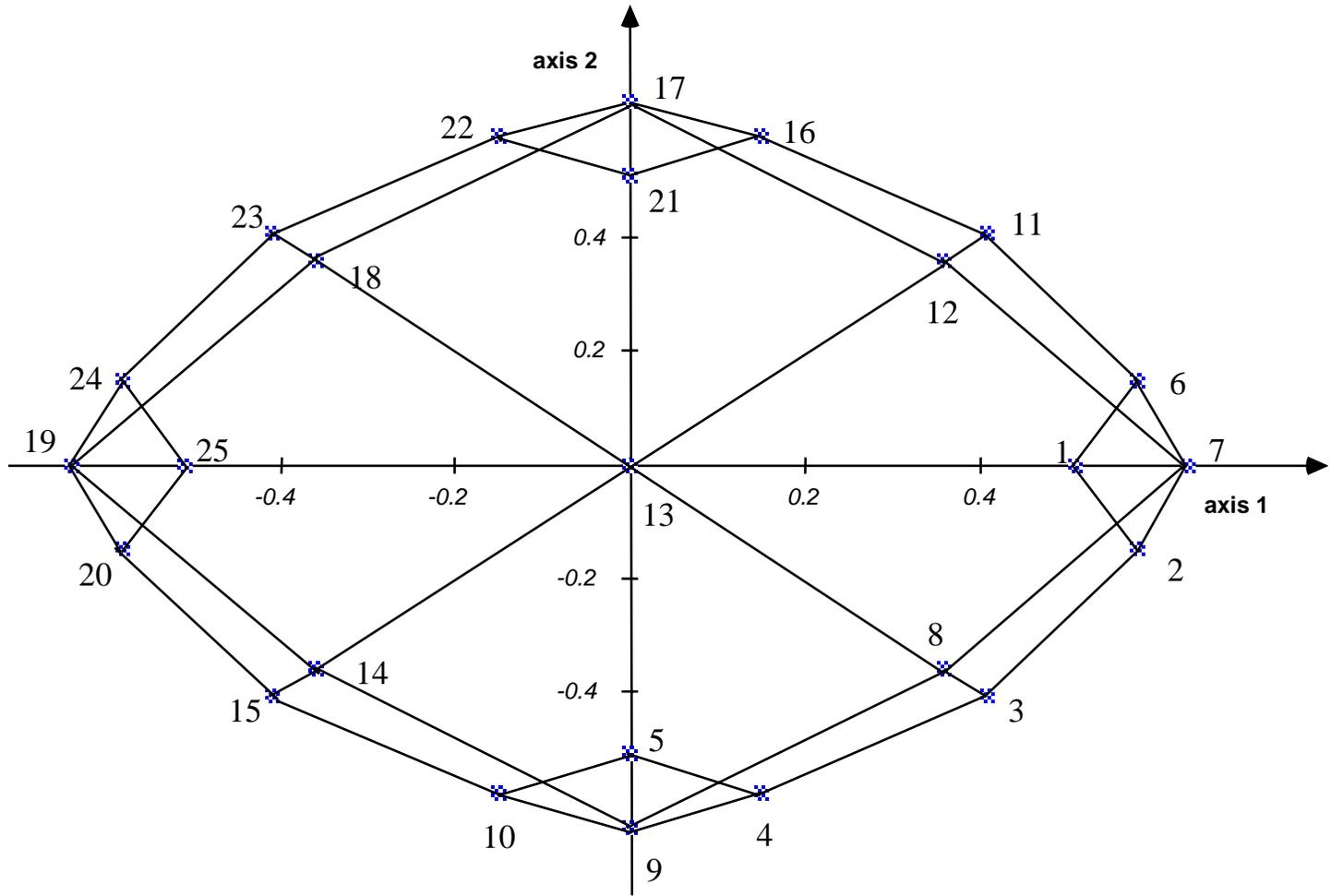
matrice:

M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
r01	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r02	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r03	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r04	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r05	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r06	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
r07	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
r08	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
r09	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
r10	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
r11	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
r12	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0
r13	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0
r14	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0
r15	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0
r16	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0
r17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0
r18	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0
r19	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0	0	1	0
r20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1
r21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0
r22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0	0
r23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1	0
r24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1
r25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1

