Corrected Version of the paper published under the same title , In:
W. Gaul, O. Opitz and M. Schader (Eds): *Data Analysis*. Springer, Berlin, 233 - 244.

Contiguity analysis and classification

Ludovic Lebart
CNRS / ENST, 46 Rue Barrault, 75013, Paris, France

Abstract
Let us consider n vertices of a symmetric graph G whose associated matrix is **M** ($m_{ii'}$ = 1 if vertices i and i' are joined by an edge, $m_{ii'}$ = 0 otherwise). These vertices are simultaneously described by p variables (x_{ij} is the value of variable j for vertex i). Such situation occurs when vertices represent time-points, geographic areas, or the vertices of a « nearest neighbours graph » derived from a particular distance between observations.
Contiguity Analysis simultaneously uses the local covariance matrix **C** and the global covariance matrix **V**. The minimization of the ratio: u'**C**u / u'**V**u (**u** being a p-vector) provides a generalization of discriminant analysis to the case of overlapping clusters. We consider here the case of graphs that are not external, but derived from the observations themselves, namely the series of nearest neighbours graphs.

Keywords: Contiguity Analysis, Graph visualization, Discriminant Analysis.

The first part of this paper considers the case of a set of multivariate observations, (n objects described by p variables, leading to a (n,p) matrix **X**), having an *a priori* graph structure. The n observations are the vertices of a symmetric graph G, whose associated matrix is **M** ($m_{ii'}$ = 1 if vertices i and i' are joined by an edge, $m_{ii'}$ = 0 otherwise). Contiguity Analysis, providing a straightforward generalization of Linear Discriminant Analysis, can then be defined. It permits to point out the levels responsible of the observed patterns (*local* versus *global* level).
The second part is devoted to the situation where the graph structure is not external, but derived from the matrix **X** itself. Some interesting possibilities of exploration of data are sketched. The idea of deriving a metric likely to highlight the existence of clusters dates back to the works of Art et al. (1982) and Gnanadesikan et al. (1982). We present here the contribution of contiguity analysis to such approaches.
The third part deals with external graphs built from a set of instrumental variables. In that context, contiguity analysis provides a powerful tool for analysing partial correlations.

1. Local principal component analysis, contiguity analysis

1.1. Local variance $v^c(y)$ of a variable y

y being a random variable taking values on each vertex i of a symmetric graph G, with m/2 edges, a first definition of the local variance $v^c(y)$ is:

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

The symbol $\Sigma^{(c)}$ means : sum for all i and j such that vertices i and i' are joined by an edge. An equivalent writing, using the binary matrix **M** = ($m_{ii'}$) associated with the graph G, is:

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

Note that if G is a complete graph (all pairs (i,i') are joined by an edge), $v^c(y)$ is nothing but v(y), the classical empirical variance. When the observations are distributed randomly on the graph, both $v^c(y)$ and $v(y)$ are estimates of the variance of y.

The contiguity ratio $c(y)$ (Geary, 1954), generalizing the *Von Neumann ratio* (1941), reads:

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

A value of the contiguity ratio $c(y)$ significantly lower than *1* indicates a positive spatial auto-correlation for the variable y. (Many more coefficients and ratios have been proposed in the same vein; see, e.g., Ripley, 1981 ; Cliff and Ord, 1981; Anselin, 1995). A review and a synthesis of various approaches can be found in Méot *et al.* (1993).
An important change will be made to the definition of the coefficient $c(y)$ to render the local variance compatible with the "within" variance when the graph describes a partition of observations (i.e. a series of complete disconnected sub-graphs, or cliques) (Mom, 1988 ; see also: Escofier, 1989 ; Benali and Escofier, 1990): the local variance is then defined as the mean squared difference between the value for a vertex and the average of its neighbouring values.

We denote by **M** the (n,n) matrix associated with G having n vertices; we denote by **N** the (n,n) diagonal matrix having the degree of each vertex i as diagonal element n_i (n_i stands here for $n_{ii'}$). y is the vector whose i-th components is y_i.
Note that: $n_i = \Sigma_{i'} m_{ii'}$.

The local variance will then be defined as:

$$v^*(y) = (1/n)\sum_{i=1}^{i=n}(y_i - m_i^*)^2, \quad \text{with:} \quad m_i^* = (1/n_i)\sum_{k=1}^{k=n_i} m_{ik} y_k$$

It is the average of the adjacent values of vertex i. Note that $v^*(y) = v^c(y)$ if G is regular (i.e. if n_i is constant).

