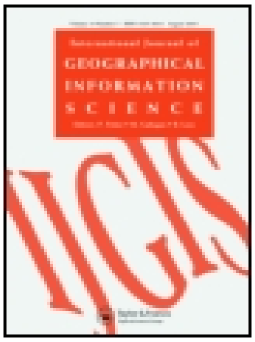

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Hierarchical community detection and functional area identification with OSM roads and complex graph theory

Ye Hong & Yao Yao

To cite this article: Ye Hong & Yao Yao (2019): Hierarchical community detection and functional area identification with OSM roads and complex graph theory, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2019.1584806

To link to this article: https://doi.org/10.1080/13658816.2019.1584806

Published online: 23 Apr 2019.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2019.1584806

RESEARCH ARTICLE

Hierarchical community detection and functional area identification with OSM roads and complex graph theory

Ye Hong [a,b] and Yao Yao [a,c]

[a]School of Geography and Information Engineering, China University of Geosciences, Wuhan, China; [b]Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zürich, Switzerland; [c]Department of Data Technology and Products, Alibaba Group, Hangzhou, China

ABSTRACT

An in-depth analysis of the urban road network structure plays an essential role in understanding the distribution of urban functional area. To concentrate topologically densely connected road segments, communities of urban roads provide a new perspective to study the structure of the network. In this study, based on OpenStreetMap (OSM) roads and points-of-interest (POI) data, we employ the Infomap community detection algorithm to identify the hierarchical community in city roads and explore the shaping role roads play in urban space and their relation with the distribution of urban functional areas. The results demonstrate that the distribution of communities at different levels in Guangzhou, China reflects the urban spatial relation between the suburbs and urban centers and within urban centers. Moreover, the study explored the functional area characteristics at the community scale and identified the distribution of various functional areas. Owing to the structure information contained in the identification process, the detected community can be used as a basic unit in other urban studies. In general, with the community-based network, this study proposes a novel method of combining city roads with urban space and functional zones, providing necessary data support and academic guidance for government and urban planners.

ARTICLE HISTORY
Received 28 July 2018
Accepted 16 February 2019

KEYWORDS
Network science; geospatial big data; road network; urban functional area; community detection

1. Introduction

The mining of information hidden in urban and regional road networks has a long-established history in the fields of traffic geography, land use planning, and economic geography (Porta *et al*. 2006). In an era of unprecedented global urbanization, the continuous improvement of the urban road system also sustains immense social mobility pressures (Batty 2008) and plays a vital role in urban development (Wang *et al*. 2012). As the artificial corridor in the urban landscape, the road network is the framework of urban development and an intrinsic factor of urban space expansion. Hence, the exploration of the road network structure not only helps to interpret the topology of the road network but also contributes to the understanding of the organization and

CONTACT Yao Yao yaoy@cug.edu.cn

spatial distribution of the urban system (Wang *et al*. 2011, Levinson and Moreno 2012), thereby analyzing its development potential

Studies on the structure of urban road networks have a longstanding history. Previous studies on urban road networks often regard the system as a graph (Garrison and Marble 1962), where each link stands for a section of the street and the nodes represent road intersections. This network abstraction model is the classic method of road network processing in traffic simulation (Rodrigue *et al*. 2009) and urban modeling (Zheng *et al*. 2014), but it is incapable of revealing the latent network structure and mode (Jiang 2007). There exist many studies that concentrate on the road network structure (Xie and Levinson 2007, Zhang and Li 2012, Barthelemy *et al*. 2013) and its impact on the social-economic environment (Porta *et al*. 2009, Zhang *et al*. 2015). For example, Duan and Lu (2014) studied the robustness of road networks at different scales (Duan and Lu 2014); Samaniego and Moses (2008) focused on measuring the accessibility of roads in different cities and revealed the evolution of roads and traffic along with urban development (Samaniego and Moses 2008); Levinson and Moreno (2012) analyzed the road network size and structure of a range of cities, and suggest that larger cities are physically more inter-connected (Levinson and Moreno 2012). Primarily based on node centrality analysis, these works tend to treat and evaluate the road nodes and segments independently, without considering the relationship between road segments.

