Network Density Estimation: Analysis of Point Patterns over a Network

Giuseppe Borruso

Università degli Studi di Trieste, Dipartimento di Scienze Geografiche e Storiche,
Piazzale Europa 1, 34127 Trieste, Italia
giuseppe.borruso@econ.units.it

Abstract. This research focuses on examining point pattern distributions over a network, therefore abandoning the usual hypotheses of homogeneity and isotropy of space and considering network spaces as frameworks for the distribution of point patterns. Many human related point phenomena are distributed over a space that is usually not homogenous and that depend on a network-led configuration. Kernel Density Estimation (KDE) and K-functions are commonly used and allow analysis of first and second order properties of point phenomena. Here an extension of KDE, called Network Density Estimation (NDE) is proposed. The idea is to consider the kernel as a density function based on network distances rather than Euclidean ones. That should allow identification of 'linear' clusters along networks and the identification of a more precise surface pattern of network related phenomena.

1 Introduction

This research is focused on examining point pattern distributions over a network, therefore abandoning the usual hypotheses of homogeneity and isotropy of space and considering network spaces as frameworks for the distribution of point patterns. Many human related point phenomena are distributed over a space that is usually not homogenous and that depend on a network-led configuration. That is the case of population or commercial facilities, which locate in space depending on road network infrastructures. The study of networks themselves can also be biased by assumptions on homogeneity and isotropy. The problem of analyzing networks and their influences over space can be reduced to analyzing their point structure, as proposed by some authors ([1]; [2]), but the consequent analyses focus on clusters in a homogenous space.

The starting point of the present research is point pattern analysis, which both in spatial analysis and GI science is commonly adopted and used. Kernel Density Estimation and K-functions (Ripley's and Diggle's) are commonly used and allow analysis of first and (reduced) second order properties of point phenomena. Authors ([3]) have proposed methods for estimating K-functions over a network structure. Here an extension of KDE, called Network Density Estimation (NDE) is proposed. The idea is to consider the kernel function as a density function based on network distances rather than Euclidean ones. One of the advantages of such estimator is that it

O. Gervasi et al. (Eds.): ICCSA 2005, LNCS 3482, pp. 126 – 132, 2005.

should allow identification of 'linear' clusters along networks and a more precise surface pattern identified of network related phenomena. NDE could find application in 'traditional' environment of KDE, as population analysis, crime studies, retail analysis, and network studies.

2 Point Pattern Analysis: Density Estimation

When dealing with a point pattern, authors like Bailey e Gatrell ([4], [5]) consider *events*, referred to observed phenomena over a point distribution, and *points*, referred to all the other places in the study area. The simple observation of events' distribution over space can provide initial information on the structure of the distribution, but more refined analytical instruments are needed for more in depth analysis, and particularly to identify clusters or regularity in the distribution.

Quadrat analysis is one of the means of ordering the pattern of a distribution of events within a region *R*. The procedure involves dividing the study region in sub-regions having equal and homogeneous surfaces, or quadrats[1]. The following step consists of counting the number of events falling in each sub-region (quadrat) in order to simplify and group the spatial distribution.

The number of events therefore becomes an attribute of the quadrat. It is then possible to represent the spatial distribution by means of homogenous and easy comparable areas. Density analyses become possible using an easy to use and to compute method ([6], [7]).

The method has some disadvantages, as the loss of information from original data, as well as different levels of arbitrariness deriving from:

- The choice of quadrat dimension;
- The orientation of the grid;
- The origin of the grid.

Different analyses could be computed, changing a grid's origin or the quadrats' dimensions. One of the solutions involves considering the number of events for each area unit within a mobile 'window'. A fixed radius is chosen and we hypothesise to centre it in a number of places in the space, where events are organised in a grid superimposed to the study region R. An estimate of the intensity in each point of the grid is therefore provided. This generates an estimate of the variation of the intensity smoother than that obtained from a fixed grid of square cells superimposed.

Such method is very close and at the basis of the procedure called Kernel Density Estimation (KDE).