The new contiguity ratio is written: $c^*(y) = v^*(y) \ / v(y)$

1.2. Bounds for c(y)

It will be reminded in this section that the eigenvectors derived from the correspondence analysis (CA) of a matrix **M** associated with a regular graph G have optimal properties with respect to the contiguity ratio.
For a variable y, the coefficient $c(y)$ is written (see Lebart, 1969):

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

where **I** is the unit (n, n) matrix, and **U** the (n, n) matrix such that $u_{ij} = 1$, for all i and j.
The transformed coefficient reads:

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

We will assume in the following that **y** is centred to simplify the denominator. Since G is supposed to be regular, let us call r the number of edges of each vertex. The matrix N is a scalar matrix that can be written $\mathbf{N} = (1/r)\,\mathbf{I}$.
Therefore : $c^*(y) \ = \ \mathbf{y'\,(I} - (1/r)\,\mathbf{M})^2\,\mathbf{y} \ / \ \mathbf{y'\,y}$

The minimum of $c^*(y)$, μ, is thus the smallest eigenvalue of $(\mathbf{I} - (1/r)\,\mathbf{M})^2$ corresponding to the eigenvector ψ such that:

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

This equation can be written :

$$(\mathbf{I} - (1/r)\,\mathbf{M})\ \psi = \varepsilon\sqrt{\mu}\ \psi$$

or: $$(1/r)\,\mathbf{M}\,\psi = (1 - \varepsilon\sqrt{\mu})\ \psi$$

If $\varepsilon = +1$, the eigenvector is said to be direct, whereas if $\varepsilon = -1$, it is said to be inverse (Benzécri, 1973). An inverse eigenvector corresponds to a negative eigenvalue of the initial symmetric matrix **M**.

Note that the *transition formulas* corresponding to the correspondence analysis of the matrix **M** are written for the first axis :

$$(1/r)\,\mathbf{M}\,\phi = \varepsilon\,\sqrt{\lambda}\,\phi$$

Since $c^*(y)$ is positive, the minimum *value* μ_0 corresponds to the maximum value λ_{max} for a direct eigenvector ($\varepsilon = +1$). Therefore, the lower bound of $c^*(y)$ is :

$$\text{Min}\,[\,c^*(y)\,] = 1 - \sqrt{\lambda_{max}}$$

That minimum is reached when ψ is the first factor ϕ derived from the correspondence analysis of matrix **X**. Then, the sequence of the first factors ϕ_r corresponds to a sequence of orthogonal variables having the property of extremal contiguity. This property accounts for the good quality of the description of graphs through the correspondence analysis of their associated matrix. It will be exemplified in the following section.

1.3 A comparison with principal components analysis

Both correspondence analysis (CA) and principal components analysis (PCA) are closely related to the general property of rectangular matrices known as *singular value decomposition*. Although it is clear that CA is appropriate for count data or binary data and PCA for real valued measurement, the user of this latter method (much more widespread) may legitimately ask what are the risks of false results when applying it to count or binary data (see: Lebart *et al.*, 1998). Since it seems natural to calibrate visualization tools on artificial data sets provided with an *a priori* structure, we present below a comparison of the two methods applied to a same binary data matrix associated with a "chessboard shaped graph", (figure 1). In this figure, a line (an *edge*) drawn between two *vertices* means that the vertices are *adjacent*.

M is the symmetric binary sparse matrix associated with the graph. Its general entry (i,j) has value of *1* if the edge (i,j) exists, and the value of *0* otherwise.

Principal components analysis of matrix M

In a first step, principal components analysis is applied to data matrix **M**. Such an analysis can be performed using either the covariance matrix or the correlation matrix. The numerical results appear to be similar in both cases, the obtained visualizations being almost identical. Thus the analysis involving the correlation matrix is presented here. Figure 2 shows a visualization of the locations of the *25* vertices in the plane spanned by the first two principal axes.

These axes correspond to two identical eigenvalues ($\lambda_1 = \lambda_2 = 3.98$), explaining together *31.86 %* of the total variance. The vertices adjacent in the original graph have been joined by an edge to highlight the initial structure.

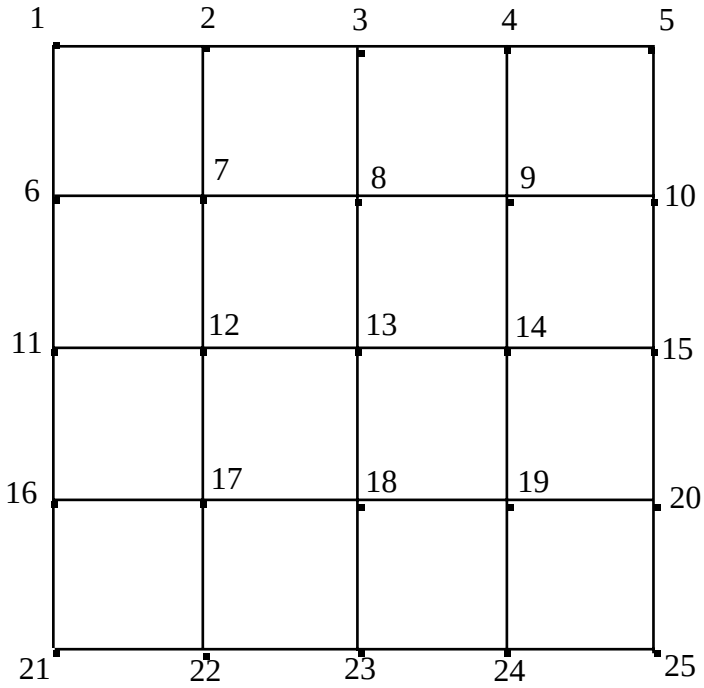

Figure 1. Graph G associated with a "chessboard" (square lattice grid)

