Approaches for Cluster Analysis of Activity Locations along Streets: from Euclidean Plane to Street Network Space

Yongmei Lu

Department of Geography

Texas State University – San Marcos

yongmeilu@txstate.edu

Telephone: 512-245-1337

Fax: 512-245-8353

Approaches for Cluster Analysis of Activity Locations along Streets: from Euclidean Plane to Street Network Space

ABSTRACT

Different from many natural phenomena (such as soil categories and vegetable species), the distributions of some social and economic activities, especially those in urban areas, are subject to location restriction imposed by urban street network. Such examples include crime offense locations, traffic accident locations, and different types of commercial service locations, just to name a few. However, many investigations of such distribution simply ignore the restriction of street network and assume that the point locations can be located anywhere across the study area. Moreover, most existing methods for analyzing and visualizing point clustering are designed to handle distributions on Euclidean plane.

This paper starts from a brief discussion of the geometrical and topological properties of street network space and their difference from Euclidean plane. The author then provides formal discussion of extending both distance approach or zoning approach for point clustering analysis from Euclidean plane into street network framework. Empirical analyses of crime clustering patterns along streets are reported in the paper. By comparing the findings from applying a street network framework versus a Euclidean plane, it is showed that possible false conclusion of point pattern is an issue if a space framework is inappropriately adopted. Emphasizing that street network space should be applied when studying activities distributed along streets, this paper promotes further studies to extend clustering analysis techniques from Euclidean plane to other types of space framework (e.g. street network space) that can better capture the distribution nature of activities under investigation.

INTRODUCTION

Spatial statistics adds a geographical dimension to data and makes integration and analysis easier (Meliskova 2000). Exploratory spatial data analysis (ESDA) is mainly based on the techniques of spatial statistics and geographical information systems (GIS); its power of spatial pattern detecting and explanation highly relies on the capability of handling spatial autocorrelation. Clustering analyses assess the spatial autocorrelation of points through evaluating the existence of clustering as well as identifying the locations of clusters. There are two major groups of methods for clustering analysis of point patterns – distance criterion or zoning technique. For the former, the metric distance separation among points of observation is compared with that of a random distribution; a clustering conclusion is reached if the observed points are closer to each other than those in a random pattern. For the latter, a zone system is defined and draped over the whole study area; the points are claimed to be clustering in certain zone(s) if there are more points in the zone(s) than there could be for a random distribution. There is an abundance of literature on clustering analysis of point patterns from both methodological (e.g. Diggle 1983; Boots and Getis 1988; Bailey and Gatrell 1995; Fotheringham *et al.* 2000) and application (e.g. Rushton 1996; Kulldorf 1998; Craglia et al. 2000; Lu and Thill 2003) perspectives.

As most GIS functions, the majority techniques of spatial statistics in general and clustering analysis in particular assumes the spatial entities under concern are referenced to the coordinatized Euclidean plan. However, "there are many situations for which representation in a Euclidean space is not the most appropriate model, nor may even be derivable" (Worboys *et al.* 1997, 35). Worboys *et al.* mentioned such examples as travel time spaces, qualitative distances and flow spaces (Worboys *et al.* 1997). Realizing this problem, a few researchers have made effort to expand the clustering analysis from Euclidean plane to other appropriate spaces. For example, Okabe and colleagues have conducted research to develop *k*-function analysis for point features distributed along street network (Okabe et al. 1995; Okabe and Yamada 2001). Yamada and Thill (2004) compared *k*-functions for planar and network spaces for the distribution of highway traffic accidents. Further, Levine and Associates, with the support from US Department of Justice, implemented the *k*-function for both Euclidean plan and grid space for the analysis of point distribution (Levine 2002).

This paper discusses the major challenges for point clustering analysis in one type of non-Euclidean space – street network space. It further explores the possible avenues of addressing these problems by pointing out that both global and local spatial statistics techniques for point pattern analysis have the potential of being extended from Euclidean plane into street network space if the representation and computation of street network distance and street zoning topology can be effectively and efficiently handled. Two empirical applications of clustering analysis of auto thefts along streets are presented to illustrate the approach being promoted by the author.

GOING BEYOND EUCLIDEAN PLANE

It has long been realized that Euclidean geometry is not the best for urban geography considering the restrictions imposed onto socioeconomic activities by urban street system (e.g. Krause 1975). Recent research has once and again recognized the importance of accounting for the difference between Euclidean and street network spaces when examining urban activities (e.g. Okabe and Okunuki 2001; Yamada and Thill 2003; Horner and Murray 2004). The major challenge for clustering analysis of point activities along street is that most existing methods are designed based on distance and spatial separation measurement on Euclidean plane, i.e. distribution can be continuous anywhere on the surface and distance between locations is straight-line distance. However, locations in streets might be far away from each other when traveling along streets while appearing close to each other in a continuous Euclidean plane. In figure 1, *P1*, *P2*, and *P3* are three locations along street network; while *P2* is closer to *P1* than *P3* is in a Euclidean plane, the street distance from *P1* to *P3* is shorter. Hence, for point pattern analysis in a non-Euclidean space, the measurement of distance should account for the continuity nature of the space.