Recently, the development of network science (Cohen and Havlin 2010, Estrada 2012) and the creation of a large number of large-scale datasets have promoted research on road networks (Barthélemy 2011, Strano *et al*. 2018). To perceive segments with specific similar characteristics in the network, the method of community detection came into being (Hric *et al*. 2014). Communities, also known as clusters or modules, are sets of nodes that are strongly linked to each other and sparsely connected to the rest of the network (Gulbahce and Lehmann 2008). With community detection algorithms, communities within the original network can be identified, and the structure of the reconstructed network can be studied, reflecting the characteristics of the original network. Significative community structures have been discovered in a range of social and information networks (Fortunato and Hric 2016). Likewise, there exist modules like traffic analysis zones (TAZ) in the network formed by urban roads, in which the arrangement of the nodes differs from other parts of the network. Previous studies indicated that urban functional zones have a close relation to the regions formed by the city road network (Yuan *et al*. 2012, Liu *et al*. 2017), and exploring this relation from the community perspective will undoubtedly bring new insights. Although some researchers have focused on studying road network structure from the community perspective (Song and Wang 2011, Duan and Lu 2013, Zhao *et al*. 2017), community studies in road networks are still insufficient, and their potential remains to be tapped.

In this study, via identified communities in the city road network, we propose a novel way to link the urban roads with the urban space and urban functional zones. Specifically, based on the urban road network model constructed with OpenStreetMap road data, the Infomap community detection method is employed to detect the hierarchical communities and analyze their spatial distribution at different levels. In addition, points-of-interest (POI) data and their indicators are used to reveal the land use and functional area density, the mixing degree and the enrichment status of the community-based network, exploring the relations between communities within road network and urban functional zone distributions.

2. Study area and data description

We selected Guangzhou, China as our study area (Figure 1). Located in the south-central part of Guangdong Province and the northern edge of the Pearl River Delta, Guangzhou is the political, economic and cultural center of Guangdong Province and one of China's most important economic development centers. The sub-provincial city has direct jurisdiction over eleven districts: Liwan, Yuexiu, Haizhu, Tianhe, Baiyun, Huangpu, Panyu, Huadu, Nansha, Conghua, Zengcheng, with a total area of 7,435 km^2.

The road network in Guangzhou has a long-developing history and is becoming increasingly mature (Liu *et al*. 2011), which makes it an ideal place for exploring the road network structure and identify the resulting communities. At the same time, Guangzhou is known as one of the most prosperous cities in China and possesses a sophisticated urban morphology and functional areas (Chen *et al*. 2017). For example, a single land parcel may be equipped with mixed types of urban functions, such as residential, commercial, and medical facilities, posing challenges for functional area identification at the community level.

The administrative boundary data used in this study was derived from the Database of Global Administrative Areas (GADM) (http://www.gadm.org/). After projection transformation and topology checking, the administrative division data of different scales are used in the preprocessing, calculation and analysis of other experimental data.

