Belgisch Tijdschrift voor Statistiek, Informatica en Operationeel Onderzoek, 4, 1980
Revue Belge de Statistique, d'Informatique et de Recherche Opérationnelle, 4, 1980

CONSTRUCTION AND THE USE OF MATRICES OF PROXIMITY IN COMMERCIAL CENTRES

THIERRY TOCK

Summary

At present, planning is being considered from various points of view, including the utilisation of space in shopping precincts planned by property developers. Utilisation of space must be based on a sound knowledge of the relationships between the amount of space required for certain types of shops or businesses. This study provides a comparison of the advantages of each of two methods which demonstrate these relationships when analysing the sites chosen for shops or businesses in precincts, selected at random for purpose of this analysis. Apart from many other advantages, the first method, known as INDSCAL, gives a clear and concise picture of the situation and the relationships between the various categories of shops, whilst the second provides a mathematical analysis of these relationships. Satisfactory results can only be obtained by analysing the results of both methods together.

Introduction

At present, planning is turning towards various new fields. Structural planning of shopping centres, which was once left to the initiative of the individual, now tends to be assumed more and more by private and public developers. If their activities are to be completely successfull, the natural laws which govern the planning and use of space in commercial enterprise must be respected as far as possible. We are not as yet fully aware of all the implications of these laws, so that more detailed research into how they work is needed.

Indeed, the positions of these shops relative to one another (see Getis & Getis, [5], p. 330) cannot presently be determined by anything more than reasoning based on intuition, knowledge of grouping characteristics of some retail establishments, differentiation of shoppers, goods from convenience goods stores, the belief that some sort of economics of agglomeration exists between certain kinds of firms or by some notions derived from empirical evidence of consumer movement. The weight of these factors are varying from town to town and sometimes from person to person. The principal steps of the shop location only stay the same. So, when a salesman is looking for a new location, the data which he can get about his neighbours are of two kinds : firstly, the nature of their activity and secondly, the standing of their shop. With respect to the way in which he thinks he will be able to increase his benefit, he will disgress or come nearer to some categories of shop. In this way, some typical associations of our towns are formed.

It will be important for the planners to identify these specific affinities or repulsions to be able to produce a model for a spatial structure. The goal of this paper is to develop a methodology more efficient than the simple nearest neighbour analysis (see Pinder & Witherick, [9] , pp. 227-288), giving a synthetic view of the associations of a town and permitting to distinguish the significant associations of those which are not.

Construction of matrices of proximity

The first step of the research will be to construct a data set concerning all the shops of a town. Information related to shop i location can be structured in the following manner :

Given x_i, y_i, k_i, l_i, D

where - x_i, y_i are the coordinates of the location of shop i

- k_i is the encoded number of the district of the town in which the shop i is situated. It is a key of the partitioning of the file based on areal domains. In fact, it is supposed that inside a district, the spatial and commercial structure stays homogeneous.
- l_i is the value of an attribute relating to the nature of the activity or to the standing. Although it is possible to carry out an analysis on the basis of several of these factors at the same time, we will confine ourselves here to developing our theories using one aspect only.
- D is the maximum distance for which two shops may be considered in close proxomity. We have chosen 50 meters. This definition comes closest to the actual situation since experience shows that even if a businessman wants to choose a site as close as possible to another shop, the land market rarely provides the opportunity of setting up

a business actually next door to the shop in question. The businessman will then have to try and find a site sufficiently close as to benefit from the advantages which he had hoped to gain by close proximity. This distance may vary depending on the type of commercial activity, the person concerned, the way in which the area between the two businesses has been used,... (see Thompson, [11], pp. 1-6).

Although land surveying may provide useful information on degree of corrective coefficients which will have to be applied to a standard and arbitrary distance, it is often preferable simply to use the standard distance. In practice, apart from the simplicity of the method used to determine the standard distance, test calculations worked out for several distances have shown that there is a linear relationship between the final matrices obtained. The task of measuring this distance on land is also a very delicate one, often subject to numerous errors. For example if we want to plot all the business units in a cadastral and measure the space between them by using their coordinates, our calculations may be based on a theory of substituting a perfect Euclidean area for the real urban one.