The kernel consists of 'moving three dimensional functions that weights events within its sphere of influence according to their distance from the point at which the intensity is being estimated' ([6]).

The general form of a kernel estimator is

$$\hat{\lambda}(s) = \sum_{i=1}^{n} \frac{1}{\tau^2} k\left(\frac{s - s_i}{\tau}\right) \tag{1}$$

[1] Other tessalations of space are also possible, as triangles, hexagons or other polygonal shapes.

where $\hat{\lambda}(s)$ is the estimate of the intensity of the spatial point pattern measured at location s, s_i the observed ith event, *k( )* represents the kernel weighting function and τ is the bandwidth.

For two-dimensional data the estimate of the intensity is given by

$$\hat{\lambda}(s) = \sum_{d_i \le \tau}^{n} \frac{3}{\pi\tau^2}\left(1 - \frac{d_i^2}{\tau^2}\right)^2 \tag{2}$$

where d_i is the distance between the location s and the observed event point s_i. The kernel values therefore span from $\frac{3}{\pi\tau^2}$ at the location s to zero at distance τ ([6]).

The kernel density estimation function creates a surface representing the variation of density of point events across an area. The procedure can be organized in three steps ([8])

- A fine grid is placed over the study region and the point distribution
- A moving three-dimensional function visits each cell and calculates weights for each point within the function's radius (threshold).
- Grid cell values are calculated by summing the values of all circle surfaces for each location.

3 Network Density Estimation (NDE): The Algorithm

Kernel Density Estimation allows finding out clusters in point pattern distributions over a study area particularly highlighting 'circular' clusters. However clusters can appear also following different distribution schemes as network-led spaces. Here a procedure for considering network spaces in a point distribution is presented.

The algorithm foresees in particular the modification of the searching kernel function from a circular to a network-based service area.

Steps in the algorithm:

1. selection of a point process (i.e., population, ATM, robberies, services' locations) ;
2. generation of a regular grid over study area;
3. generation of centroids of cells belonging to regular grid overlapped to study area;
4. definition of a bandwidth; (same process as Kernel Density Estimation)
5. selection of a network;
6. computation of service area analysis from cells' centroids over network
7. overlay (= spatial intersection) of service areas and point process;
8. count of point (events) per service area;
9. assignment of count (= weight; relative density, etc.) to cell centroid;

9a. if necessary, before visualising density surface a further interpolation could be performed between cell's centroids in order to smooth the density surface]

10. Visualisation of density surface.

The derived density function reflects the network structure of the space. Point processes are therefore analysed rejecting the hypotheses of homogeneity and isotropy of space, considering the network-driven structure of the pattern. The density function is therefore not the result of a circular search radius but of a network-shaped one. Doing so allows evaluating more precise densities of network based phenomena, as ATM locations, burglaries, etc. One of the advantages in using such modified kernel density estimation is that clusters can be more easily detected when phenomena group along a street or a road, what is not always perceivable in traditional, Euclidean KDE.

4 Network Density on Point Patterns: Applications

4.1 Kernel Density Estimation on Networks

An application of the Network Density Estimation procedure was carried out using GIS software and spatial statistical packages considering different phases and steps. The example considered starts from a research on network structures, where KDE algorithm was used to highlight network spaces starting from the spatial distribution of nodes ([2]). In that occasion a network density analysis was performed over a point distribution consisting of nodes of a road network. It was assumed that the network structure could be simplified by using nodes instead of arcs and therefore compute a density analysis. The aim of the network density analysis was the identification of network spaces and particularly centres of urban areas and settlements. As KDE allowed the identification of peak areas corresponding to higher concentration of nodes - junctions in the road network – centres and subcentres in urban areas and peri-urban settlements could be highlighted.

