CSIS Discussion Paper No.59

Cell Count Method on a Network with SANET

Atsuyuki Okabe* and Shino Shiode**
Center for Spatial Information Science, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
*****atsu@ua.t.u-tokyo.ac.jp
******shino@ua.t.u-tokyo.ac.jp

Aug, 2003

Abstract

This paper first proposes a cell count method defined on a network, called the network cell count method, as an extension of the ordinary cell count method defined on a plane. Second, the paper develops a user-friendly tool for the network cell count method in conjunction with a general toolbox for spatial analysis on a network, called SANET. Third, the paper shows an actual application of the network cell count method to the distribution of retail stores in Shibuya, Tokyo.

Key words: cell count method, network, toolbox, software, SANET, GIS

1. Introduction

This paper has two objectives. The first objective of this paper is to propose a new cell count method, called the *network cell count method*, as an extension of the ordinary cell count method (or the quadrat method). The second objective is to develop a user-friendly tool for the network cell count method, and incorporate it into a toolbox called SANET, a general GIS software package for spatial analysis on a network.

The cell count method is one of the most classic methods for point pattern analysis (Bailey and Gatrell, 1995; Cressie, 1993; Upton and Fingleton, 1985). This cell count method ordinarily makes the following assumptions.

Assumption i: Geographical space is represented by a plane.

Assumption ii: The plane is homogeneous, i.e., the probability of a point being placed on a unit area is invariant regardless of the location of the unit area on the plane.

There may be some actual cases in which these assumptions hold or approximately hold. For example, (Matui, 1932) examined the 'scattered village' (houses are scattered) on the Tonami plain, which appears to be a fairly homogeneous plane. However, when we examine the distribution of features in an urbanized area, the above assumptions are hard to accept. For example, consider the distribution of beauty parlors in Shibuya, one of the sub-centers in Tokyo, which is shown in Figure 1. As is seen in this figure, the beauty parlors are located along streets, and the distribution of beauty parlors seems to be strongly constrained by the street network. This suggests us to replace Assumption i with the following assumption.

Assumption i': Geographical space is represented by a network.

As a consequence of this assumption, we assume a heterogeneous space in the sense that the space is not isotropic. As a matter of fact, direction is restricted along a path of a network. However, we assume a homogeneous space in the following sense (i.e. a network version of Assumption ii).

Assumption ii': The network is homogeneous in the sense that the probability of a point being placed on a unit line segment is invariant regardless of the location of the unit line segment on the network.

Obviously this assumption is unrealistic to deal with the distribution of features in an urbanized area, but this assumption is easily relaxed in a network space through the 'uniform network transformation'. This transformation is the topic of the next section.

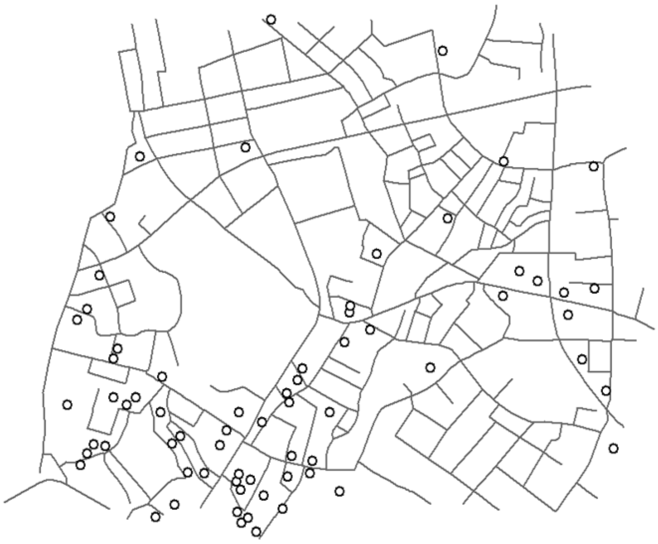

Figure 1: Beauty parlors in Shibuya, Tokyo

2. Uniform network transformation

A heterogeneous network, N, is represented by a density function defined on links of the network. To be explicit, let D_i be the density of the *i*th link of the network, and L_i be the length of the *i*th link. Then the heterogeneous network means that D_i varies from link to link; the homogeneous network (Assumption ii') is written as

$$D_i = c \text{ for } i = 1, \ldots, n .$$

Note that the location of a point, *p*, on the *i*th link can be identified by the path length, *t*, from an end point of the *i*th link to the point *p* along the *i*th link.

