Journal of the Economic Geographical Society of Korea
Vol.11, No.3, 2008(492~507)

An Application of Network Autocorrelation Model Utilizing Nodal Reliability

Youngho Kim*

Abstract : Many classical network analysis methods approach networks in aspatial perspectives. Measuring network reliability and finding critical nodes in particular, the analyses consider only network connection topology ignoring spatial components in the network such as node attributes and edge distances. Using local network autocorrelation measure, this study handles the problem. By quantifying similarity or clustering of individual objects' attributes in space, local autocorrelation measures can indicate significance of individual nodes in a network. As an application, this study analyzed internet backbone networks in the United States using both classical disjoint product method and Getis-Ord local G statistics. In the process, two variables (population size and reliability) were applied as node attributes. The results showed that local network autocorrelation measures could provide local clusters of critical nodes enabling more empirical and realistic analysis particularly when research interests were local network ranges or impacts.

Keywords : network autocorrelation, disjoint product method, reliability, Getis-Ord G statistics, GIS

1. Introduction

It is well known that network reliability is presented by ability of nodes, links, and their stable connectivity (Agrawal and Barlow 1984). In a telecommunication network, where the nodes are interconnected through links, the nodal ability to communicate to each other is considered as the most important factor for evaluating network performance. The nodal ability is commonly presented as reliability in a network and can be considered as a valuable measure to assess the significance of network components such as city nodes or linkages (Barlow and Proschan 1975; Shier 1991). The reliability provides a probability that network components are vulnerable or tolerant to various types of failures. In general, higher reliabilities presents more critical nodes and vice versa. Overall, the measure of reliability can be used to identify critical components in a given

* Reseacher, The Habibie Center, Indonesia

network or any flow system (Agrawal and Barlow 1984; Colbourn 1999). For example, a node with high reliability can be considered as a critical node in a network.

Even though literature uses a measure of node reliability to represent critical node in a given network, this study identifies critical nodes in an alternative way by adopting local spatial (network) autocorrelation measures. Being applied to a network, a term, network autocorrelation measure, is used instead of spatial autocorrelation in this study. This study applies network autocorrelation measures, and the use of network autocorrelation in finding critical nodes has contribution to literature. First, network autocorrelation enables spatial network analyses. Classical methods in finding critical nodes are can be considered as aspatial network analyses because they consider only network connection topology (Black and Thomas 1998). However, network autocorrelation accounts for spatial components, such as location attributes and edge distance, in finding critical nodes. Many real-world networks such as electricity, internet, communication, and transportation networks have various quantifiable node attributes (e.g. size of population) and edge distances. Network autocorrelation measures can take these node and edge attributes into account enabling realistic spatial analyses.

Second, the use of local network autocorrelation measures enables to find individual critical nodes, which can make significant local impact on a network. The network autocorrelation concerns the influence on values associated with a node on other nodes, which are inter-connected. Network autocorrelation (as a subcategory of spatial autocorrelation) can be divided into global and local autocorrelation. Global autocorrelation measures include Moran's I (Moran 1948) and its variants, which provide a single value of the overall spatial patterns of a study area. Local autocorrelation measures quantify similarity or clustering of individual spatial objects in study area (for detail, see Anselin 1995; Fotheringham and Brunsdon 1999). Naturally, local autocorrelation values point out significance of individual nodes in a network indicating which nodes are more critical than others rather than providing a single value for a study area.

Third, given the lack of literature about network autocorrelation, this study supplements literature. Even though the concept of network autocorrelation has potential applicability for network analyses, its use has been quite limited in literature. In particular, some studies applied network autocorrelation in the analyses (Black 1992; Black and Thomas 1998; Neville et al. 2004; Xu and Sui 2007). However, these studies used global spatial autocorrelation measures rather than local measures causing limitations in providing critical nodes in network. By applying local measures, this study overcomes the limitation and contributes to literature.

This study aims to present relative advantages (and disadvantages) of using network autocorrelation models in finding critical nodes in a network analysis. For this purpose, this study applied internet backbone networks in the U.S. for an analysis. In particular, this study identified critical nodes in two ways. First, classical disjoint product method (Ball and Provan 1988) was used to measure node reliability as a proxy of critical nodes. Second, Getis-Ord local G statistics (Getis and Ord 1992; Ord and Getis 1995) was used to measure network autocorrelation in finding critical node. The node reliability, measured by disjoint product method, presents significance of nodes in overall network connectivity. However, local G statistic presents clusters of high or low values among nodes connected by one linkage. Thus, applied to network topology, local G statistic values indicate significance of local nodes within its inter-connected nodes. In addition, two variables were applied to network autocorrelation measure and compared; population and node reliability.

Literature review

A network is defined as a system of nodes (or vertices) with connecting links (or edges) (Amaral and Ottino 2004). Haggett et al. (1977) indicated the importance of network analyses in geography because networks are widely used to represent the underlying structure of a variety of systems. In fact, network systems are observed everywhere in nature and even society such as information, technological, biological, and biological networks (Newman 2003)

Initiated from the earliest discoveries of Euler's 1736 solution of the Seven Bridges of Konigsberg, various types of spatial networks have been studies in geography where vertices spread over geographic space and have specific location information. For example, (Kansky 1963) investigated measures of transportation network structure in relationship with its regional characteristics, and classic work of Haggett and Chorley (1969) reviewed studies on a variety of networks in geography. These early network studies are characterized by their primary interests on network connection topology, where studies focus on the number of nodes and links in a network and how they are connected to each other. Recently, because of rapid development in networks such as internet or telecommunications, it becomes enormously complicated to measure the number of network components and the level of network connections. Further, studies based on topological structures of a network required new effective models for analyzing constantly developing complex networks.