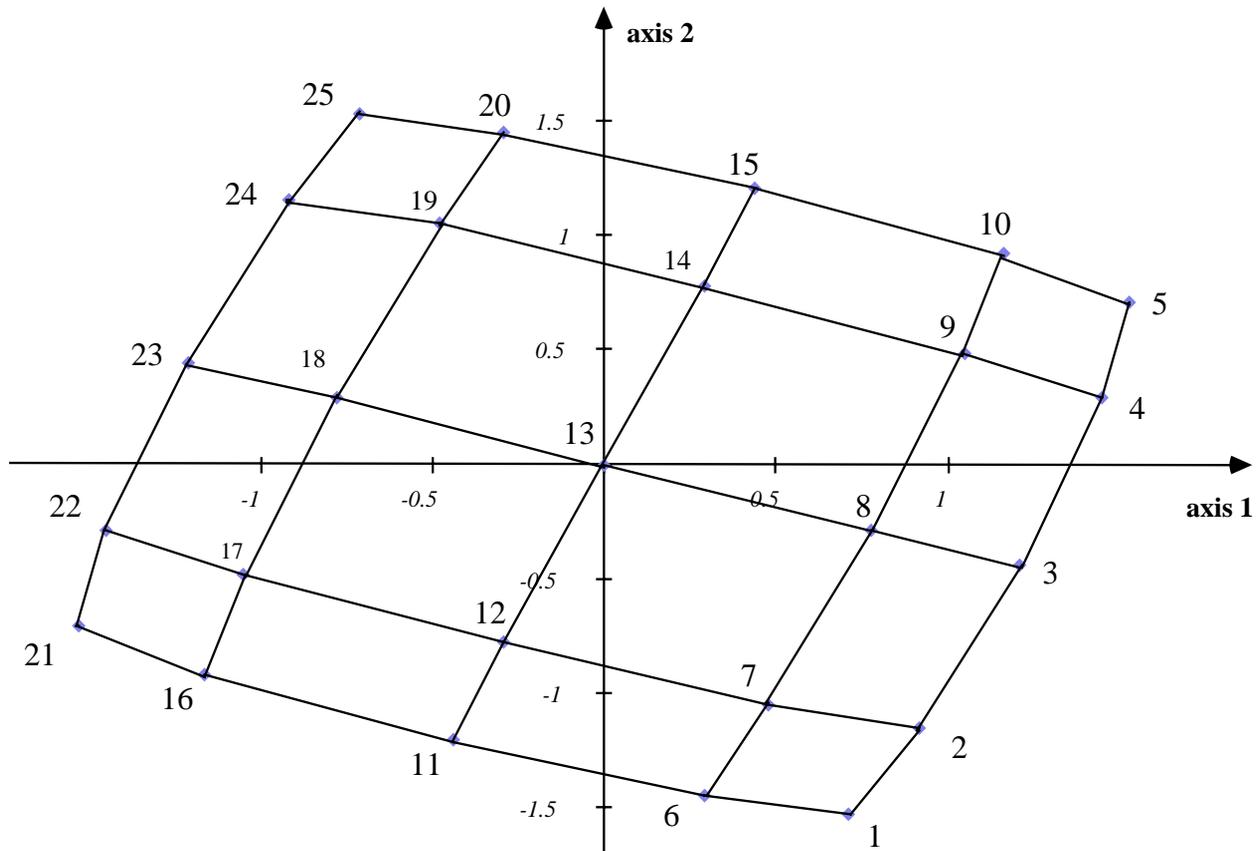
Codage
réduit
de **M**

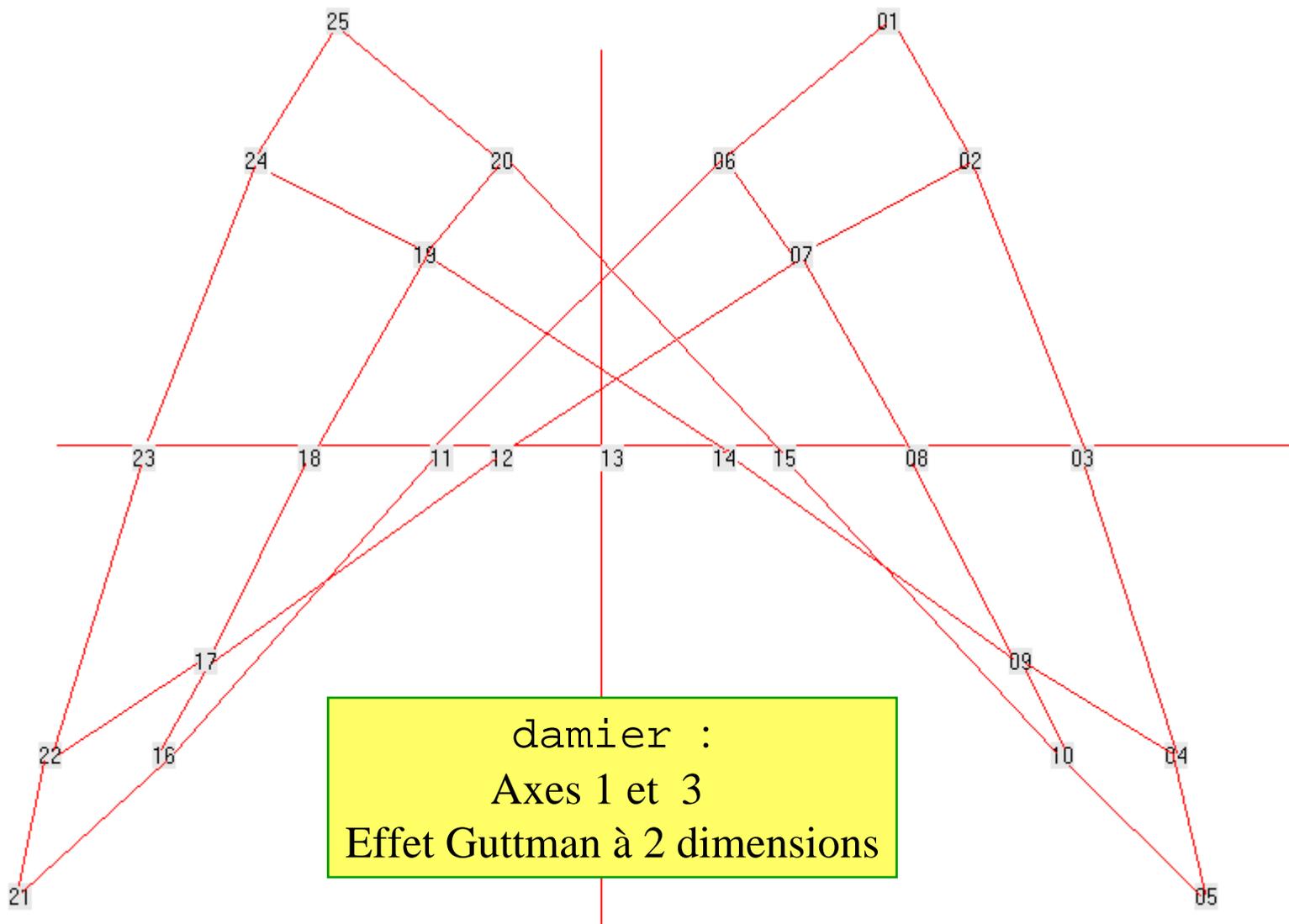
<i>Vertex</i>	<i>Adjacent vertices</i>				
1	1	2	6		
2	2	1	7	3	
3	3	2	8	4	
4	4	3	9	5	
5	5	4	10		
6	6	1	7	11	
7	7	2	6	8	12
8	8	3	7	9	13
9	9	4	8	10	14
10	10	5	9	15	
11	11	6	12	16	
12	12	7	11	13	17
13	13	8	12	14	18
14	14	9	13	15	19
15	15	10	14	20	
16	16	11	17	21	
17	17	12	16	18	22
18	18	13	17	19	23
19	19	14	18	20	24
20	20	15	19	25	
21	21	16	22		
22	22	21	17	23	
23	23	18	22	24	
24	24	19	23	25	
25	25	24	20		

