4I1-4

次数混合パターンに基づく世界42都市道路網の類似構造分析

Analyzing Similarity Structure of Spatial Networks
Based on Degree Mixing Patterns on 42 cities of the world

アリフ マウラナ[*1] Arief Maulana　斉藤 和巳[*1] Kazumi SAITO　池田 哲夫[*1] Tetsuo IKEDA　湯瀬 裕昭[*1] Hiroaki YUZE　渡邉 貴之[*1] Takayuki WATANABE
大久保 誠也[*1] Seiya OKUBO　武藤 伸明[*1] Nobuaki MUTO

[*1]静岡県立大学大学院 経営情報イノベーション研究科
Graduate School of Management and Information of Innovation, University of Shizuoka

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. For this purpose, we extend our previous methods by introducing a new measure to identify characteristic discriminative patterns. In our expanded experiments using spatial networks constructed from urban streets of 42 cities, we confirmed that the identified discriminative patterns play a substantially important role to interpret regional characteristics of these cities.

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 [Newman 03]. 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 or three-dimensional Euclidean space, and whose links are real physical connections [Crucitti 06].

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 analyses [Milo 02]. In this paper, we mainly consider spatial networks constructed from urban streets by mapping the intersections of streets into nodes and the streets between the nodes into links, although our approach is potentially applicable to a wider range of spatial networks.

In order to analyze given spatial networks in terms of local connection patterns of node degrees, we have proposed two methods referred to as the bi- and tri-mixing methods [Maulana 16]. More specifically, we first enumerate and count the combinations of node degrees with respect to connected pair or triple nodes for each of given networks, just like network motifs in [Milo 02]. Second, we calculate feature vectors of these networks expressing which mixing pattern appears with a significantly high (or law) frequency, just like computation of assortative mixing coefficients [Newman 02]. Finally, we construct a dendrogram of these networks based on a cosine similarity between these feature vectors [Ward 63].

連絡先: 〒 422-8526 静岡市駿河区谷田 52-1 静岡県立大学

In this paper, we extend our previous methods by introducing a new measure to identify characteristic discriminative patterns. In our expanded experiments using spatial networks constructed from urban streets of 42 cities, we confirm that the identified discriminative patterns play a substantially important role to interpret regional characteristics of these cities.

2. Proposed Method

In this section, for the sake of readers' convenience, we first revisit our previous methods referred to as the bi- and tri-mixing methods [Maulana 16]. Then, we propose a new measure to identify characteristic discriminative patterns.

2.1 Bi- and tri-mixing methods

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 directional networks. For each node $u \in \mathcal{V}$, we denote its degree by $r(u)$. Then, we can consider a degree mixing matrix $\mathbf{C}^{(2)}$ whose i-jth element $c(i, j)$ is calculated by

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

where $|A|$ means a number of elements in a set A. By setting a marginal probability defined as $p(i) = \sum_j c(i, j)/|\mathcal{E}|$ for each degree i, 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 $c(i, j)$,

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

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 the bi-mixing method, we calculate a feature vector $\mathbf{x}^{(2)}$ from the network by suitably arranging each Z score $z(i,j)$ such that $i \le j$, i.e., $\mathbf{x}^{(2)} = (z(1,1), z(1,2), \cdots)^T$, where $\mathbf{a}^T$ means a transposed vector of $\mathbf{a}$. 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.

Next, we explain the tri-mixing method which utilizes the connected triple nodes, just like network motifs analyses based on triad patterns [Milo 02], instead of connected pairs. Let $\mathcal{F}$ be the set of the connected triples defined by $\mathcal{F} = \{(u,v,w) \mid (u,v) \in \mathcal{E}, (v,w) \in \mathcal{E}\}$. Then, we can consider a three-dimensional array $\mathbf{C}^{(3)}$ whose i-j-kth element $c(i,j,k)$ is calculated by

$$c(i,j,k) = |\{(u,v,w) \in \mathcal{F} \mid r(u) = i, r(v) = j, r(w) = k\}|.$$

Thus, by setting a marginal probability defined as $p(i) = \sum_{j,k} c(i,j,k)/|\mathcal{F}|$ for each degree i, we can also obtain the following Z score $z(i,j,k)$ with respect to the observed value $c(i,j,k)$,

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

Again, we calculate a feature vector $\mathbf{x}^{(3)}$ from the network by arranging each Z score $z(i,j,k)$ such that $i \le k$ and $j \ge 2$, i.e., $\mathbf{x}^{(3)} = (z(1,2,1), z(1,2,2), \cdots)^T$.

