DEIM Forum 2017 G6-2

Characterizing Areas of Spatial Networks Based on Mixing Patterns

Arief MAULANA†, Kazumi SAITO†, Tetsuo IKEDA†, and Hiroaki YUZE†

† Graduate School of Management and Information of Inovation, University of Shizuoka
52–1 Yada, Suruga-ku, Shizuoka 422–8526 Japan
E-mail: †{j15501,k-saito,t-ikeda,yuze}@u-shizuoka-ken.ac.jp

Abstract In this paper, we propose a new method for classifying a road network automatically into some areas and characterizing them based on the tri-mixing patterns of node degrees and the K-medoids clustering algorithm. In our experiments, using the three road networks of three cities in Shizuoka prefecture collected from OpenStreetMap (OSM), we show that our method can produce some promising results.

Key words spatial network, degree mixing pattern, area classification and characterization.

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 road networks, whose nodes occupy precise positions in two or three-dimensional Euclidean space, and whose links are real physical connections [2].

In this paper, we address the problem of classifying and characterizing spatial networks in terms of local connection patterns of node degrees [3], 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 [4]. For such spatial networks, we explore techniques for extracting areas with some coherent characteristics. Although this task is somehow relating to community extraction ones widely studied in complex network analysis [5], unlike these general techniques only using network structure, we focus on exploiting the local connection patterns of node degrees.

In this paper, we propose a new method for classifying a road network automatically into some areas and characterizing them based on the tri-mixing patterns of node degrees [3] and the K-medoids clustering algorithm. The proposed method first computes a feature vector for each node, consisting of the normalized frequency of the ego-centric tri-mixing patterns, then classifies these vectors into some areas by the use of the greedy K-medoids clustering algorithm, and finally provide each of the classified areas with the normalized frequency curve of the corresponding medoid vector as its characteristics. In our experiments, using the three road networks of three cities in Shizuoka prefecture collected from OpenStreetMap (OSM), we show that our method can produce some promising results.

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 directional networks. For each node $u \in \mathcal{V}$, we denote its degree by $r(u)$. Now, we consider connected triple nodes, just like network motifs analyses based on triad patterns [4]. 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, for each node $s \in \mathcal{V}$, we can define the egocentric connected triples of the node s by $\mathcal{F}_s = \{(u, v, w) \in \mathcal{F} \mid s \in \{u, v, w\}\}$, and consider a three-dimensional array $\mathbf{C}_s$ whose i-j-kth element $c_s(i, j, k)$ is calculated by $c_s(i, j, k) = |\{(u, v, w) \in \mathcal{F}_s \mid r(u) = i, r(v) = j, r(w) = k\}|$. Thus, for each node $s \in \mathcal{V}$, by arranging these three-dimensional array elements such that $i \leqq k$ and $j \geqq 2$, we can construct a feature vector of the node s expressed as $\mathbf{x}_s = (c_s(1, 2, 1), c_s(1, 2, 2), \cdots)^T$, where $\mathbf{x}_s^T$ means the transposed vector of $\mathbf{x}_s$. Table 1 shows our arrangement of tri-mixing patterns in case that the maximum degree is 5, $max_{v \in \mathcal{V}} r(v) = 5$. Hereafter, each tri-mixing patter is also referred to as its associated ID.

Let $\mathcal{X} = \{\mathbf{x}_u, \mathbf{x}_v, \cdots\}$ be the set of feature vectors computed by the above procedure. Then, we consider classifying $\mathcal{V}$ into K areas, $\mathcal{V}_1, \cdots, \mathcal{V}_K$, by the K-medoids algorithm. More specifically, we select the set of medoids (representative vectors) $\mathcal{R} \subset \mathcal{X}$ whose number of elements is K, i.e., $|\mathcal{R}| = K$, so as to maximize the following objective function