Description du damier **G** par
ACP de **M**

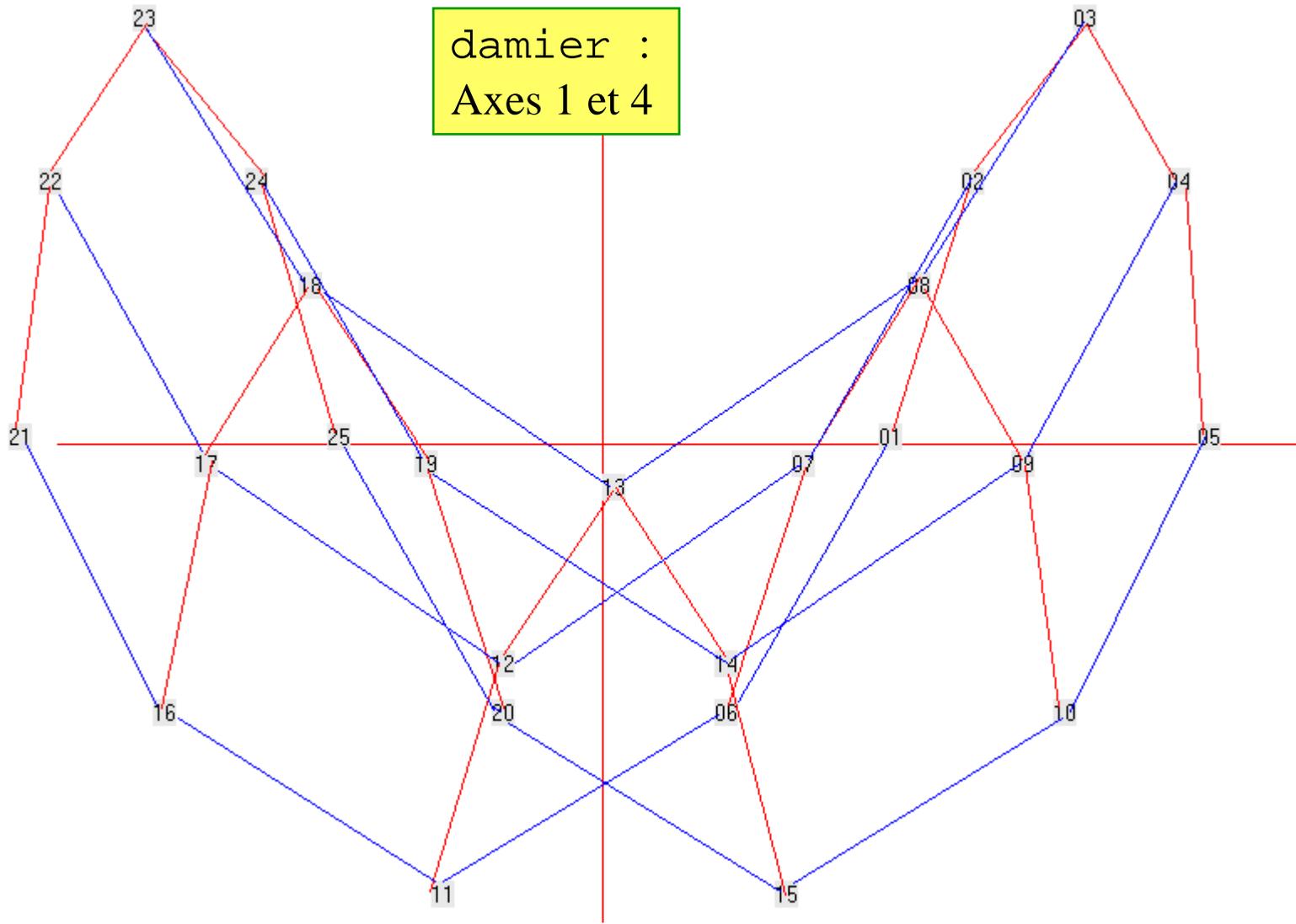


Description du damier **G** par AC de **M**

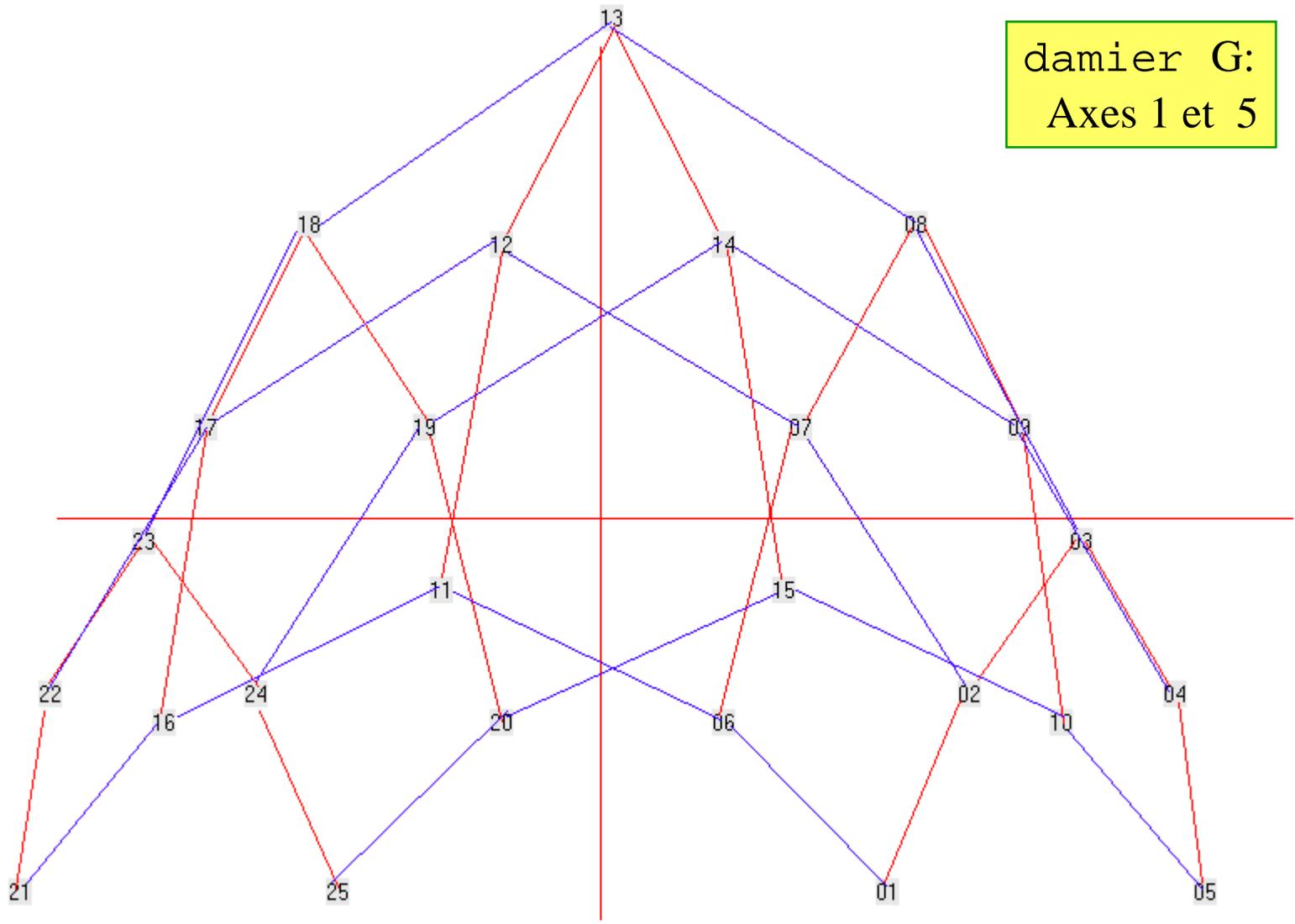


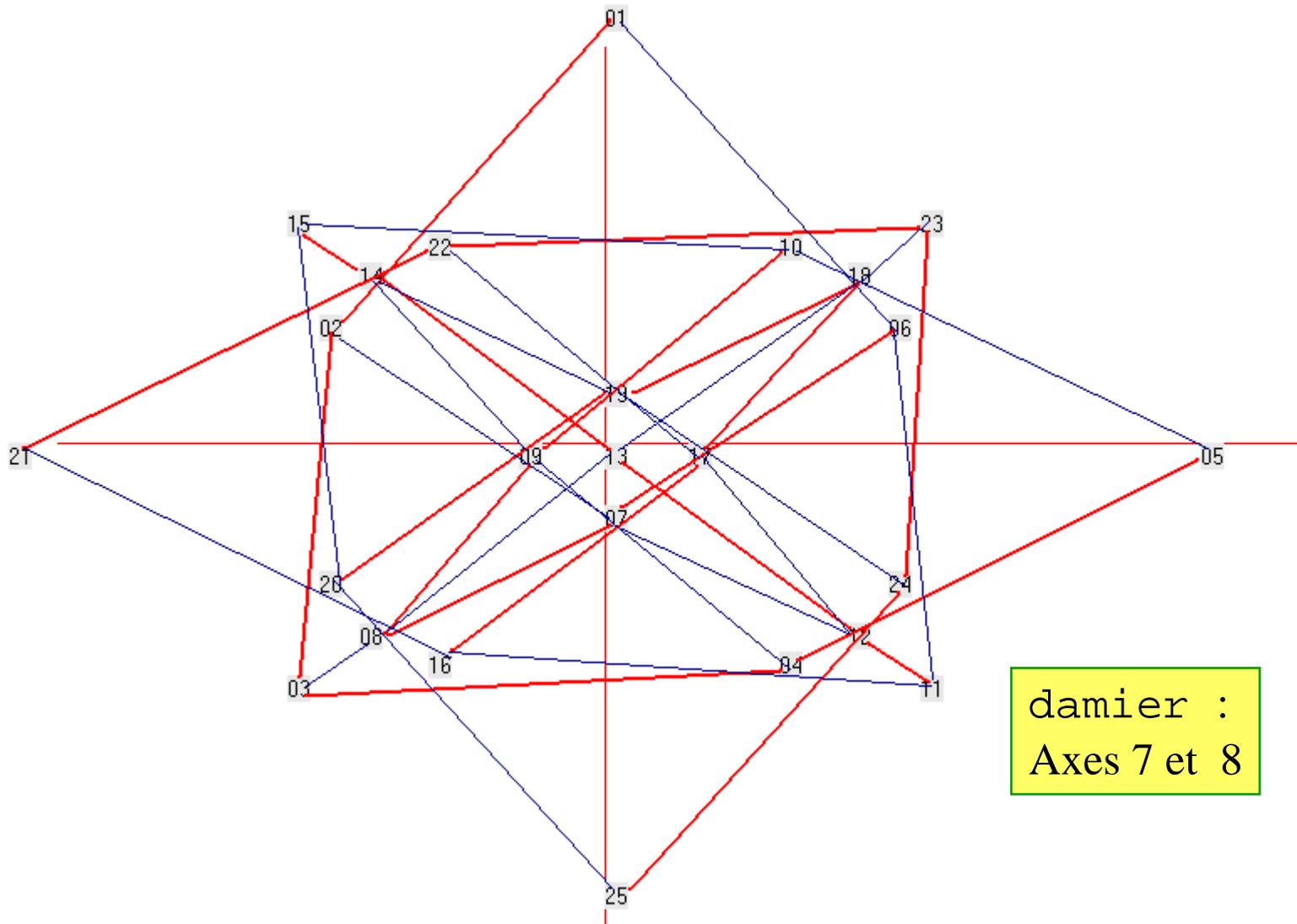


damier :
Axes 1 et 4



damier G:
Axes 1 et 5





damier :
Axes 7 et 8

Notons que les calculs peuvent être faits analytiquement..