Let $\mathcal{G} = \{G_1, \cdots, G_N\}$ be a set of given networks; then we can calculate a normalized feature vector $\mathbf{y}_n^{(h)} = \mathbf{x}_n^{(h)}/||\mathbf{x}_n^{(h)}||$ from each network G_n based on connected pairs or triples, where $h \in \{2,3\}$. Here, $||\mathbf{x}||$ means the standard L2 norm defined by $||\mathbf{x}|| = \sqrt{\mathbf{x}^T\mathbf{x}}$. Then by using the following dissimilarity measure based on a cosine similarity between these feature vectors,

$$d^{(h)}(G_m, G_n) = \sqrt{1 - (\mathbf{y}_m^{(h)})^T \mathbf{y}_n^{(h)}}.$$

we can construct a dendrogram of these networks based on Ward's minimum variance method [Ward 63]. Finally, by using an adequate cut-off point, we can classify the set of networks $\mathcal{G}$ into S groups denoted by $\{\mathcal{G}_1, \cdots, \mathcal{G}_s, \cdots, \mathcal{G}_S\}$

2.2 Discriminative pattern identification

Let $lp^{(h)}$ be a local mixing pattern defined by $lp^{(2)} = i$-j and $lp^{(3)} = i$-j-k for $h = 2$ and 3, respectively. Now, with respect to a local pattern $lp^{(h)}$, we can calculate the difference of the average feature value over group $\mathcal{G}_s$ from the average over $\mathcal{G}$ as follows:

$$\delta(lp^{(h)}, \mathcal{G}_s) = \frac{1}{|\mathcal{G}_s|}\sum_{n \in \mathcal{G}_s} y_n^{(h)}(lp^{(h)}) - \frac{1}{|\mathcal{G}|}\sum_{n \in \mathcal{G}} y_n^{(h)}(lp^{(h)}) \quad (1)$$

Thus, we can calculate the deviation of the difference $\delta(lp^{(h)}, \mathcal{G}_s)$ for each local pattern $lp^{(h)}$ as follows:

$$\Delta(lp^{(h)}) = \sum_{s \in \mathcal{S}}\sum_{t \in \mathcal{S}} (\delta(lp^{(h)}, \mathcal{G}_s) - \delta(lp^{(h)}, \mathcal{G}_s))^2 \quad (2)$$

Table 1: Basic statistics as network.

No	Cont.City Name	$\mathcal{V}$	$\mathcal{E}$	No	Cont.City Name	$\mathcal{V}$	$\mathcal{E}$
1	AF Accra	49420	57354	22	EU London	3022742	3266643
2	EU Amsterdam	628606	724497	23	NA Los Angeles	2736534	2959818
3	EU Barcelona	352243	386908	24	WA Mecca	44818	52262
4	EU Berlin	1587916	1775479	25	SA Mexico	527929	632761
5	EU Bologna	267433	286126	26	EU Munich	373660	434936
6	SA Brasilia	201969	244748	27	AF Nairobi	115447	123497
7	SA Buenos Aires	496813	665626	28	WA New Delhi	289725	341004
8	AF Cairo	202934	232834	29	SA New York	1305256	1454084
9	AU Canberra	317425	344936	30	EA Okinawa	295178	319716
10	NA Washington DC	119352	133336	31	SA Panama	56341	61831
11	WA Dubai	828468	920915	32	EU Paris	1486406	1654543
12	WA Hong Kong	1039557	1113520	33	NA Richmond	375390	398417
13	NA Houston	1371412	1516435	34	EU Rome	545411	593206
14	EU Istanbul	656041	760031	35	NA San Francisco	502051	551028
15	EA Jakarta	358610	403490	36	EA Seoul	851584	917112
16	WA Jerusalem	163105	172776	37	EA Singapore	311928	348560
17	AF Johannesburg	361192	421997	38	AU Sydney	585675	638197
18	EA Kathmandu	1007859	1020498	39	EA Tokyo	6571077	7312007
19	EA Kolkata	301891	322059	40	EU Venice	7359358	8632841
20	SA Lima	240066	298826	41	EU Vienna	9195477	10456807
21	EU Lisbon	774922	832978	42	EU Zurich	613848	670858

where $\mathcal{S} = \{1, \cdots, S\}$. In this paper, we propose to evaluate each local pattern $lp^{(h)}$ by $\Delta(lp^{(h)})$ and to identify those pattern with large values as some candidates of discriminative local patterns.