This theory is difficult to put into practice since buildings may obstruct direct access to any predetermined destinations. Even if a distance of only 30 meters, as the crow flies, separate two shops, customers may still have to walk twice this distance either through having to go round a row of houses, or along an unavoidable pedestrian route or any obstacles which naturally tend to make the journey longer. Obviously, not all these obstacles are immediately apparent on the map and the distances between shops or businesses will have to be measured on the spot if any degree of accuracy is to be obtained. Although objectivity would seem to be the major factor in any calculations made by this method, one consideration is however left out : the shopper will find the actual distance long or short depending on how much shopping he or she is carrying, the degree of attraction which window displays hold for him, how crowded the route is, ... (see Meyer, [8], pp. 355-361). Since measurements made in situ are not always too efficient as has been shown, these distances are usually calculated by using maps. Furthermore, this system has been more competitive with the development of data processing (see Baxter, [1], pp. 176-179). Instead of determining new directions of axes - along an perpendicular to the main streets considered - for each shop, it has been preferred only to consider the shops of a same homogeneous urban section in which there exists no spatial interruption.

Algorithm

Known $N = i = 1, 2, 3, \ldots, n$ where n equals the total number of shops in a given town

N_k the total number of shops in the district k

N_{kl} the total number of shops in district k and in category l

K the total number of districts in the same given town

L the total number of categories into which the various features have been divided

V the number of neighbouring shops within which close proximity to another unit has been sought

$T_{lm}^{(k)}$ is the number of shops of type m in a district k which lie within distance D of a shop of type l in the same district

$U_{lm}^{(p)}$ is the number of shops of type l for which there are p shops of type m among the V nearest shops this index is calculated globally for the town.

- Construction of the first index

1) Initialization

$$T_{lm}^{(k)} = 0 \qquad \text{for } l = 1, 2, 3, \ldots, L \quad m = 1, 2, 3, \ldots, L \quad k = 1, 2, 3, \ldots, K$$

2) $\forall\, i \in N$, read x_i, y_i, k_i, l_i

3) $\forall\, i \in N$, $\forall\, j \in N$, $\forall\, i \neq j$, $\forall\, k_i = k_j$

calculation of $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$

4) Choice of a reference distance D within which two shops may be considered to be in close proximity (for ex. 50 meters).

5) $\forall\, d_{ij} \leqslant D$, then $T_{l_i l_j}^{(k_i)} = T_{l_i l_j}^{(k_i)} + 1$

6) $$T^{*\,(k)}_{lm} = \frac{T_{lm}^{(k)}}{N_{kl}} \qquad \text{if } N_{kl} \geqslant N_{km}$$

$$T^{*\,(k)}_{lm} = \frac{T_{lm}^{(k)}}{N_{km}} \qquad \text{if } N_{km} < N_{kl}$$

$\left[T^{*(k)}_{lm}\right]$ form a tridimensionnal matrix constituted of K tables and L lines and L colums

- Construction of the second index

1) Initialization

$$U^{(v)}_{lm} = 0 \qquad \text{for } l = 1, 2, 3, \ldots, L \quad m = 1, 2, 3, \ldots, L \quad v = 1, 2, 3, \ldots, V$$

2) and 3) see above

4) $\forall\, i \in N, \forall\, i \neq j,$

$\forall\, k_i = k_j$

. $d_{ir_1} = \min_{j \in N} d_{ij} \rightarrow$ calculation of associated l_{r_1}

. for $v = 2, 3, 4, \ldots, V$

$$d_{ir_v} = \min_{\substack{j \in N \\ j \neq r_1, r_2, r_3, \ldots, r_{v-1}}} d_{ij} \rightarrow \text{calculation of associated } l_{r_v}$$

. for $v = 1, 2, 3, \ldots, V$

$$s_{l_{r_v}} = s_{l_{r_v}} + 1$$

. Among the V nearest shops, selection and incrementation of the categories which are not represented

$$\forall\, m = l_{r_v} \qquad U^{(0)}_{l_i m} = U^{(0)}_{l_i m} + 1$$

. for $l_{r_v} = l_{r_{v'}}$ where $v = 1, 2, 3, \ldots, V$, $v' = 1, 2, 3, \ldots, v-1$

if $p = s_{l_{r_v}}$ where $s_{l_{r_v}}$ is an integer $[1,V]$

$$U^{(p)}_{l_i l_{r_v}} = U^{(p)}_{l_i l_{r_v}} + 1$$

$\left[U^{(v)}_{lm}\right]$ form a tridimensionnal matrix constituted of V tables at L lines and L colums

5) End

The formula set above clearly shows that there are two ways of looking at the question of proximity of shops.