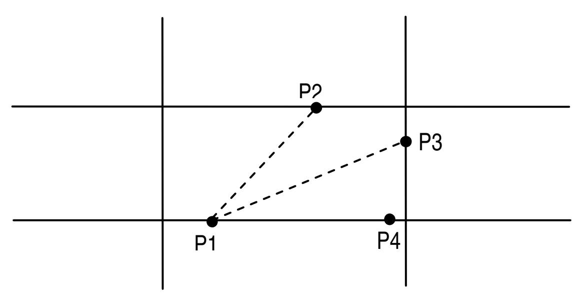


Figure 1. Distance measurement in Euclidean plane versus street network

Another major issue of concern for point pattern analysis is related to topology of the operational space. Non-metric measure of proximity in a street network, such as the number of service nodes or stops between locations (Fabrikant *et al.* 2004), is an example. As raised by Fabrikant *et al.* (2004), to evaluate the spatial separation between two locations along a subway route, it is worth investigating if the length of the route network more meaningful or the number of stops between the two locations more indicative. It is easy to imagine for anyone who has experience driving in urban streets that, more often than not, it is the street intersections not the length of street segment (i.e. section of street that is between two neighboring street intersections) that makes difference for the traveling speed and time. Hence, the connection / proximity between two locations in a street network might be more sensitive to the number of street intersections than the metric distance between them. In another word, two locations that are not so close along one street (e.g. *P1* and *P4* in Figure 1) might be more connected to each other than two other locations that are close to each other metrically but on different streets (e.g. *P3* and *P4* in figure 1). Jiang and Claramunt (2004) adopted a "named-street-oriented" view to describe the topology of urban street network and developed a *k*-clustering coefficient to measure the connectivity of streets. Similarly, street network distance or other non-metric measurement (e.g. number of subway stops between locations, as proposed by Fabrikant *at al.* (2004)) rather than Euclidean distance should be used for the assessment of spatial separation and connection between locations in streets.

CLUSTERING OF LOCATIONS IN STREET NETWORK SPACE

Given the special geometry and topology properties of street network and locations along streets, it is inappropriate to directly apply point pattern analysis techniques from Euclidean plane to a space of street network. Okabe and colleagues (Okabe et al. 1995; Okabe and Yamada 2001) developed a global scale measurement of point clustering in street network – network *k*-function. Instead of using direct Euclidean distance to measure the separation between neighbor points, network *k*-function uses distance along street network. Yamada and Thill (2004) applied both the traditional *k*-function and a network *k*-function to assess the clustering of traffic accidents along highway; the study showed that false positive is a problem with traditional *k*-functions. For a comprehensive evaluation of applying traditional planar *k*-function to analyze locations distributed in a street network, Lu and Chen (2004) links the propensity of false alarm from traditional *k*-function with the properties of urban street network and point location distribution. To summarize, existing studies have proved that it is dangerous to simply extending the point pattern analysis techniques from Euclidean space to street network, since these techniques can not account for the geometric and topological properties of street network space. But there is a definite need for a systematic development of the counterpart in street network space for the point clustering analysis techniques in Euclidean space.

For the clustering analysis of activity locations in Euclidean space, two groups of approaches are commonly adopted. One group uses continuous function of distance metric to measure spatial separation between locations; the other group implements a zoning system onto the study area and measures separation using topological relationship between zones. Anselin (1996) named them distance view and neighborhood view respectively. Boots and Getis (1988), on the other hand, classified them into distance method and quadrat method. As discussed above in the paper, both the geometric distance measurement and the topology property between locations along street network are different from those in traditional

Euclidean plane. The techniques for clustering analysis of point patterns in a street network need to account for the difference and treat distance and connection between locations accordingly. Put another way, the following two aspects need to be taken care of before clustering analysis techniques for point locations along street network can be fully developed –

Distance approach: substitute Euclidean surface distance with network distance

Let Ω^2 denote a 2-D Euclidean plane, K denote a street network space, and ζ denote a set of points under assessment. Also, let $d_\Omega(i, j)$ denote the distance between point i and point j in space Ω^2, and $d_K(i, j)$ denote the network distance between point i and point j in space K. The network distance, $d_K(i, j)$, is defined as a route distance, i.e. the shortest-path distance along the street network between point i and point j. For all point clustering analysis in K, not only the distance between points of concern, i.e. $d_K(i, j)$, should be network distance, but also the estimated "expected distance", $d'_K(i, j)$, to which the observed distance is compared to in order to judge the degree of clustering should be generalized using network distance. The capability of efficiently capturing network distance and connectivity measurement is very sensitive to the size of street network and point locations under concern.