3. Experiments

We used a dataset of OSM (OpenStreetMap) obtained from Metro Extracts[*1]. In order to evaluate the reliability and consistency of our proposed methods we performed our experiments by newly adding 25 cities for our former experiments using 17 cities [Maulana 16], i.e., the total number of cities is 42.

3.1 Dataset

From the OSM dataset of each city, we extracted all highways and all nodes, and constructed each spatial network by mapping the ends, intersections and curve-fitting-points of streets into nodes and the streets between the nodes into links. Table 1 shows the basic statistics of the networks for the 42 cities. where each continent of North America, South America, Europe, East-Asia, West-Asia Africa, and Australia is abbreviated by NA, SA, EU, EA, WA, AF, and AU. We can see that the numbers of nodes and links ($|\mathcal{V}|$, $|\mathcal{E}|$) are substantially different from each other, and the covered regions spread worldwide.

3.2 Bi-Mixing

Figure 1 shows the dendrogram constructed by our bi-mixing method, where the total number of mixing patterns from 1-1 to 5-5 amounts to 15. In this figure, the cities of NA, SA, EU, EA, WA, AF, and AU are depicted by green, magenta, blue, red, brown, yellow and violet, respectively. As shown in Fig. 1, we can classify these cities into the three groups, i.e., $\mathcal{G}_1^{(2)}$, $\mathcal{G}_2^{(2)}$, and $\mathcal{G}_3^{(2)}$, by using the cut-off point around 0.69 drawn by a yellow dotted line. Here note that we employ the same cut-off point for the case of experiments using the tri-mixing method as shown in Sec. 3.3. We can observe that these groups $\mathcal{G}_1^{(2)}$, $\mathcal{G}_2^{(2)}$, and $\mathcal{G}_3^{(2)}$, are

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

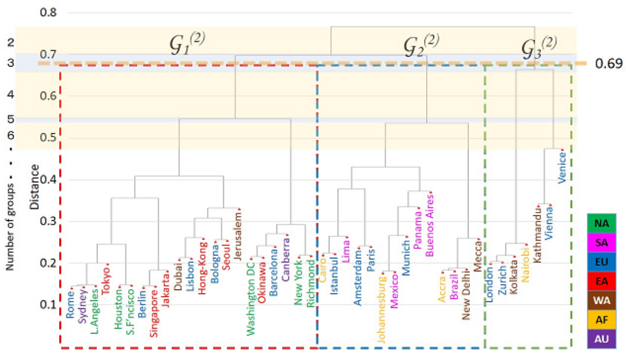

Figure 1: Dendogram constructed by bi-mixing method

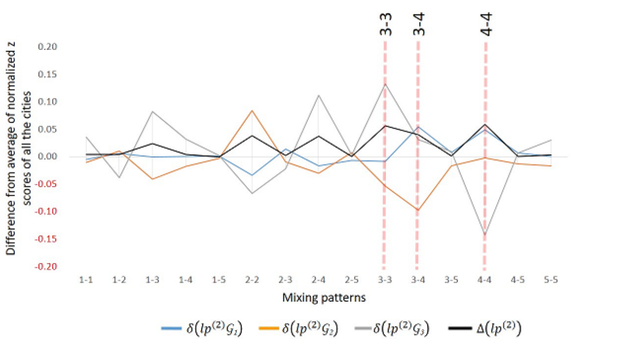

Figure 2: Difference of average z score $\delta(lp^{(2)}, \mathcal{G}_s)$ and the deviation $\Delta(lp^{(2)})$ over each group $\delta(lp^{(2)}, \mathcal{G}_s)$

mainly characterized by cities of NA for $\mathcal{G}_1^{(2)}$, and AF, SA for $\mathcal{G}_2^{(2)}$, and WA, EA for $\mathcal{G}_3^{(2)}$, respectively, although a few number of the other continent cities are included. Thus, we consider that our method could produce naturally interpretable results which reflect regional characteristics of these cities. Here we should emphasize that these results are consistent to our former experiments using 17 cities except that only Jakarta and London were classified into some different groups [Maulana 16].

