Point pattern analysis: an application to the loyalty networks of chain stores

Giovanni A. RABINO* and Lorenzo MASTRANGELO†

Abstract

At the 41st congress of E.R.S.A. Staufer-Steinnocher (2001) proved as a point pattern analysis technique, the kernel density estimation approach, can be usefully applied to geomarketing. Accordingly, the aim of this paper is to show how this tool, like other point pattern analysis techniques (centrographic statistics, nearest neighbour index, Ripley's K, etc.) can be exploited in marketing and micromarketing researches, allowing to explicit the "geographical knowledge" embedded in available information. Here we investigate this approach, using data on locations of the outlets of a big italian retail group and its partners.

1 The problem

Modern retail market is characterized by an high degree of uncertainty and complexity.
In this context the patrimony of immaterial resources, consumer's loyalty *in primis*, becomes the most important target for companies and their marketing or, more correctly, micromarketing actions.
The main activity in which the micromarketing gets down to facts is the realization and management of the loyalty program: through which these the companies aim to take advantage of the behavioural heterogeneity of the customers.
A number of innovations were gradually introduced in the loyalty programs, respect to the original outline, in which only a reduction in price according to the expense carried out was considered. In particular more complex mechanics, loyalty cards, prizes brochure, a greater coherence of the program with the company position, change in the duration of the promotional sales and development of a partners network were introduced.
The latter is surely the most interesting among these innovations. It consists of an integration or coordination of the loyalty programs of the chain store with one or more loyalty programs which other companies had already developed as a stand alone

* Architecture and Planning Department, Polytechnic of Milan, Piazza Leonardo da Vinci 32, 20133 Milan (Italy). Phone: (+39) 02.2399.4102, Fax: (+39) 02.2399.4105.
Internet: http://digilander.iol.it/rabinogiovanni, E-mail: giovanni.rabino@polimi.it
† E-mail: lore.mast@tin.it

initiatives. The choice of the partners is obviously not accidental, but it is performed among all the companies that share an identical target of consumers.

The aim of this paper is to investigate the features of these initiatives. The organization of these networks could be indeed the symptom of important strategies of diversification of the activity of the great distribution with respect to the core business (Mauri, 2001).

This hypothesis has been suggested from the continuous presence of one particular kind of partner, i.e. those of the fuel distribution companies in this kind of initiatives, all over the Europe and USA. These enterprises offer a type of products analogous to the chain stores: their sale points have one land spread comparable to that one of supermarkets and hypermarkets and the frequency of visits is comparable approximately weekly for both.

Although in a qualitative way, the geographic variable suggests a closest relation between these partners, compared to others composing the network, which are normally represented by companies of services.

Accordingly to two specific italian real cases we ask ourselves whether geomarketing and actually point pattern analysis techniques, may help us to clarify this matter. We also aim to investigate whether point pattern synthetic indexes might give us a more precise indication of the effectiveness and real aims of these promotional initiatives.

2 Point Pattern Analysis

Point pattern analysis deals with the simplest type of spatial data, a point pattern. By this we mean a data set consisting of a series of point locations in a study region at witch events of interest have occurred. A spatial point pattern is simple because the data comprise only the coordinates of events at least at the most basic level. This doesn’t mean that analysis is easier than for the other types of spatial data; from a statistical perspective point patterns can be mathematically more complex to handle (Bailey and Gatrell, 1995).

Generally point pattern analysis techniques allow to understand when and where, events of a distribution are organized in the space either randomly or regularly or in cluster of different size corresponding to certain locations.

From a statistical point of view this occurs analyzing the so-called first and second order properties of the distribution. Some techniques, as the kernel density estimation, allow to analyze the first order properties, that characterize a considering the intensity, or density, of the events in the region of study R. Briefly the determination of the medium number of events for each zone A, in which R can be arbitrarily decomposed, allows to estimate the global trend or wider scale of the distribution.

On the other hand, techniques such as the K-function allow to estimate the second order properties of the distribution, i.e. relations and interactions of the events on a local scale. As any distribution is characterized from a mix of these properties, the two kind of techniques must be used therefore in complementary way.