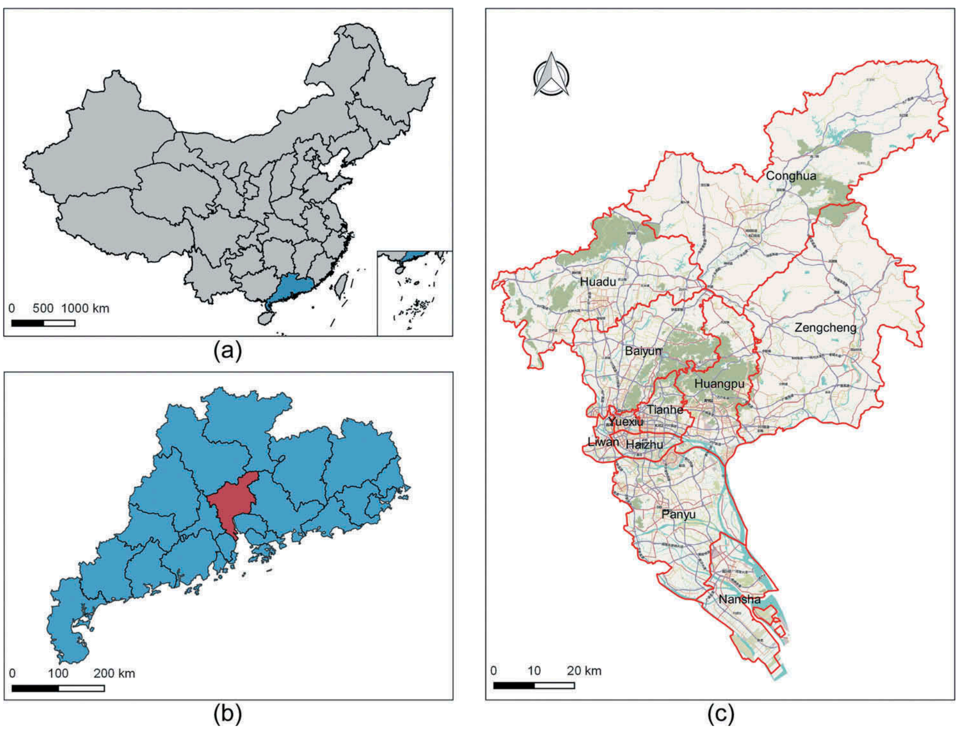

Figure 1. The study area and its location in Guangdong Province and China: (a) the geographic location of Guangdong Province in China; (b) the geographic location of Guangzhou City in Guangdong Province; and (c) the study area Guangzhou and its 11 jurisdictions.

The primary dataset for this study is the road data obtained from OSM, which was acquired in January 2018 (http://www.openstreetmap.org). OSM is an open source map provider that aims to provide users with free and easily accessible digital map resources (Haklay 2010) and is considered to be the most successful and prevailing volunteered geographic information (VGI) at this stage (Fan *et al*. 2014). The accuracy of the positioning and the topological relationship of the OSM roads in the study area is exceptionally high (Yao *et al*. 2018). In addition to the fundamental spatial information such as latitude and longitude, the road data obtained from the OSM contains additional attribute information, including road type, maximum travel speed, and one-way streets, which contribute to preprocessing and establishing the road network. After preprocessing operations, such as simplification, merging, and topology inspection, 81,377 roads and 60,756 nodes were extracted (Figure 2).

As one of the fundamental data types derived from location-based services (LBS), POIs are used to ascertain functional areas at the community level in the study. POI data, which record the geospatial and attribute information of points, can reflect the land use pattern at a fine scale and capture the types of land functions that cannot be represented in urban master planning, thus effectively presenting the functional zones, human mobility and economic development features of the region (Yuan *et al*. 2012, Jiang *et al*. 2015). As shown in Figure 3, the POIs used in this study were acquired from the Gaode map, one of the most popular map providers in the Chinese market (http://lbs.amap.com/) (Meng *et al*. 2017). Through the application programming interface (API) provided by Gaode Map, this study obtained a total of 1,181,475 POIs records across 20 categories in Guangzhou.

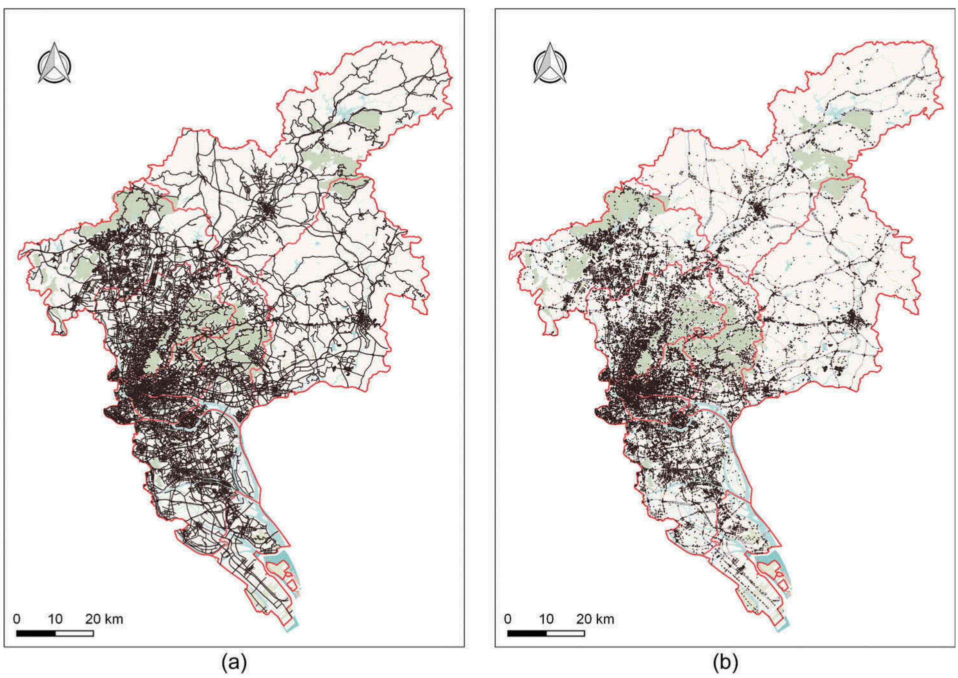

Figure 2. Roads (a) and road nodes (b) delivered from OSM data.