In Fig. 2, based on Equation (1), we plot the difference of average z score $\delta(lp^{(2)}, \mathcal{G}_s)$ from the average of all the 42 cities with respect to each local mixing pattern $lp^{(2)}$ of group $\mathcal{G}_s$. We can see that the local mixing patterns of 3-3, 3-4 and 4-4 exhibit relatively larger differences. Moreover, based on Equation (2), we also plot the deviation value $\Delta(lp^{(2)})$ over each group $\delta(lp^{(2)}, \mathcal{G}_s)$ and calculate by black line. Here, as shown on the left hand side of table 2, we can confirm that the top three characteristic mixing pattern are 3-3, 3-4 and 4-4, which are referred to as discriminative mixing patterns and used as useful measure in order to characterize each group in our method.

Table 3 shows the rankings of the cities according to their normalized z scores with respect to the three discriminative mixing patterns, 3-3, 3-4 and 4-4, respectively. From these tables, we can see that the groups $\mathcal{G}_1$ and $\mathcal{G}_3$ are individually characterized by relatively larger values at the 4-4 and 3-3

Table 2: Rank by $\Delta(lp^{(h)})$

	Bi-Mixing		Tri-Mixing	
Rank	$\Delta(lp^{(2)})$	$lp^{(2)}$	$\Delta(lp^{(3)})$	$lp^{(3)}$
1	0.0587717	4-4	0.059665	1-3-1
2	0.0566443	3-3	0.03812	3-4-3
3	0.0403719	3-4	0.034149	5-5-5
4	0.0382356	2-2	0.024923	3-3-3
5	0.0370605	2-4	0.020593	2-2-2

Table 3: Rank by bi-mixing

	3-3			3-4			4-4		
Rank	Grp	Cities	z	Grp	Cities	z	Grp	Cities	z
1	$\mathcal{G}_3$	Vienna	0.726	$\mathcal{G}_2$	Buenos Aires	-0.101	$\mathcal{G}_1$	Richmond	0.778
2	$\mathcal{G}_3$	Kathmandu	0.661	$\mathcal{G}_2$	Mexico	-0.056	$\mathcal{G}_1$	New York	0.763
3	$\mathcal{G}_3$	Venice	0.627	$\mathcal{G}_2$	Brasilia	-0.035	$\mathcal{G}_1$	Barcelona	0.745
4	$\mathcal{G}_1$	Seoul	0.553	$\mathcal{G}_2$	Mecca	-0.001	$\mathcal{G}_1$	Canberra	0.699
5	$\mathcal{G}_3$	Nairobi	0.537	$\mathcal{G}_2$	Accra	0.004	$\mathcal{G}_1$	San Francisco	0.691
6	$\mathcal{G}_2$	Mecca	0.528	$\mathcal{G}_2$	Lima	0.010	$\mathcal{G}_1$	Okinawa	0.684
7	$\mathcal{G}_1$	Dubai	0.519	$\mathcal{G}_2$	Munich	0.011	$\mathcal{G}_1$	Hong Kong	0.682
8	$\mathcal{G}_1$	Lisbon	0.496	$\mathcal{G}_2$	Johannesburg	0.015	$\mathcal{G}_2$	Panama	0.68
9	$\mathcal{G}_3$	Zurich	0.493	$\mathcal{G}_2$	Panama	0.024	$\mathcal{G}_2$	Paris	0.678
10	$\mathcal{G}_1$	Berlin	0.461	$\mathcal{G}_2$	New Delhi	0.031	$\mathcal{G}_1$	Washington DC	0.670
11	$\mathcal{G}_3$	Kolkata	0.460	$\mathcal{G}_2$	Paris	0.041	$\mathcal{G}_2$	Mexico	0.667
12	$\mathcal{G}_1$	Tokyo	0.458	$\mathcal{G}_2$	Amsterdam	0.042	$\mathcal{G}_1$	Houston	0.664
13	$\mathcal{G}_1$	Bologna	0.457	$\mathcal{G}_1$	Houston	0.059	$\mathcal{G}_2$	Johannesburg	0.657
14	$\mathcal{G}_2$	New Delhi	0.453	$\mathcal{G}_1$	New York	0.059	$\mathcal{G}_2$	Cairo	0.655
15	$\mathcal{G}_1$	Jakarta	0.448	$\mathcal{G}_1$	Dubai	0.068	$\mathcal{G}_1$	Jerusalem	0.647

patterns, while the group $\mathcal{G}_2$ is characterized by relatively smaller value at 3-4 pattern respectively.