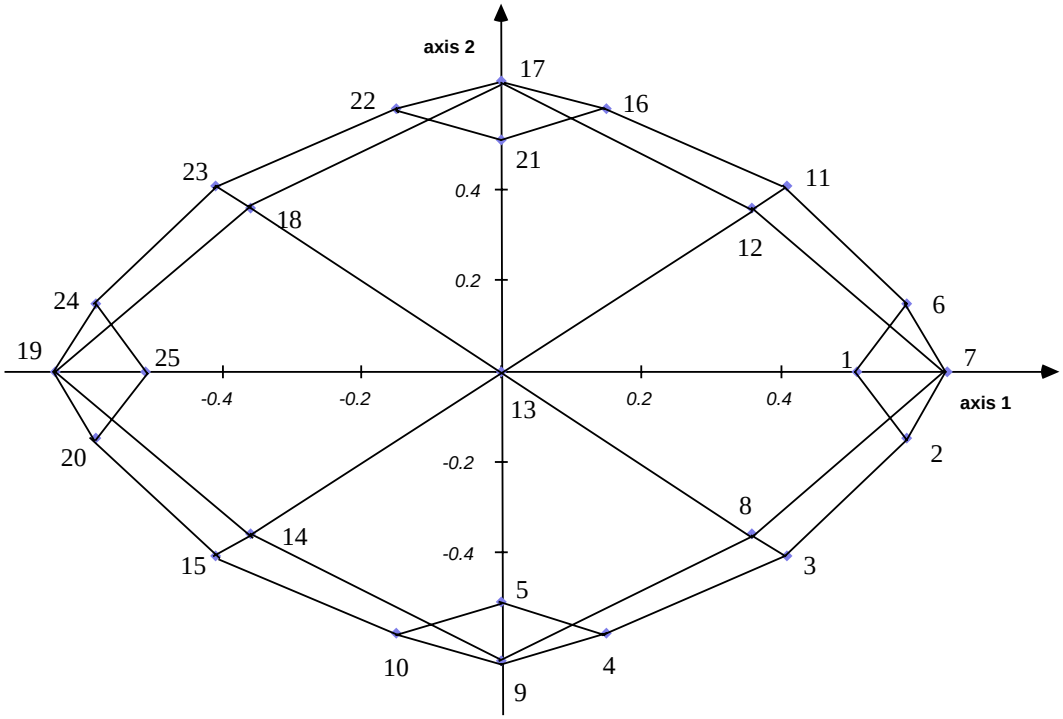

Figure 2. Visualization of graph G through principal components analysis (plane spanned by the first two principal axes)

The symmetry with respect to vertex number 13 is reconstituted. The relative locations of the vertices vis-à-vis their neighbours is generally taken into account by the display, with the exception of the four vertices corresponding to the corners of the rectangle (vertices 1, 5, 21,

25) that are folded back toward the center. The changes in the lengths of some edges are noticeable. They are characterized by a dilation of the four most central cycles of the graph.

Correspondence analysis of matrix M

Correspondence analysis is then applied to the same data matrix **M**. Figure 3 shows a visualization of the locations of the *25* vertices in the plane spanned by the first two principal axes. These axes also correspond to two identical eigenvalues ($\lambda_1 = \lambda_2 = 0.814$), explaining together *32.24 %* of the total variance.

Although the graph in figure 1 is somewhat conventional (it can be drawn in several possible ways), the display in figure 3 satisfactorily reconstitutes both the relative positions of the vertices and an acceptable order of magnitude for the lengths of the various edges. This ability of CA to produce legible maps out of such data matrices can be extended to binary matrices describing various planar graphs.

Note that the calculations involved in the CA of such typical graphs could be carried out directly, without the help of a computer. In the case of a simpler graph (a chain) equation (1) above leads to a simple finite difference equation. A chessboard can then be defined as a « tensorial sum of chains », and the final results analytically derived from those of the chain (see : Benzecri, 1973).