Okabe and Yamada's (2001) network k-function and its implementation is a good example of extending point pattern analysis technique from Ω^2 to K while successfully taking care of distance measurement. Given a set of points, ζ, that are distributed on a plane, Ω^2, the original Ripley's planar K compares the observed K value, $K(d_\Omega)$, with the expected K value, $K'(d_\Omega)$.

$$K(d_\Omega) = \lambda^{-1} \times n^{-1} \times \sum_{i=1}^{n} p_i(d_\Omega) \qquad \text{........... (1)}$$

$$K'(d_\Omega) = \lambda^{-1} \times E(d_\Omega, \zeta, \Omega) \qquad \text{........... (2)}$$

where, d_Ω indicates the distance for which the K value is calculated, λ is an unbiased estimation of the distribution density of the points in ζ, E represents the expected number of points within distance d_Ω to a random point in set ζ over plane Ω^2, $P_i(d_\Omega)$ denotes the number of points observed within distance d_Ω of point i, and n is the total number of points in ζ. Through increasing d_Ω from small to large, typically with 50-100 small increments to cover the extent of point set ζ on plane Ω^2, a series of $K(d_\Omega)$ can be obtained and plotted against d_Ω. Monte Carlo simulation is usually conducted to generate a series of random distributions of n points on the same plane Ω^2 in order to estimate the expected $K'(d_\Omega)$. For each simulated distribution, a series of $K'(d_\Omega)$s are calculated for different d_Ωs. The simulation is usually conducted to generate 100 or more distributions. The observed $K(d_\Omega)$ is then ranked together with the set of $K'(d_\Omega)$s from simulated distributions for certain distance. By comparing $K(d_\Omega)$ with $K'(d)$, one can tell if the observation set, ζ, is clustered at certain distances and dispersed at others.

For network k-function, given a set of points ξ distributed on street network K, the total link of the network is L_T and the total length of L_T is l_T, the expected network k-function under complete spatial randomness (CSR) can be defined as

$$K'(d_k) = \omega^{-1} \times E(d_k, \xi, L_T) \qquad \text{........... (3)}$$

where ω is the density of points in K, calculated as $\omega = n/l_T$, ξ represents the set of points (p_1, p_2,…, p_n) in K, E is the expected number of points within network distance d_K to a point in ξ on

L_T. The expected number of points within d_K to a point in ξ over network L_T is estimated under a Poisson process. Then the observed network $K(d_K)$ can be estimated as

$$K(d_k) = \omega^{-2} \times l_T^{-1} \times \sum_{i=1}^{l_i} l_{T_i}(d_k) \quad \text{........... (4)}$$

where $l_{Ti}(d_K)$ denotes the total number of points on a subset of L_T, $L_{Ti}(d_K)$. The network distance between point i and any other point on $L_{Ti}(d_K)$ is less than or equal to d_K. Similar as planar k-function, Monte Carlo simulation is performed to obtain an estimation of expected network K value, $K'(d_K)$. When a series of $K'(d_K)$s are obtained, the $K(d_K)$ is ranked together with and compared to the distribution of $K'(d_K)$s to decide if the observed points are clustered or dispelled in $\mathcal{K}$ at different network distances.

Similar extension can be made for other distance approach techniques (e.g. nearest neighbor method or p-means technique) to develop point clustering analysis method from Ω^2 to $\mathcal{K}$. The major technical challenge is to handle the computation when calculating d_ks and when simulating distribution in $\mathcal{K}$ under certain conditions (Lu and Chen 2004). Note that, in addition to compute network distance d_k rather than straight-line distance d_Ω for any location pairs (see figure 1), calculation in $\mathcal{K}$ has to dynamically keep track of the links and length of the whole or partial street network according to the location pairs being included. Hence, the functioning of related extension from Ω^2 to $\mathcal{K}$ for point pattern analysis is, to a large degree, depending on the computation capability and algorithm efficiency to simulate point distribution in $\mathcal{K}$ and to measure the geometry and topology relationships between locations in $\mathcal{K}$ while keeping track of the whole or part of the network under investigation.

Zoning approach: define 1-D zone in a network space rather than 2-D plane zone

Again, let Ω^2 denote a 2-D Euclidean plane, $\mathcal{K}$ denote a street network space, and ζ denote a set of n points under investigation, $\{P_i(x_i, y_i), i=1, 2, \ldots n\}$. Further let ψ denote a set of m service points, $\{Q_u(x_l, y_l), u=1, 2, \ldots m\}$, for which a zone system in $\mathcal{K}$ need to be defined. Different from distance approach, zoning approach assumes a catchment or impact zone for each of the point in ψ. Let $Z_\Omega(u)$ indicates the zone for service point u in Ω^2 and $Z_{\mathcal{K}}(v)$ represents the defined zone for service point v in $\mathcal{K}$. Since zoning methods represent spatial proximity by number of zones, the topology (i.e. contiguity relationship) between zones is critical. Hence, there are two aspects for the zoning approach – definition of zone and description of zone contiguity.

Zones in Euclidean plane, i.e. $Z_\Omega(u)$ in Ω^2, are usually defined as either a regular grid or irregular catchment for Q_u in ψ. Let $F(d_\Omega(P_0, Q_u))$ be a zone definition function based on the distance between any point $P_0 \in \Omega^2$ and $Q_u \in \psi$. The zone of service point u is a polygon that can be described as

$$Z_\Omega(u) = a_{u,\Omega} F(d_\Omega(P_0, Q_u)) \quad \text{........... (5)}$$

where $a_{u,\Omega}$ is a parameter describing the impact of service point Q_u to its surroundings in Euclidean space, Ω^2; $d_\Omega(P_0, Q_u)$ is the Euclidean continuous distance between Q_u and P_0. When $a_{u,\Omega}$ is a constant for all u, then the zoning system is a regular grid; otherwise, an irregular zone system presents. Moreover, if distance deterrence function, $F(d_\Omega(P_0, Q_u))$, has anisotropy property, the zones defined will be less compact and probably will show as irregular polygons.

