空間重み行列と都市空間構造分析への応用

<table>
<thead>
<tr>
<th>著者</th>
<th>ZHANG Changping</th>
</tr>
</thead>
<tbody>
<tr>
<td>雑誌名</td>
<td>国際地域学研究</td>
</tr>
<tr>
<td>巻</td>
<td>□</td>
</tr>
<tr>
<td>号</td>
<td>□</td>
</tr>
<tr>
<td>ページ</td>
<td>□</td>
</tr>
<tr>
<td>発行年</td>
<td>2012年</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://id.nii.ac.jp/1060/00003664/">http://id.nii.ac.jp/1060/00003664/</a></td>
</tr>
</tbody>
</table>
Spatial Weights Matrix and its Application

Changping ZHANG

1 Introduction

A spatial weights matrix $W$ what is indicates the strength of the potential interaction between spatial areal units. The choice of $W$ is a critical step in spatial analysis and can influence the significance of testing procedures (Tiefelsdorf, Griffith, and Boots, 1999). To think about the type of spatial weights matrix $W$, Getis (2009) indicated that at least three ways exist that could be theoretical, topological, and empirical points of view. According to theoretical point of view, the $W$ matrix is exogenous to any system, and should be based on a preconceived matrix structure. Usually this structure is based on a theory of distance decay, however defined. The theory of distance may come from notions of spatial interaction and gravity models. The topological viewpoint arose from a need to depict the actual configuration of the areal units contained within a study region. For example, a long, narrow areal unit would be represented in differently than a short, wide unit. In this case the matrix $W$ might be specified in different ways that might be number of neighbors, length of side in common, and proportion of perimeter in common. Cliff and Ord (1969) said, “With a flexible system of weights, the researcher can highlight those features of a study area which he believes to be important”. The empirical viewpoint of approach implies that one detects spatial dependence in variables under study. One must use a local autocorrelation point of view to identify the exact level of spatial autocorrelation surrounding any given observation then creates a $W$ matrix.

Since 1960s, most researchers have made every effort to attempt to create proper dependence representation in spatial weights matrix $W$. When data are in a raster model, $W$ was constructed in the rook’s case or queen’s case definition of neighbors. Using vector data, popular were two types of $W$ to be used. The first is called a simple binary connectivity definition and uses a discrete function where spatial entities are assumed to be adjacent if and only if they share a common boundary. The nature of interaction for the spatial phenomenon under study cannot always be captured by a simple binary proximity measurement. In this case, based on theoretical and conceptual considerations, a generalized weight function can be used as advanced by Cliff and Ord (1981) which offers flexibility in defining spatial proximity. Typical functions incorporate the distance between the geographical centroids of areal units and/or the length of common boundary between areal units (Can, 1996).

Getis and Aldstadt (2004) summarized previous attempts to create a spatial weights matrix and identified many different types of weighting schemes. That are
1. spatially contiguous neighbors;
2. inverse distances raised to some power;
3. length of shared borders divided by the perimeter;
4. bandwidth as the $n$th nearest neighbor distance;
5. ranked distances;
6. constrained weights for an observation equal to some constant;
7. all centroids within distance $d$; and
8. $n$ nearest neighbors, and so on

Some of the newer schemes are:
1. bandwidth distance decay;
2. Gaussian distance decline; and
3. “tri-cube” distance decline function.

Furthermore, Getis and Aldstadt (2004) also proposed a spatial weights matrix based on the $G_i^*$ local statistic (Ord and Getis, 1995) which accounts for the spatial association extant within any region that has been divided into its constituent parts. In a series of simulation experiments the matrix is compared to well-known spatial weights matrix specifications — two different contiguity configurations, three different inverse distance formulations, and three semi-variance models and performed best according to the AIC (Akaike Information Criterion) and the autocorrelation coefficient evaluation.

Zhang and Murayama (2000) suggested a concept and algorithm of $k$-order neighbours based on Delaunay’s triangulated irregular networks (TIN) and redefined Getis & Ord’s (1992) local spatial autocorrelation statistic as $G_i(k)$ with weight coefficient $w_i(k)$ based on $k$-order neighbours for the study of local patterns in spatial attributes.