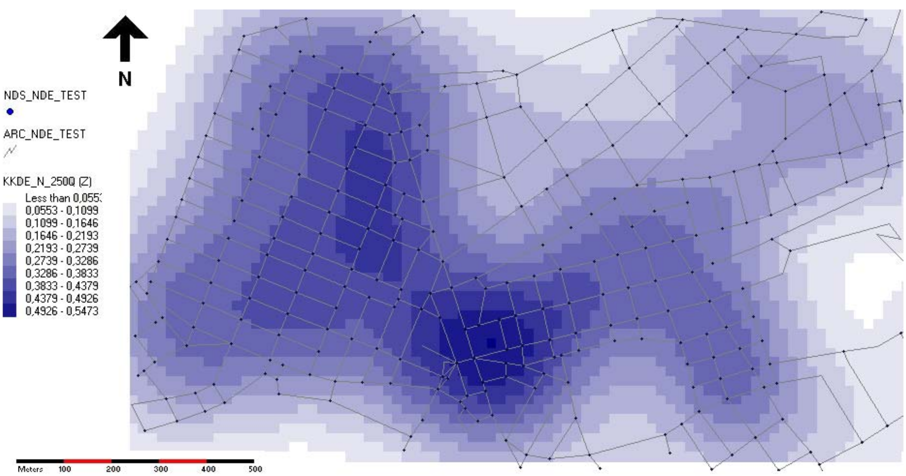

Fig. 1. Kernel Density Estimation computed over road network's nodes distribution in the Trieste (Italy) city centre. 250 m Bandwidth (Density surface produced using CrimeStat 1.1 – [9])

Figure 1 shows the results for a KDE performed over the road network of Trieste city centre (Italy). A 250 m bandwidth was chosen and nodes were weighted by the number of arcs converging to each node. The bandwidth of 250 m generates a surface that presents several peaks and therefore seems more suitable to highlight local interaction phenomena. Darker areas represent clusters of highly connected junctions in the road network. The density decreases as one moves from the ‘centres’, characterized by high values of road transport network density.

Although interesting results can be found in the analysis of network structure and shape, particularly when they are compared with the built environment’s distribution, some limitations in the use of a ‘pure’ KDE to network point datasets exist. When dealing with network structures in fact the spatial configuration of point distribution over a network cannot be considered as lying onto a homogeneous and isotropic space.

Such a limitation become particularly evident when linear clusters appear. Using KDE, which relies on a circular searching function, such kind of clusters are not always detected. A network density estimator therefore appears as a more suitable solution for highlighting a network’s structure and orientation.

4.2 Network Density Estimation

A initial version of the Network Density Estimation algorithm was implemented to adapt density analysis over network spaces. As suggested in paragraph 3, one of the steps of NDE involves the creation of service areas for each grid cell’s centroid. Service areas were therefore used to simply count the number of nodes falling inside and then assigning such value to cell’s centroid. In this initial application of the algorithm the proximity of the events to centroids into the service area was not considered.

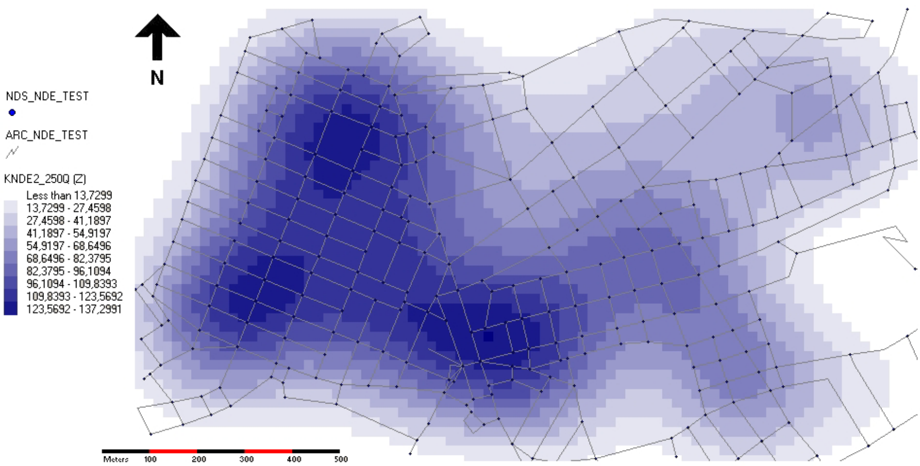

Fig. 2. Network Density Estimation computed over road network’s nodes distribution in the Trieste (Italy) city centre. 250 m Bandwidth. (Density surface produced using CrimeStat 1.1 – [9])

Service areas were computed in a GIS environment, as well as the nodes' count and assignment to service areas' centroid. As not all of the cells' centroids fall onto the network or close to it, not all of the grid cells were filled with density values. To avoid the presence of 'holes' in the final graphical representation of the results, the values from the density analysis were interpolated to obtain a KDE-like density surface. Interpolation was performed using a spatial analytical package.