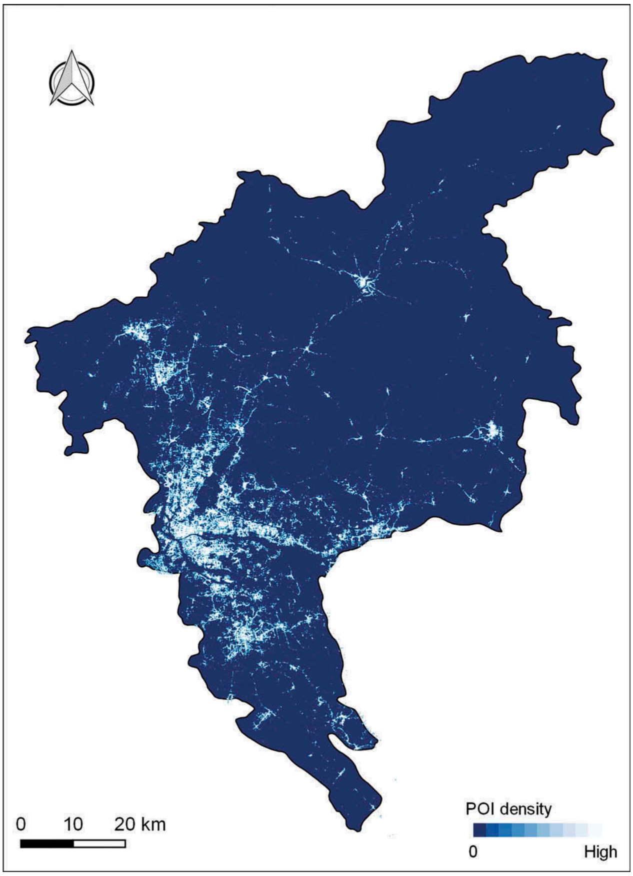

Figure 3. The spatial distribution of Gaode POIs in Guangzhou.

The POI data were reclassified into eight categories according to their category code (Table 1): POIs that belong to life services (LIS) (e.g. life service, shopping sites, catering services) were the most prevalent, followed by office building/space (OBS), other facilities (OTH), medical and education (MED), entertainment (ENT) (e.g. tourist sites, hotels), government (GOV), residence communities (RES), and financial services (FIN). Although POIs belonging to OTH were taken into account in the analysis of land use density, they

Table 1. The type and number of gaode POIs after reclassification.

Type	Abbreviation	Counts
Life services	LIS	691,083
Office building/space	OBS	196,555
Other facilities	OTH	136,256
Medical/Education	MED	51,228
Entertainment	ENT	37,579
Government	GOV	26,153
Residence communities	RES	26,117
Financial services	FIN	13,828

were excluded from the mixed land use analysis because they do not reflect the specific functional characteristics of the area (Liu and Long 2015).

3. Methodology

Based on the theory and method of complex graph theory, this study aims to describe the structure of the road network in Guangzhou quantitatively. We explore the community formed by the urban roads at different levels and identify the functional areas from a community perspective. This study can be divided into the following steps: 1. Establish an urban road network model using the processed OSM roads. 2. Divide the road network model into multilevel communities based on the Infomap method. 3. Depict the functional density, mixed and enrichment degree of each community with Gaode POIs. The flow chart of the study is shown in Figure 4.

3.1. *Urban road network model construction*

The processed OSM roads are abstracted into a weighted directed graph (Corcoran *et al.* 2013, Yao *et al.* 2018). A weighted directed graph $G \equiv (V, E, W)$ consists of a set of vertexes V, their connecting edges E and the weight of each edge W. In this study, the starting points, the ending points and the intersection of the roads constitute the vertexes V; the road segment connecting these vertexes is regarded as the edge E; the weight W of each edge is set to the geographical length of the road; and the direction of the edge is determined by the 'one-way street' attribute of the OSM road. If the attribute indicates that the road is unidirectional, we will only consider the travel direction of the road; otherwise, both directions of the road will be added to the graph. In order to store data more efficiently and concisely, this study employes an adjacency list to store the topology of the network (Gao *et al.* 2016).