Driven by rapidly growing availability of cheap and powerful computers and large-scale electronic dataset, researchers have made substantial progress in network analyses reformulating old ideas, introducing new techniques (Watts 2004). The result brought the "new science of networks " (Barabasi 2002; Buchanan 2002; Watts 2003, 2004). In particular, the "small world networks" , introduced by Watts and Strogatz (1998), deserve attention for its early and major contribution to the new science of networks. The simple background of small world networks is that real-world network is not represented by major assumptions in network connection topology. Given that network connection topology is commonly assumed to be either completely regular or completely random, random network models are characterized by high connectivity, and regular networks shows relatively higher clustering with lower connectivity (Newman et al. 2006). Watts and Strogatz (1998), however, indicates that real-world networks are neither regular nor completely random, but rather exhibit important properties of both. Thus, these networks can be highly clusters, like regular lattice, but have small path lengths, like random graphs. The name, small world networks, was made by analogy with small world phenomenon in which two random strangers on earth are connected by only six chains of intermediate acquaintances (Milgram 1967; Newman 2000),.

In describing small-world network, Watts and Strogatz (1998) presented regular by a uniform one-dimensional lattice, where each node was connected to its k nearest neighbors, and random by a tunable probability parameter p that specified the fraction of randomly rewired links as shown in Figure 1. In particular, the randomness was introduced by rewiring regular network adjusting each edge by a probability p. For 'p=0' indicates that the original ring is unchanged. However, for 'p=1' all edges are rewired randomly.

In identifying small-world network, *average path length* (L) and clustering *coefficient* (C) are used to measure structural properties of a network. L presents the number of edges in the shortest path between two vertices, calculated by the average of overall pairs of vertices in the network. C is the ratio between the existing edges among neighbors of a node and the possible edges in the neighborhood. For example, when a node d has k neighbors, then at most $k(k - 1)/2$ edges can exist between them. If C_d denotes the ratio between the possible and the existing edges of node d, C is defined as the average of C_d over all d. This method has been used to identify a lot of currently known small-world networks; for example, scientific collaboration network (Newman 2001), the World Wide Web (Albert et al. 1999), and the electronic power grid (Watts and Strogatz 1998).

Although the small-world network model is considered as an appropriate model to represent real-world complex networks (Watts 2003), Xu and Sui (2007) criticize that many studies in the literature so far focus on aspatial, in particular, topology-based network, only accounting for connectivity in network. Many real-world networks are spatial having specific locations of nodes, and well-defined length (or distance) of edges such as transportation and computer networks. Connection topology alone cannot represent the multiple

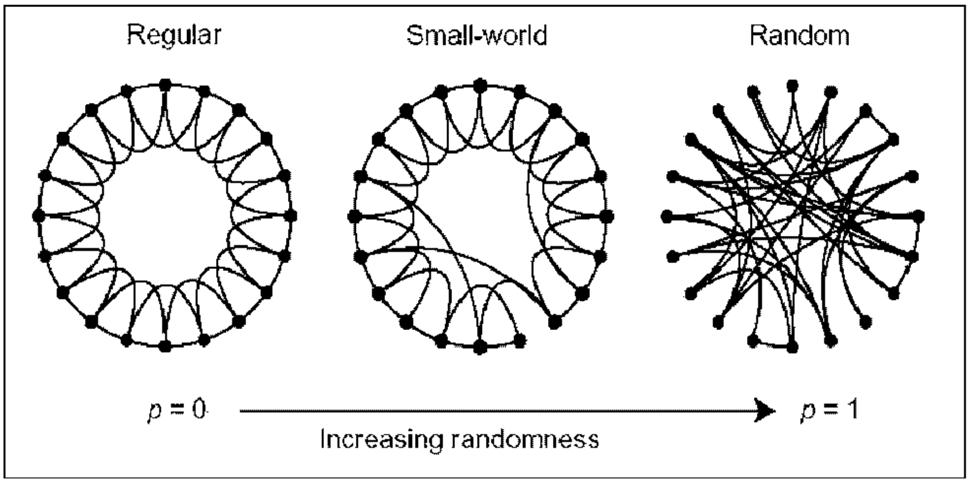

Figure 1. Regular, random vs. small-world networks (Source: Watts and Strogatz 1998)

dimensions of spatial networks, composed location and distance aspects (Gorman and Kulkarni 2004).

Small number of studies focused on modeling spatial aspects of networks. In particular, spatial analytical measures were transformed to be applied for network analyses. The network autocorrelation been studied transportation network and flow systems (Black 1992, 2003; Black and Thomas 1998). As a recent work, Xu and Sui (2007) applied network autocorrelation measures to small-world networks to account for locations and distances of spatial network. They account for location and distances applying network autocorrelation measures such as Moran's I (Cliff and Ord 1981) and Getis-Ord's G (Getis and Ord 1992) statistics. Their applications of network autocorrelation measures to network analyses are characterized by the use of global autocorrelation measures, which provide a single measure of the overall interdependence of the spatially distributed variable for the entire area (Anselin 1995). Therefore, although they could present overall connectivity patterns in a network, they could not indicate local critical nodes.

Methodology: classical reliability measure and application of local spatial autocorrelation to network analyses

Reliability measure: Disjoint product methods

Although there are many methods are available to measure exact reliability, exact reliability evaluation technique in the literature would belong to one of following three categories (Rai et al. 1995); decomposition or factoring, inclusion-exclusion, and disjoint products (for detail, see Barlow and Proschan 1975; Colbourn 1987; Shier 1991). This study uses disjoint product method. The method is known as more efficient way to reduce the computation time compared to other exact methods such as inclusion-exclusion (Misra 1993).