When extending the zone definition from Euclidean plane to street network space, zones become line segments instead of polygons. $Z_{K}(v)$ is the zone for service point v in network K. The zones do not necessarily be of same size, i.e. street zones might not be equal in length. The definition of $Z_{K}(v)$ can be described as

$$Z_{K}(v) = a_{v,K} F(d_{K}(P_0, Q_v)) \qquad \text{…………..} \quad (6)$$

where $a_{v\,K}$ is a parameter denotes the impact of service point Q_v to its surroundings in network space K; $d_{K}(P_0, Q_v)$ is the network distance between any point $P_0 \in \zeta$ and service point $Q_v \in \psi$; $F(d_K(P_0, Q_v))$ is a distance deterrence zone definition function. With $a_{v,K}$ being a constant, the zone system in K will be made up of equally-sized streets; otherwise, some street zones will be longer than others, depending on the service point. Also, if showing a property of anisotropy, the distance function will define zones with point Q_v not being the center of street zones.

Contiguity in Euclidean plane is a 2-D property, meaning that zones are neighboring to each other by sharing common boundaries for the most cases (with one point connection occasionally). GIS software that handles topological information has the capability of identifying neighboring polygons. Furthermore, a neighbor's neighbors can also be identified for a zone so that high-order neighbors for a service location's catchment can be tracked. Based on the definition of zones, evaluation of the clustering of observed points $\{P_i\}$ in the catchment of service points Q_u (or Q_v) can be conducted by comparing the count or density of P_is in $Z_{\Omega}(u)$ (or $Z_{K}(v)$) with that of expectation under random distribution. Point clustering of P_is in higher order catchment for each service point (which can be built based on the zone contiguity pattern) can be evaluated similarly. For Euclidean plane analysis, it is hard to estimate the number of neighbors that one polygon might have; but for street network, a street line commonly has between one and six first order neighbors.

It can be seen from the above discussion that a formal extension of point clustering analysis techniques from traditional Euclidean plane to street network space is not complicated at the conceptual level. But capturing and measuring geometry and topology properties between locations in a street network space might be challenging. Plus, different application scenarios call for different zoning frameworks, which might need different specifications of coefficient $a_{v\,K}$ to measure the impact of service point and different distance deterrence function $F(d_K(P_0, Q_v))$. Moreover, the distance measurement and topological relationship tracking in street network space is computational intensive, as discussed by Okabe and Yamada (2001). But it is undoubted that activity locations along street network need to be handled differently from those in Euclidean plane when evaluating location clustering. Both geometry and topology property of street network space are different from those of Euclidean plane. The following section reports two empirical implementations of the above discussed extension from Euclidean plane into street network space. The first one is based on distance approach while the second one is a zonal approach for auto thefts clustering analysis along street network.

EMPIRICLE ANALYSES: CLUSTERING OF AUTO THEFTS ALONG STREETS

Being a major means for transporting people and goods, vehicles are usually parked or stored at places in or along streets in urban area, or at places close to and having good access to streets, such as garages or parking lots. Hence, auto thefts are usually showing a pattern following the layout of street network (e.g. Copes 1999; Lu 2003; Lu and Thill 2003). Although there are discussions in crime literature about the difference between Euclidean space and street network space and its implication for understanding crime (e.g. Groff and

Alexander 1998) and criminal's travel behavior (e.g. Rossmo 2000), studies have rarely explored the possibility or challenges of adopting a street network framework when analyzing the patterns of crime along streets. Lu and Chen (2004) are among the earliest to take a distance approach and extend the *k*-function from Euclidean plane to street network (following Okabe and Yamada 2001) when examining the clustering of crime locations.

This section of the paper is designed to demonstrate two examples of extending point clustering analysis from Euclidean plane to street network space. It contains two parts. The first part follows Lu and Chen (2004) and shows the difference between the findings from applying a distance approach with different space frameworks to the same set of auto theft locations in a part of a big city, San Antonio, Texas. The second part reports an extension of zoning approach into street network space when analyzing auto theft hot spots in the city of Buffalo, New York. The case studies are selected considering data availability and data quality. Readers need to keep in mind that these two empirical analyses reported here are for the purpose of showing the necessity of accounting for the restrictions and properties of street network space when analyzing activities distributed along streets. It is not the main goal of these two analyses to accurately evaluate the auto theft clustering in either of the study area. In another word, these examples are included here to promote the idea that the analyses of crimes along street should be conducted using a street network framework rather than simply assuming that they are features distributed in a Euclidean space. It is believe that when related extensions of clustering analysis techniques are matured and widely adopted by crime analysis, the analysis and reporting of crime clustering pattern will be improved significantly. Crime policy and management will surely benefit from this movement.