Let us now consider a new network, N^*, by replacing the *i*th link with the new *i*th link and we correspond a point *p* at *t* on the *i*th link to the point *p'* at *s* on the new *i*th link in such a way that the path length, *s*, from an end point of the new *i*th link to the point *p'* along the new *i*th link is given by

$$s = \int_0^t D_i dx.$$

Note that the integration is done with respect to x along the ith link. This transformation implies that the ith link of length L_i is replaced with the new ith link of length $D_i L_i$.

The new network obtained from the above transformation has a nice property. Noticing that the total quantity on the ith link is given by

$$\int_0^{L_i} D_i dx = D_i L_i ,$$

and that the length of the new ith link is $D_i L_i$, we obtain the density of the new ith link is $D_i L_i /(D_i L_i) = 1$. This equation hold for all links. This means that the transformed network N* is homogeneous as in Assumption ii'. We refer to the above transformation as the *uniform network transformation* (Okabe, 2002).

We can easily deal with a heterogeneous network by transforming it into a homogeneous network through the uniform network transformation.

3. Cell count method

Let us consider a network, L, consisting of line segments that are connected. We decompose the network L into a set of sub-networks (called *cells*) satisfying that the length of each sub-network is the same, c, and each sub-network is connected. In practice, however, this decomposition is impossible except for very special cases, because the total length divided by c is not integer in general.

Noticing this fact, we consider two types of cells: *proper cells,* $L^{PC} = \{L_1, \ldots, L_n\}$, and *improper cells*, $L^{IC} = \{L_{n+1}, \ldots, L_{n+n^*}\}$. The former cells are used for cell counting, but the latter cells are peripheral cells not used for cell counting. We suppose that these cells satisfy the following conditions.

(i) L_i is connected in the sense that for any pair of points on L_i, there exists a path between these points that is included in L_i $(i = 1, \ldots, n + n^*)$.

(ii) Proper cells $L_1, \ldots, L_n$ have the same size, i.e.

$$|L_i| = c \quad \text{for} \quad i = 1, \ldots, n,$$

where $|L_i|$ denotes the length of L_i.

(iii) The size of the improper cells $L_{n+1}, \ldots, L_{n+n^*}$ is smaller than that of the proper cells, i.e.

$$|L_i| < c \quad \text{for} \quad i = n+1, \ldots, n+n^*.$$

(iv) The union of the proper cells and the improper cells cover the whole network L, i.e.

$$L = L^{PC} \cup L^{IC} = [\bigcup_{i=1}^{n} L_i] \cup [\bigcup_{i=n+1}^{n+n^*} L_i].$$

(v) Cells do not intersect each other except at boundary points, i.e.

$$|L_i \cap L_j| = 0 \quad \text{for} \quad i \neq j, i, j = 1, \ldots, n+n^*.$$

We make five remarks on this decomposition. First, this decomposition is not unique. There are many ways of decomposition satisfying the above conditions. Second, the shape of cells is not the same, because a given network L is not regular. This contrasts to the planar case where the shape is the same (usually squares) for all cells. Third, it is very difficult to control the shape of cells, but the shape should be as 'compact' as possible. Forth, the length of the improper cells should be as small as possible. Last, improper cells should not appear inside a network. In the planar case, improper cells appear only on the periphery, but in the network case, improper cells may appear inside a network. We call such cells *hole cells*. It is not easy to avoid hole cells in decomposing a network, and we shall discuss this problem in Section 5.

The idea of the network cell count method is the same as that of the planar cell count method. Suppose that we observe m points distributed on the network $L^{PC} = [\bigcup_{i=1}^{n} L_i]$, and we want to test the null hypothesis that the m points are randomly distributed on the network $L^{PC} = [\bigcup_{i=1}^{n} L_i]$ according to the uniform distribution,

$$f(x) = \frac{1}{|L^{PC}|}, \quad x \in L^{PC}.$$

We can test this hypothesis using the goodness-of-fit test. To be explicit, let $\overline{N}_j$ be observed number of cells that contains exactly j points, except for $\overline{N}_k$. $\overline{N}_k$ is the observed number of cells that contain k or more points. Let P_j be the probability that j points are placed in a cell under the above null hypothesis, which is given by the binominal distribution or the Poisson distribution for a large number of points. Thus we can test the null hypothesis with the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(nP_i - \overline{N}_i)^2}{nP_i}.$$