Once the first part of the algorithm has been completed as demonstrated above, we are left with a tridimensionnal matrix comprising as many tables L x L as there are districts in the town. The elements in the diagonal comprise the indices of proximity within a given category and when all the factors remain constant, their values are twice the value of other indices since there are equal to the contraction of 2 symmetrical elements of a table.

For a given category, notwithstanding this rule, the index of the diagonal is not always the highest one, and this will give rise to major problems when the statistics are analysed later.

Once the second stage of the algorithm has been completed, another 3-dimensionnal matrix is obtained, made up of V + 1 square tables with L lines and L columns, where V is the number of neighbours selected for the analysis. These tables show the number of times a category x was associated 0, 1, 2, ..., V times to a category y within the period during which V neighbours of n businesses in a town were observed.
This technique does not fully resolve the problems pertaining to the distance. In practice, because of the buildings and pedestrian precincts which the shopper must pass, the i^{th} neighbour may sometimes be farther away then the $i + 3$ or $i + 4^{th}$ neighbour. Here, the issue is not arbitrarily determining a distance within which proximity can be said to exist. However, it is still necessary to determine which businesses specifically chose a certain site in view of proximity and to what extent ; in any event, V will depend on the density of shops and businesses in a given area as well as the subjective apprehension of it and the layout of cadastral plots. The ideal solution would be to determine V for each homogeneous section. However, this method, which is an attractive proposition from the point of view of the quality of the results, does not lend itself to the use of statistics and will therefore not be applied.

Statistical use of matrices

The matrices obtained from the first part of the algorithm are symmetrical. However, the absolute value of the elements obviously depends on the symmetrical value of the two categories which correspond to this element. In order to work out a table where the elements are less dependent on the number of shops per category, each element is divided by the highest number of one of the two categories affecting each element. The symmetry of the table is thus also obtained.

This method implies that the affinities will show a certain degree of overlapping in the categories. This overlapping, however, is sometimes deceptive : whilst certain categories of shops do not go well together, other types of shops have such a high degree of dependance that there is no deed to take the proximity factor into consideration. Exceptions to the rule arise therefore if the number of shops varies greatly from one category to another. It is necessary to work on the assumption that any such distorsions are minimal if the table is to be kept symmetrical.

The table may be analysed by using various statistical techniques. However, we consider that the best method is to look at the question from the point of view of proximity. The aim is to examine the information contained in one or several matrices - when exceptionally an analysis is being made of proximity factors in several centres at the same time - whose elements show by district the extent of the relationships linking one category to another. These indices not only provide a system of affinities but also permit these affinities to be measured and thus constitute a viable formula for our analysis. Among the possibilities offered by these indices, i.e. the analysis of either agregate or individual statistics, the latter aspect is particulary advantageous since it provides the opportunity of analysing the proximity factors of several centre or several sections at the same time in cases where centres or sections are too heterogeneous to have given rise to the same type of proximity factors over the total area accidentally.

Several aspects are important here :

- Firstly, there is always the maximum degree of proximity between one shop and other shops from the same category. This constraint is merely the result of the original aim of the INDSCAL formula which is to analyse the information contained in a table of similarities between units where, by definition, each unit is most like itself. Since we have shown that this condition is not always fulfilled in the proximity matrices of commercial units, the elements on the diagonal will be systematically ignored. It will not therefore be possible to obtain any information on the relative position of shops of the same tyme by the INDSCAL method.

- Secondly, the categories of shops have intrinsic economic or other characteristics which affect the siting of commercial units.
- When arranging the positions of each category, the number of proximity factors to be taken into account may be determined a priori. In general, there will be two or three factors at play so that the results may be presented diagrammatically.
- The principles involved in choosing a site are assumed to be constant, especially with regard to the perimeter of the area studied. For example, if a low degree of commercial activity in a district gives rise to a situation where food shops are not in proximity to clothes shops, then this principle must be assumed to be valid for any urban district where there is a low concentration of commercial activity.
- Finally, only the various properties of the categories themselves affect the actual siting of these categories. The features of the districts alone are not determining factors and may only be taken into account in an analysis if they show a general classification of categories for all centres.