The computation of reliability utilizing disjoint products is to simply sum up probabilities of all disjoint events that are mutually exclusive to each other. The standard mathematical expression is as follows:

$$R_{OD}(G, p)=\sum_{i=1}^{n} \Pr\{\delta_i\} \quad (1)$$

$$\Pr\{\delta_i\}=\prod_{j=1}^{m} \Pr\{e_{ij}\} \quad (2)$$

Where,

R_{OD} is the reliability for two selected nodes, origin O and destination D

G is a graph of a network with known probability p for edges

$\Pr\{\delta_i\}$ is the probability of the disjoint event δ_i

e_{ij} are the edges j constituting δ_i of states

p operational probability of e_{ij} (p='0.9')

The algorithm used in this study is a Boolean algebra. The main idea of this algorithm is to work forward from an initial successful event by depth first search and to find another disjoint event by utilizing its complement and other unused viable paths based on Boolean logic. This process continues until all disjoint events are found. General expression is represented as:

$$R_{OD}(G)$$
$$=P(\delta_1)+P(\delta_1\delta_2)+P(\delta_1\delta_2\delta_3)+\cdots+P(\delta_1\delta_2\cdots\cdots\delta_{n-1}\delta_n) \quad (3)$$

Where,

$P(\delta_n)$: Probability of disjoint event

$P(\delta_n)$: Probability of complement event of δ_n

If a node shows a higher reliability than others, it implies that the node would have better connection since reliability of a node is influenced by the connection degree of the node as well as the composition of what structures are involved to the node. The reliability can be computed in terms of individual nodes by computing mean value of a node in a reliability OD matrix. Therefore, by introducing this measure to network autocorrelation measure, models can be extended, particularly telecommunication network.

Network Autocorrelation measure: Getis-Ord G statistics

Network autocorrelation measures are subdivided into global and local measures. Global network autocorrelation statistics provide a single value of the overall network patterns of a study area. Local network autocorrelation statistics, however, provide estimates disaggregated to the level of the spatial analysis units, allowing assessment of the dependency relationships across space (Anselin 1995). Spatial heterogeneity, which is an intrinsic features of spatial data, derives emphasis on local spatial statistics (Fotheringham 1997; Fotheringham and Brunsdon 1999) because the level of spatial dependency relationships in space are different among sub-areas. Local spatial statistic measures, such as Local Indicators of Spatial Autocorrelation (LISA) (Anselin 1995) and Local Statistics Model (LSM) (Getis and Aldstadt 2004; Getis and Ord 1996; Ord and Getis 1995) have been introduced in literature as major local spatial autocorrelation measures.

This study applied Getis-Ord G_i statistics (Getis and Ord 1992; Ord and Getis 1995). The application of G_i statistics to network setting is identical to area setting in its original literature. The only difference is coming from the setting of spatial weight matrix, in which the spatial weight matrix in network is formed by node connectivity by edges. However, the spatial weight matrix in area data is formed by connectivity from neighboring areas.

While local Moran's I statistic provides similarity in an area, G_i statistic provides local clusters of low or high values compared to a global mean in the study area making G_i statistic more appropriate for critical node analyses. In addition, as pointed out by Zhang and Lin (2006), the use of local Moran's I requires additional Moran scatter plot to distinguish between low- and high-values. A high G_i value in a network indicates a network clustering of high values. A small value indicates clustering of small values. Figure 2 present conceptual background of G_i statistics in network applications. Node 'A' surrounded by neighboring large values results in large G_i statistic values and vice versa for node 'B'.

In its general formulation, given a network weight matrix , the statistic is defined as (Ord and Getis 1995),

$$G_i(W)=\frac{\sum_j w_{ij}y_j - w_i y}{s\{(n-1)S_1 - W_i^2/(n-2)\}^{1/2}} \qquad (4)$$

where

$$y=\frac{1}{n-1}\sum_{j,j\neq i} y_j,\quad s^2=\frac{1}{n-2}\sum_{j,j\neq i}(y_j y)^2,$$