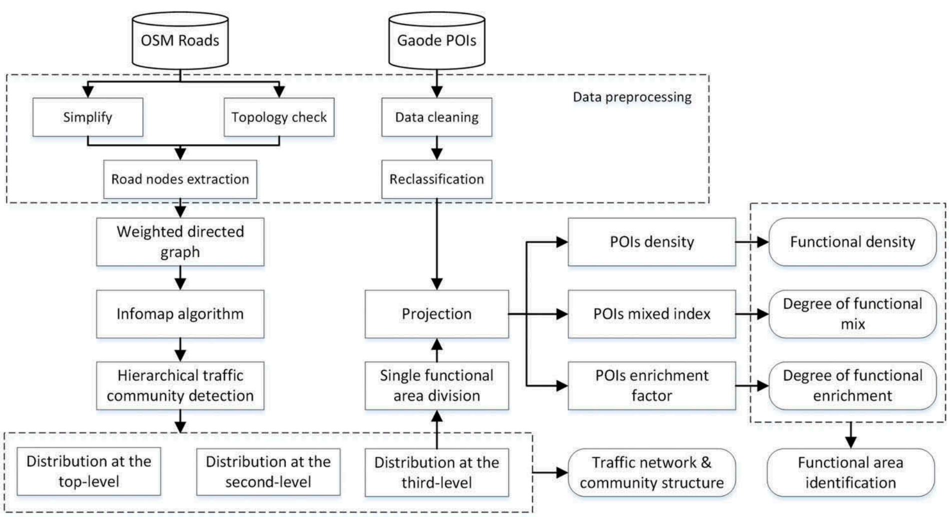

Figure 4. The flowchart of community detection and functional area identification.

3.2. *Community detection with Infomap algorithm*

To comprehensively understand the structure of the road network, in addition to omitting some secondary information and abstracting it into a graph, we also need to choose an appropriate method to detect the hidden structure in the network (Rosvall and Bergstrom 2008, Leskovec *et al*. 2010). For different types of networks and application scenarios, previous studies have proposed community detection methods based on different principles (Tang and Liu 2010). Among them, the Infomap method has been used in various networks due to its rigorous structure and high availability (Rosvall and Bergstrom 2008). Also, among all the hierarchical clustering models, the Infomap method has proven to be one of the best performing nonoverlapping clustering models (Lancichinetti and Fortunato 2009). In this study, we adopt the Infomap method to identify the hidden communities in the road network.

Based on a combination of information-theoretic techniques and random walks, Infomap treats the probability flow of random walks as a proxy for flows in the real system. Infomap introduces Huffman coding (Huffman 1952) and describes each node in the network with a two-level description – a prefix and a suffix. The prefix indicates the module in which the node is located, and the suffix represents the number of the nodes in the module. In the whole network, each module is assigned with unique names, but a different Huffman code is used to name the nodes within each module. Based on this, Infomap can combine the finding of community structure and the coding problem; the algorithm looks for a partition M into m modules to minimize the expected description length of a random walk (Rosvall and Bergstrom 2008).

The average description length $L(\mathrm{M})$, given a partition M, is defined as:

$$L(\mathrm{M}) = q_{\curvearrowright} H(\zeta) + \sum_{i=1}^{m} p^{i}_{\circlearrowright} H(\mathcal{P}^{i}) \tag{1}$$

This formulation consists of two parts: the entropy of movements between modules and the entropy of movements within each module. $q_{\curvearrowright}$ is the probability of the random walk switching modules at each step; $H(\zeta)$ represents the entropy of the module's encoding (i.e. the prefix); $H(\mathcal{P}^{i})$ is the entropy of each movement's encoding (i.e. the suffix); and $p^{i}_{\circlearrowright}$ is the probability of being accessed for module i, including the access probability of each node in module i and the probability of exiting module i (Rosvall and Bergstrom 2008).

Through a deterministic greedy search algorithm, the study can efficiently solve Formula (1), thus obtaining an optimal two-layer network community division (Wakita and Tsurumi 2007). On this basis, by extending the two-level description to the multi-level description, we can reveal the hierarchical structure in the network and its relationship (Rosvall and Bergstrom 2011).

In this study, based on the built urban road network, the Infomap algorithm is employed to divide the road network into multiple levels to reveal the rich hierarchical features in the network. In particular, the Infomap method does not require preset parameters, so that the result of network segmentation is wholly determined by the nature and topological relationship of the network.