3 Esselunga-TotalFina network: a geomarketing analysis

Here is shown the analysis of the partners network constituted by Esselunga chain stores and the TotalFina fuel stations placed in the geographic area under investigation. A second analysis, carried out on the network of an other important italian retail group (Coop), gave similar results, thus it is here omitted for brevity.

The data employed are the following:

A. Esselunga chain stores in the area under investigation (78 sale points)
B. TotalFina fuel stations that belong to the network in the area under investigation (161 sale points)
C. TotalFina fuel stations globally present in the area under investigation (280 sale points)

The visualization of these three distributions on the map of the area represents the first obvious but precious passage of the analysis. The observation of both total and local course of the events and the mutual position of the same events, lead us to formulate the first hypothesis of the geographical characteristics of network.

The first impression reflects the substantial likeness in the course of the distributions as for the previous points A) and B). TotalFina fuel stations that belong to the promotional network (57% of the fuel stations in the area) are usually the closest to the Esselunga chain stores. This seems to be the most conceivable result. Spatial closeness among chain stores and fuel station makes easier the transit of the same consumer from a point to the other and therefore the enjoyment of the prizes for his fidelity to the network.

Next paragraphs will show main results obtained by some point pattern analysis techniques implemented by CrimeStat freeware software and used in order to confirm, or reject, our first observations.

4 Centrographic statistics

The most basic type of descriptors for a spatial distribution of events is œntrographic statistics. These indexes represent basic parameters of the distribution, and therefore its first order properties. They are called centrographic in that they are two dimensional correlates to the basic statistical moments of a single variable distribution: mean, standard deviation, skewness and kurtosis (Bachi, 1957; Levine, 1999).

By applying the concept of mean center, standard deviation of the coordinates, standard distance deviation and standard deviational ellipse to the three distributions under analysis, we obtain easily interpretable outcomes, well coherent with our expectations (Fig.1).

Central indexes as well as dispersion measures around to the mean value show clearly how distribution B) (TotalFina fuel stations joining the promotion) tends to events of distribution A) (Esselunga chain stores). This trend is quite different compared distribution C) (TotalFina fuel stations globally considered).

Fig. 1 – Standard distances

5 Distance analysis

Centrographic statistics provide useful indications on the general behaviour of the distributions in the studied area. However anything is known about the local behaviour of these distributions. To this aim we can use some of the spatial distance based analysis techniques like K-Order Nearest Neighbours Index and K-function.

In figure 2 the K-Order Nearest Neighbours Index for “A”, “B” and “C” distributions are compared. The graph shows the nearest neighbour indices up to the 50^{th} order. Edge effects corrections have not been considered.