Dans le cas le plus simple (une chaîne) la description par AC conduit à la relation : $\mathbf{Ma} = \lambda \mathbf{a}$, qui correspond à l'équation aux différences.

$$a_{i-1} + a_{i+1} = \lambda a_i$$

→

$$\varphi_\alpha(j) = \cos\left(\frac{2j\alpha\pi}{n}\right) \quad \phi_\alpha(j) = \sin\left(\frac{2j\alpha\pi}{n}\right)$$

Un damier peut être défini comme une somme tensorielle de chaînes et le résultat final s'en déduit (cf : Benzécri, 1973). On peut noter l'analogie avec les éléments propres de l'opérateur de Laplace.

Quelques références à propos du « Laplacien » (matrix: $N - M$)
d'un graphe :

CHUNG F.R.K., *Spectral Graph Theory*. CBMS Reg. Conf. Ser. Math. 92, American Mathematical Society, 1997.

KOREN Y., CARMEL L., HAREL D., ACE: a Fast Multiscale Eigenvectors Computation for Drawing Huge Graphs, *Proceedings of IEEE Information Visualization, 2002*, p 137-144.

MOHAR B., The Laplacian Spectrum of Graphs, *Graph Theory, Combinatorics and Application*, 2, 1991, p 871-898.

MOHAR B., Some Applications of Laplace Eigenvalues of Graphs, *Graph Symmetry, Algebraic Methods and Application*, Hahn G., Sabidussi G., NATO Ser. C., 497, Kluwer, 1997, p 225-275.

Pourquoi obtient-on des visualisations de bonne qualité ?

$$\begin{aligned} \text{variance locale} &= \mathbf{y}'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \mathbf{y} \\ \text{variance globale} &= \mathbf{y}'\mathbf{y} \end{aligned}$$

Bornes pour $c(\mathbf{y}) = \text{coefficient de contigüité}$.

$$c(\mathbf{y}) = \mathbf{y}'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \mathbf{y} / \mathbf{y}' \mathbf{y}$$

le minimum of $c(\mathbf{y})$, μ , est la racine de la plus petite valeur propre μ^2 de :

$$(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \boldsymbol{\psi} = \mu^2 \boldsymbol{\psi}$$

Si le graphe est régulier, $\mathbf{N}^{-1} = (1/r) \mathbf{I}$

$$(\mathbf{I} - (1/r)\mathbf{M})^2 \boldsymbol{\psi} = \mu^2 \boldsymbol{\psi}$$

$$(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \boldsymbol{\psi} = \mu \boldsymbol{\psi}$$

Ou encore : $\mathbf{N}^{-1}\mathbf{M} \boldsymbol{\psi} = (1 - \mu) \boldsymbol{\psi}$

On a obtenu la relation : $\mathbf{N}^{-1}\mathbf{M} \psi = (1 - \mu) \psi$

La formule de transition de l'AC de \mathbf{M} s'écrit, (**la et non les !**)

$$\mathbf{N}^{-1}\mathbf{M} \phi = \varepsilon \sqrt{\lambda} \phi$$

si $\varepsilon = +1$, facteur direct,

si $\varepsilon = -1$, facteur inverse .

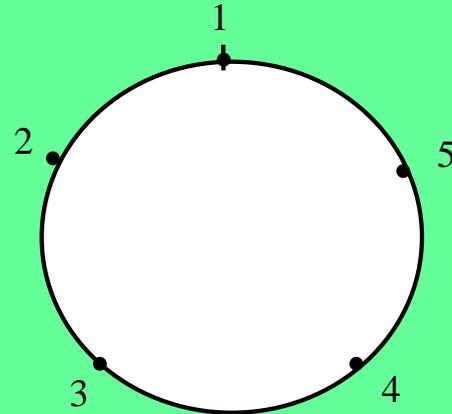
(Cf., e.g., l'analyse des matrices de confusion).

Alors: $\text{Min} [c(y)] = \text{Min} \mu = \text{Max} \lambda$, $[= \lambda_{max}$ if ($\varepsilon = 1$).]

Finalemment : $\text{Min} [c(y)] = 1 - \sqrt{\lambda_{max}}$

Mesure d' *information* paradoxale: cas du *cycle*

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$



$$\varphi_\alpha(j) = \cos\left(\frac{2j\alpha\pi}{n}\right)$$

$$\phi_\alpha(j) = \sin\left(\frac{2j\alpha\pi}{n}\right)$$

$$\lambda_\alpha = \cos^2\left(\frac{2\alpha\pi}{n}\right)$$

$$\tau_\alpha = \frac{2}{n} \cos^2\left(\frac{2\alpha\pi}{n}\right)$$

Le coefficient de contiguité peut être généralisé :

a) à différentes distances entre sommets sur le graphe,

b) à des observations multivariées

(Pour le coefficient de Geary, cf.: Lebart, 1969).

a) Le graphe correspondant à la distance définie comme “ le plus court chemin de longueur k joignant deux sommets ” est associée à la matrice $\mathbf{M}^{(k)} - \mathbf{M}^{(k-1)}$, où $\mathbf{M}^{(k)}$ désigne la k -th puissance booléenne de la matrice $(\mathbf{I} + \mathbf{M})$.

Ceci permet de tester une éventuelle auto-corrélation spatiale..

Cette approche est une variante, pour les graphes, du variogramme de la *géostatistique*, présentée dans les travaux de pionniers de Matheron (1963, 1965).

b) Généralisation aux observations multivariées

Soit $\mathbf{Y}'\mathbf{u}$ le vecteur des n valeurs de la combinaison linéaire u des p variables.

Son coefficient de contiguïté vaut alors :

$$c(u) = \mathbf{u}' \mathbf{V}^* \mathbf{u} / \mathbf{u}' \mathbf{V} \mathbf{u}$$

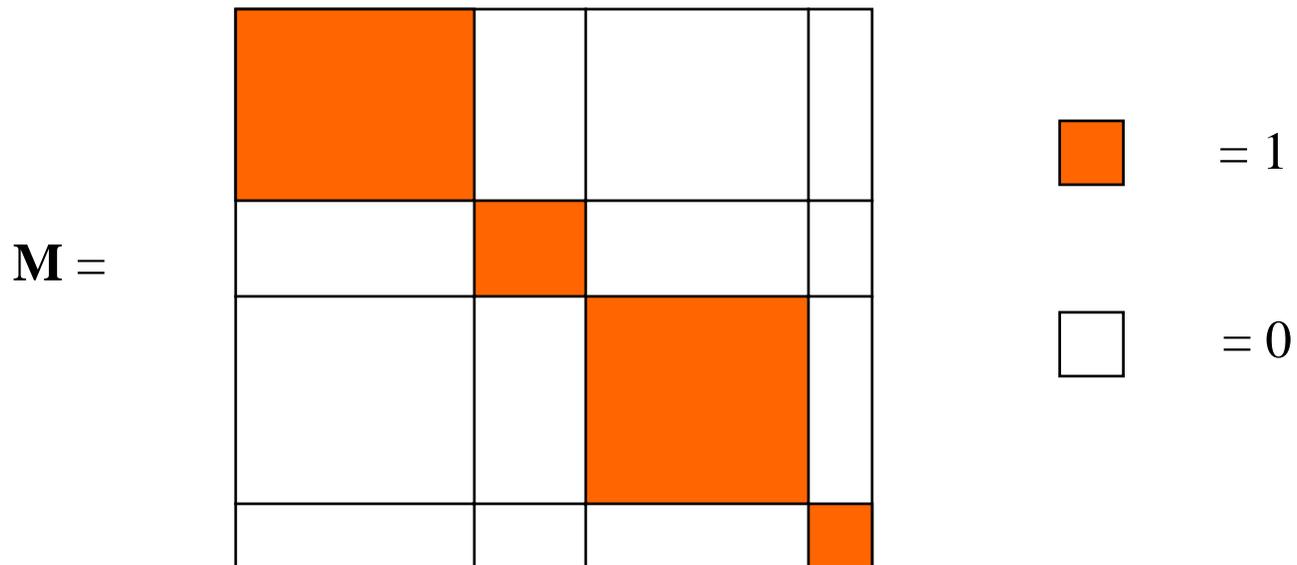
... où : $\mathbf{V}^* = (1/n) \mathbf{Y}'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \mathbf{Y}$

est la matrice (p, p) de covariance locale.