4. SANET

We develop a user-friendly tool for achieving the network cell count method defined above and include it in a general toolbox for <u>s</u>patial <u>a</u>nalysis on a <u>net</u>work, called SANET (Okabe, Okunuki and Funamoto, 2002). SANET consists of two components. The first component is a software package for computing methods for spatial analysis on a network. We can interface this package with a viewer of GIS through input and output files. The computation in this software package is independent of a viewer of GIS, and so we can use any viewer. The second component is an interface with a viewer of GIS, and this interface depends on a viewer of GIS. We use ArcView8.x., and we have developed the interface that commute data between the network computation software package and ArcView8.x.

SANET is under development but the first version was released in November, 2002. The first version provides the following seven tools:

Tool 1: Construction of dataset for SANET.
Tool 2: Access point assignment.

Tool 3: Table calculation.

Tool 4: Generation of the network Voronoi diagram (Okabe, Boots, Sugihara and Chiu, 2000).

Tool5: Generation of random points on a network (for Monte Carlo simulations).

Tool 6: Cross *K*-function method (Okabe and Yamada, 2002).

Tool 7: *K*-function method (Okabe and Yamada, 2002; Yamada and Thill, 2002).

SANET is open to non-profit users without charge and it can be downloaded from the SANET site: http://okabe.t.u-tokyo.ac.jp/okabelab/atsu/sanet/sanet-index.html

This paper adds the cell count method to SANET as the eighth tool. The cell count method uses Tool 1 for converting the network data format in the network computation software package to data format of ArcView8.x. It also uses Tool 2 for assigning the representative points of polygon-like features, such as beauty parlors, to the nearest points on a network (which may be regarded as entrances of the facilities). By this assignment, we can use the cell count method that assumes that points are distributed over the network.

5. Computational method for the cell count method

The computational method is fairly complex and technical and so we only outline the algorithm in this section.

In the first step, we choose a node of a network around the center of the network. We call the node the *root node*. In the second step, we construct the "extended" shortest-path tree rooted at the root node. The "extended" shortest-path tree is obtained from the ordinary shortest-path tree in the following manner.

The ordinary shortest-path tree does not cover the whole network. We call the links that are not included in the ordinary shortest-path tree *collision links*. On each collision link, we can find such a point, c_i, that the shortest-path distance from the point to the root node through one end node of the collision link is the same as the shortest-path distance from the point c_i to the root node through the other end node of the collision node. We call the point c_i a *collision point*. We cut the network at all collision points and add nodes on both cut ends. We call the resulting tree the *extended shortest-path tree*. Once we obtain the shortest-path tree, we can easily obtain the shortest-path from any point on the network to the root node.

In the third step, we construct a tree rooted at the root node whose total length is a given distance. We next construct a tree rooted at each end node of the tree whose total length is a given distance. If we cannot construct such a tree, we leave it. We continue this procedure until when we cannot construct trees. Then we obtain a set of trees that are included in the network and the length of each tree is the same. We may use the resulting trees as cells.

However, these cells are likely to produce the holes cells in the network. The last step is to put out these hole cells to the periphery of the network. This part is too lengthy to explain, and so it is omit here.

6. An Application

Having established the network cell count method and its software package, let us show an actual example. The study region is part of Shibuya, one of the sub-centers in Tokyo, and the street network is shown by colored lines in Figure 2.

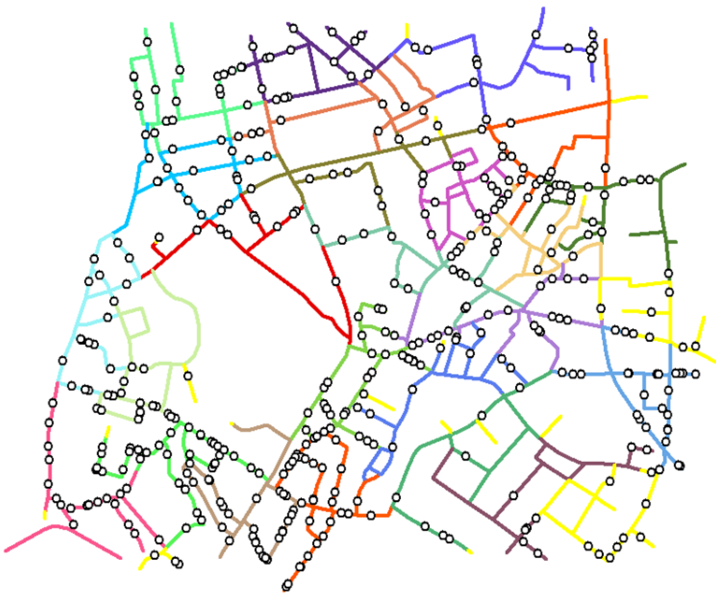

Figure 2: Retail stores assigned to the street network in Shibuya, Tokyo (cells are indicated by different colors)