3.3 Tri-Mixing

Figure 3 shows the dendrogram constructed by our tri-mixing method, where the total number of mixing patterns from 1-2-1 to 5-5-5 amounts to 60, and each color of these cities is depicted by the same one as shown in Fig. 1. Similar to our experiment using the bi-mixing method, we can classify these cities into the three groups, i.e., $\mathcal{G}_1^{(3)}$, $\mathcal{G}_2^{(3)}$ and $\mathcal{G}_3^{(3)}$, by using the same cut-off point around 0.69 drawn by a yellow dotted line.

In Fig. 4, based on Equation (1), we also plot the difference of average z score $\delta(lp^{(3)}, \mathcal{G}_s)$ from the average of all the 42 cities with respect to each local mixing pattern $lp^{(3)}$ of group $\mathcal{G}_s$. We can see that the local mixing patterns of 1-3-1, 3-4-3 and 5-5-5 exhibit relatively larger differences. Based on Equation (2), we also plot the deviation $\Delta(lp^{(3)})$ over each group $\delta(lp^{(3)}, \mathcal{G}_s)$ by black line. Again, as shown on the right hand side of table 2, we can confirm that the top three characteristic mixing patterns are 1-3-1, 3-4-3 and 5-5-5, which are used as discriminative mixing patterns.

Tables 4 shows the rankings of the cities according to their normalized z score with respect to the three discriminative mixing patterns, 1-3-1, 3-4-3 and 5-5-5, respectively. From these tables, we can see that the groups $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$ are individually characterized by relatively larger values at the 3-4-3, 1-3-1 and 5-5-5 patterns, respectively. Namely, these results also support our claim that the characteristics of these cities can be reasonably described in terms of a relatively small number of selected discriminative mixing patterns, as building blocks of given spatial networks.

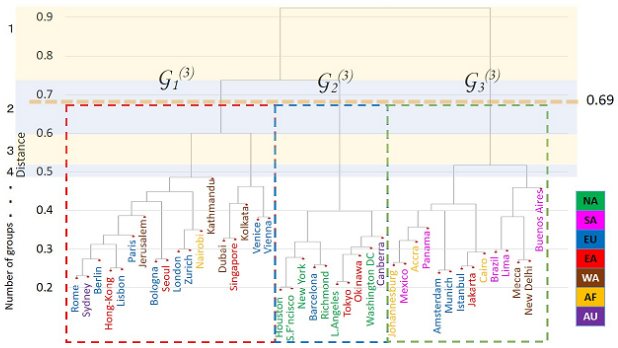

Figure 3: Dendogram constructed by tri-mixing method

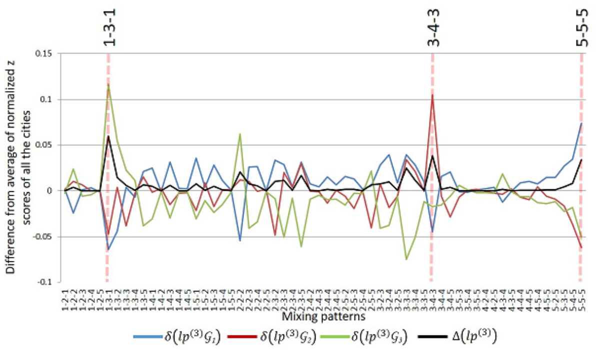

Figure 4: Difference of average z score $\delta(lp^{(3)}, \mathcal{G}_s)$ and the deviation $\Delta(lp^{(3)})$ over each group $\delta(lp^{(3)}, \mathcal{G}_s)$