L'analyse de contiguïté est la recherche du minimum de $c(u)$:

$$c(u) = \mathbf{u}' \mathbf{V}^* \mathbf{u} / \mathbf{u}' \mathbf{V} \mathbf{u}$$

Elle se régit à une Analyse Discriminante de Fisher quand G is associé au graphe d'une partition.



$C(u)$ nous permet de travailler avec des classes empiétantes, des partitions floues.

Graphes externes : corrélation partielle

La matrice des corrélation locale permet de définir un “contexte local” à partir d’un ensemble Z de variables instrumentales, pour étudier Y à Z constant.

$$V^* = (1/n) Y'(I - N^{-1}M)' (I - N^{-1}M) Y$$

M peut vérifier : $m(i,j) = 1$ iff $d^2(z_i, z_j) < t_0$

(sans “effet de bord”).

- Cf. Kendall (1941), Mantel, (1967), Quade (1974), Hubert (1985).

Comparaison avec $V^* = (1/n) Y'(I - Z(Z'Z)^{-1}Z') (I - Z(Z'Z)^{-1}Z') Y$

- « régression Quasi-Polynomiale » au lieu de régression linéaire.

Graphes internes

L'idée est de résumer la structure locale par un graphe

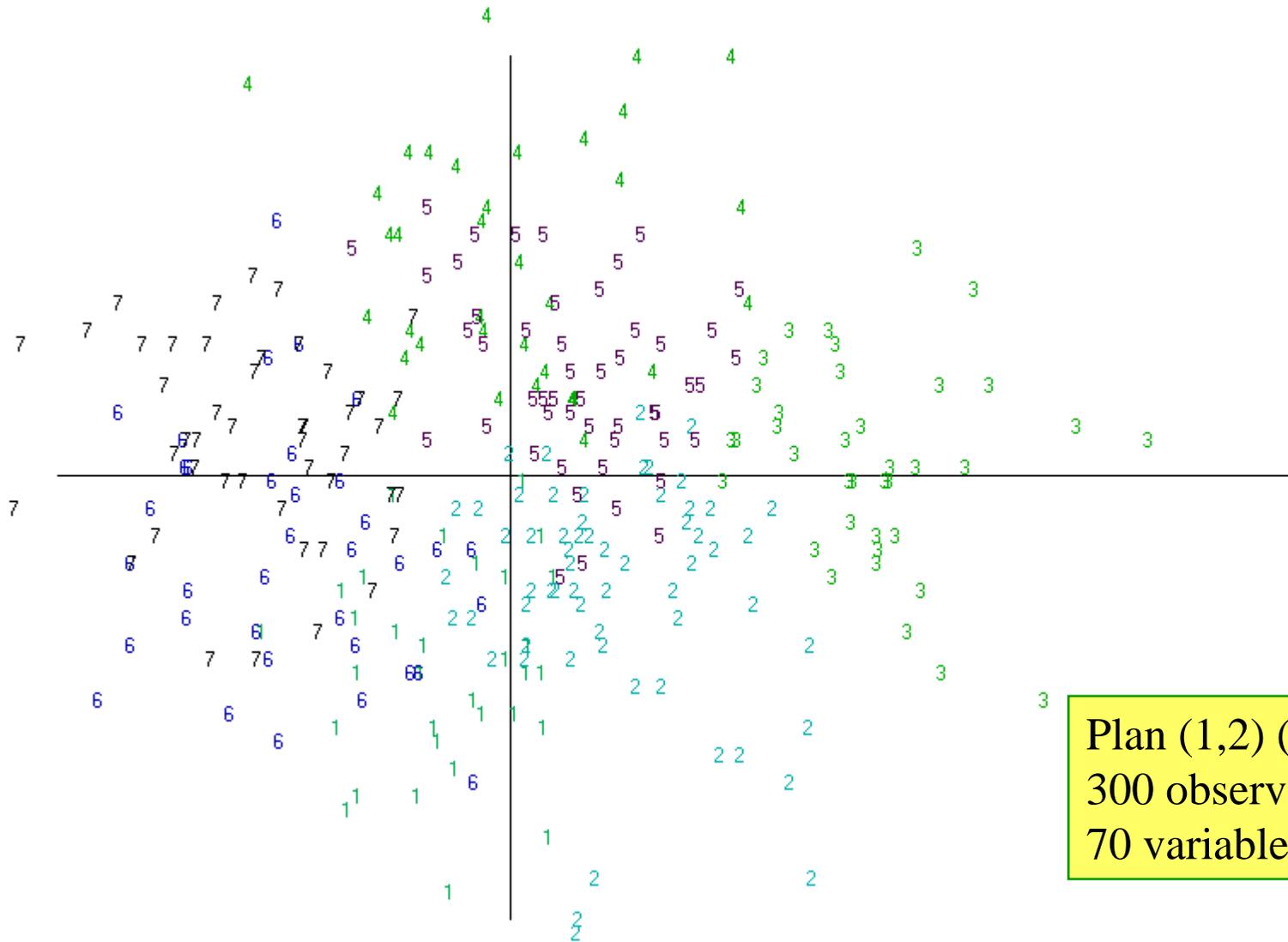
→ graphe de PPV ($1 \leq k \ll n$)