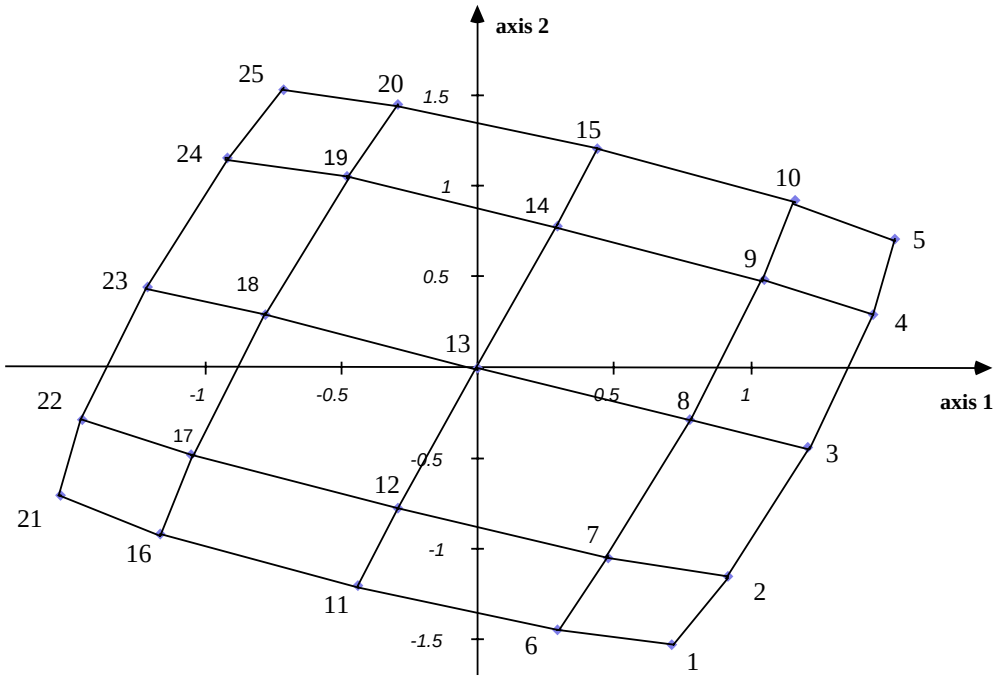

Figure 3. Visualization of graph G through correspondence analysis (plane spanned by the first two principal axes)

We note that the percentage of explained variance (*32.24 %*) is relatively modest if confronted with the quality of the reconstitution of the original structure in the corresponding plane. In CA (and in PCA as well), this phenomenon often occurs when dealing with binary data. In this context, the percentages of variance explained by principal axes always give a pessimistic view of the extracted information. Such empirical evidences favour the use of CA to visualize regular planar graphs known through their associated matrices.

1.4. *Local Principal Component Analysis*

The contiguity ratio can be generalized :

i) to different distances between vertices in the graph,

ii) to multivariate observations (both generalizations are dealt with in: Lebart, 1969).

The graph corresponding to the distance defined as « the shortest path of length k between two vertices » is associated to the matrix $\mathbf{M}^{(k)} - \mathbf{M}^{(k-1)}$, where $\mathbf{M}^{(k)}$ designates the k-th booleean power of the matrix (**I** + **M**) (**I** is the identity matrix, **M** the matrix associated with the graph, with zeros as diagonal elements). Therefore, it is easy to test the significance of spatial autocorrelation, so long as these distances on the graph remain meaningful. This approach provides a variant, in the discrete case, of the variogram used in geostatistics, as presented in the seminal papers of Matheron (1963, 1965).

This section is devoted to the second generalization: the analysis of sets of multivariate observations having an *a priori* graph structure. Such situation occurs frequently in geography, ecology, geology. The multivariate analogue of the local variance is now the local covariance matrix, whose elements *cov(j,j')* are given by (using the previously defined notation):

$$cov^*(y_j, y_{j'}) = (1/n)\sum_{i=1}^{i=n}(y_i - m_i^*)(y_{j'} - m_{j'}^*)^2$$

If **X** designates the (*n*,*p*) data matrix giving the values of the *p* variables for each of the *n* vertices of the graph described by its associated matrix **M**, the local covariance matrix can be written :

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

The diagonalization of the corresponding local correlation matrix (Local Principal Component Analysis) produces a description of the local correlations, which can be compared to the results of a classical PCA performed with the global correlation matrix. Comparisons between covariance or correlation matrices (local and global) is usually done through Procustean Analysis (Tucker, 1958; Shönemann, 1968; Gower, 1984; Lafosse, 1985).

If the graph is made of k disjoined complete subgraphs, **V*** coincide with the classical "within covariance matrix" used in linear discriminant analysis.

1.5. *Contiguity Analysis*

Let **u** be a vector defining a linear combination *u(i)* of the *p* variables for vertex i:

$$u(i) = \Sigma_j\, u_j\, y_{ij} = \mathbf{u}'\mathbf{y}_i$$

The local variance of the artificial variable u(i) is then :

$$v^*(\mathbf{u}) = \mathbf{u}'\,\mathbf{V}^*\,\mathbf{u}$$

The Geary coefficient of this linear combination can be written :

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

where **V** is the classical covariance matrix of vector **y**.

The search for **u** that minimizes c*(**u**) produces functions having the properties of "minimal contiguity": these functions are, in a sense, the linear combinations of variables the more continuously distributed on the graph.