Although the choice of a spatial weights matrix specification for spatial statistical analysis is not clear-cut and seems to be governed primarily by convenience or convention, Griffith (1996) proposed some explicit guidelines on specification of the spatial weights matrix and concluded that relative large numbers of areal units should be employed ($n>60$) in a spatial statistical analysis and low-order spatial models should be given preference over higher-order ones.

2 Definitions of spatial weights matrix

In this section, four types of weight coefficient are defined and the spatial weight matrices formed by these coefficients are created for a simple hypothetical region.
2.1 Binary weight

If we are studying spatial interrelationships distributed over a set of areal units, the spatial structure of units might be defined as the spatial contiguity which is treated as a $n \times n$ spatial weights matrix $W$ with binary variable (Cliff & Ord, 1981; Zhang and Murayama, 2003). The binary weight is represented as follows:

$$w_{ij} = \begin{cases} 1 & \text{if area } i \text{ and } j \text{ are contiguous;} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

The binary spatial weights matrix $W$ is a symmetric form. If the $W$ is row-standardized:

$$w_{ij}^* = \frac{w_{ij}}{\sum_{j \in J} w_{ij}}$$

where $J$ is the set which includes all areal units that are contiguous with $i$. The $w_{ij}^*$ is satisfying non-negative $w_{ij}^* \geq 0$ and $\sum_{j \in J} w_{ij}^* = 1$ (for all $i = 1, \cdots, n$). The binary weight is often used in spatial autocorrelation analysis with regular and irregular areal units. A hypothetical example will clarify the use of this index. Imagine a region that is divided into four units (Figure 1), unit 1 and 2, 1 and 3, 1 and 4, 2 and 3 are contiguous.

![Figure 1. A hypothetical region](image)

The binary spatial weights matrix and its standardized form of figure 1 are

$$W_b = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad W_b^* = \begin{pmatrix} 0 & 0.3333 & 0.3333 & 0.3333 \\ 0.5 & 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$
2.2 Distance decay weight

Tobler (1970) has referred to as the “first law of geography: everything is related to everything else, but near things are more related than distant things.” In researches on the interrelation and interaction, the phenomenon in which the contiguity or interaction will decrease with increasing distance between two areal units is called as distance decay. The distance decay weight is usually expressed as a reciprocal of distance.

\[ w_{ij} = \frac{1}{d_{ij}} \]  

(2)

where \( d_{ij} \) is the distance from the center of units \( i \) to the neighboring units \( j \). Generally the distance can be straight-line distance, road distance, time distance, cost distance, mental distance, or some other way. The distance decay spatial weights matrix and its standardized form of figure 1 are

\[
\begin{bmatrix}
0 & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\
\frac{2}{\sqrt{5}} & 0 & \frac{2}{3} & \frac{2}{}\sqrt{29} \\
\frac{1}{\sqrt{2}} & \frac{2}{3} & 0 & \frac{1}{\sqrt{2}} \\
\frac{2}{\sqrt{29}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\
\end{bmatrix} \quad \begin{bmatrix}
0 & 0.4256 & 0.3365 & 0.2379 \\
0.4256 & 0 & 0.3450 & 0.1923 \\
0.3365 & 0.3450 & 0 & 0.3398 \\
0.2379 & 0.1923 & 0.3398 & 0 \\
\end{bmatrix}
\]

2.3 Generalized weight

A generalized weight definition based on distance and length of boundary between areal units. In this case, the elements of spatial weights matrix \( W \) are:

\[ w_{ij} = \frac{l_{ij}}{d_j \sum_{j \in J} l_{ij}} \quad (i = 1, \ldots, n) \]

(3)

where \( J \) is the set which includes all units that are contiguous with unit \( i \). \( l_{ij} \) is the length of join between units \( i \) and \( j \). \( \sum_{j \in J} l_{ij} \) is the length of all joins for unit \( i \), i.e. its perimeter; and \( d_{ij} \) is the distance from the center of units \( i \) to the neighboring units \( j \). \( \sum_{j \in J} l_{ij} \) is simply the proportion of the perimeter of unit \( i \) which is contiguous with unit \( j \). This weighting function assumes that the interaction between two units will increase with increasing length of common boundary and will decrease with increasing distance between their geographical centers. The generalized spatial weights matrix and its standardized form of figure 1 are
2.4 k-order neighbors weight

