7R-07

Characterizing Similarity Structure of Spatial Networks Based on Degree Mixing Patterns

Maulana ARIEF　Kazumi SAITO　Tetsuo IKEDA　Hiroaki YUZE
Takayuki WATANABE　Seiya OKUBO　Nobuaki MUTO
University of Shizuoka　Graduate School of Management and Information

1 Introduction

Studies of the structure and functions of large complex networks have attracted a great deal of attention in many different fields such as sociology, biology, physics and computer science [1]. As a particular class, we focus on spatial networks embedded in the real space, like urban streets, whose nodes occupy a precise position in two-dimensional Euclidean space, and whose links are real physical connections [2].

In this paper, we address a problem of classifying and characterizing spatial networks in terms of local connection patterns of node degrees, by especially focusing on the property that the maximum node degrees of these networks are restricted to relatively small numbers. Such characteristic connection patterns that appear frequently in some networks can be regarded as their main building blocks, just like network motifs in [3].

2 Proposed Method

Let $G = (\mathcal{V}, \mathcal{E})$ be a given spatial network, where $\mathcal{V} = \{u, v, w, \cdots\}$ and $\mathcal{E} = \{(u, v), \cdots\}$ mean sets of nodes and links, respectively. In this paper, we only consider undirected networks such that $(u, v) \in \mathcal{E}$ implies $(v, u) \in \mathcal{E}$, but we can straightforwardly extend our approach to deal with directed networks. For each node $u \in \mathcal{V}$, we denote its degree by $r(u)$. Then, we can consider a mixing matrix $\mathbf{C}^{(2)}$ whose i-jth element $c(i, j)$ is calculated by

$$c(i, j) = |\{(u, v) \in \mathcal{E} \mid r(u) = i, r(v) = j\}|, \qquad (1)$$

By setting a marginal probability defined as $p(i) = \sum_j c(i, j)/|\mathcal{E}|$ for each degree i, where $|\mathcal{E}|$ means a number of elements in a set $\mathcal{E}$. we can calculate the expected value for the i-jth element of $\mathbf{C}$ as $|\mathcal{E}|p(i)p(j)$ after $|\mathcal{E}|$ independent trials assuming a binomial distribution. Thus, we can obtain the following Z score $z(i, j)$ with respect to the observed value $c(i, j)$,

$$z(i, j) = \frac{c(i, j) - |\mathcal{E}|p(i)p(j)}{\sqrt{|\mathcal{E}|p(i)p(j)(1 - p(i)p(j))}}. \qquad (2)$$

where $|\mathcal{E}|p(i)p(j)(1 - p(i)p(j))$ is the variance of $|\mathcal{E}|$ trials for the binomial distribution with the probability of $p(i)p(j)$. Evidently, when $z(i, j)$ is large (or small), we can conjecture that there exist a significantly large (or small) number of links between nodes with degrees i and j. In our proposed method, we calculate a feature vector $\mathbf{x}^{(2)}$ from the network by suitably arranging each Z score $z(i, j)$ such that $i \leq j$, i.e., $\mathbf{x}^{(2)} = (z(1, 1), z(1, 2), \cdots)^T$, where $\mathbf{x}^T$ means a transposed vector of $\mathbf{x}$. Recall that since the maximum node degree of spatial networks is restricted to relatively small numbers, the dimensionality of the feature vector $\mathbf{x}^{(2)}$ does not become too large. Let $\mathcal{G} = \{G_1, \cdots, G_N\}$ be a set of given networks; then we can calculate a feature vector $\mathbf{x}_n$ from each network G_n. Based a cosine similarity between these feature vectors, we consider the following dissimilarity measure,

$$d(G_m, G_n) = \sqrt{1 - \frac{(\mathbf{x}_m)^T \mathbf{x}_n}{\|\mathbf{x}_m\| \|\mathbf{x}_n\|}}, \qquad (3)$$

where $\|\mathbf{x}\|$ means the standard L2 norm defined by $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}}$. Finally, we construct a dendrogram of these networks based on Ward's minimum variance method [4] using the above dissimilarity measure. We call this proposed method the bi-mixing method. Below we summarize the algorithm of our proposed method.