$$W_i=\sum_j w_{ij},\quad S_1=\sum_{j,j\neq i} w_{ij}^2,$$

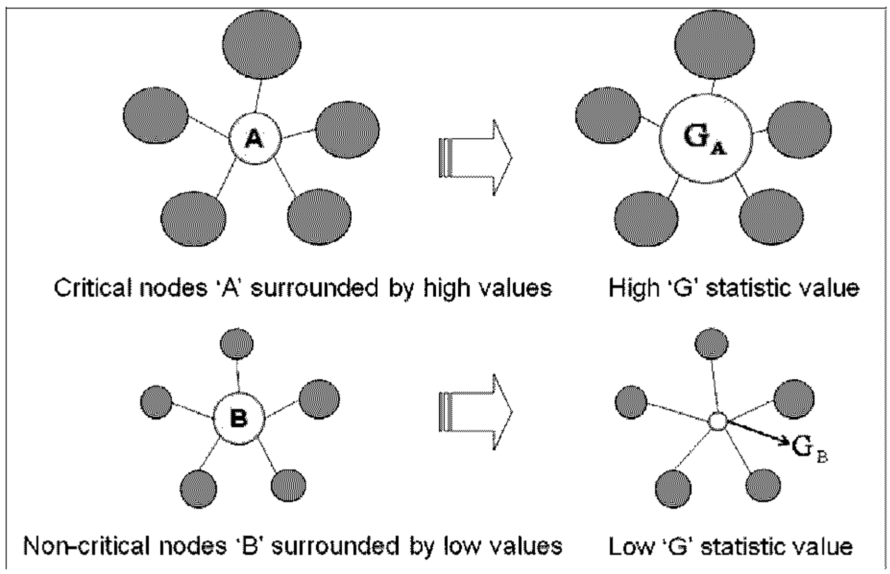

Figure 2. Conceptual background of statistics in network applications

In equation (4), y_j indicates a value for each node, and W is the common binary network weight matrix. The generalization of network weight matrix is well defined in Black's work (1992). In n by n matrix for n numbers of nodes, two nodes connected by a link are coded as 1 and unconnected nodes are coded as 0. Even though very simplistic form of network weight matrix are applied in this study to relieve computational load, it should be noted that the capability of local spatial statistic measures can be extended by applying appropriate spatial weight matrix (Getis and Aldstadt 2004; Leenders 2002). Regarding coding schemes for spatial weight matrix, this study applied W-coding scheme, which standardizes matrix elements using row-sum values. W-coding scheme facilitates the interpretation of the underlying models and calculation process (Hordijk 1979; Ord 1975). Alternatively, globally standardized C-coding scheme or variance stabilizing S-coding scheme (Tiefelsdorf et al. 1999) can be considered.

Data and Application results: Internet backbone network and applied variables

Reliability measure: Disjoint Product method

Two Internet backbone dataset from U.S Internet service providers, specifically, SAVVIS and SERVINT were tested to measure network. Both networks are characterized by their relatively large sizes and different nodal reliabilities due to their different topologies. In particular, SERVINT is composed of 22 different nodes, and SAVVIS is composed of 28 nodes. Each node is located in the major city in the U.S. The dataset provides the location of the nodes and their connectivity to other nodes through linkages. Since the numbers and locations of nodes of each service provider are different, we combined the two-network dataset and formed a large network. The resulting network covers total 33 city nodes. Therefore, in the result, the combined dataset provides overall Internet nodes reliability, and two initial dataset does individual companies' node reliability. Figure 3 shows the reliability calculated by equation (1), (2), and (3). In particular, Denver is 1st ranked in SERVINT, but the 12th in SAVIS, and 4th in combined network. New York ranked the lowest in SERVINT, but the 2nd in SAVIS, and 6th in combined. These differences in nodal reliabilities among networks are coming from different network topologies.

Network autocorrelation : Getis-Ord's G_i statistics

Classical reliability measures such as disjoint products have limitations in analyzing spatial network because they approach spatial network analyses in aspatial perspectives. Many real-world spatial networks such as transportation, telecommunication, and electricity networks have spatial features in their nodes and edges. For instance, many networks have size and location of a node and edge distances. And these network components can be quantitatively specified for realistic spatial network analyses. In applying G_i statistic, this study quantifies nodes using population size and node reliability. Because the applied network is internet backbone, where the effect of physical distance is negligible, the edge was specified by binary network weight matrix.

Among two variables applied, population size of a node implies relative significance of the node in geographical perspectives. In an internet backbone network, population size of a city can be a proxy of potential internet use. In addition, seeing that

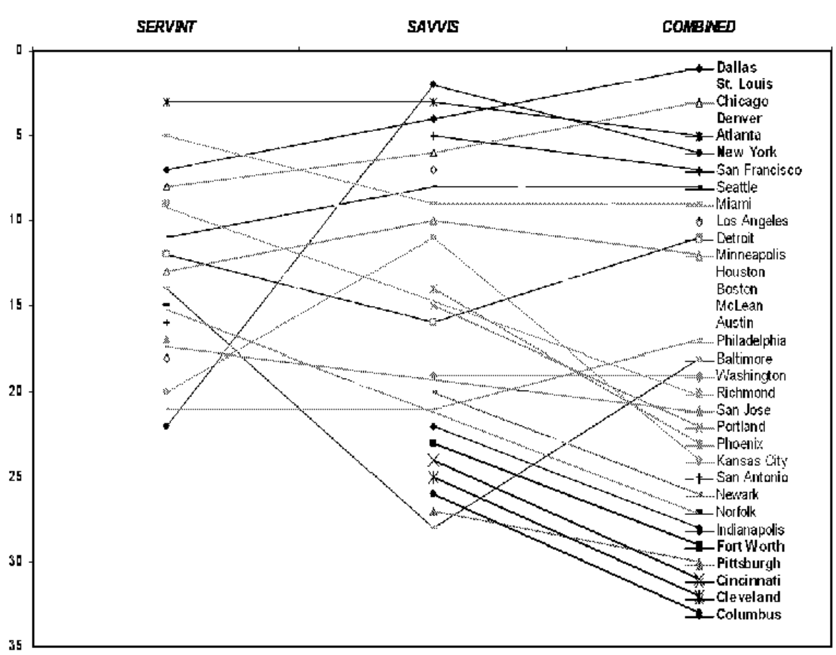

Figure 3. City node reliability in each network

many critical national infrastructures for power, finance, transportation, and other basic resources rely on information and telecommunication networks, population size in network analyses would be far beyond simple internet use. In general, a node with large population would be more significant compare to a node with small population considering the amount of network use and corresponding impact under network failure. Since G_i statistics only considers directly connected local nodes, significance of a node is decided by neighboring nodes' population sizes rather than multi-connected source and terminal links of whole network. Similar to reliability measure above, G_i statistic results for individual nodes are different in the 3 networks because of different network topologies.

Node reliability, calculated by disjoint products, provides significance of nodes taking all routing links between source and terminal nodes into account. Since G_i statistic provides local clusters of high or low reliability values, high G_i statistic values indicate relatively high and stable communication ability in a network.