The principle of the INDSCAL method due to Caroll and Chang, [4], pp. 283-319, is as follow (see Bouroche, [3], p. 19 and Marion, [7], pp. 15-35) :

From a 3-dimensional table, the method proposes to find out

1) a configuration of n categories in R^p of smaller dimension ; this configuration must be common for each district (see graphic I where p=2) ;

2) a configuration of points situating the K districts in R^p, so the projection of the district i on the axis j pointing out the importance in the district i of the characteristic measured by the axis j in the configuration of the categories (see graphic II).

Given $k = 1, 2, 3, \ldots, K$	all urban districts
$\Delta^{(k)} = (\partial_{lm}^{(k)})$	the table of indices of proximity between L categories of district k
$= T^{*(k)}_{lm}$	
r where $r < n$	a given integer

A matrix of scale products $B^{(k)}$ is calculated by the equations of Torgerson [13], for each urban district on the basis of the matrix of distances

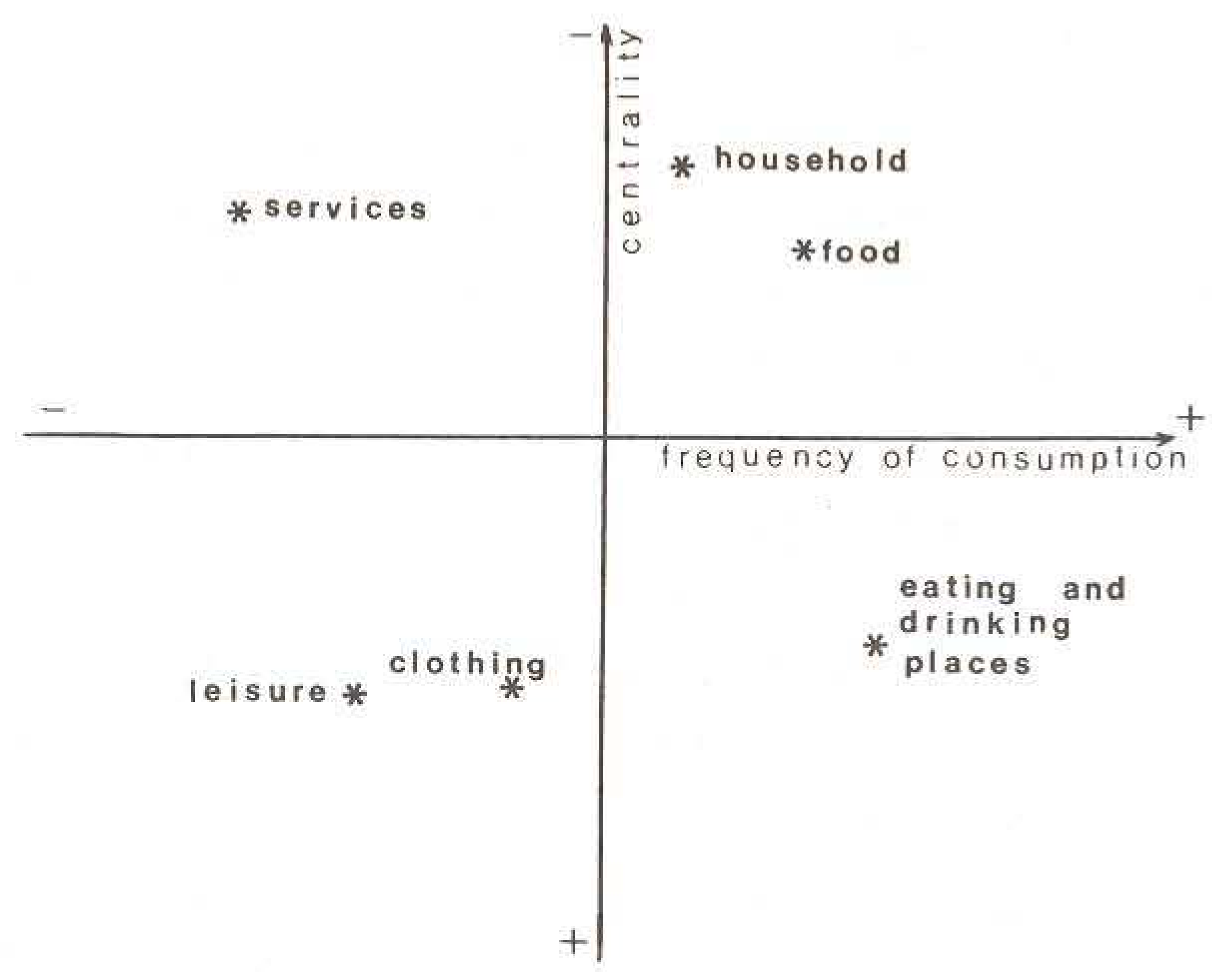

Figure I : Configuration of the categories given by the INDSCAL model

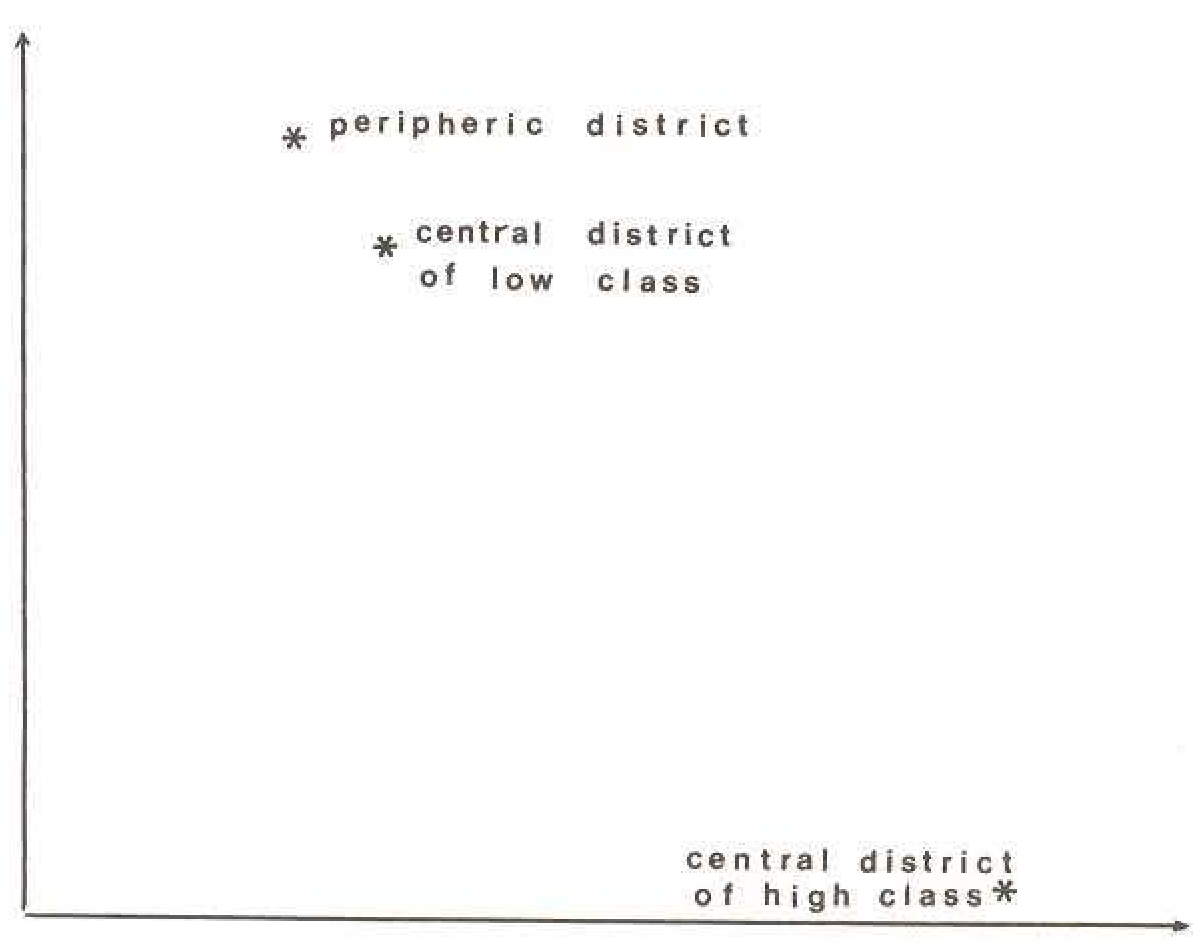

Figure II : Configuration of the districts

$\Delta^{(k)}$.