Table 1 List of tri-mixing patterns

ID	Pattern	ID	Pattern	ID	Pattern	ID	Pattern
1	1-2-1	16	1-3-1	31	1-4-1	46	1-5-1
2	1-2-2	17	1-3-2	32	1-4-2	47	1-5-2
3	1-2-3	18	1-3-3	33	1-4-3	48	1-5-3
4	1-2-4	19	1-3-4	34	1-4-4	49	1-5-4
5	1-2-5	20	1-3-5	35	1-4-5	50	1-5-5
6	2-2-2	21	2-3-2	36	2-4-2	51	2-5-2
7	2-2-3	22	2-3-3	37	2-4-3	52	2-5-3
8	2-2-4	23	2-3-4	38	2-4-4	53	2-5-4
9	2-2-5	24	2-3-5	39	2-4-5	54	2-5-5
10	3-2-3	25	3-3-3	40	3-4-3	55	3-5-3
11	3-2-4	26	3-3-4	41	3-4-4	56	3-5-4
12	3-2-5	27	3-3-5	42	3-4-5	57	3-5-5
13	4-2-4	28	4-3-4	43	4-4-4	58	4-5-4
14	4-2-5	29	4-3-5	44	4-4-5	59	4-5-5
15	5-2-5	30	5-3-5	45	5-4-5	60	5-5-5

based on the cosine similarity:

$$f(\mathcal{R}) = \sum_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{r}\in\mathcal{R}} \left\{ \frac{\mathbf{x}^T\mathbf{r}}{\sqrt{\mathbf{x}^T\mathbf{x}}\sqrt{\mathbf{r}^T\mathbf{r}}} \right\}.$$

Then, the greedy K-medoids algorithm can be summarized as follows: after initializing $k \leftarrow 1$ and $\mathcal{R} \leftarrow \emptyset$, we repeatedly select and add each medoid by

$$\mathbf{r}_k \leftarrow \max_{\mathbf{x}\in\mathcal{X}\setminus\mathcal{R}} \{f(\mathcal{R}\cup\{\mathbf{x}\}) - f(\mathcal{R})\},\ \mathcal{R} \leftarrow \mathcal{R}\cup\{\mathbf{r}_k\},$$

during $|\mathcal{R}| \leqq K$ together with increment $k \leftarrow k+1$. From the obtained set of the K medoids, $\mathcal{R} = \{\mathbf{r}_1, \cdots, \mathbf{r}_K\}$, we can compute each classified area as

$$\mathcal{V}_k = \left\{ v \in \mathcal{V}; \mathbf{r}_k = \max_{\mathbf{r}\in\mathcal{R}} \{\rho(\mathbf{x}_v, \mathbf{r})\} \right\}.$$

Finally, we provide each classified area $\mathcal{V}_k$ with the normalized frequency curve of the corresponding medoid vector $\mathbf{r}_k/\sqrt{\mathbf{r}_k^T\mathbf{r}_k}$ as its characteristics.

3. Experiments

In this section, we first explain the datasets used in our experiments, and then show the results obtained by our proposed methods.

3. 1 Datasets

We used OSM data of three cities, Hamamatsu, Numazu and Shizuoka, in Shizuoka prefecture, We extracted all highways and all nodes appearing in them, 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, where the largest connected component of each spatial network was used as our road network. Here we regarded the degree of each node whose degree is greater than 5 simply as 5 because the number of such nodes was quite small, Namely, we used feature vectors as shown in Tab. 1. Table 2 shows the basic statistics of the networks for the three cities.

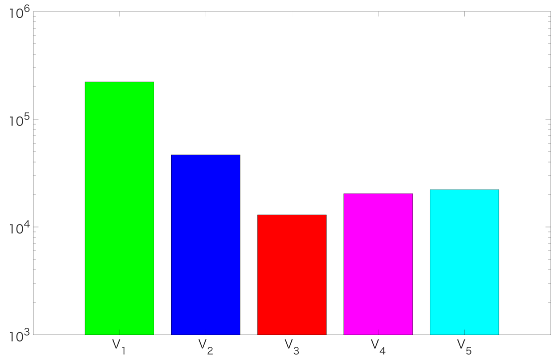
Table 2 Basic statistic as network

No	City	$\|V\|$	$\|E\|$
1	Hamamatsu	323,457	346,294
2	Numazu	44,980	48,550
3	Shizuoka	195,786	208,329