Table 1 shows the results of node significance calculated by G_i statistics with population and Table 2 with reliability. These two tables present quite different results among each other. Though there are many, the main reason of the difference is explained by ranges of nodes values in each network. In particular, population sizes of city nodes have large variance level making G_i statistic values dependent on neighboring cities' population sizes. Given that G_i statistic represents local mean values of neighboring nodes, a node would have significantly large G_i statistic value when its neighboring nodes have large population size.

In combined network in Table 1, reversing common expectations, Newark, Boston, Dallas nodes are the most critical nodes with highest G_i statistic values respectively. In particular, it should be noted that Newark and Boston do not have outstanding population size. Their neighboring cities are, however, characterized with large population size in the network. For instance, Newark is directly linked to New York and Philadelphia, and Boston has Chicago and New York as neighboring node cities. This result implies that a critical node with large G_i value may not be significant for itself when its population size is considered. However, in interpreting local network autocorrelation, a node should be considered together with neighboring nodes. For instance, under serious information service outage in a node and its diffusion, local impact from the outage would be the largest in the critical nodes because it covers neighboring nodes,. In addition, it should be noted that critical nodes themselves are vulnerable since they are easily influenced by neighboring nodes. Node cities with the lowest G_i statistic values are Denver, Richmond, and McLean respectively. These cities are characterized by their relatively large number of neighboring nodes with small population sizes. Since the node and its neighbors do not have significantly large population size, their local impact from the outage and vulnerability are not as large as that of critical nodes.

Figure 4 shows spatial patterns of G_i statistic values of Table 1. The spatial pattern shows that high G_i statistic values coincide with population clustering areas such as L.A., New York, and Dallas. The use of population size for G_i statistic may give an impression that population size decides the G_i statistic results. However, it should be noted that the edge connections among nodes shows high level of spatial autocorrelations. Generally, nodes with high G_i statistic values are interconnected among each others, and vice versa. This spatial pattern confirms that clusters of large population sizes of nodes and corresponding high G_i statistic values as indicated above.

The results of combined network in Table 2 show that New York, Atlanta, and St. Louis have the highest G_i statistic values. Similarly to population application, the high G_i values represent local (neighboring) clusters of high reliability. This G_i statistic results show overall similarities to the classical reliability measures calculated by disjoint product methods. In combined network, top 6 out of 8 nodes are redundant in both G_i statistics and disjoint product method results with slightly a different order. Top 8 nodes of reliability measures by disjoint product method are Dallas, St. Louis, Chicago, Denver, Atlanta, New York, San Francisco, and Seattle by order. Compared to G_i statistic result in Table 2, only Miami and Los Angeles are added to top 8 nodes of high G_i statistic values replacing Dallas and Chicago. This similarity between the results is explained by small variance in reliability values calculated by disjoint product method. 27 out of 33 reliability values are ranging between 0.966 to 0.977. Consequently, this small variance in reliability values make G_i statistic values similar among nodes because there are very few differences between global and local mean values. In addition, very small variance among nodal reliability values makes G_i statistic values close to zero. Seeing the G_i statistic is calculated by comparing local mean values to global mean,

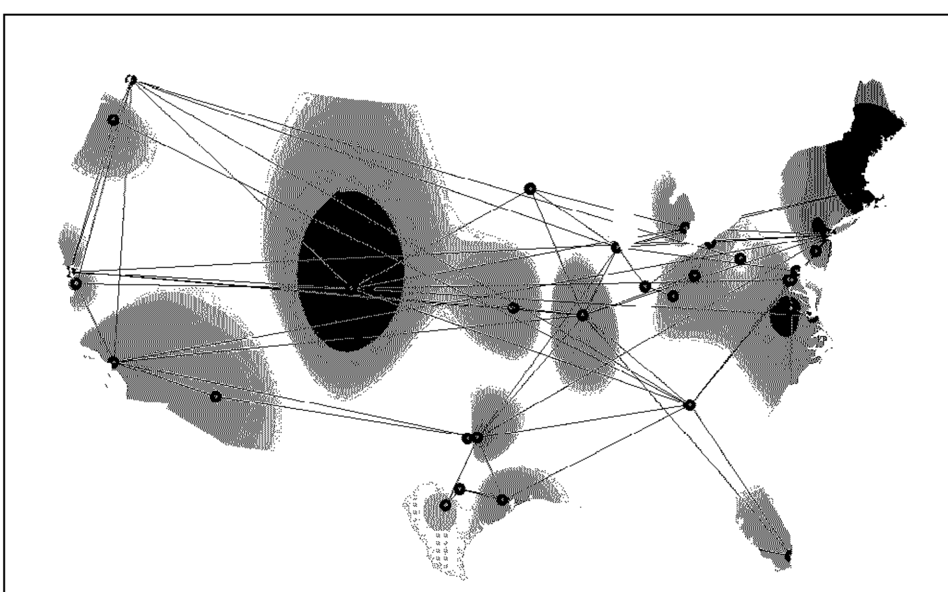

Figure 4. Spatial pattern of statistic values by population

similarity between these two make the resulting G_i statistic close to zero as expected.