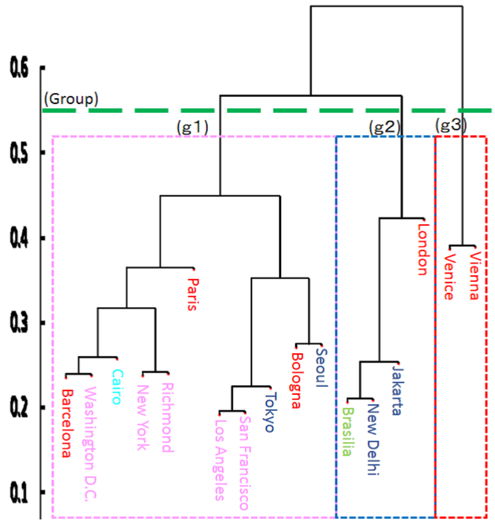

Figure 1: Dendogram constructed by bi-mixing method.

3 Experience

In this section, we show our experimental results obtained by applying the bi-mixing methods. We used OSM (OpenStreetMap) data of seventeen cities in our experiments, In August, 2015, we obtained the OSM data of these seventeen cities from Metro Extracts*. Each continent of Europe, North America, South America, Asia,

*https://mapzen.com/data/metro-extracts

Table 1: Ranking by 2-2, 3-3, 4-4 mixing pattern

Rankk	City Name	nz(2-2)	G	C	City Name	nz(3-3)	G	C	City Name	nz(4-4)	G	C
1	Brasilia	.5612	g2	SA	Vienna	.7258	g3	EU	Richmond	.7781	g1	NA
2	New Delhi	.5421	g2	AS	Venice	.6273	g3	EU	New York	.7627	g1	NA
3	Jakarta	.4733	g2	AS	Seoul	.5533	g1	AS	Barcelona	.7445	g1	EU
4	Cairo	.4116	g1	AF	Tokyo	.4584	g1	AS	San Francisco	.6905	g1	NA
5	London	.4089	g2	EU	Bologna	.4566	g1	EU	Paris	.6782	g1	EU
6	Tokyo	.4028	g1	AS	New Delhi	.4529	g2	AS	Washington DC	.6703	g1	NA
7	Washington DC	.3874	g1	NA	Jakarta	.4483	g2	AS	Cairo	.6552	g1	AF
8	Los Angeles	.3725	g1	NA	London	.4267	g2	AS	Los Angeles	.6328	g1	NA
9	Bologna	.3657	g1	EU	Los Angeles	.4245	g1	NA	Tokyo	.5976	g1	AS
10	San Francisco	.3602	g1	NA	Brasilia	.4075	g2	SA	Seoul	.5470	g1	AS
11	Paris	.3565	g1	EU	San Francisco	.4005	g1	NA	Bologna	.5423	g1	EU
12	New York	.3470	g1	NA	Cairo	.3996	g1	AF	Jakarta	.5321	g2	AS
13	Barcelona	.3221	g1	EU	Richmond	.3890	g1	NA	Brasilia	.5208	g2	SA
14	Seoul	.3155	g1	AS	Washington DC	.3709	g1	NA	London	.4760	g2	EU
15	Vienna	.2888	g3	EU	Paris	.3378	g1	EU	New Delhi	.4684	g2	AS
16	Richmond	.2733	g1	EU	Barcelona	.3301	g1	EU	Vienna	.3996	g3	EU
17	Venice	.2432	g3	EU	New York	.3194	g1	NA	Venice	.2882	g3	EU