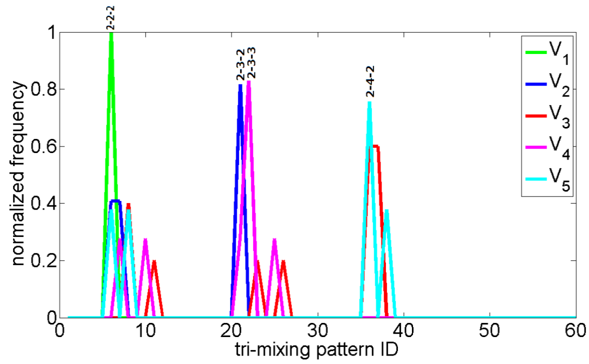
3. 2 Results

We applied our proposed method to road networks of three cities by setting the number of the classifying areas to five, i.e., $K = 5$. Figures 1, 2, and 3 show our experimental results for the three cities, Hamamatsu, Numazu, and Shizuoka, where Figs. 1(a), 2(a), and 3(a) show the numbers of nodes in the classified areas, $\mathcal{V}_1$, $\mathcal{V}_2$, $\mathcal{V}_3$, $\mathcal{V}_4$, and $\mathcal{V}_5$, and Figs. 1(b), 2(b), and 3(b) show the normalized frequency curves for these areas. From these figures, we can observe the following three similar characteristics. First, as for the $\mathcal{V}_1$ areas colored in green in all the three cities, they contains the largest numbers of nodes as shown in Figs. 1(a), 2(a), and 3(a), and their normalized frequency curves have only one peak at the 2-2-2 pattern (pattern ID 6) as shown in Figs. 1(b), 2(b), and 3(b), Second, as for the the $\mathcal{V}_2$ area colored in blue in Hamamatsu, the $\mathcal{V}_4$ area colored in magenta in Numazu, and the $\mathcal{V}_5$ area colored in cyan in Shizuoka, they contains the second largest numbers of nodes, and their normalized frequency curves have the only one largest peak at the 2-3-2 pattern (pattern ID 21), Third, as for the $\mathcal{V}_3$ areas colored in red in all the three cities, although their orders in terms of the numbers of nodes in these areas are different, their normalized frequency curves are clearly characterized by containing many peaks in comparison to the other areas.

Figure 4, 5, and 6 show the visualization results by plotting each node at its position with a color of its classified area as shown in Figs. 1, 2, and 3 for Hamamatsu, Numazu, and Shizuoka, respectively, We consider that these visualization results are reasonably interpretable in the sense that several areas such as urban, mountainous and intermediate regions were clearly classified by using different colors. Below we explain three examples for clarifying characteristics of the classified areas. First, as for the $\mathcal{V}_1$ areas colored in green in all the three cities, which were characterized by both the largest numbers of nodes and the only one peak at the 2-2-2 pattern as described above, we can see that each of these areas spread all over the city as shown in Figs. 4, 5, and 6. Thus, we can naturally interpret characteristics of these areas as containing a large numbers of curve-fitting-points, which typically appear in mountainous regions. Second, as for the the $\mathcal{V}_2$ area colored in blue in Hamamatsu, the $\mathcal{V}_4$