$$B^{(k)} = F(\Delta^{(k)})$$

$$= \{ b_{lm}^{(k)} \}$$

$$= X'W^{(k)}X$$

where the $B^{(k)}$ are known ; so X and the $W^{(k)}$ (see Bertier and Bouroche, [2], p. 205) have to be estimated by the Niles - Non linear iterative least square) procedure proposed by H. Wold, [14] : for any value of X, it is possible to estimate the $W^{(k)}$ by the method of least square. After, we consider that those $W^{(k)}$ found at the preceding step, are fixed and a new value of X is estimated. This procedure is recommenced supposing at each step whether X or the $W^{(k)}$ are fixed at their precedent estimation. The residual error is decreasing and generally, there is convergence.

$$W^{(k)} = \begin{bmatrix} w_1^{(k)} & & & 0 \\ & \ddots & & \\ & & w_t^{(k)} & \\ & & & \ddots \\ 0 & & & w_r^{(k)} \end{bmatrix}$$

It is assumed that each district "weights" the axes of the represented space in order to bring the distances between points.

The adequation of the formula is verified by the percentage of explained variance.

This method has been used in view to summarize the informations included in a file concerning the shops of Liège (see Tock, [12], pp. 236-238). In the graphic I, the categories are located so that their affinities are the most obvious in a two-dimensionnal space of which the significance may be determined. Indeed, if we take a series of values for more or less bipolar variables from each category (degree of concentration, size number, degree of attraction, etc...), the significance of the axes may be determined on the basis of the position of the categories in the configuration. For example, this may be done by calculating the correlation between the coordinates of the categories on an axis together with the corresponding values of the variable. The criterion of division into classes is often found among the first two or three dimensions, i.e. status of the shops, turnover, nature of activity (this is less evident since it is not bipolar to such a large extent), etc...

In the second configuration, the position of points depends on the degree of importance attached to each dimension. Apart from some exceptional cases, this weighting will always be higher than 0 and thus only the higher quadrant law will contain the whole range of points. This configuration is interesting in that it leads to the discovery of a new classification of districts based on the spatial arrangement of shops within these ones.

The matrices adopted in the second part of the algorithm are no longer symmetrical. However, since the theoretical number of neighbours is unknown, mathematical statistics make it possible to determine the real meaning of associations. The principle is to compare the observed frequencies of neighbouring to the theoretical ones. The number of times where throughout a town out of V neighbours, each shop of the category l has v ($0 \leqslant v \leqslant V$) shops of the category m as neighbours is equal to (see last part of the algoritm) $U_{lm}^{(v)}$.

Because there is more than one neighbour, it is no longer possible to use the binomial law like it was done by Getis & Getis, [5], pp. 218-230 and by Shepherd & Rowley, [10], p. 234 and instead, the hypergeometric law is used for the calculus of the theoretical frequencies. Indeed, if the spatial distribution of shops was random, the probability to find v shops of the category m out of the V nearest neighbourgs of a shop l may be compared to the probability to obtain v bowls marked "m" at the time of a simultaneous random drawing of 20 bowls out of an urn which contains N ones.

If $P_{lm}^{(v)} \quad | \quad o \leqslant v \leqslant V$

$$= \frac{V!\ N_m!\ (N-N_m - 1)\ !\ (N-V-1)\ !}{v!\ (V-v)!\ (N_m-v)!\ (N-N_m-1-V+v)!\ (N-1)!}$$

where N is the total number of shops of the town, N_m the number of shops of the category m.
We can substract 1 from N because we neglect the shop from which we calculate the neighbours.

Then, $${}_{lm}X_V^2 = \frac{\sum_{v=0}^{V} (U_{lm}^{(v)} - N_l \times p_{lm}^{(v)})^2}{(N_l \times P_{lm}^{(v)})}$$

where ((V+1)-1) is the number of degree of freedom.

As a result of this probability, we know whether the position of a shops l in relation to shops m is a chance element or not. A measure of affinity can eventually be deduced.

However, although this method of finding the V^{th} neighbour seems attractive, it does nevertheless have certain disavantages. For example, the $(V-v)^{th}$ neighbour is sometimes separated by such a large distance that it is obvious no attempt has been made to establish the shop as near as possible to another shop. This means that if we want to keep to the test conditions listed above, then there is no question of fixing a limit to distance.

On the other hand, even though this method enables us to single out two types of association, i.e. A's relation to B and B's to A, it is not however possible to deduce in real terms the extent to which the site of the neighbouring shop has been chosen for proximity.. Concerning this problem, an interesting method has been developped by LEE, [6], pp. 172-173 :

If ρ_A, ρ_B are the mean-density factors of the shops of the categories A and B

n_A, n_B are the respective proportions of the two types

$n_A + n_B = 1$.