→ graphe de seuil de distance

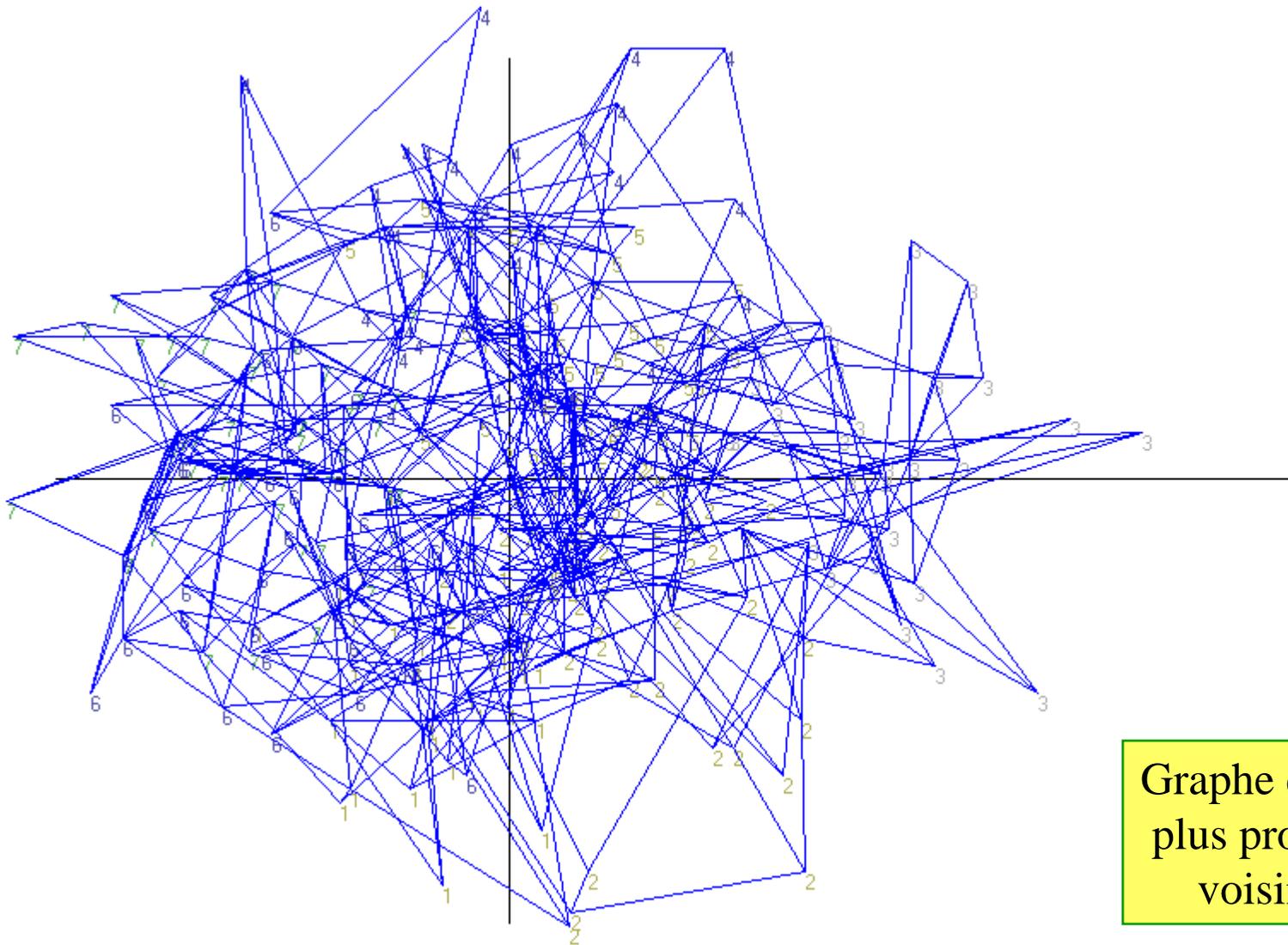
La matrice des covariances locale

$$\mathbf{V}^* = (1/n) \mathbf{Y}'(\mathbf{I} - \mathbf{N}^{-1}\mathbf{M})' (\mathbf{I} - \mathbf{N}^{-1}\mathbf{M}) \mathbf{Y}$$

permet de **délinéariser** la recherche des axes principaux ...

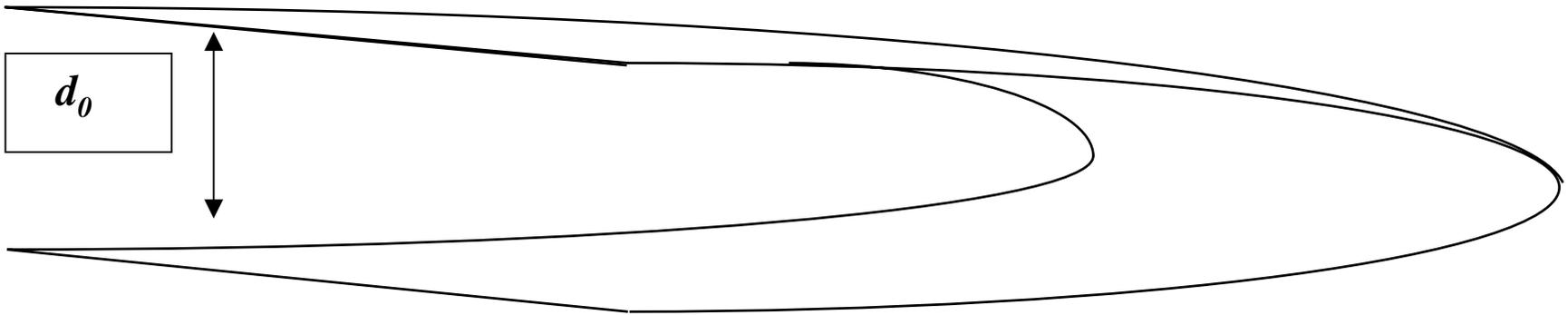


Plan (1,2) (PCA)
 300 observations
 70 variables



Graphe des 3 plus proches voisins

Exemple de dépliage (unfolding)



Le seuil de distance d doit être $< d_0$ pour permettre à l'analyse de contiguïté de déplier le nuage.

Exemple 1: Iris d' Anderson -Fisher :

Les 3 catégories a priori peuvent être ignorées...

1	5.10	3.50	1.40	.20	1
2	4.90	3.00	1.40	.20	1
3	4.70	3.20	1.30	.20	1
4	4.60	3.10	1.50	.20	1
.....					
51	7.00	3.20	4.70	1.40	2
52	6.40	3.20	4.50	1.50	2
53	6.90	3.10	4.90	1.50	2
.....					
101	6.30	3.30	6.00	2.50	3
102	5.80	2.70	5.10	1.90	3
103	7.10	3.00	5.90	2.10	3

Traitement des données « Iris »

ACP

Visualisation (V)

Graphe Externe (partition) : discrimination selon Fisher (V)

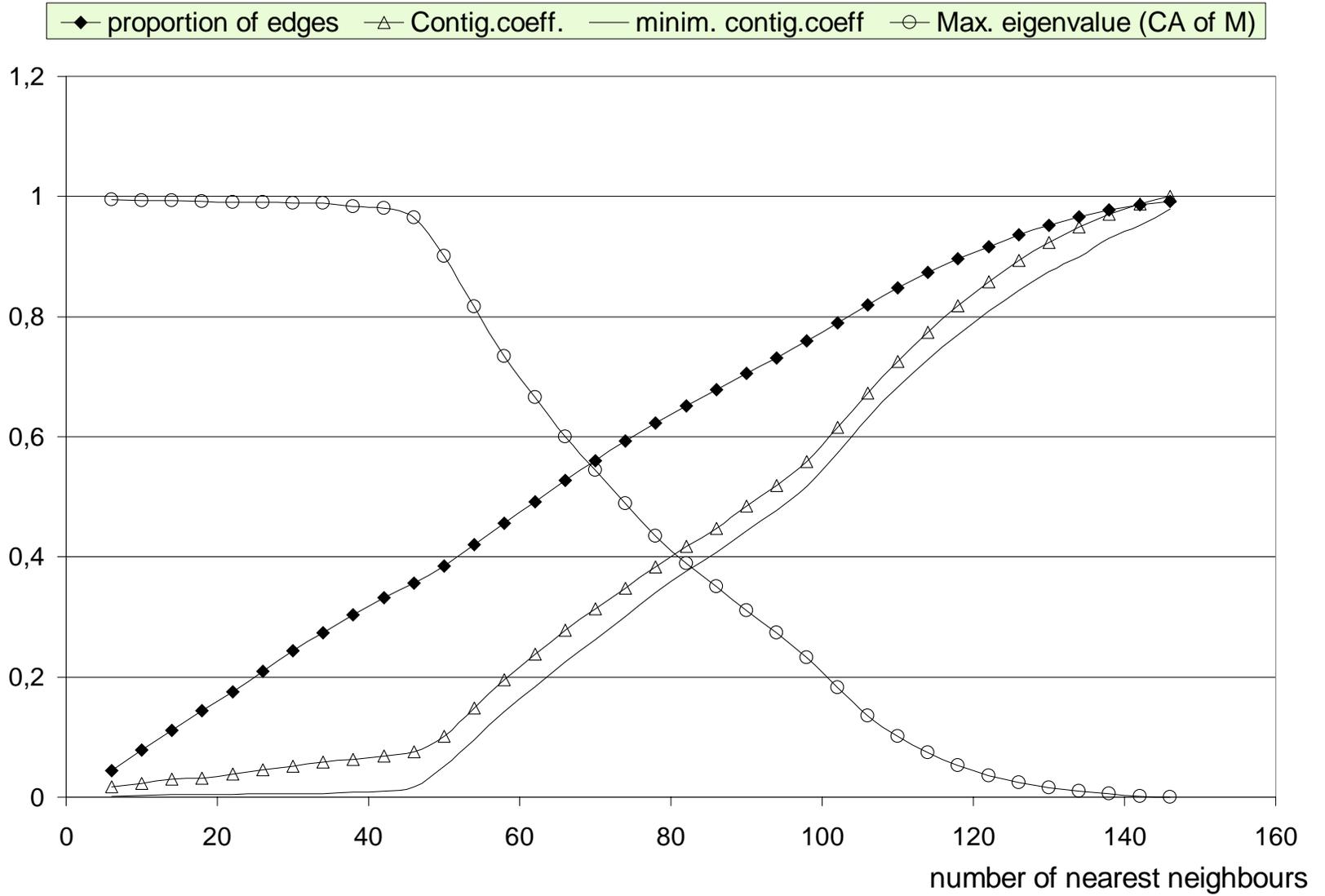
Graphes de PPV (V)