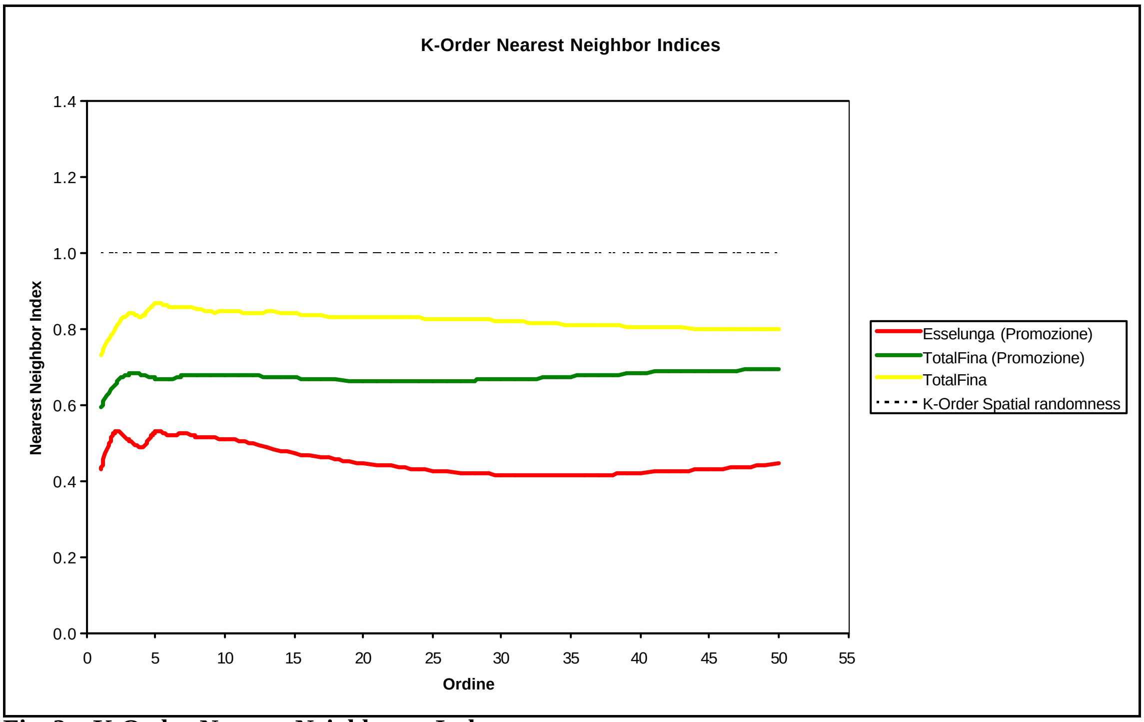

Fig. 2 – K-Order Nearest Neighbours Index

The information we obtain from this graph are coherent with those of the previous paragraph but also slightly deeper.

The nearest neighbour index is scaled from 0 (extreme clustering) up to 1 (extreme dispersion), equal to the expected value under complete spatial randomness condition. All the distributions are much more concentrated than K-order spatial randomness, which is the scale (order) considered.

Distribution “A” (Esselunga chain stores) are more concentrated than distribution “B” (TotalFina fuel station joining the network). Moreover the most interesting point for our analysis is that events of the distribution “B” are more concentrated than those belonging to the distribution “C” (TotalFina fuel stations globally considered).

We can't demonstrate at the moment whether the different behaviour of "B" events comes from the presence of the "A" distribution events, that is whether events keep to concentrate around "A" events. However previous results seem to confirm this expectation. Similar results come from the K-function based analysis.

L(d) statistics values for "B" and "C" distributions are reported in the graphs of figure 3 and 4, with the envelopes that delimit the region of significance obtained through 1000 simulations in conditions of complete spatial randomness.