Then the probability that there are X shops B inside a circle of radius r of which a shop A is the centre is

$$P = n_A \frac{(\rho_B \pi r^2)^X e^{(-\rho_B \pi r^2)}}{X!} + n_B \frac{(\rho_A \pi r^2)^X e^{(-\rho_A \pi r^2)}}{X!}$$

From that, it may be possible to deduce the estimated radius of the circle* , that is to say the theoretical distance from a shop A to the nearest shop B. This distance is compared with the observed distance which is equal to the

---

* If P is the probability that the circle contains 0 shop B is thus the proportion of distance to the nearest B

$$= n_A e^{-\rho_B \pi r^2} + n_B e^{-\rho_A \pi r^2}$$

(1-P) is the proportion of distance from the centre A to the nearest other shop A

Then, $f(r) = \frac{\partial (1-P)}{\partial r}$

$E(r) = \int_0^\infty r f(r)\, dr$

average nearest neighbour distances between a certain point and its nearest neighbour points from the other distribution.
So, this method permits to determine the way of the relation although $r_e^{AB} = r_e^{BA}$, nevertheless $r_a^{AB} \neq r_a^{BA}$ and the greatest of the ratios r_a/r_e will show which category has attracted the other one.

Conclusion

The affinities between categories of commerce may no longer be overlooked at a time when town planning is being developped to such a large extent. In order to study these affinities, the ideal solution is to analyse the results of the three methods together. INDSCAL provides the possibility of analysing proximity factors in a chosen number of dimensions and brings the elements which lie behind the affinities to the fore, since a set of data is available for each category. INDSCAL highlights the frequency of commercial relationships observed in the urban environment. This method is very much influenced by the number of shops in each category. The statistical tests carried out in the second method have the advantage of giving results which are totally independant of the number of shops. Such tests quantify the relationship!

The two first methods of analysing the problem therefore complement each other since each concentrates on factors which are not covered by the other method. Finally, it should be stressed that neither of the two methods lends itself to analysing the layout of shops in too great a number of categories. Other methods such as the cluster analysis would seem more appropriate.

Literature cited

[1] Baxter, R.S. "Computer and Statistical Techniques for Planners". London : Methuen and Co Ltd, 1976.

[2] Bertier, P. & Bouroche, J.M. "Analyse des données multidimensionnelles". Paris : Presses Universitaires de France, 1975.

[3] Bouroche, J.M. "Analyse des données en Marketing". Paris : Masson, 1977.

[4] Caroll, J.D. & Chang, J.J. "Analysis of Individual Differences in Multidimensionnal Scaling via N-way Generalization of Eckart-Young Decomposition". Psychometrica, 35, 1970, 283-319.

[5] Getis, A. & Getis, J.M. "Retail Store Spatial Affinities". Urban Studies, 5, 1968, 317-322.

[6] Lee, Y. "A nearest-neighbor spatial-association measure for the analysis of firm interdependance". Environment and Planning, 11, 1979, 169-176.

[7] Marion, C. "Analyse multidimensionnelle des proximités. Modèle INDSCAL". Unpublished M.A. Thesis, Université Catholique de Mons, 1977.

[8] Meyer, G. "Distance perception of consumers in shopping streets". Tijdschrift voor Economische en Sociale Geografie, 67, 1977, 355-361.

[9] Pinder, D.A. & Witherick, M.E. "The principles, practice and pitfalls of nearest neighbour analysis". Geography, 57, 1972, 227-288.

[10] Shepherd, McL.P. & Rowley, G. "The association of retail functions within the city centre". Tijdschrift voor Economische en Sociale Geografie, 69, 1978, 233-237.

[11] Thompson, D.L. "New concept : Subjective distance". Journal of Retailing, 39, 1963, 1-6.

[12] Tock, Th. "Affinités et structures spatiales des commerces". Bulletin de la Société Belge d'Etudes Géographiques, 2, 1979, 234-254.

[13] Torgerson, W.S. "Theory and methods of scaling". Wiley, 1958.

[14] Wold, H. "Estimation of principal components and related models by iterative least squares". Krishnaiah PR, Multivariate Analysis, New York Academic Press, 1966, 391-420.