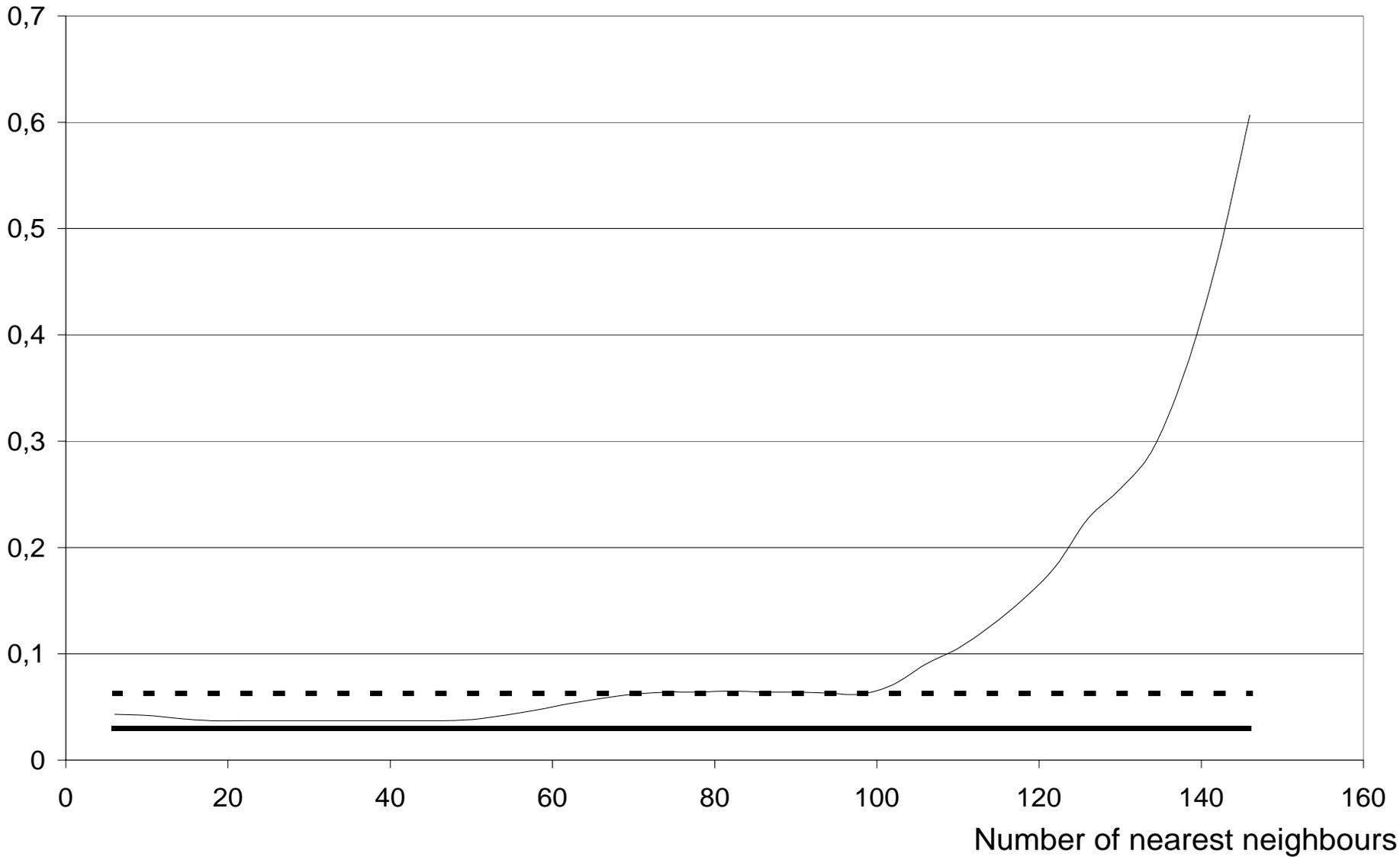
Analyse de Contiguïté (V)

Graphes à seuil

Analyse de Contiguïté(V).

Synthèse numérique ...





Example 2 : Visualization in Semiometry

The basic idea of “Semiometry” is to insert in a questionnaire a series of questions consisting uniquely of words.
(a list of 210 words is currently used, but some abbreviated lists containing a subset of 70 words could be used as well).

The interviewees must rate these words according to a seven levels scale, the lowest level (mark = 1) concerning a "most disagreeable (or unpleasant) feeling about the word", the highest level (mark = 7) concerning a "most agreeable (or pleasant) feeling" about the word.

Questionnaires in 5 languages

FRENCH	ENGLISH	GERMAN	SPANISH	ITALIAN
l'absolu	absolute	absolut	el absoluto	l'assoluto
l'acharnement	persistence	hartnaeckig	el empeno	l'accanimento
acheter	to buy	kaufen	comprar	comprare
admirer	to admire	bewundern	admirar	ammirare
adorer	to love	anbeten	adorar	adorare
l'ambition	ambition	der ehrgeiz	la ambicion	l'ambizione
l'âme	soul	die seele	el alma	l'anima
l'amitié	friendship	die freundschaft	la amistad	l'amicizia
l'angoisse	anguish	die angst	la angustia	l'angoscia
un animal	animal	ein tier	un animal	un animale
un arbre	tree	ein baum	un arbol	un albero
l'argent	silver	das geld	el dinero	il denaro
une armure	armour	die ruestung	una armadura	un'armatura
l'art	art	die kunst	el arte	l'arte

122	La modestie	-3	-2	X	0	+1	+2	+3
133	Mcelleux	-3	-2	-1	X	+1	+2	+3
124	La mort	-3	X	-1	0	+1	+2	+3
100	Une muraille	-3	-2	-1	0	+1	+2	+3
085	Un mystère	-3	-2	-1	0	+1	+2	+3
105	Nager	-3	-2	-1	0	+1	+2	+3
043	Une naissance	-3	-2	-1	0	+1	+2	+3
025	Un nid	-3	-2	-1	0	+1	+2	+3
106	La nudité	-3	-2	-1	0	+1	+2	+3
071	Obéir	-3	-2	-1	0	+1	+2	+3
173	L'océan	-3	-2	-1	0	+1	+2	+3
086	Un orage	-3	-2	-1	0	+1	+2	+3

Facsimile of a questionnaire

The processing of the filled questionnaires (*mainly through Principal Components Analysis*) produces a stable pattern (*up to 8 stable principal axes*).

Very similar patterns are obtained in ten different countries, despite the problems posed by the translation of the list of words.

Surveys carried out since 1990

Country	Year	Sample Size
France	1990	2764
France	1996	2764
Spain	1996	2983
U.K.	1996	1849
Italy	1996	2606
Germany	1996	3065
France	1998	2764
Greece	2000	1062
France	1999	2763
Canada	1998	1865

Ludovic Lebart
Marie Piron
Jean-Francois Steiner

LA SÉMIOMÉTRIE

Essai de statistique structurale

Au-delà de leur signification, les mots, par les souvenirs qu'ils mobilisent, ont le pouvoir de provoquer en nous des sensations agréables ou désagréables.