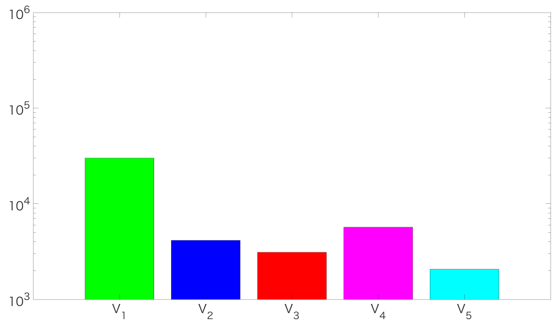


(a) Numbers of nodes in classified areas

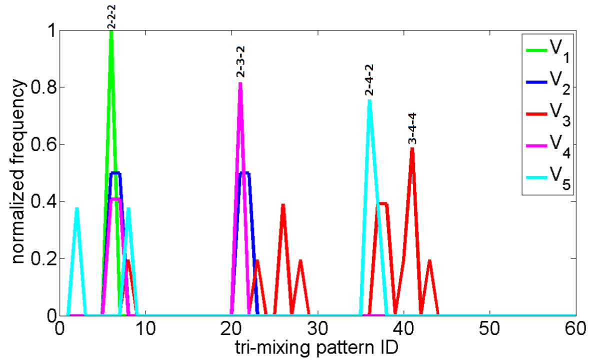


(b) Normalized frequency curves

Figure 1 Hamamatsu City

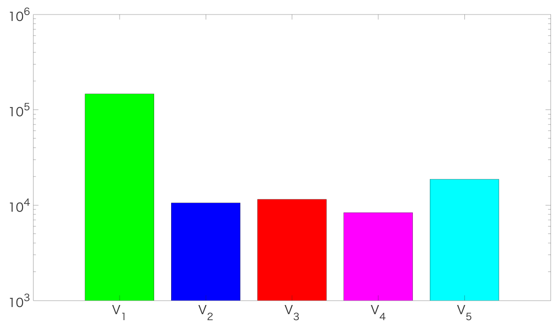


(a) Numbers of nodes in classified areas

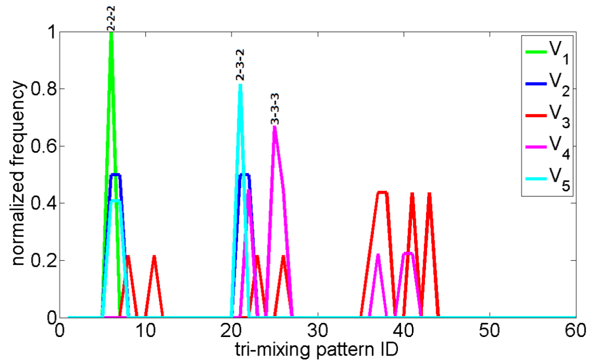


(b) Normalized frequency curves

Figure 2 Numazu City



(a) Numbers of nodes in classified areas



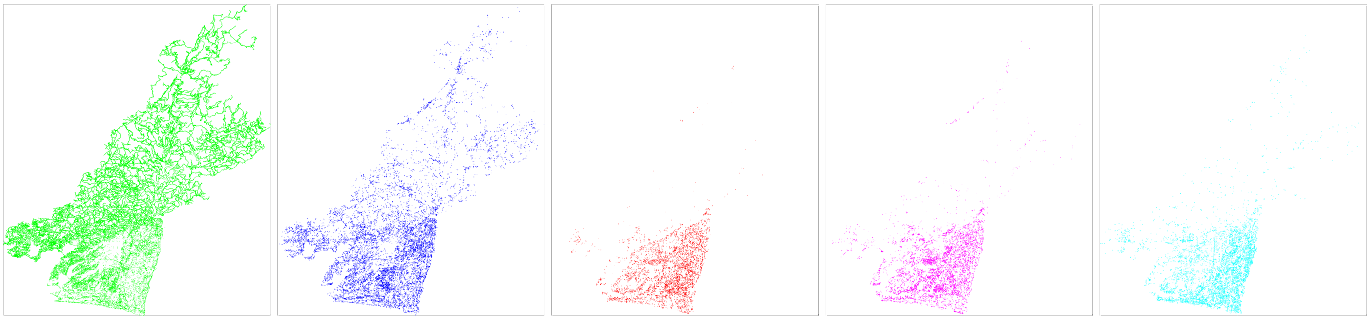
(b) Normalized frequency curves

Figure 3 Shizuoka City

area colored in magenta in Numazu, and the $\mathcal{V}_5$ area colored in cyan in Shizuoka, which were characterized by both the second largest numbers of nodes and the only one largest peak at the 2-3-2 pattern, we can see that each of these areas also spread all over the city. Thus, we can naturally interpret characteristics of these areas as containing a large numbers of crossroads like T-junctions. which typically appear in intermediate regions. Third, as for the $\mathcal{V}_3$ areas colored in red in all the three cities, which were characterized only by the many peaks in comparison to the other areas, we can see that each of these areas spread relatively small regions of the city. Thus, we can naturally interpret characteristics of these areas containing wide variety of junctions as local connection patterns. which typically appear in urban regions. In short, from these experimental results, we can confirm that our proposed method can be a vital and useful tool for reasonably measuring the characteristics of cities.