Zhang and Murayama (2000) proposed a concept of $k$-order neighbors based on Delaunay's triangulated irregular network (TIN) with centroids of areal units and a definition of $k$-order neighbors weight. For example, figure 2(B) shows the result of drawing a TIN based on the point distribution in figure 2(A). The $k$-order neighbors in the Delaunay triangulation network are defined as follows. Those points directly connected to a point $v$ by the edges of the TIN are called the "nearest neighbors", or "1-order neighbors" of this point (figure 2 (C)). Then, those points which are directly connected to the first-order neighbors, of that are not 1-order neighbors themselves, are called "2-order neighbors" (figure 2 (D)). By continuing this process we can define $k$-order neighbors for any point, $v$.

The weight coefficient $w_{ij}(k)$, which measures proximity of $k$-order neighbors in a point distribution, is determined as follows. $w_{ij}(k)$ is binary coefficient set to 1 if point $j$ is a $k$-order neighbour of point $i$, otherwise, it is 0. Hence,

$$w_{ij}(k) = \begin{cases} 1 & \text{point } j \text{ is a } k \text{- order neighbour of point } i; \\ 0 & \text{otherwise.} \end{cases}$$

The defined spatial weight matrix ($w_{ij}(k)$) is a symmetrical binary matrix.
Figure 2. $k$-order neighbors based on Delaunay triangulation. (A) point distribution, (B) Delaunay triangulation, (C) nearest neighbors, (D) 2-order neighbors.

The 1-order neighbors spatial weights matrix and its standardized form of figure 1 are

$$W_k = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \quad W_k^* = \begin{pmatrix} 0 & 0.3333 & 0.3333 & 0.3333 \\ 0.5 & 0 & 0.5 & 0 \\ 0.3333 & 0.3333 & 0 & 0.3333 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

### 3 Prominence of areal units

In spatial statistical analysis of geographical phenomena, a region or a city under study might be divided into some small areal units such as a regular square tessellation or administrate units emerged in irregular shape and have different spatial characteristics. If we are using GIS to support the analysis, irregular areal unit such as cho in Japan is usually represented as one such polygon with geometric attributes. In spatial statistics, an areal unit that has special geometric attributes and maintains significant spatial correlation and spatial interaction close to adjacent units is called a prominent areal unit or important areal unit. The prominence of areal units can be measured by a prominence or influence-centrality index which is obtained by using eigenfunction or Markov chains method from a spatial weights matrix (Tinkler, 1972; Griffith and Jones, 1980; Boots, 1982; Bavaud, 1998, Zhang and Murayama, 2003).
In this Study, we will use four types of weight function definition to create different measures of prominent areal units and use them to analyze the urban spatial pattern in Matsudo City, Chiba Prefecture.

### 3.1 Eigenfunctions method

In several geographical literatures it is suggested that the principal eigenvector can be used to measure the prominence of irregular areal units (Tinkler, 1972; Griffith and Jones, 1980; Boots, 1982). The eigenfunctions of spatial weights matrix is

\[
\phi(\lambda) = \det(\lambda I - W) = \sum_{i=0}^{n} a_i \lambda^{n-i} = 0
\]

The eigenvalues are the roots of \( W \). Since spatial weights matrix \( W \) is real, the \( \lambda_i \) are real and may be ordered as \( \lambda_1 \geq \cdots \geq \lambda_n \). Corresponding to each \( \lambda_i \) is an eigenvector \( v_i \), which satisfies

\[
Wv_i = \lambda_i v_i \quad (i = 1, 2, \ldots, n)
\]

The elements of \( v_i \) corresponding the largest eigenvalue \( \lambda_1 \) provide a measure of the relative position of each areal units since the magnitude of the element is related to the centrality or prominence. Generally, a unit located in central region of a city possesses a larger element of \( v_1 \), vice versa a unit with smaller element might be located near the city limit.