False alarm when applying planar k-function to the analysis of auto thefts in San Antonio

The city of San Antonio is covered by six San Antonio Police District (SAPD). To ensure that the data set is large enough to show statistical patterns but not too large to burden the computer with excessive computation (especially when network K-function is calculated), the analyses of this part examine auto thefts reported to the East and Central SAPDs in January, February and March of 2002. There were 330 offenses in these two SAPDs during this period. The data set was acquired from the Police Department of San Antonio. Whilst the quality of geocoding of the police data is worth investigating, it is not a major concern here as the purpose is to compare different types of *k*-functions for crime pattern analysis.

The planar *k*-function was calculated through modifying the *k*-function routine in CrimeStat2.0 (Levine, 2002). The network *k*-function was calculated with SANET, a software package developed by Okabe's group (Okabe and Yamada 2001). Detailed explanation of the two *k*-functions can be found in section 3 of this paper. One hundred times of Monte Carlo simulation were conducted at each distance band for each *k*-function to obtain the expected distributions. Figures 2 and 3 report the results of the *k*-function analyses for auto thefts in the East and Central SAPD. Since street network distance is believed to be a better description of the distance between theft locations than Euclidean distance, the analyses using network *k*-function is supposed to provide an overall better evaluation about the clustering of theft locations than those based on planar *k*-function. Therefore, the performance of planar *k*-function can be assessed by comparing to that of network *k*-function for analyzing crimes along streets. A false positive of planar *k*-function refers to the scenario when a planar *k*-function indicates clustering or higher degree of clustering while a network *k*-function points to non-clustering or low degree of clustering. Similarly, when a planar *k*-function shows non-clustering or low degree of clustering but a network *k*-function designates clustering or high degree of clustering, a false negative problem exists for planar *k*-function.

It can be seen from figures 2 and 3 that planar *k*-function is subject to both false positive (for East Police District) and false negative (for Central Police District) problem. The planar *k*-function shows that auto thefts in East SAPD cluster through out the study distance (up to 20,000 feet) while network *k*-function indicates that the clustering of these auto thefts goes up to about 16,500 feet and they disperse beyond that. For Central SAPD, while planar *k*-function reports the clustering of auto theft only up to about 5,000 feet distance, network *k*-function reports that the clustering pattern exists clearly for the whole investigated distance (up to 15,000 feet). It is beyond the scope of this paper to discuss at what situation a false positive presents and what situation is favored by a false negative. Nevertheless, the case study shows clearly that it is necessary to put the analysis of activity patterns that are distributed along streets (e.g. auto thefts) into the framework of street network space. There is a potential danger of false conclusion if techniques for Euclidean space analysis are adopted directly (e.g. applying planar *k*-function on a street network distribution analysis).

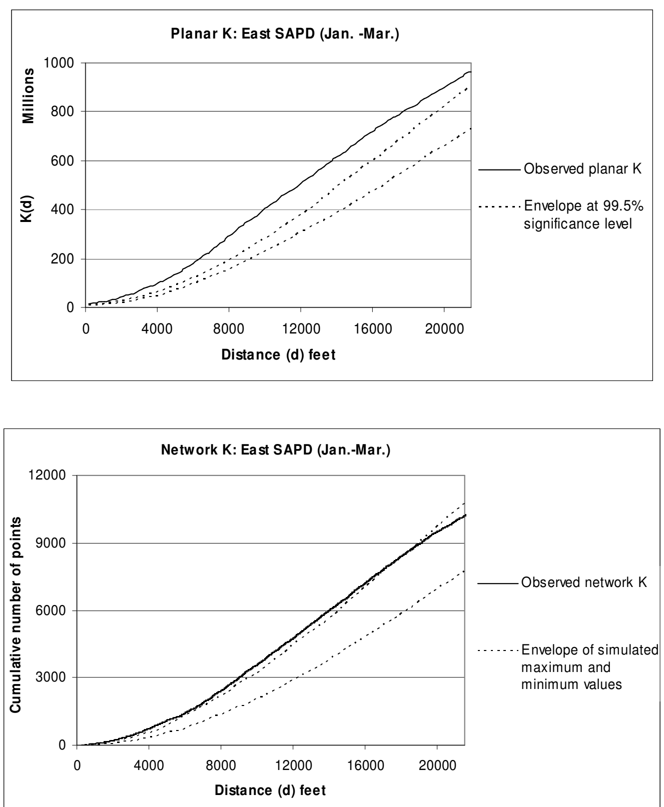
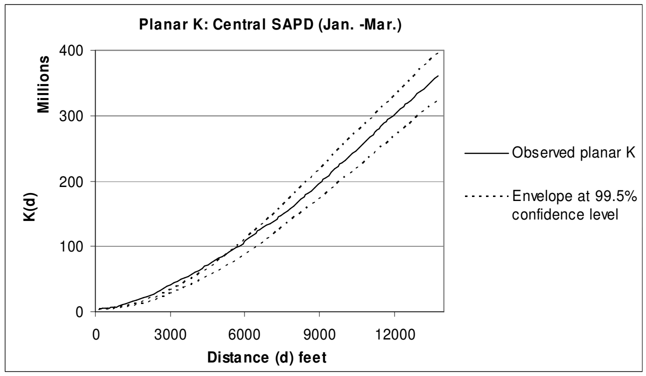