3.3. *Functional area identification with POI data*

The Infomap algorithm hierarchically divides each node in the network. The Thiessen polygon method is employed to determine the coverage area of the modules at each level, thus continuing the exploration of the module's functional space.

In this section, we explore the urban function types of each community. The study first calculates the POI density, which reflects the density of all urban functions (Li *et al.* 2011). In a community i, the ratio between all POIs $Count_i$ and the community area $Area_i$ is defined as the $Density_i$:

$$Density_i = \frac{Count_i}{Area_i} \tag{2}$$

To reveal the mixing degree of urban functions in a specific community, the study used the mixed index that portrays mixed land use (Frank *et al.* 2004). The mixed index $Mixed_i$ for the community i is given by:

$$Mixed_i = -\sum_{i=1}^{n}(p_i \times \ln p_i) \tag{3}$$

where n stands for the number of POI types (we excluded POIs that belong to 'Other facilities', thus $n = 7$); and p_i is the proportion of type i POIs in the total POIs. The index is widely used to identify mixed urban land use, understand changes in travel patterns, and community emotional migrations (Manaugh and Kreider 2013, Yue *et al.* 2016).

We also introduced the POI enrichment factor as auxiliary discriminant information to determine the aggregation degree of a particular functional area in the community (Verburg *et al.* 2004). The enrichment factor $E_{i,j}$ of the POIs of type j in the i-th community is expressed as:

$$E_{i,j} = \frac{n_{i,j}/n_i}{N_j/N} \tag{4}$$

where $n_{i,j}$ is the number of POIs type j in the i-th community; n_i is the total POI number within the i-th community; N_j represents the total number of POIs of type j; and N is the total number of POIs in the study area. A higher $E_{i,j}$ represents a higher aggregation degree of POIs type j in the i-th community. In order to measure the spatial autocorrelation of the different functional areas, we calculate the Moran's I index and its corresponding z-scores and p-values (Li *et al.* 2007).

By calculating these indicators, the study can quantitatively identify and analyze the spatial distribution of urban functional areas on the community-based network.

4. Results

4.1. *Hierarchical community detection*

The analyses detect the multilayered community structure of the original road network in the study area, and the statistical results after the division are shown in Table 2. The Infomap solution divides the 60,756 original nodes of the Guangzhou road network into 15,617 lowest-level communities. The average depth of a single node is 6.60, showing that the nodes are divided into highly fine scales, which also represent the characteristics of a complex network structure and low regional connectivity. The average number of nodes in a single lowest-level module is 3.89, which is less than those found in other common networks (Rosvall and Bergstrom 2011). Moreover, the network compression ratio is 21.14%. Compared with the amount of encoded information in the original network, the amount of data after the hierarchical division is significantly reduced, which intuitively reflects the significance of using Infomap to divide the road network.

Since road networks represent real geographical entities, and the geographical constraints hinder the fast and direct connection between the remote parts of the network (Strano *et al*. 2018), the system possesses a higher per-node average depth and significant hierarchical features (Rosvall and Bergstrom 2011). This limitation makes the spatial location the dominant factor in the community detection of the road network, i.e. nearer nodes in space are more easily partitioned into the same community than those with a certain distance. At the same time, the direction and interconnection of the roads will also affect the results of community detection, which is more significant at fine levels.

Figure 5 illustrates the communities at the top and second level, reflecting the community division of the Guangzhou road network at a higher scale. Table 3 compares the community divisions at different levels and the administrative districts of the subdistrict level. Compared with the subdistrict level, the proposed community detection method can generate more modules with a smaller average area and produce more fine-scaled urban division results.

From the statistical results of the top-level community, 93.91% of the nodes are concentrated in the first ten modules. Figure 5(a1) maps the classification result of the nodes in this level, where a small number of communities can represent most of Guangzhou's area. Except for Nansha District, the communities in other places are inconsistent with the division of the district-level administrative zones, indicating that there is no apparent road robustness and connectivity difference between the districts, and using district-level administrative areas cannot divide the existing roads in Guangzhou. In addition, the downtown area of Guangzhou (Yuexiu, Liwan, Tianhe, and Haizhu) (Figure 5(a2)) is divided into different communities along with the neighboring suburbs. The majority of Haizhu forms the south community with Panyu; the eastern part of Yuexiu and the southern part of Tianhe constitute the east community together with the southern part of Huangpu and Zengcheng; Liwan, the western part of Haizhu and Yuexiu constitutes the west community with the southern part of Baiyun District.