### 3.2 Markov Chains method

The standardized spatial weights matrix \( W^* \) with elements of row-standardized variable \( w_{ij}^* \) is identical to the Markov chains transition matrix. Assuming the chain is ergodic \( (w^*)_{ij} > 0 \) for some \( n > 0 \) and all \( i, j \), we can obtain a unique stationary distribution \( p_j > 0, (\sum_j p_j = 1.0) \) solution:

\[
W^T p = p
\]

or

\[
\sum_j p_j w_{ij} = p_j
\]

The vector \( p, p^T = (p_1, \cdots, p_n) \) is an eigenvector of \( W^T \) with a corresponding eigenvalue of 1. Whereas \( w_{ij} \) is a measure of the relative influence of unit \( j \) on unit \( i \), \( p_i \) can be interpreted as the total influence of unit \( i \) on the total region of city. \( p_i \) will be further referred to as a prominence index (Bavaud, 1998). We also call the index as spatial prominence \( p_i \) here because the spatial attributes are necessary to calculate it.

### 3.3 Examples

As mentioned above, different prominences can obtained from different spatial weight matrices. We use the standardized spatial weight matrices to calculate the prominence of units of figure 1. The elements of eigenvector \( p \) of binary spatial weights matrix \( W_b^* \) are
\[ p_1 = 0.375, p_2 = 0.250, p_3 = 0.250, p_4 = 0.125. \]

Areal unit 1 has the largest value of prominence, because it is contiguous to three other units. The prominence of unit 2 is just the same value as unit 3, because they are all contiguous to two units. As the unit 4 is only contiguous to unit 1, it has the least prominence.

The elements of eigenvector \( p \) of distance decay \( W_d^* \) are
\[ p_1 = 0.273, p_2 = 0.251, p_3 = 0.270, p_4 = 0.205. \]

The value of prominence of unit 1 is similar unit 3, because their distances linked to other units are almost equivalent. The prominences of unit 2 and unit 4 are all smaller than unit 1 and unit 3 since they are quite separate from each other. As unit 2 is nearer to unit 1 than unit 4, its prominence is larger than unit 4.

The elements of eigenvector \( p \) of generated \( W_g^* \) are
\[ p_1 = 0.409, p_2 = 0.335, p_3 = 0.188, p_4 = 0.068. \]

Areal unit 1 has the largest value of prominence, because it possesses the largest unit and longest common boundary joined with other units. The prominence of unit 2 is larger than unit 3 since unit 2 has a longer boundary and a shorter distance joined to unit 1 than unit 3. The prominence of unit 4 is smallest within 4 units.

The elements of \( p \) of 1-order neighbors \( W_k^* \) are
\[ p_1 = 0.3, p_2 = 0.2, p_3 = 0.3, p_4 = 0.2. \]

As unit 1 and unit 3 are all directly connected to other 3 units by the edges of the TIN, their prominences are same as 0.3. Similarly, unit 2 and unit 4 are all only connected to other 2 unit, their prominences are smaller than unit 1 and unit 3 as same as 0.2. Therefore, the variation of prominence based on 1-order neighbors weight is relatively smaller and its sensitivity is lower comparing with other types of weight.

As a result of above analysis, the variation of prominence of areal units is depending on the definitions of spatial weights matrix. The binary weight is affected by the topological attributes and the generated weight is impacted by the geometric attributes. The distance decay is completely determined by the distance between tow units. However the influence of geometric attributes is not definitely reflected in 1-order neighbors weight.
4 Application

Matsudo is a midsize city located in north-eastern part of Tokyo metropolitan area. Its administrative division was formed when the municipality was established in 1953. The cho is the fundamental administrative areal unit and the variation of its geometric attributes is very large. For example, the area of the biggest cho is 185 times as large as the smallest, and the longest distance from the center of a cho to the neighboring cho is 22 times as long as the shortest.

The standardized binary, generated, distance decay and 1-order neighbors spatial weight matrices $W_b^*$, $W_d^*$, $W_g^*$, $W_k^*$ can be derived from topological attributes of the polygons of chos and TIN data which are built by GIS e.g. ArcGIS and digital map 2500 constructed by Japanese Geographical Survey Institute (Can, 1996; Zhang, 1999; Zhang and Murayama, 2000). Then the spatial prominences $p_i$ can be calculated according to the equation (7). Figure 3 shows 4 distributions maps of classified prominence of chos in Matsudo city.