Figure 2. Planar and network *k*-function analyses of auto thefts in East Police District



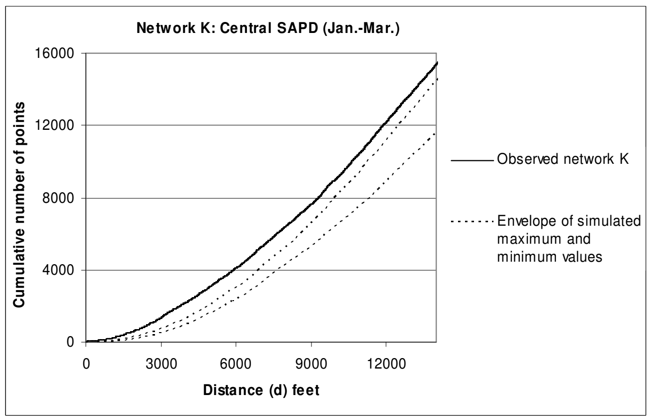


Figure 3. Planar and network *k*-function analyses of auto thefts in Central Police District

A zoning approach in street network space for the analysis of auto thefts in Buffalo

The auto theft data for 1998 in the city of Buffalo was obtained from the Buffalo Police Department. According to the police records, there were a total of 3,271 auto thefts in Buffalo in 1998, out of which 2,284 stolen vehicles were recovered. Using the case ID for each reported offense as key filed, a subset of the auto theft cases were matched with corresponding vehicle recoveries. Further, after geocoding the theft and recovery locations, a total of 1600 auto theft and recovery location pairs were matched and mapped. Readers can refer to Lu (2003) for details of data processing. This part of analysis demonstrates a street network zoning approach to the clustering study of the1600 auto theft locations.

Couple points need to be addressed before further discussion of the approach and related findings. Firstly, this example of auto theft clustering analysis aims at identifying the locations of clusters. In another word, different from the previous example and other existing studies that stayed at the level of assessing the existence of point clustering (in either street network space or grid space) (e.g. Okabe et al. 1995; Okabe and Yamada 2001; Levine 2002; Yamada and Thill 2004), this is a local clustering analysis designed to reveal the locality of auto theft clusters. Secondly, this approach takes into consideration the distribution restriction of street network – locations of activities are all distributed along streets and their patterns are compared with expected location distributions restricted to streets as well. Hence, the clusters of auto thefts will show a linear pattern rather than the commonly identified circle or other convex polygons (e.g. Craglia *et al.* 2000; Fotheringham and Zhan 1996; DeLima and Lu 2004). Put in another way, the finally identified local clusters of auto thefts are expected to show a pattern of linear "hot street" rather than the commonly adopted round "hot spot". Considering the locating and moving nature of vehicles, a linear clustering pattern might be closer to the reality, and is easier for police officers to design and adjust their patrol routes.

The first challenge is to build a zonal system in the street network using zoning function as described in equation 6. Since the city of Buffalo is a relatively big and geometrically compact city with matured urban grid-style streets dominating its transportation network, it is assumed for this analysis that the connection between locations within the same street segment is much smoother than that between locations in different street segments. Hence, each street segment is treated as a nature zone. Traffic controls at street intersections are believed to be the major deterrence for travels along streets and thus form the boundary for zones in street network. While we admit that this is an arbitrary decision that might oversimplify the connection situation along streets, we leave it for further investigation to evaluate different ways of defining zonal system along streets, since the focus here is to conduct an exploratory study to extend zonal approach to street network space for location clustering analysis. After cleaning and rebuilding the topology of the street network database, there were a total of 7720 street segments in the city of Buffalo. The 1600 auto thefts were found to be located 1189 street segments, among which 895 segments received only one auto theft. Figure 4 and table 1 reports the distribution of the 1600 auto thefts in the streets of Buffalo.

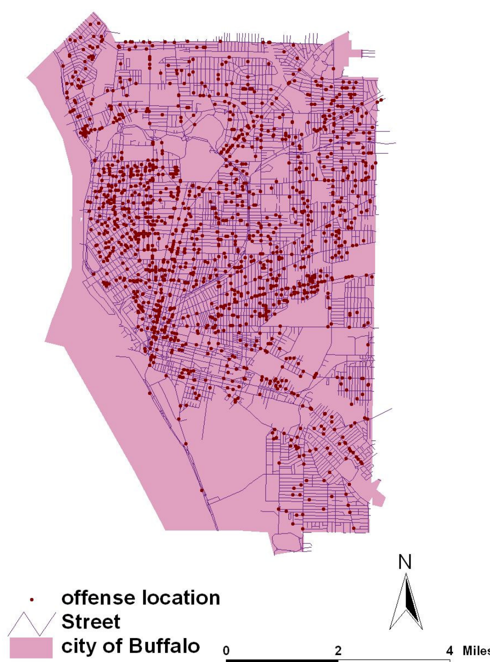


Figure 4. Distribution of the 1600 auto thefts in the city of Buffalo

No. of auto thefts each street zone received	0	1	2	3	4	5	13
No. of street zones	6537	895	218	49	21	5	1