Instead of assigning an observation to a specific class, (as it is done in classical discriminant analysis) these functions allows one to assign it in a specific area of the graph. Therefore, this technique (designated as Contiguity Analysis) can be use to discriminate between overlapping classes, provided that the relationships between

observations are described by a graph. Faraj (1993) has proposed to use it to discriminate simultaneously between several categorical variables, and Chateau (1999) suggests to use it when the classes have an *a priori* structure. Computer programs and examples of contiguity analysis can be found in Lebart and Tabard (1973).

2. Contiguity analysis and nearest neighbours graphs

The preceding results can be applied and enriched in several manners.
It is easy to derive a contiguity matrix from the basic data array itself: any threshold applied to the set of *n(n-1)/2* distances or similarities between observations allows one to define a binary relationship, and, as an immediate consequence, a graph.
Another series of contiguity matrices can be derived, for instance, from the k nearest neighbours of each observation (k varying from 1 to n-1).

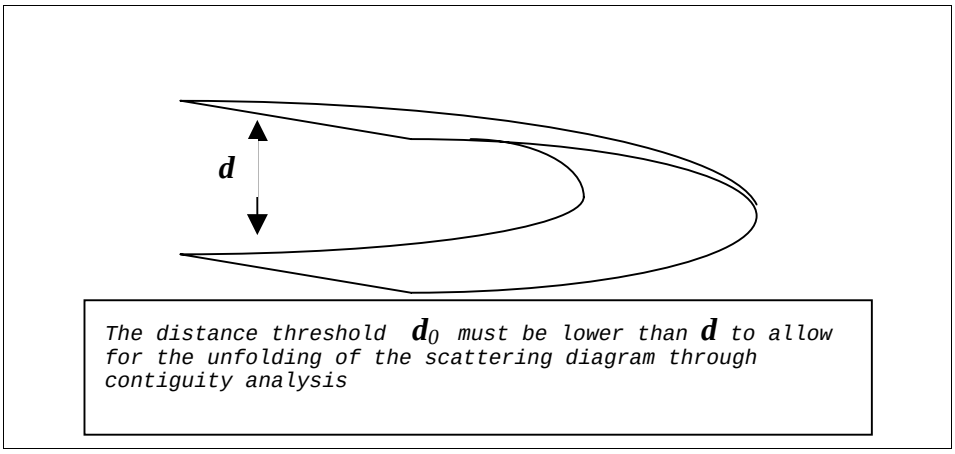

Figure 4 : Unfolding through contiguity analysis

If the scattering diagram of *n* points described by *p* variables is concentrated in the *p*-dimensional space around a folded hypersurface as shown in Figure 4, a graph G can be derived with associated matrix **M** such as $m_{ii'} = 1$ if the observations (vertices of the graph) i et i' are at a distance less than *d*, $m_{ii'} = 0$ otherwise.

Section 1.2 suggests that the correspondence analysis of such matrix **M** will unfold the diagram since there are no edge joining the two main branches of the horseshoe. Contiguity analysis performs a similar unfolding, since the distant observations are ignored in computing the local covariance matrix. It can be considered as a particular projection pursuit algorithm (Burtschy and Lebart (1991).

The term "projection pursuit" was coined by Friedman and Tukey (1974) to name a technique for revealing structure in the original data by offering selected low-dimensional subspaces for inspection. In section 1.5, an algorithm for attempting this goal was presented. This algorithm can be adapted as follows, **X** designating the matrix of standardized observations:

a) Determine the contiguity matrix **M** either by applying a threshold to the set of n (n-1) similarities or distances between observations (n rows of matrix **X**), or by selecting a fixed number of nearest neighbours for each observation.

b) Compute the local covariance matrix **V*** and the global covariance matrix **V** (the matrix **N** and the integer m being defined as above, section 1.2)

c) Find the set of **V**-orthogonal vectors u_α that maximize the criterion c such as:

$$c = u_\alpha' \mathbf{V}^* u_\alpha / u_\alpha' \mathbf{V} u_\alpha$$

This algorithm is based, as the usual projection pursuit algorithms, on the search for the most interesting q-dimensional linear projections of the initial p-dimensional data.

Another technique consists of using the complete graph G* (each pair of vertices are joined by an edge) whose edges are weighted by a decreasing function of the distances between the corresponding vertices. The edge (i,i'), for example, could be given the weight $m_{ii'} = \text{Log}\ (1 + d_{ii'}/d) / \text{Log2}$ (d= max $\{d_{ii'}\}$ for every i and i'). However, it is no more possible, then, to use the sparsity of the matrix **M** to reduce the amount of calculation (Enyukov 1988, Caussinus and Ruiz, 1990)

2.1 Selecting the best contiguity graph

Contiguity Analyses have been performed on the classical Fisher's IRIS data set, using different graphs according to the number of kept nearest neighbours.
The IRIS data set contains 150 individuals corresponding to three species, each species being represented by 50 observations.