Table 2. The statistical results of community detection in Guangzhou's road network.

Indexes	Guangzhou roads
Nodes	60,756
Links	81,377
Total number of modules	15,617
Per-node average depth	6.60
Per-node average size of the lowest-level module	3.89
Compression ratio	21.14%

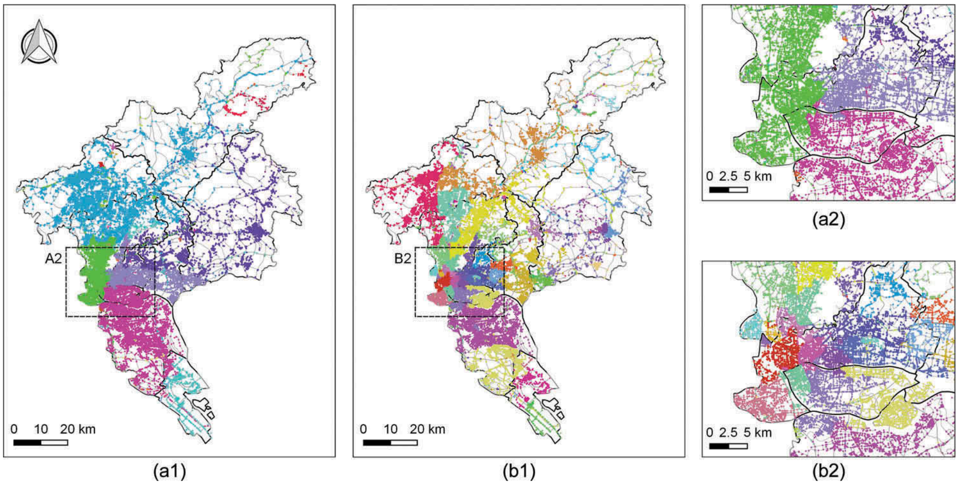

Figure 5. The spatial distribution of communities at high levels: (a1) top-level and (b1) second-level results of Guangzhou, (a2) top-level and (b2) second-level results of central districts in Guangzhou.

Table 3. Comparison of different levels of the community and the administrative district of the subdistrict level.

Partition method	Number of modules	Mean area (km^2)	Proportions of nodes in top 10 modules
Top-level community	963	7.46	93.91%
Second-level community	2,227	3.24	47.97%
Third-level community	3,712	1.95	6.11%
Subdistrict level	166	43.57	-

Meanwhile, with a more detailed division, the modules' mean area of the second-level community is reduced to 3.24 km^2. Nodes in the top ten modules account for 47.97%, which implies that nodes are scattered among different communities (Figure 5(b1)). Compared with the top-level area, the entire study area is divided more meticulously, and the community better reflects the relation between the interior of the central city and the various parts of the suburban areas (Figure 5(b2)).

Table 3 shows that more modules can be identified in the third-level area, and the road nodes are further dispersed in different communities. Figure 6 illustrates the results of the third-level community in Guangzhou, the central city area, and Yuexiu District. Although the community areas are inconsistent in regions with different road densities, the partitioning results can represent typical functional areas of the city to a certain extent. For example, Figure 6(a1) is the Baiyun International Airport; B2 is Shamian, a famous historical tourist attraction; B3 is Lujiang and Datang Village, urban villages at the southern end of Guangzhou's central axis; B4 is the South China University of Technology and the South China Agricultural University; C5 is the Beijing road business district, a bustling commercial center located in the heart of Yuexiu district; and C6 is Ersha island, known as a senior housing district in Guangzhou. Communities detected by Infomap method represent areas that are internally well-connected but externally less so, which indicate land blocks with certain urban functions. Hence, the identified communities have the potential to assist in understanding the spatial distribution of typical functional blocks in the