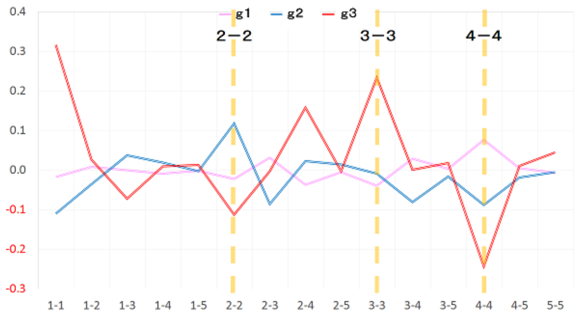

Figure 2: Differences of normalized z score of bi-mixing

Africa is abbreviated by EU, NA, SA, AS, and AF. Here note that $p(i)$ is a ratio of nodes with degree i defined by $p(i) = |\{u \in \mathcal{V} \mid r(u) = i\}|/|\mathcal{V}|$, and since $p(5), p(6), \cdots$ were very small values for each city, we treated them as one variable defined by $p(5) \leftarrow \sum_{i>4} p(i)$. From this OSM data, we can see that although the area and the numbers of nodes and links, $|\mathcal{V}|$ and $|\mathcal{E}|$ are substantially different, the degree distributions defined by $p(i)$ are quite similar as common characteristics of these spatial networks. We consider that these results are intuitively interpretable.

Figure 1 shows the dendrogram constructed by our bi-mixing method, where these cities are depicted by magenta (NA), red (EU), blue (AS), green (SA), and cyan (AF), respectively. As shown in this figure, we can classify these cities into the three groups, i.e., $g1$, $g2$ and $g3$, by using the cut-off point around 0.55 depicted by a green dotted line.

Figure 2 shows the difference of each average of the normalized Z scores (hence abbreviated by nz) devided into three groups based on dendogram over the seventeen selected cities, where the nz of each city is defined by $nz(i, j) = z(i, j)/\|\mathbf{x}^{(2)}\|$, and they are directly used for calculating the cosine similarity between any pair of the cities. From this figure, we can see that the average normalized Z scores for the three groups substantially differ in the three characteristic mixing patterns, 2-2, 3-3 and 4-4, which are referred to as the discriminative mixing patterns.

Tables 1 show the rankings of the cities according to their nz with respect to the three discriminative mixing patterns, 2-2, 3-3 and 4-4, respectively. From these tables, we can see that the groups $g1$, $g2$ and $g3$ are individually characterized by relatively larger values at the 4-4, 2-2 and 3-3 patterns, respectively. Thus, we can consider that the characteristics of these cities can be reasonably described in terms of a relatively small number of selected mixing patterns, as building blocks of given spatial networks.

4 Conclusion

We addressed the problem of classifying and characterizing spatial networks in terms of local connection patterns of node degrees, by especially focusing on the property that the maximum node degrees of these networks are restricted to relatively small numbers. In our experiments using spatial networks constructed from urban streets of seventeen cities, we confirmed that our method can produce intuitively interpretable results which reflect regional characteristics of these cities. Moreover, we showed that these characteristics can be reasonably described in terms of a relatively small number of selected mixing patterns, as main building blocks of given spatial networks. In future, we plan to evaluate our method using various spatial networks, and attempt to establish more useful techniques for uncovering degree mixing patterns, as building blocks of given spatial networks.

Acknowledgements This work was supported by JSPS Grant-in-Aid for Scientific Research (No.15K00311) and SCOPE (No.142306004).

References

[1] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, Vol. 45, pp. 167–256, 2003.

[2] P. Crucitti, V. Latora, and S. Porta. Centrality measures in spatial networks of urban streets. *Physical Review E*, Vol. 73, No. 3, p. 036125, 2006.

[3] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science (New York, N.Y.)*, Vol. 298, No. 5594, pp. 824–827, 2002.

[4] J.H.Jr. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, Vol. 58, pp. 236–244, 1963.