We can consider Figure 3(A), which shows that the prominent chos with larger than 0.006 of $p_i$ are concentrated on the triangular region formed by the eastern side along the JR Jouban Train Line, the western side along the JR Musashino Train Line, and the northern and southern sides along the Keisei Subway Line. There are a large number of large and non-compact chos in the region. In contrast, small and compact chos show lower value of prominence and are located on the periphery of Matsudo city.

Figure 3(B) shows a distribution map of the prominence based on distance decay weight. As shown in this figure, chos with larger than 0.004 of $p_i$ tend to be concentrated in central region and the northern region of the city. There are many small chos with relatively regular shape. In reverse, many larger chos with less than 0.004 of $p_i$ are distributed in the southern and western parts of the periphery of the city. The distribution of prominence appears more clustered than other three distributions.
The distribution of prominence based on generated weight of chos is represented in figure 3(C), and it can be seen that chos with larger prominence value are distributed along the main train lines in which a lot of large and non-compact chos are located. Reversely, in the region with smaller prominence value, a lot of small chos with relatively regular shape are located there. Another wide region with smaller prominence value is located at southern sides along the Keisei Subway Line.

As can be seen from distribution of prominence based on 1-order neighbors weight (figure 3(D)), some clustered regions are not formed and chos with different value of prominence are widely distributed in entire
Matsudo city.

5 Conclusion and future prospects

In regional science, an areal unit which has special geometric attributes and keeps significant spatial correlation and spatial interaction closely to adjacent units is called prominent zone. The prominence of irregular units can be measured by prominence index which is a stationary distribution of Markov chains transition matrix identical to a spatial weight matrix.

In this study, to identify the strength of the potential interaction and correlation between spatial units, we reviewed previous researches to create the spatial weights matrix and discussed some explicit guidelines on specification of the spatial weights matrix and many different types of weighting schemes. Then four definitions of spatial weights are showed as binary, distance decay, generalized, and \( k \)-order neighbors weight and prominences are obtained from these weights matrix. Finally we used them to analyze the spatial pattern in Matsudo City. The result of this analysis is shown as though different prominences can get from different weights matrix, generalized weight matrix is more appropriate to measure prominence of units than the distance decay and \( k \)-order.

As we know that the spatial structure of city is determined by not only geometric attributes and topological attributes, but also social and economic thematic attributes of areal units in city, it is necessary to create some definitions and measures for prominent unit and to use them to analyze the urban spatial structure. The definitions propose begin with confirming the relationship between the prominence index and geometric attributes of areas, expand to include all spatial attributes of areas: geometric attributes, topological attributes and thematic attributes. Furthermore, an approach to implement the definitions will be tested and evaluated to analyze the spatial structure of a city.

References


空間重み行列とその都市空間構造分析への応用

張 長 平

アブストラクト：地域分析において、都市や地域を規則な形状の地域メッシュか不規則な形状の行政区域に細分することから出発する。従来、小区域の空間的相互作用、空間的相互関係を表す空間重み行列は、地理的事象の空間的自己相関検定や空間的自己回帰モデルの構築に適用されている。本研究は、空間重み行列の構築に関する従来の研究を概述した上で、不規則な形状の小区域が地域全体のなかでどの程度の「顕著さ」（prominence）を有し、都市内部から顕著な地域をいかに抽出するかを明らかにすることを課題としている。この目的を達成するために、代表的な4つの空間重み行列（2進的、距離減衰、一般化、k階近隣）に基づき区域の顕著度を求め、これらの顕著度指標の有効性を検討する。松戸市の町丁界を事例とする分析の結果、空間重み行列をいかに定義するかによって、顕著度とそれによって抽出された顕著な地域は異なることが判明した。すなわち、幾何属性と位相属性に基づく小区域の顕著度の評価には、一般化空間重み行列の適用が、2進的、距離減衰、k階近隣の空間重み行列より有効であると考えられる。

キーワード：空間重み行列、顕著度、都市空間構造、松戸市