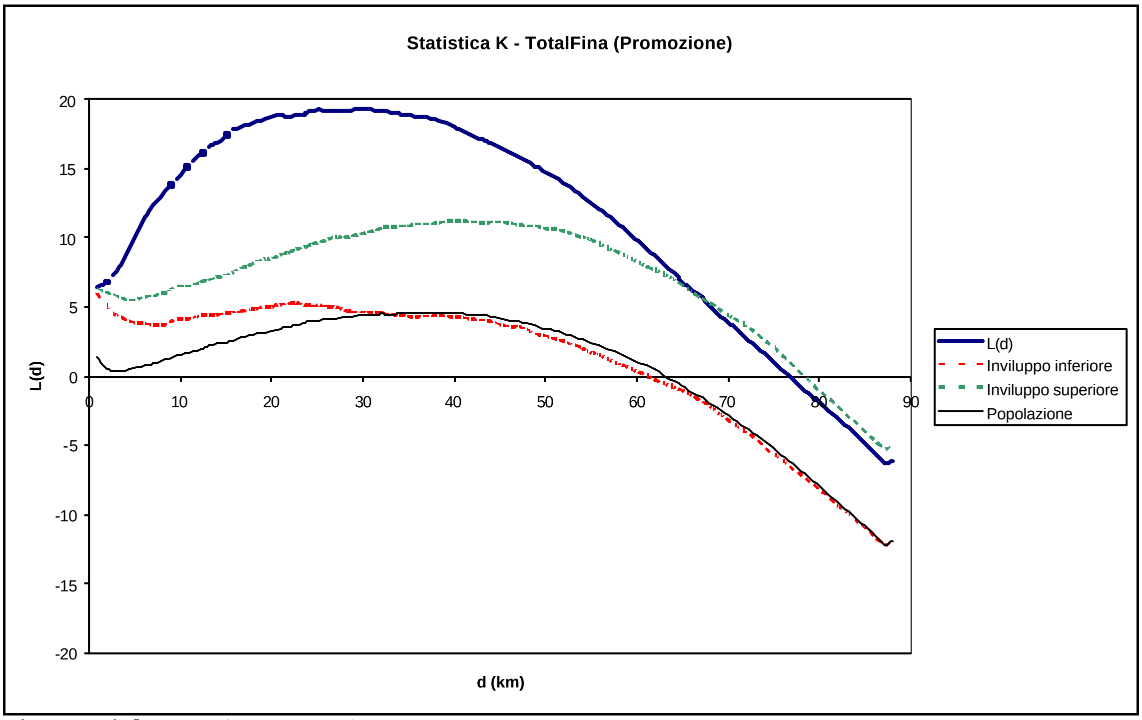

Fig. 3 – Ripley's K ("B" events)

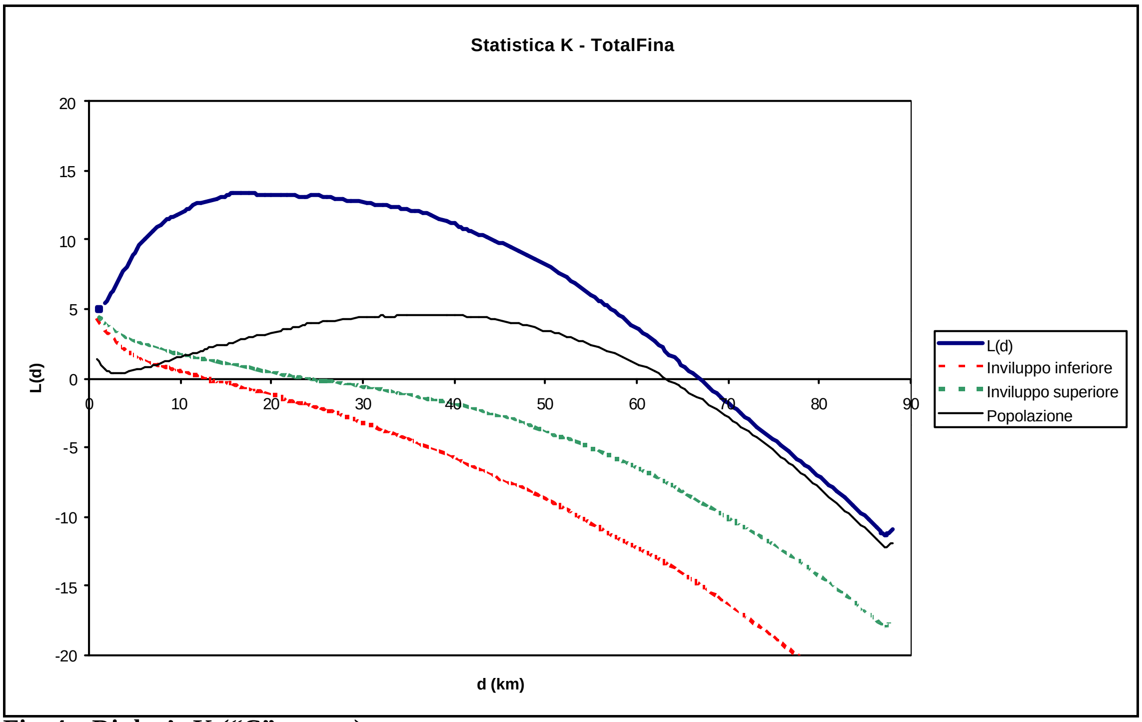

Fig. 4 – Ripley's K ("C" events)

Figure 3 shows a graph of L against distance for "B" events and compares this to both the envelope produced from 1000 random simulations as well as the L distribution from the 2001 population; this latter variable was obtained by taking the centroid of Lombardia's boroughs and using population as the intensity variable. As can be seen, the amount of clustering is much greater than both the random envelope as well as the distribution of population. In other words "B" events are more clustered together than what would be expected basing on population distribution and this holds for distance up to 68 kilometres. Figure 4 shows the distribution of "C" events, compared both to a random envelope and the distribution of population. We find that "C" events are more clustered than population distribution, but less than "B" events for identical distances. Thus, the distribution of L confirms the result that "B" events tend to spread over a much larger geographical area in smaller clusters than "C" events, which tend to be more concentrated in large clusters.