In spite of high level of similarity, the differences between G_i statistic and disjoint product methods should be considered also. Although the difference is not significant in the reliability analysis, the difference comes from the different notions of connectivity between G_i statistics and disjoint product method. While disjoint product considers all routing links between source and terminal nodes based on disjoint path sets, G_i statistics only considers links that connect a node and its neighboring nodes. Therefore, in G_i statistics, significance of a node is more dependent on surrounding nodes' values rather than multi-connected source and terminal links of disjoint paths. For example, as Figure 5 shows, Chicago surrounded by non-critical nodes such as Detroit, Indiana, Cincinnati, Cleveland, and Columbus, should (and actually) have a relatively low G_i statistic value though it has high reliability in disjoint product method. In this perspective, highly autocorrelated nodes present not only network reliability, but also give implication of its vulnerability to neighboring nodes. Compared to this G_i measure, the reliability measure of disjoint product methods provides a node's influence to over all nodes in the network rather neighboring nodes. Consequently, when research interests are more focused on local network ranges, G_i statistics performs better than disjoint product method

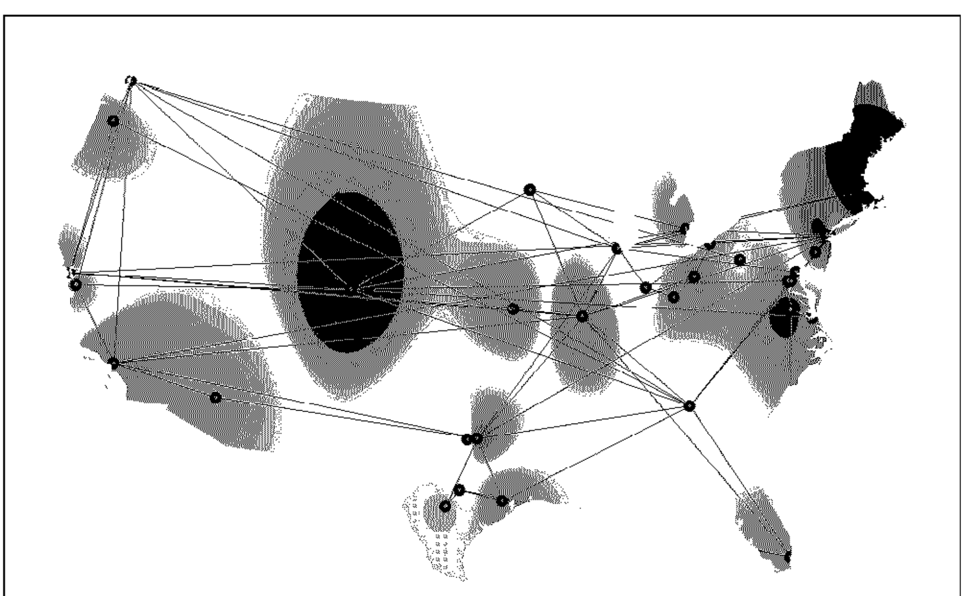

Figure 5. Spatial pattern of statistic values by reliability

Table 1. Node significance calculated by G_i statistics using population

G Values by Population							
		Rank:		Rank:		Rank:	
ID	CITY	Combined	Combined	SAVVIS	SAVVIS	SERVINT	SERVINT
30	Newark	1	0.6035	2	0.6035	NA	NA
6	Boston	2	0.5287	1	0.7175	8	0.0294
13	Dallas	3	0.3889	3	0.3924	10	-0.0162
43	St. Louis	4	0.3323	6	0.3323	3	0.124
32	Philadelphia	5	0.3084	27	-0.0983	1	0.5341
16	Detroit	6	0.3071	4	0.3729	7	0.0607
23	Los Angeles	7	0.305	7	0.305	15	-0.0818
40	San Francisco	8	0.2973	8	0.2973	NA	NA
26	Miami	9	0.246	5	0.3724	13	-0.0629
33	Phoenix	10	0.2301	9	0.2301	NA	NA
8	Chicago	11	0.1784	10	0.2179	18	-0.0922
2	Atlanta	12	0.1564	11	0.2075	12	-0.041
29	New York	13	0.1386	13	0.1198	5	0.0801
42	Seattle	14	0.1039	12	0.1569	2	0.1319
21	Indianapolis	15	0.0373	16	0.0373	NA	NA
27	Minneapolis	16	0.0292	15	0.0675	4	0.1084
3	Austin	17	0.0254	14	0.0846	9	0.0202
4	Baltimore	18	0.0233	22	-0.0554	6	0.0677
18	Fort Worth	19	0.0153	17	0.0153	NA	NA
34	Pittsburgh	20	-0.0128	18	-0.0128	NA	NA
38	San Antonio	21	-0.0187	NA	NA	11	-0.0187
44	Washington	22	-0.0386	19	-0.0386	NA	NA
10	Cleveland	23	-0.0508	20	-0.0508	NA	NA
9	Cincinnati	24	-0.0515	21	-0.0515	NA	NA
35	Portland	25	-0.063	23	-0.063	NA	NA
41	San Jose	26	-0.0793	NA	NA	14	-0.0793
22	Kansas City	27	-0.081	26	-0.081	16	-0.082
20	Houston	28	-0.0822	25	-0.0775	17	-0.0822
11	Columbus	29	-0.1054	28	-0.1054	NA	NA
31	Norfolk	30	-0.1125	NA	NA	19	-0.1125
25	McLean	31	-0.1424	NA	NA	20	-0.1424
36	Richmond	32	-0.1709	NA	NA	22	-0.1709
15	Denver	33	-0.1712	24	-0.0749	21	-0.1542

Table 2. Node significance calculated by G statistics using reliability
Rank* presents G stat rank and (Disjoint product rank)