First we decompose the street network into cells. In Figure 2, the cells are indicated by different colors. The yellow lines are improper cells.

Second, we assign retail stores to the nearest points on the network (which may be regarded as the entrances of the stores). The points in Figure 2 show the assigned retail stores on the street network.

Third, the tool counts the number of retail stores in each cell and shows the numbers of cells having *i* points as in Figure 3.

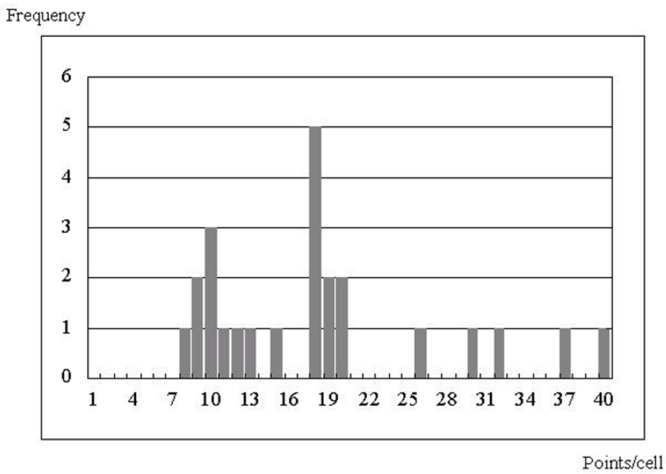

Figure 3: The number of cells having *i* points

Last we apply the goodness-of-fit test to the data shown in Figure 4. The test rejects the null hypothesis, implying that the retail stores are non-randomly distributed in Shibuya.

7. Concluding remarks

In this paper we have proposed the network cell count method as an extension of the ordinary cell count method defined on a plane and developed a tool for the network cell count method.

Compared with the ordinary cell count method, the proposed network cell count method has a few advantages. First, the network space is more natural to deal with the distribution of features in an urbanized area than a homogeneous plane. Second, it is easy to treat heterogeneity with the uniform network transformation. For example, it is easy to treat the distribution of stores in relation to the distribution of population.

One might consider that the results obtained from the ordinary cell count method are not so different from those obtained from the network cell count method. We compared these results in a few actual examples and found that this difference was significant. We will further examine this comparison using many cases and we will report the results in another occasion.

Acknowledgements

We express our thanks to Yasushi Asami, Yukio Sadahiro and Kei-ichi Okunuki for their comments on earlier drafts. This study was part of the outcomes of the Grant-in-aid project "Spatial Information Science for Human and Social Sciences" supported by the Japanese Ministry of Education and Science.

References

Bailey, T. and Gatrell, A. (1995) *Interactive Spatial Analysis*, London: Longman.

Cressie, N. (1993) *Statistics for Spatial Data*, New York: Wiley.

Matui, I. (1932) "Statistical study of the distribution of scattered villages in two regions of the Tonami plain, Toyama prefecture", *Japanese Journal of Geology and Geography*, 9, 251-266.

Okabe, A. (2002) "Uniform network transformation for spatial analysis on a heterogeneous network", *CSIS Discussion Paper* #51.

Okabe, A., Okunuki, K. and Funamoto, S. (2002) *SANET: A Toolbox for Spatial Analysis on a Network*, Center for Spatial Information Science, University of Tokyo.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N. (2000) *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, (2nd edition), Chichester: John Wiley.

Okabe, A. and Yamada, I. (2001) "The *K*-function method on a network and its computational implementation", *Geographical Analysis*, Vol.33, No.3, pp.271-290.

Upton, G. and Fingleton, B. (1985) *Spatial Data Analysis by Example*, Chichester: Wiley

Yamada, I. and Thill, J-C. (2002) "An Empirical Comparison of Planar and Network K-function Analyses", *Abstract*, AAG Annual Meeting, Los Angeles, CA.