Now we have to explore the features of these clusters and the mutual interactions between the distributions using the kernel density estimation approach.

6 Kernel Density Estimation

Kernel density estimation is an interpolation technique suitable for individual point locations. Interpolation is a technique for generalizing occurrences in an entire area.

Applying this approach to our distributions we obtain density maps that easily demonstrate our previous issues: zones characterized by an high density of “B” events (TotalFina fuel stations joining the network) parallel those in which “A” events have the highest density. On the other hand “C” distribution areas with not null density which are not present in density map of “B” distribution appear to overlap with zones in which supermarket are not located (Fig. 5).

Therefore the previous paragraph hypothesis was right: the presence of fuel stations joining the promotional network strongly depends on presence or absence of the chain stores in the same zones.

Fig. 5 – Kernel Density Estimation

In the next paragraph our distributions will be finally analyzed with an interaction analysis perspective. Although next results will not deal with point pattern analysis, they might be to definitively and easily confirm the previously reported outcomes.

7 Spatial interaction analysis

Gravitation areas or attraction's poles are aggregations of elementary territorial units, as boroughs, that show systematic commuting phenomena from an origin zone - usually the residential area, to a destination area which exerts attraction force on the first one.

So gravitation areas outline social groups travelling daily for job, entertainment, and so on. The zones that attract these groups are cities, districts, extra city areas where goods and services are particularly concentrated and characterized by high qualitative standards, more convenient prices, and so on.

The so called SOMEA (an italian research group) gravitational areas seem to be suitable for our specific matter referring to the retail framework (Fig. 6).

First of all we consider the distribution of the Esselunga chain stores. If gravitation areas hypothesis is correct we expect to find the greater part of the outlet in those boroughs which exert an attraction force on the other boroughs belonging to the same gravitational area. By this way chain stores would be potentially able to intercept consumers of its own city and consumers of the zones which compose the gravitation area.

By GIS software we can carry out a simple spatial selection query of the chain stores located in these boroughs. The number of outlets fulfilling this condition is equal to 51 and represents 91% of the outlets of the study region.

Fig. 6 – Gravitation areas for retail market

Now we consider the distribution of the fuel station joining the promotional network. Carrying out another spatial selection query by GIS software it is easy to see how the great part of the sale points (71.4%) are in the same areas of the chain stores. The percentage is equal to 91. 3% considering immediately adjacent zones.

So, consumers moving towards central areas localities for shopping can take advantage by the partner network collecting promotional scores not only in the chain stores but also in the next fuel stations and in the fuel stations they find in their trips.

According to previous paragraph results we can therefore assert that partners promotional networks enable retail chain stores to strengthen and amplify their own influence and indirectly their presence on the territory.

8 Summary and future research

The aim of this paper was characterize the nature of fidelity network of one important italian retail group on a geographical perspective. To address this issue we have employed some point pattern analysis techniques. Using affordable algorithms and simple output these techniques have allowed a deeper understanding of this matter.

The application of such techniques to our particular distributions, i.e. very frequent in marketing and micromarketing problems shows their great potentialities in such promising discipline as geomarketing.

Finally we believe that point pattern analysis tools should be surely further investigated for their applicability to other important geomarketing problems.