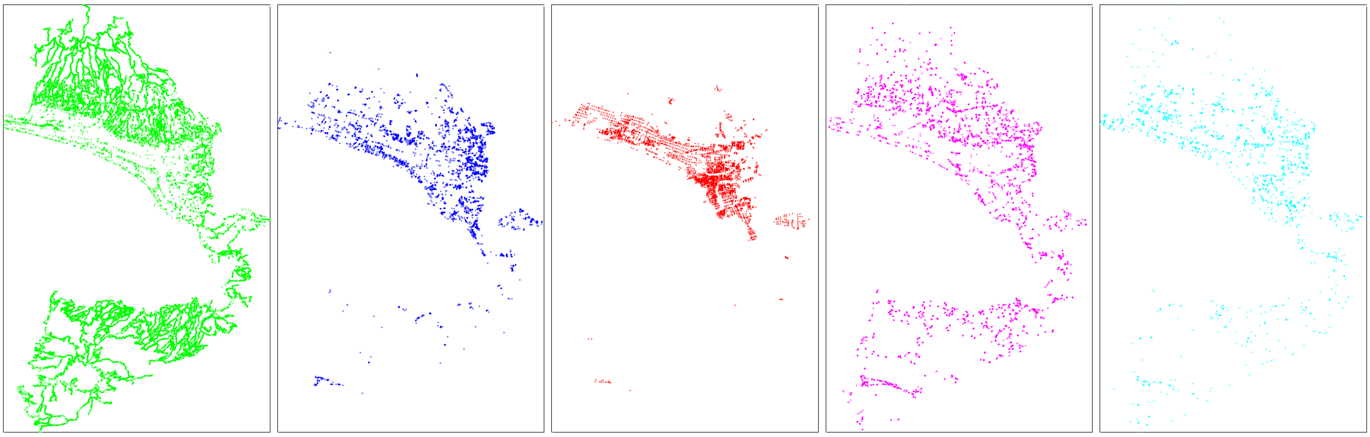
4. Conclusion

In this paper, we propose a new method for classifying a road network automatically into some areas and characterizing them based on the tri-mixing patterns of node degrees and the K-medoids clustering algorithm. In our experiments using the three road networks of three cities in Shizuoka pre-



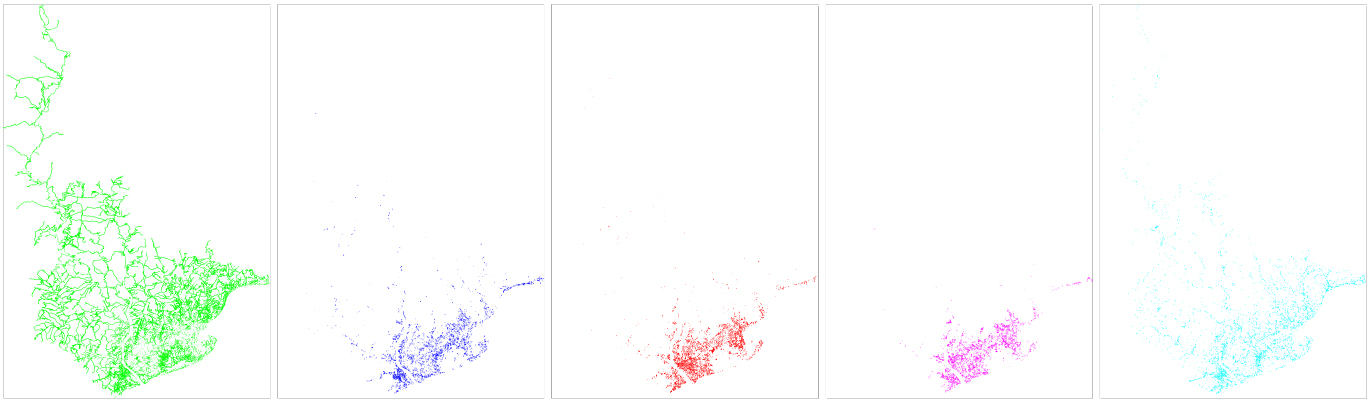
(a) $\mathcal{V}_1$ (b) $\mathcal{V}_2$ (c) $\mathcal{V}_3$ (d) $\mathcal{V}_4$ (e) $\mathcal{V}_5$

Figure 4 Cluster distribution in Hamamatsu



(a) $\mathcal{V}_1$ (b) $\mathcal{V}_2$ (c) $\mathcal{V}_3$ (d) $\mathcal{V}_4$ (e) $\mathcal{V}_5$

Figure 5 Cluster distribution in Numazu



(a) $\mathcal{V}_1$ (b) $\mathcal{V}_2$ (c) $\mathcal{V}_3$ (d) $\mathcal{V}_4$ (e) $\mathcal{V}_5$

Figure 6 Cluster distribution in Shizuoka

fecture collected from OpenStreetMap (OSM), we showed that our method can be a vital tool for producing promising results. In future, we plan to evaluate our method using various road networks.

Acknowledgements

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

References

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

[2] 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).

[3] 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).

[4] 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 (online), 10.1126/science.298.5594.824 (2002).

[5] Newman, M. E. J. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol. 69, p. 026113 (2004).