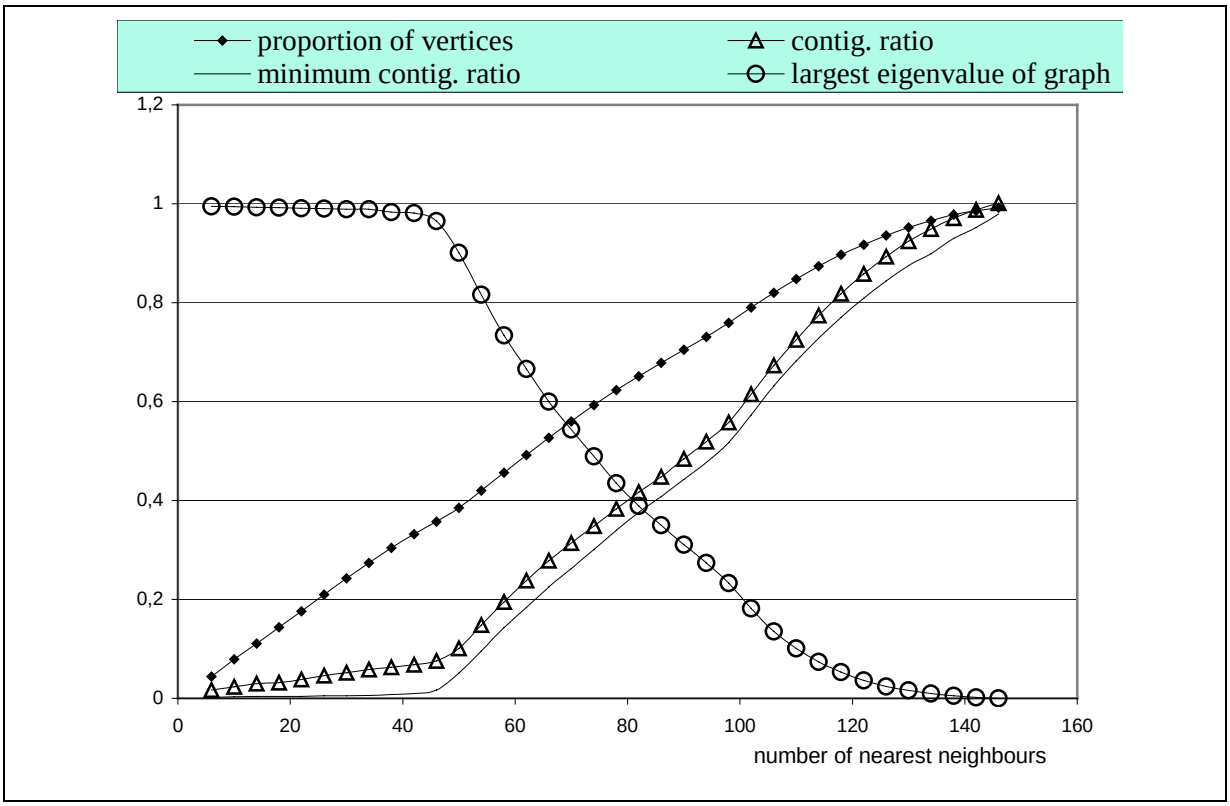

Figure 5 : Contiguity graphs and ratios as functions of the number of N.N.

Comments about Figure 5
The horizontal axis represents the number k of nearest neighbours kept in the contiguity graph (varying from 4 to 149). Four curves are shown.
- The black diamond shaped symbol curve, close to the first diagonal of the rectangular frame, describes the proportion of vertices of the graphs, with reference to a complete graph, having n(n-1) vertices. This proportion is a function approximately linear of the number of nearest neighbours.

- The curve having white triangular symbols, below the preceding one, represents the smallest eigenvalue of c*(u) = u'V*u/u'Vu, that is the smallest contiguity ratio of a linear combination of the 4 original variables. An angle in the trajectory is discernable for about 50 nearest neighbours, as well as around 100 nearest neighbours. The fact that, in the Iris data set, a group of 50 observations is located far apart from the others accounts for the observed angle in that curve.
- The curve without symbol immediately below the previous one is the trajectory of the minimum contiguity ratio Min [c*(u)], as given by the CA of the matrix **M** associated to the k nearest neighbours graph.
- Finally, the only decreasing curve of the display represents the first eigenvalue λ_{max} of the CA of **M** (each point is then produced by the diagonalization of a (150, 150) matrix). This information is equivalent to that given by the curve Min [c(y)] , since Min [c(y)] = $1-\sqrt{\lambda_{max}}$, but the isolation of one group of observations is all the more evident here: the eigenvalue 1 appears in CA when the graph is broken down into non-connex components (see, e.g., Lebart and Mirkin (1993)). Thus, the marked decrease of the curve beyond approximately 50 neighbours pinpoints the existence of an isolated group.

A criterion using the a priori knowledge (labels) about the groups

Figure 6 shows the trajectory of a criterion, designated by W/T (variance within groups divided by total variance).