9 References

Bachi R. (1957), *Statistical Analysis of Geographical Series*, Central Bureau of Statistic, Kaplan School, Jerusalem

Bailey T.C., Gatrell A.C. (1995), *Interactive Spatial Data Analysis*, Longman, Essex

Mauri C. (2001), *Le promozioni fedeltà della distribuzione moderna nel quinquennio 1996-2000*, Università Bocconi, Milano

Staufer-Steinnocher P., Leitner M. (2001), *Kernel Density Estimation: The Application of Point pattern analysis Techniques to Geomarketing*, Vienna University of Economics and Business Administration, Vienna

10 Further reading

Anselin L. (1992), *Spatial Analysis with GIS: An Introduction to Application in the Social Sciences*, Technical Report 92-10, University of California, Santa Barbara

Anselin L. (1995), "Local Indicators of Spatial Association – LISA", *Geographical Analysis*, n.2, pg. 93-115

Anselin L., Griffith D.A. (1993), *Operational Methods of Spatial Data Analysis*, Oxford University Press, Oxford

Boots B.N., Getis, A. (1988), *Point Pattern Analysis*, Sage Publications, Newbury Park

Burt J.E., Barber G.M. (1996), *Elementary Statistic for Geographers*, The Guilford Press, New York

Buxton T. (1992), "GIS Expected to Meet Micromarketing Challenge", *GIS World*, n. 5, pg. 70-72

Clark P., Evans F.C. (1954), “Distance to the nearest neighbor as a measure of spatial relationship in populations”, *Ecology*, n. 35, pg. 445-453

Cressie N. (1991), *Statistics for Spatial Data*, Wiley & Sons, New York

Ding Y., Fotheringham A.S. (1992), “The integration of spatial analysis and GIS”, *Computers, Environment and Urban Systems*, n. 16, pg. 3-19

Ebdon D. (1988), *Statistic in Geography*, Blackwell, Oxford

Fischer M.M., Nijkamp P. (1992), “Geographic information systems and spatial analysis”, *The Annals of Regional Science*, n. 26, pg. 3-17

Getis A., Boots B.N. (1978), *Models of Spatial Processes*, Cambridge University Press, Cambridge

Getis A., Ord K.J. (1992), “The analysis of spatial association by use of distance statistics”, *Geographical Analysis*, n. 24, pg. 186-206

Griffith D.A., Amrhein C.G. (1991), *Statistical Analysis for Geographers*, Prentice-Hall, Englewood Cliffs

Grimshaw D.J. (1994), *Bringing Geographical Information Systems into Business*, Longman, London

Haining R. (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge

Hammond R., McCullagh P. (1978), *Quantitative Techniques in Geography : An Introduction*, Clarendon Press, Oxford

Isaaks E.H., Srivastava R.M. (1989), *An introduction to applied geostatistics*, Oxford University Press, Oxford

Isard W. (1962), *Methods of Regional Analysis: An Introduction to Regional Science*, The Technology Press of The Massachusetts Institute of Technology and Wiley & Sons, New York

Levine N. (1996), “Spatial statistics and GIS”, *Journal of the American Planning Association*, n. 62, pg. 381-391

Levine N. (1999), *CrimeStat. A Spatial Statistic Program for the Analysis of Crime Incident Locations*, Ned Levine and Associates, Annadale, VA

Longley P., Clarke G.P. (1995), *GIS for Business and Service Planning*, Geoinformation, Cambridge

Moran, P. (1948), “The interpretation of statistical maps”, *Journal of the Royal Statistical Society*, n.10

Okabe A., Yomono H., Kitamura M. (1995), “Statistical Analysis of the Distribution of Points on a Network”, *Geographical Analysis*, vol. 27, n. 2, pg. 152-175

Okabe A., Kitamura M. (1996), “A Computational Method for Market Area Analysis on a Network”, *Geographical Analysis*, vol. 28, n. 4, pg. 330-349

Okabe A., Yamada I. (2000), *The K-function Method on a Network and its Computational Implementation*, University of Tokio, Tokio

Ripley B.D. (1976), “The second-order analysis of stationary point process”, *Journal of Applied Probability,* n.13

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

Silverman B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London

Upton G., Fingleton B. (1985), *Spatial Data Analysis by Example – Volume 1: Point Pattern and Quantitative Data*, Wiley & Sons, New York

Venables W.N., Ripley B.D. (1997), *Modern Applied Statistics with S-Plus*, Springer, New York