G Value by Mean Reliability							
		Rank:		Rank:		Rank:	
ID	CITY	Combined	Combined	SAVVIS	SAVVIS	SERVINT	SERVINT
29	New York	1 (6)	0.2947	1	0.2703	22	0.0696
2	Atlanta	2 (5)	0.2767	3	0.2389	3	0.2639
43	St. Louis	3 (2)	0.264	2	0.2606	2	0.3077
15	Denver	4 (4)	0.2301	12	0.1231	1	0.3175
26	Miami	5 (9)	0.2234	9	0.1511	5	0.2633
42	Seattle	6 (8)	0.2126	8	0.1933	11	0.2239
40	San Francisco	7 (7)	0.2126	5	0.2261	NA	NA
23	Los Angeles	8 (10)	0.2094	7	0.2122	18	0.1855
27	Minneapolis	9	0.1853	10	0.1484	13	0.1862
20	Houston	10	0.1849	17	0.1171	4	0.2636
6	Boston	11	0.1848	13	0.1231	19	0.1412
25	McLean	12	0.1789	NA	NA	7	0.2619
3	Austin	13	0.1631	18	0.1171	9	0.2262
32	Philadelphia	14	0.1607	21	0.1073	21	0.0995
8	Chicago	15	0.1576	6	0.2157	8	0.2591
4	Baltimore	16	0.1571	25	0.0749	14	0.1861
44	Washington	17	0.1569	20	0.1101	NA	NA
36	Richmond	18	0.1324	NA	NA	10	0.2262
22	Kansas City	19	0.129	11	0.1231	20	0.1307
41	San Jose	20	0.129	NA	NA	17	0.1855
35	Portland	21	0.129	15	0.1231	NA	NA
33	Phoenix	22	0.129	14	0.1231	NA	NA
38	San Antonio	23	0.1284	NA	NA	16	0.1858
30	Newark	24	0.1283	19	0.1125	NA	NA
13	Dallas	25	0.1274	4	0.2266	6	0.262
31	Norfolk	26	0.1231	NA	NA	15	0.1858
18	Fort Worth	27	0.0789	23	0.0857	NA	NA
34	Pittsburgh	28	0.0789	24	0.081	NA	NA
16	Detroit	29	0.0486	16	0.1215	12	0.2103
21	Indianapolis	30	-0.3868	22	0.0957	NA	NA
9	Cincinnati	31	-0.488	26	0.0608	NA	NA
10	Cleveland	31	-0.488	26	0.0608	NA	NA
11	Columbus	33	-0.537	28	0.0557	NA	NA

Conclusion

So far this research discussed network autocorrelation in the context of spatial network analyses and applied Getis-Ord's G_i statistics as a local network autocorrelation measure in comparison to classical disjoint product method. In the process, two variables, population size and reliability, were applied to represent node attributes. The use of local network autocorrelation measures extended existing network analyses in finding critical nodes. In particular, by providing individual local level of network autocorrelation, the applied method showed local clusters of significant nodes and its possible vulnerability under local network outages. Therefore, if research interests were local network ranges or impacts, local network autocorrelation measures would be more appropriate compared to classical network analysis methods.

In application perspectives, this study quantified only node values because an edge value, such as distance, was not significant in the analysis of internet backbone network. However, depending on network feature, nodes can be specified in different ways, and edges can be quantified also. In addition, network autocorrelation analysis can be applied to find critical links by adjusting weight matrix (Berglund and Karlstrom 1999). For instance, in analyses of transportation or migration networks, edges can be specified by distance or differently weighted by modes of transportations. While this study used G_i statistics in applying local network autocorrelation, local Moran's I statistics can be applied when research interest is similarity among nodes values rather than clustering of high or low values.

While this is not the first, the use of network autocorrelation in network analysis has potentials for other applications and deserve interests for further studies. This study hopes to stimulate further studies of network autocorrelation in the field.

Reference

Agrawal A. and Barlow R. E. 1984. A survey of network reliability and domination theory. Operation Research 32, pp. 478-492.

Albert R., Jeong H. and A.L. B. 1999. Diameter of the world wide web. Nature 401, pp. 130-131.

Amaral L. A. N. and Ottino J. M. 2004. Complex network: augmenting the framework for the study of complex systems. European Physics Journal 38.

Anselin L. 1995. Local Indicators of Spatial Association - LISA. Geographical Analysis 27, pp. 93-115.

Ball M. O. and Provan J. S. 1988. Disjoint products and efficient computation of reliability. Operations Research 36, pp. 703-715.

Barabasi A. L. 2002. Linked: how everything is connected to everything else and what it means. . Cambridge: Perseus Publishing.

Barlow R. E. and Proschan F. 1975. Statistical Theory of Reliability and Lifetesting: Probability Models. New York: Rinehart and Winston.

Berglund S. and Karlstrom A. 1999. Identifying local spatial association in flow data. Journal of Geographical System, pp. 219-236.

Black W. R. 1992. Network Autocorrelation in Transport Network and Flow Systems. Geographical Analysis 24, pp. 207-222.

Black W. R. 2003. Transportation: a geographical analysis. New York: The Guilford Press.

Black W. R. and Thomas I. 1998. Accidents on Belgium' s motorways: a network autocorrelation analysis.

Journal of Transportation Geography 6, pp. 23-31.

Buchanan M. 2002. Nexus: small worlds and the groundbreaking science of networks. New York: W.W. Norton.

Cliff A. D. and Ord J. K. 1981. Spatial Process: Models and applications. London: Pion Limited.

Colbourn C. J. 1987. The Combinatories of Network Reliability. Oxford: Oxford University Press.

Colbourn C. J. 1999. Reliability isues in telecommunications network planning. in P. Soriano and B. Sanso (eds.) Telecommunications network planning. Kluwer Academic Publishers. 135-146.

Fotheringham A. S. 1997. Trends in quantitative methods I: stressing the local. Progress in Human Geography 21, pp. 88-96.