Figure 2 shows the results of the NDE performed over the road network of Trieste. The 250 m bandwidth was used to compute distances over the network rather than in the Euclidean space. Nodes – junctions – were counted and weighted by the number of arcs converging to each node.

Centres highlighted by higher values of network densities can be spotted in Figure 2. If we compare the results from the NDE with those previously obtained with the KDE we can notice different peak areas in the NDE analysis particularly in the western part of the map and a lighter one on the north-eastern part. Such peaks seem to be more consistent with the pattern drawn by the real road network.

5 Conclusions

The first applications of the Network Density Estimation procedure provided interesting results in highlighting network-driven distributions of events in space and therefore reducing the hypothesis of homogeneity and isotropy of space that to some extent is contained in the Kernel Density Estimation algorithm. NDE performed over road network's junction distribution highlights in particular clusters following the orientation of the network, showing network spaces and centres in a more proficient way.

Further research will involve tackling different open issues, both regarding network density analysis and concerning the study of other point-like phenomena distributed over a network space.

With reference to the 'pure' network density analysis there is the need to implement the NDE's steps within an algorithm in a complete GIS or spatial analytical environment. A distance-weighting function should be inserted to weight events according to the distance from the service areas' centroid.

With reference to the applications of the algorithm, when continuing the experiments on networks' spaces minor, linear settlements should be considered to test clustering along main roads. An application of the algorithm to real transport networks should also involve the directional constraints an urban network usually has, as one ways and limited access arcs. That should interestingly affect the shape of service areas obtained and offer different network density areas in an urban environment.

Finally the algorithm should be performed over other point datasets, as ATM, post office, retails, etc. in order to study their distribution along networks in urban and extra-urban areas.

References

1. Borchert J R.: The twin cities urbanized areas: past present and future. Geographical Review, (1961) 51: 47-70.

2. Borruso G.: Network Density and the Delimitation of Urban Areas. Transactions in GIS, (2003), 7 (2), pp. 177 – 191.
3. Okabe, A., Yamada, I.: The K-function method on a network and its computational implementation. Geographical Analysis (2001), Vol.33, No.3, pp.271-290.
4. Gatrell A., Bailey T, Diggle P., Rowlingson B.: Spatial Point Pattern Analysis and its Application in Geographical Epidemiology. Transactions of the Institute of British Geographers (1996) 21: 256-74.
5. Bailey T. C., Gatrell A. C.:.Interactive Spatial Data Analysis. Longman Scientific & Technical, Essex (1995)
6. Holm T.: Using GIS in Mobility and Accessibility Analysis, ESRI Users Conference, San Diego, http://www.esri.com/library/userconf/proc97/proc97/to450/pap440/p440.htm (1997)
7. Matti W.: Gridsquare Network as a reference system for the analysis of small area data. Acta Geographica Lovaniensia, (1972) 10: 147-163.
8. Chainey S., Reid S and Stuart N., When is a hotspot a hotspot? A procedure for creating statistically robust hotspot maps of crime, in Kidner D., Higgs G. and White S. (eds.), *Socio-Economic Applications of Geographic Information Science*, Innovations in GIS 9, Taylor and Francis, (2002), pp. 21-36
9. Levine Ned, CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2.0). Ned Levine & Associates, Houston, TX, and the National Institute of Justice, Washington, DC. May 2002.
10. Ratcliffe, J. H., McCullagh, M. J.: Hotbeds of crime and the search for spatial accuracy, Journal of Geographical Systems (1999), Volume 1 Issue 4: 385-398.