Table 1. Distribution of the 1600 auto thefts in streets of Buffalo

The next step is to evaluate the concentration of auto thefts in each zone, i.e. to assess the clustering of events in each street segment. The observed number of auto thefts in each street segment is compared to the expected number of auto thefts in the same zone. Since auto theft is a type of rare event and does not happen with equal likelihood everywhere in streets. However, due to limited knowledge about the distribution of offense opportunities and acknowledging that it is not a major concern for this analysis to describe the opportunity patterns, the theoretical distribution of auto thefts is modeled as complete spatial random pattern. The probability for a street segment with a length of l o receive n auto thefts can be described as

$$p_l(n) = \frac{\exp(-f) * f^n}{n!} \qquad \text{……………} \quad (7)$$

$$f = \frac{N}{L} * l \qquad \text{…………….... } \quad (8)$$

Where L is the total length of the street network, l is the length of the street segment under concern, N is the total number of auto thefts in the distribution, and n is the number of auto thefts to be expected in l. $P_l(n)$ denotes the probability of observing n auto thefts in a street segment with length of l, and it is a Poisson distribution (Getis and Boots 1978; Beseg and Newell 1991).

For this analysis, a street segment i is claimed to have clustered auto thefts in it if the probability of observing less than n_i auto thefts is 99%, where n_i is the number of auto thefts observed in i. The street segment is then named “hot street” for auto theft, corresponding to the commonly used term, “hot spot”. Figure 5 shows the analysis results. The highlighted streets segments are those zones in the Buffalo street network that received significantly higher number of auto thefts than they are theoretically expected to. For crime control and analysis purpose, these are the streets that warrant further examination and frequent police patrol.

Compared to applying Euclidean plane as the distribution space for auto thefts, this approach takes into consideration the locating restriction imposed by street network. Hence, it can more accurately reflect the spatial distribution of auto thefts and can capture the true clustering pattern and locality. Furthermore, for any analyses of auto thefts using Euclidean plane, the results would show polygon areas of hot spots. The hot spots might cover offenses from more than one street while the true street distance between offenses in the hot spot might be far. Plus, one hot spot might cover multiple streets depending on analysis and visualization scales, even though their connection is actually much more complicated than appeared in Euclidean plane. This might confuse the interpretation of the results and puzzle the police when deciding which streets should be patrolled more. A better control of crime is related to the ability to allocate resources efficiently based on a good understanding of the connection between crime locations and their concentration along streets.

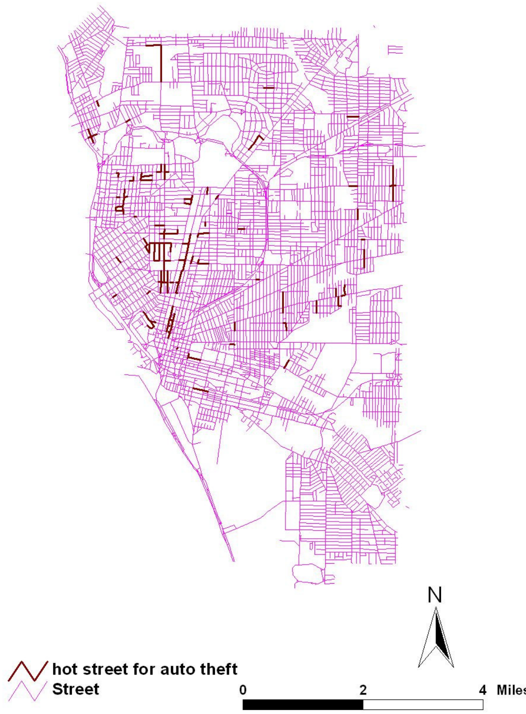


Figure 5. Auto theft hot streets identified using zoning approach in street network space

CONCLUSION

This paper is designed to promote the investigation of activity patterns along streets in a street network space. Many social and economic activities, especially those in urban area are restricted to the spatial layout of urban street framework. It is risky to simple ignore the distribution limits imposed by street network to the locations of these activities. For both researchers and practitioners, it is important to fully comprehend the patterns of urban activities when a street network is used as a distribution background. For the former, a better understanding of the patterns in an urban street network can help to build more valid hypothesis regarding how the observed location patterns are formed and sustained; for the latter, an accurate description of the patterns under investigation should help with resource allocation (e.g. for police to allocate manpower) and business planning and operation (e.g. for retailers to identify service area, as discussed by Okabe and Okunuki (2001)).

To extend the spatial statistics techniques from Euclidean plane to street network space, it is essential to accommodate two groups of properties – geometry and topology measurement and representation. The appropriate description of these properties has significant implication for the measurement of spatial separation between locations. Correctly measuring spatial separation is essential for the evaluation of clustering among point locations, no matter a distance or a zonal approach is adopted. The two empirical examples reported above show how the extension of spatial statistics from Euclidean plane to street network can be implemented by either distance approach or zoning approach for both global and local statistics. These applications are also good examples to demonstrate the necessity for accounting for the distribution restriction imposed by street network to certain street activities. However, they are mainly reported here as exploratory approach on the journey of moving-to-street-network-space for analysis of activity patterns along streets. Major challenges along this journey are related but not limited to the computational aspect and topological modeling of the complexity of street network.