Fotheringham A. S. and Brunsdon C. 1999. Local Forms of Spatial Analysis. Geographical Analysis 31, pp. 340-358.

Getis A. and Aldstadt J. 2004. Constructing the spatial weights matrix using a local statistic. Geographical Analysis 36, pp. 90-104.

Getis A. and Ord J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 24, pp. 189-206.

Getis A. and Ord J. K. 1996. Local spatial statistics: and overview. in P. Longley and M. Batty (eds.) Spatial Analysis: modelling in a GIS environment. Cambridge: Geoinformation International, pp. 261-277.

Gorman S. P. and Kulkarni R. 2004. Spatial small worlds: new geographic patterns for an information economy. Environment and Planning B 31, pp. 273-296.

Haggett P. and Chorley R. J. 1969. Network analysis in geography. New York: ST Martin's Press.

Haggett P., Cliff A. D. and Frey A. 1977. Locational Analysis in Human Geography. London: Edard Arnold.

Hordijk L. 1979. Problems in estimating econometric relationships in space. Papers of the Regional Science Association 42, pp. 99-115.

Kansky K. J. 1963. Structure of transportation networks. Chicago: University of Chicago Press.

Leenders R. 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. Social Networks 24, pp. 21-47.

Milgram S. 1967. The small world problem. Psychology Today 2, pp. 60-67.

Misra K. R. 1993. New trends in system reliability evaluation. Amsterdam: Elsevier.

Moran P. A. P. 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B 10, pp. 245-251.

Neville J., Simsek O. and Jensen D., 2004, Autocorrelation and relational learning: Challenges and opportunities, ICML Statistical Relational Learning Workshop.

Newman M. 2000. Models of the small world. Journal of Statistical Physics 101, pp. 819-841.

Newman M. 2001. The structure of scientific collaboration networks. Proceedings of National Academy of Sciences of the USA 98, pp. 404-409.

Newman M. 2003. The structure and function of complex networks. Society for Industrial and Applied Mathematics Review 45, pp. 167-256.

Newman M., Barabasi A. L. and Watts D. J. 2006. The structure and dynamics of networks. Princeton: Princeton University Press.

Ord J. K. 1975. Estimation methods for medels of spatial interaction. Journal of American Statistical Association 70, pp. 120-126.

Ord J. K. and Getis A. 1995. Local spatial autocorrelation statistics: Distributional issues and an application. Geographical Analysis 27, pp. 286-306.

Rai S., Veeraraghavan M. and Trivedi K. S. 1995. A survey of efficient reliability computation using disjoint products approach. Networks 25, pp. 147-163.

Shier D. R. 1991. Network Reliability and Algebraic Structures. Oxford: Oxford University Press.

Tiefelsdorf M., Griffith D. A. and Boots B. 1999. A Variance-stabilizing Coding Scheme for Spatial Link Matrices. Environment and Planning A 31, pp. 165-180.

Watts D. J. 2003. Six degrees: the science of a connected age. New York: W.W. Norton.

Watts D. J. 2004. The "New" science of network. Annual Review of Sociology 30, pp. 243-270.

Watts D. J. and Strogatz S. H. 1998. Collective dynamics of 'small-world' networks. Nature 393, pp. 440-442.

Xu Z. and Sui D. Z. 2007. Small-world characteristics on transportation networks: a perspective from network autocorrelation. Journal of Geographical System 9, pp. 189-205.

Zhang T. and Lin G. 2006. A supplemental indicator of high-value or low-value spatial clustering. Geographical Analysis 38, pp. 209-225.

Correspondence : Youngho Kim, The Habibie Center, JL Kemang Selatan No. 98, Jakarta Selatan 12560, Indonesia, E-mail: kempffkim@gmail.com

최초투고일 2008년 8월 21일
최종접수일 2008년 9월 15일

한국경제지리학회지 제10권 제4호 2007(492~507)

집합점의 신뢰성을 이용한 네트워크 자기상관 모델의 연구

김영호*

요약 : 일반적으로 사용되는 많은 네트워크 분석방법들은 비공간적인 측면에서 네트워크를 인식하는 경향이 있다. 가령 네트워크의 신뢰성(reliability)을 측정하고 중요한 집합점(node)을 찾는 기본적인 문제에 있어서도, 이러한 분석방법들은 집합점의 속성이나 연결선(edge)의 거리와 같은 네트워크의 공간적인 요소를 배제한 채, 네트워크요소들의 위상적인 접합 여부만을 고려한다. 그에 따라 이러한 네크워크 분석은 실제 공간에서의 네트워크 특성을 반영하지 못하는 제한적인 결과만을 도출하게 한다. 그러나 본 연구는 국지네트워크의 자기상관 (local network autocorrelation measure) 값을 이용하여 이러한 문제의 해결을 시도하였다. 국지자기상관값은 공간객체들의 유사성이나 군집성을 개별적으로 측정하여 각 객체들의 중요도를 나타낸다. 본 연구는 미국의 주요 인터넷 네트워크를 disjoint product method와 Getis-Ord의 G 수치를 이용하여 분석하였으며 그 과정에서 인구와 신뢰도를 변수로서 이용하였다. 그 결과 국지네트워크의 자기상관값은 주요한 집합점들의 국지적인 군집정도를 나타냈고, 이러한 연구 결과는 연구 촛점이 국지네트워크의 범위나 그 영향일 경우, 국지자기상관값의 이용이 더 실용적이고 현실적임을 보여준다.

주요어 : 네트워크 자기상관, disjoint product method, 신뢰성, Getis-Ord G 수치, GIS

* 연구원, The Habibie Center