Table 4: Rank by tri-mixing

	1-3-1			3-4-3			5-5-5		
Rank	Grp	Cities	z	Grp	Cities	z	Grp	Cities	z
1	$\mathcal{G}_3$	Buenos Aires	0.744	$\mathcal{G}_2$	Richmond	0.800	$\mathcal{G}_1$	Vienna	0.537
2	$\mathcal{G}_3$	Lima	0.696	$\mathcal{G}_2$	New York	0.793	$\mathcal{G}_1$	Kolkata	0.373
3	$\mathcal{G}_3$	New Delhi	0.655	$\mathcal{G}_2$	Barcelona	0.725	$\mathcal{G}_1$	Dubai	0.349
4	$\mathcal{G}_3$	Brasilia	0.636	$\mathcal{G}_2$	San Francisco	0.723	$\mathcal{G}_1$	Singapore	0.279
5	$\mathcal{G}_3$	Mecca	0.607	$\mathcal{G}_3$	Panama	0.710	$\mathcal{G}_1$	Venice	0.228
6	$\mathcal{G}_3$	Mexico	0.591	$\mathcal{G}_1$	Kathmandu	0.705	$\mathcal{G}_1$	Paris	0.193
7	$\mathcal{G}_3$	Jakarta	0.577	$\mathcal{G}_2$	Houston	0.690	$\mathcal{G}_1$	Lisbon	0.188
8	$\mathcal{G}_3$	Istanbul	0.558	$\mathcal{G}_3$	Johannesburg	0.682	$\mathcal{G}_1$	Jerusalem	0.174
9	$\mathcal{G}_3$	Munich	0.527	$\mathcal{G}_2$	Okinawa	0.668	$\mathcal{G}_1$	Hong Kong	0.161
10	$\mathcal{G}_3$	Johannesburg	0.525	$\mathcal{G}_2$	Canberra	0.668	$\mathcal{G}_3$	Brasilia	0.145
11	$\mathcal{G}_1$	Berlin	0.520	$\mathcal{G}_3$	Cairo	0.661	$\mathcal{G}_3$	Cairo	0.141
12	$\mathcal{G}_1$	Singapore	0.519	$\mathcal{G}_2$	Los Angeles	0.661	$\mathcal{G}_2$	Canberra	0.111
13	$\mathcal{G}_3$	Amsterdam	0.514	$\mathcal{G}_3$	Mexico	0.654	$\mathcal{G}_1$	Sydney	0.110
14	$\mathcal{G}_3$	Accra	0.512	$\mathcal{G}_3$	Accra	0.650	$\mathcal{G}_1$	London	0.107
15	$\mathcal{G}_2$	Washington DC	0.491	$\mathcal{G}_1$	Rome	0.649	$\mathcal{G}_1$	Seoul	0.105

4. Conclusion

We addressed the problem of classifying and characterizing spatial networks in terms of local connection patterns of node degrees, In this paper, we extended our previous methods referred to as the bi- and tri-mixing methods by introducing a new measure to identify characteristic discriminative patterns. In order to evaluate the reliability and consistency of our proposed methods, We performed our experiments using spatial networks constructed from urban streets of 42 cities, by adding 15 cities to our former experiments. From our experimental results, we confirmed that our methods could produce naturally interpretable results which reflect regional characteristics of these cities. Moreover, we showed that the characteristics of these cities could be reasonably described in terms of a relatively small number of selected discriminative mixing patterns, as 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.

Acknowledgments This work was supported by JSPS Grant-in-Aid for Scientific Research (No.15K00311).

References

[Crucitti 06] Crucitti, P., Latora, V., and Porta, S.: Centrality measures in spatial networks of urban streets, *Physical Review E*, Vol. 73, No. 3, pp. 036125+ (2006)

[Maulana 16] Maulana, A., Saito, K., Ikeda, T., Yuze, N., Watanabe, T., Okubo, S., and Muto, N.: Analyzing Similarity Structure of Spatial Networks Based on Degree Mixing Patterns, *Advanced Information Networking and Applications*, Vol. 30, pp. 1–8 (2016)

[Milo 02] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U.: Network motifs: simple building blocks of complex networks., *Science (New York, N.Y.)*, Vol. 298, No. 5594, pp. 824–827 (2002)

[Newman 02] Newman, M. E. J.: Assortative mixing in networks, *Structure*, Vol. 2, No. 4, p. 5 (2002)

[Newman 03] Newman, M. E. J.: The structure and function of complex networks, *SIAM Review*, Vol. 45, pp. 167–256 (2003)

[Ward 63] Ward, J.: Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, Vol. 58, pp. 236–244 (1963)