De cette observation simple est née une méthode, la Sémiométrie, largement utilisée en marketing et dans les études psychosociologiques. Mais, grâce à la puissance des outils statistiques actuels, son pouvoir d'investigation va, bien au-delà de ces applications pratiques, jusqu'aux confins de la psychanalyse et de la linguistique.

Il semblait indispensable qu'un ouvrage fasse le point sur les principes de cette méthode, les travaux réalisés et les applications potentielles. Le lecteur peut maintenant découvrir, au fil des chapitres, l'étendu du travail d'expérimentation, la sévérité des épreuves de validation, la profondeur et la finesse des résultats obtenus, enfin les promesses de cet outil transdisciplinaire.

Plusieurs niveaux de lecture sont possibles selon les connaissances mathématiques et statistiques du lecteur ; les développements plus techniques sont en effet regroupés dans une annexe.

Ce livre s'adresse aux spécialistes du marketing et de la communication, aux socio-économistes, aux statisticiens, aux psychosociologues, aux linguistes. Il intéresse un large public, allant des sociétés d'études et des instituts de sondage aux étudiants, professeurs et chercheurs des universités ou des grandes écoles.



ISBN 2 10 008105 5

<http://www.dunod.com>



L. LEBART
M. PIRON
J.-F. STEINER

LA SÉMIOMÉTRIE

La Sémiométrie

LUDOVIC LEBART
MARIE PIRON
JEAN-FRANCOIS STEINER

DUNOD

Processing Semimetric data

PCA + bootstrap

Visualization (V)

Clustering 1

Clustering 2 (Kohonen)

Some nearest neighbours graphs (V)

Contiguity analysis (V)

Conclusion

- Transparence du traitement (solutions analytiques)
- Extension de l 'analyse lineaire disriminante à des situations plus complexes
- Possibilité d 'ACP robuste et de corrélation partielle non-linéaire.
- Possibilités de validation.

- ALUJA T. and LEBART L. (1984): Local and Partial Principal Component Analysis and Correspondence Analysis, *COMPSTAT Proceedings*. Physica Verlag, Vienna, 113-118.
- ANSELIN L. (1995): Local indicators of spatial association . *Geog. Anal.*, 27, 2, 93-115.
- ART D., GNANADESIKAN R., KETTENRING J.R. (1982): Data Based Metrics for Cluster Analysis, *Utilitas Mathematica*, 21 A, 75-99.
- BENZECRI, J.P. (1973): *Analyse des Données: Correspondances*. Dunod, Paris.
- BURTSCHY B., LEBART L. (1991): Contiguity analysis and projection pursuit. In: *Appl. Stoch. Mod. and Data Anal.* R. Gutierrez et al., Eds, World Scientific, Singapore, 117-128.
- CAZES P. (1986) Correspondance entre deux ensembles et partition de ces deux ensembles, *Les Cahiers de l'Analyse des Données*, vol.XI, no.3, 335-340.
- CHUNG F.R.K., *Spectral Graph Theory*. CBMS Reg. Conf. Ser. Math. 92, American Mathematical Society, 1997.
- CLIFF A.D. et ORD J.K. (1981): *Spatial Processes: Models and Applications*. Pion, London.
- COTTRELL M., ROUSSET P. (1997): The Kohonen Algorithm: a powerful tool for analysing and representing multidimensional qualitative and quantitative data. In: *Biological and Artificial Computation : From Neuroscience to Technology*. J. Mira, R. Moreno-Diaz, J. Cabestany, (eds), Springer, 861-871.
- ESCOFIER B. (1989): Multiple correspondence analysis and neighboring relation. In: *Data Analysis, Learning Symbolic and Numeric Knowledge*. Diday E. (ed.), Nova Science Publishers, New York, 55-62.
- FARAJ A. (1993): Analyse de contiguïté: une analyse discriminante généralisée à plusieurs variables qualitatives. *Revue Statist. Appl.*, 41, (3), 73-84.

- GEARY R.C. (1954): The Contiguity Ratio and Statistical Mapping, *The Incorporated Statistician*, 5, 115-145.
- GNANADESIKAN R., KETTENRING J.R. et LANDWEHR J.M. (1982): Projection Plots for Displaying Clusters, In: *Statistics et Probability*. G. Kallianpur *et al.*, eds, North-Holland.
- GOWER J. C. (1984): Procrustes analysis. In: *Handbook of Applicable Mathematics*. 6, Lloyd E.H. (ed.), J. Wiley, Chichester, 397-405.
- HUBERT L. (1985): Combinatorial data analysis: association and partial association. *Psychometrika*, 50, 4, 449-467.
- KLEIWEG P. (1996): *Een inleidende cursus met practica voor de studie Alfa-Informatica*. Master's thesis, Rijksuniversiteit Groningen, 1996.
- KOHONEN T.(1989): *Self-Organization and Associative Memory*. Springer-Verlag, Berlin, 3rd edition, 1989.
- KOREN Y., CARMEL L., HAREL D., ACE: a Fast Multiscale Eigenvectors Computation for Drawing Huge Graphs, *Proceedings of IEEE Information Visualization, 2002*, p 137-144.
- LAFOSSSE R. (1985): *Analyse Procustéenne de deux tableaux*. Thèse, Université de Toulouse.
- LE FOLL Y. (1982): Pondération des distances en analyse factorielle. *Statist. et Anal. des Données*. 7, 13-31.
- LEBART L. (1969): Analyse Statistique de la Contiguïté, *Publ. de l'ISUP*. XVIII, 81-112.
- LEBART, L. (2000): Contiguity Analysis and Classification, In: W. Gaul, O. Opitz and M. Schader (Eds): *Data Analysis*. Springer,Berlin, 233--244.
- LEBART, L. (2001): Representing words and texts through contiguity analysis. In: *ASMDA 2001, 10th International Symposium on Applied Stochastic Models and Data Analysis*. G. Govaert, J. Janssen, N. Limnios (eds), UTC, Compiègne, 654-659.

- LEBART, L., SALEM, A. and BERRY, L. (1998): *Exploring Textual Data*. Kluwer, Dordrecht.
- LEBART L., MIRKIN B. (1993): Correspondence Analysis and Classification. In: *Multivariate Analysis: Future Directions 2*. Cuadras C.M. and Rao C.R., (eds), North-Holland, 341-357.
- LEBART L., MORINEAU A. PIRON M., L. (1998): *Statistique Exploratoire Multidimensionnelle*,. Dunod, Paris.
- LEBART L., PIRON M., STEINER J.F. (2003): *La Sémiométrie*. Dunod, Paris.
- MATHERON G. (1963): Principles of geostatistics. *Economic Geology*. 58, 1246-1266.
- MEOT A., CHESSEL D. et SABATIER R. (1993): Opérateur de voisinage et analyse des données spatio-temporelles. In *Biométrie et environnement*, Lebreton J.-D., Asselain B., (eds), Masson, Paris, 45-71.
- MOM A. (1988): *Methodologie Statistique de la Classification des reseaux de transport*. Thèse, Université des Sciences et Techniques du Languedoc, Montpellier.
- RIPLEY B. D. (1981): *Spatial Statistics*. J. Wiley, New York.
- SCHONEMANN P. H. (1968): On two-sided orthogonal procrustes problems. *Psychometrika*. 33, 19-33.
- STEINER J.-F. and AULIARD, O. (1992): La sémiométrie: un outil de validation des réponses. In *Qualité de l'Information dans les Enquêtes*, ASU (ed) Dunod, Paris, 241-274.
- TUCKER, L. R. (1958): An inter-battery method of factor analysis. *Psychometrika*. 23, (2).
- VON NEUMANN, J.(1941): Distribution of the ratio of the mean square successive differences to the variance. *Ann. of Math. Statistics*. 12, 367-395.

Danke

Thank You

Obrigado

Grazie

Gracias

Domo Arigato

Choukrane

Merci