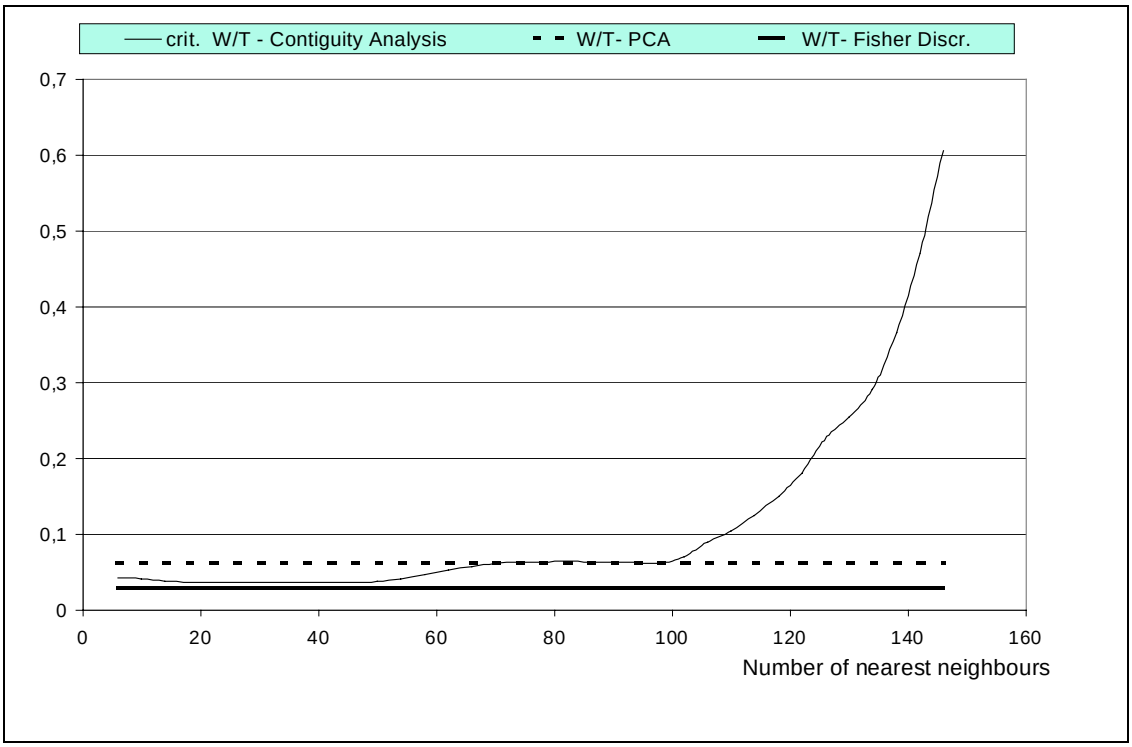

Figure 6 : Criterion W/T (first axis) as a function of the number of N.N.

This criterion takes into account the labels of the observations. It has been computed for the first principal axis derived from each Contiguity Analysis. It is shown here as a function of the number of nearest neighbours kept.
The bold horizontal straight line corresponds to the value of the criterion (0.030) derived from a Fisher Discriminant Analysis on the same data set. Evidently, the first principal axis of a contiguity analysis that ignores the labels of the observations cannot compete with the

first discriminant function that makes use of these labels and that aims precisely at minimizing the criterion W/T.
The dotted horizontal straight line corresponds to the value of the criterion (0.063) derived from a Principal Component Analysis of the data set. It can be seen that the trajectory of the criterion is located below this dotted line for numbers of nearest neighbours comprised within the range [4, ..., 70]. A conservative estimate of this range [4, ..., 40] is given by the values preceding the angles in the curves of the contiguity ratios in Figure 5. The minimum value is 0.0365, closer to the theoretical minimum 0.030 than the value of the criterion given by the PCA.
In fact, this theoretical minimum is abnormally small, being a resubstitution estimate (i.e. an estimate computed on the training sample) that gives an overly optimistic view of the quality of the discrimination. Thus the smaller values of the criterion derived from Contiguity Analysis are all the more acceptable.

Discussion
One must keep in mind that each point of the displays in Figure 5 and 6 corresponds to a low-dimensional display of the data (1, 2 or 3 dimensions). These displays are not exhibited here for lack of space. They constitute however the most interesting output for the user. Those representations that relate to the lowest values of the criterion W/T are similar to the display derived from Fisher discriminant analysis.
Such representations are associated to a number of nearest neighbours selected from the trajectory either of the contiguity ratio or of the minimum contiguity ratio (Figure 5), without using the a priori information about the classes. In terms of learning theory, we are dealing with an unsupervised method.
Why using nearest neighbours graphs instead of graphs derived from distance thresholds? Several experiments show that the latter graphs are often disconnected (except for large values of the threshold, that may not favour the discovery of small clusters). Moreover, external criterions such as W/T are generally less close to the theoretical minimum than those provided by nearest neighbours contiguity graphs (for the Iris data set, the minimum value of W/T is 0.044 when using a distance threshold contiguity matrix).
In conclusion, the non-parametric approach involving a local covariance matrix derived from the series of nearest neighbours graphs allows us to:
- detect potential clusters, after the selection of the appropriate number of nearest neighbours,
- obtain simultaneously a visualization of observations and of potential clusters.
- deal with a classical linear discriminant analysis when the contiguity graph is external (graph associated with an a priori partition).