ACKNOWLEDGEMENT

The author is grateful to Jean-Claude Thill for related discussion regarding street segment and order of street segments. Xuwei Chen's assistance with SANET for the calculation of network *k*-functions is greatly appreciated. The author also wants to thank San Antonio Police Department and Buffalo Police Department for providing auto theft data. All views and opinions remain those of the author.

REFERENCES

Anselin, L., 1996, The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In: Fisher, M., H. Scholten, and D. Unwin (eds.) *Spatial Analytical Perspectives on GIS*. London: Taylor & Francis. Pp. 111-125.

Bailey, T. C., and A. C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Harlow, UK: Longman Scientific and Technical.

Besag, J., and J. Newell. 1991. The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A* 154, Part 1, 143–55.

Boots, B. and A. Getis, A. 1988. *Point Pattern Analysis*. Newberry Park, CA: Sage Publication.

Copes, H. 1999. Routine Activities and Motor Vehicle Theft: A Crime Specific Approach *Journal of Crime and Justice* 22(2): 125-146.

Craglia, M., Hanning, R. & Wiles, P. 2000. A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies* 37(4): 711-729.

DeLima, I. and Yongmei Lu. 2003. Investigating the Spatial Relationship between Drug Possession Arrests and Residential Burglaries in Austin. *Southwestern Geographer* 7: 1-25.

Diggle, P. J. 1983. *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.

Eugene F. 1975. *Taxicab Geometry*. Philippines: Addison-Wesley Publishing Company, Inc.

Fabrikant, Sara I., Daniel R. Montello, Marco Ruocco, and Richard S. Middleton. 2004. The distance-similarity metaphor in Network-display spatializations. *Cartography and Geographic Information Science* 31(4): 237-252.

Fotheringham, A. S., and F. B. Zhan. 1996. A Comparison of Three Exploratory Methods for Cluster Detection in Spatial Point Patterns. *Geographical Analysis* 28: 200–218.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis.* Thousand Oaks, CA: Sage Publications.

Getis, A., and B. Boots. 1978. *Models of Spatial Processes. An Approach to the Study of Point, Line, and Area Patterns*. Cambridge: Cambridge University Press.

Groff, Elizabeth, Monica Alexander. 1998. A Comparison of Using Radial Versus Street Distance for Characterizing Crime around a Given Location. *Paper presented at the 1998 ESRI International User Conference*, July 27-31. San Diego, CA.

Horner, Mark, and Allan Murray. 2004. Spatial representation and scale impacts in transit service assessment. *Environment and Planning B: Planning and Design* 31(5):785-798.

Jiang B. and Claramunt C. 2004. Topological Analysis of Urban Street Networks. *Environment and Planning B: Planning and Design* 31(1): 151- 162.

Kulldorf, M. 1998. Statistical Methods for Spatial Epidemiology: Tests for Randomness. In: Loytonen M. and A. Gatrell (eds), *GIS and Health*. London: Taylor & Francis. Pp:49-62.

Levine, Ned, 2002, *CrimeStat: 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 D.C.

Lu, Yongmei. 2003. Getting Away with the Stolen Vehicle: an Investigation of Journey-after crime. *The Professional Geographer* 55(4): 422-433.

Lu, Yongmei. and Jean-Claude Thill. 2003. Assessing the cluster correspondence between paired point locations. *Geographical Analysis* 35(4), 290-309.

Lu, Yongmei. and Xuwei Chen. 2004. False alarm of planar k-function for analyzing crime distributed along streets. *Manuscript available from the author. Department of Geography, Texas State University, San Marcos, TX 78666. (yongmeilu@txstate.edu).*

Meliskova, Jana. 2000. Integration of Geography and Statistics. *Statistics Journal of the United Nations ECE* 17(2000):189-193.

Okabe, A., H. Yomono, and M. Kitamura. 1995. Statistical analysis of the distribution of points on a network. *Geographical Analysis* 27(2): 152-75.

Okabe, Atsuyuki, and Kei-ichi Okunuki. 2001. A Computational method for estimating the demand of retail stores on a street network and its implementation in GIS. *Transactions in GIS* 5(3): 209-220.

Okabe, A., and I. Yamada. 2001. The K-function method on a network and its computational implementation *Geographical Analysis* 33(3): 271-90.

Rossmo, Kim. 2000. *Geographic profiling*. New York: CRC Press.

Rushton, G., and P. Lolonis. 1996. Exploratory Spatial Analysis of Birth Defect Rates in an Urban Population. *Statistics in Medicine* 7:717-726.

Worboys, Michael, Keith Mason, and Jenny Lingham. 1997. Computational techniques for non-Euclidean planar spatial data applied to migrant flows. In: Carver, Steve (eds.) *Innovations in GIS 5: Selected Papers from the Fifth National Conference on GIS Research UK (GISRUK).* London, UK: Taylor & Francis.

Yamada, Ikuho, and Jean-Claude Thill, 2004. Comparison of Planar and Network K-functions in Traffic Accident Analysis. *Journal of Tansport Geography* 12(2): 149-158.