REFERENCES

Aluja Banet T., Lebart L. (1984) Local and Partial Principal Component Analysis and Correspondence Analysis, *COMPSTAT Proceedings*, 113-118, Physica Verlag, Vienna.

Anselin L. (1995) Local indicators of spatial association - LISA. *Geographical Analysis*, 27, 2, 93-115.

Art D., Gnanadesikan R., Kettenring J.R. (1982) Data Based Metrics for Cluster Analysis, *Utilitas Mathematica*, **21** A, 75-99.

Benali H., Escofier B. (1990) Analyse factorielle lissée et analyse des différences locales. *Revue Statist. Appl.* **38**, 2, p 55-76.

Benzecri J.P. (1973) *Analyse des Données* : *Correspondances*, Dunod, Paris.

Burtschy B., Lebart L. (1991) Contiguity analysis and projection pursuit. In : *Applied Stochastic Models and Data Analysis,* R. Gutierrez and M.J.M. Valderrama, Eds, World Scientific, Singapore, 117-128.

Caussinus H., Ruiz A. (1990) Interesting projections of multidimensional data by means of generalized principal component analysis. In :*COMPSTAT 90*, Physica Verlag, Heidelberg, 121-126.

Cazes P., Moreau J. (1991) Contingency table in which the rows and columns have a graph structure. In : E.Diday, Y.Lechevallier (Eds) *Symbolic-Numeric Data Analysis and Learning*, Nova Science Publishers. New York, p 271-280.

Chateau, F. (1999) Structured Discriminant Analysis. *Communication in Statistics, Theory-Meth.*, 255-256.

Cliff A.D. and 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.

Enyukov I. S. (1988) Detecting structure by mean of projection pursuit. *COMPSTAT Proceedings*, Physica Verlag, Heidelberg, 47-58.

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, p 55-62.

Faraj A. (1993) Analyse de contiguïté : une analyse discriminante généralisée à plusieurs variables qualitatives. *Revue Statist. Appl.* 41, (3), p 73-84.

Friedman J.H., Tukey J.W. (1974) A Projection Pursuit Algorithm for Exploratory Data Analysis, *IEEE Transactions on Computers*, Ser. C, 23, 881-889.

Friedman J. H. (1987) Exploratory Projection Pursuit, *Journal of the Amer. Stat. Assoc.*, 82, 249-266.

Geary R.C. (1954) The Contiguity Ratio and Statistical Mapping, *The Incorporated Statistician*, 5, 115-145.

Gnanadesikan R., Kettenring J.R., Landwehr J.M. (1982) Projection Plots for Displaying Clusters, *in Statistics and Probability, Essays in Honor of C.R. Rao,* G. Kallianpur, P.R. Krishnaiah, J.K.Ghosh, eds, North-Holland.

Gower J. C. (1984) Procrustes analysis. In : *Handbook of Applicable Mathematics*, **6**, Lloyd E.H. (ed.), J. Wiley, Chichester, p 397-405.

Lafosse R. (1985) *Analyses Procustéennes de deux Tableaux*. Thèse, Univ. Paul Sabatier, Toulouse.

Le Foll Y. (1982) Pondération des distances en analyse factorielle. *Statist. et Anal. des Données*, **7**, p 13-31.

Lebart L. (1969) Analyse Statistique de la Contiguité, *Publications de l'ISUP,* XVIII, 81-112.

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, Amsterdam, 341-357.

Lebart L., Tabard N. (1973) *Recherches sur la description automatique des données socio-économiques.* Rapport Credoc / Cordes, CR n° 13.

Lebart L. (1984) Correspondence analysis of graph structure, *Bulletin technique du CESIA*,2, 5-19.

Lebart L., Morineau A., Warwick K. (1984) *Multivariate Descriptive Statistical Analysis,* J. Wiley, New-York.

Lebart L., Salem A., Berry L.(1998) *Exploring Textual Data*, Kluwer, Dordrecht.

Lebart, L. (2000): Contiguity Analysis and Classification, In: W. Gaul, O. Opitz and M. Schader (Eds):*Data Analysis*. Springer,Berlin, 233--244.

Matheron G. (1963) Principles of geostatistics. *Economic Geology*, 58, 1246-1266.

Matheron G. (1965) *Les variables régionalisées et leur estimation.* Masson, Paris.

Meot A., Chessel D., 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, USTL, Montpellier.

Ripley B. D. (1981) *Spatial Statistics.* J. Wiley, New York.

Schönemann P. H. (1968) On two-sided orthogonal procrustes problems. *Psychometrika,* **33**, p 19-33.

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.