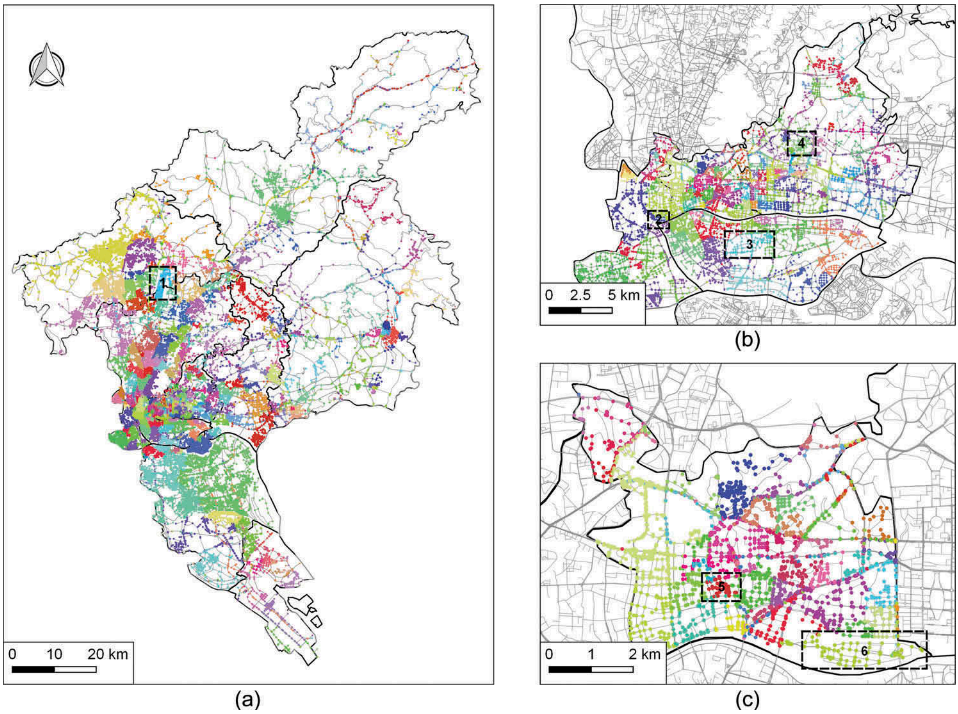

Figure 6. The spatial distribution of communities at the third-level: (a) Guangzhou, (b) the Guangzhou downtown area, and (c) the Yuexiu District. Some identified functional blocks are also shown: (a1) Baiyun International Airport, (b2) Shamian, (b3) Lujiang and Datang Village, (b4) the South China University of Technology and the South China Agricultural University, (c5) Beijing road business district, and (c6) Ersha island.

city. This feature lays the foundation for the functional area identification using the community-based network.

By observing the obtained community detection results (e.g. Figure 6), we can find that high-grade roads, such as the Guangzhou Ring Expressway, Xinguang Express Road, and Airport Express, play a crucial role in shaping the community. Some communities are distributed along high-grade roads, while more distant areas may also constitute the same community because of the connection of high-grade roads. Moreover, the emergence of high-grade roads also produces fine-grained community patches around larger communities, mainly because high-grade roads link communities with a specific spatial distance and simultaneously bring spatial heterogeneity to the original uniform communities. These minor community patches surrounded by other communities play an essential role in understanding regional heterogeneity.

In general, by using the Infomap method to reveal the hierarchical characteristics of the Guangzhou's road network, the study explores the hidden communities of the existing roads and their role in dividing the urban space at different levels. The proposed results at higher levels can reflect not only the connection between the city center and the suburbs but also the regional division within the city center, providing a reference for understanding urban space and internal activities.

4.2. Urban functional pattern recognition at the community level

Using the Thiessen polygon method to estimate the influence range of network nodes, the study can generalize the hierarchical node division to the entire research area. With POI data, this section quantitatively describes the relevant indicators of the urban function in each community. The conclusions of the previous section show that the third-level community can identify typical functional blocks of the city, so we regard the third level's result as the primary study unit in this section.

Figure 7 shows the POI density calculated using community as the basic unit. It can be found that the POI density in Guangzhou gradually decreases from the city center, showing distinct annular layer characteristics. The central districts possess the highest functional density, which steadily declines in the suburban area (Figure 7(b)). The discrete high-value POI density zones (Figure 7(c)) the center of Huadu District and Figure 7(d) the center of Zengcheng District) represent functional gathering areas away from the core of downtown and have a particular role in sharing and decentralizing the function of the city's main center.

We calculate the mixed index of POIs to reflect the functional mix degree of the various communities (Figure 8). The study only considers communities that contain more than 50 POIs in order to reduce the impact of data volume differences on the results. As a result, 856 communities were considered, accounting for 75.95% of the total area of the study area. Table 4 shows the statistical results of the mixed index in various districts of Guangzhou. From the results of the central city, as the downtown and

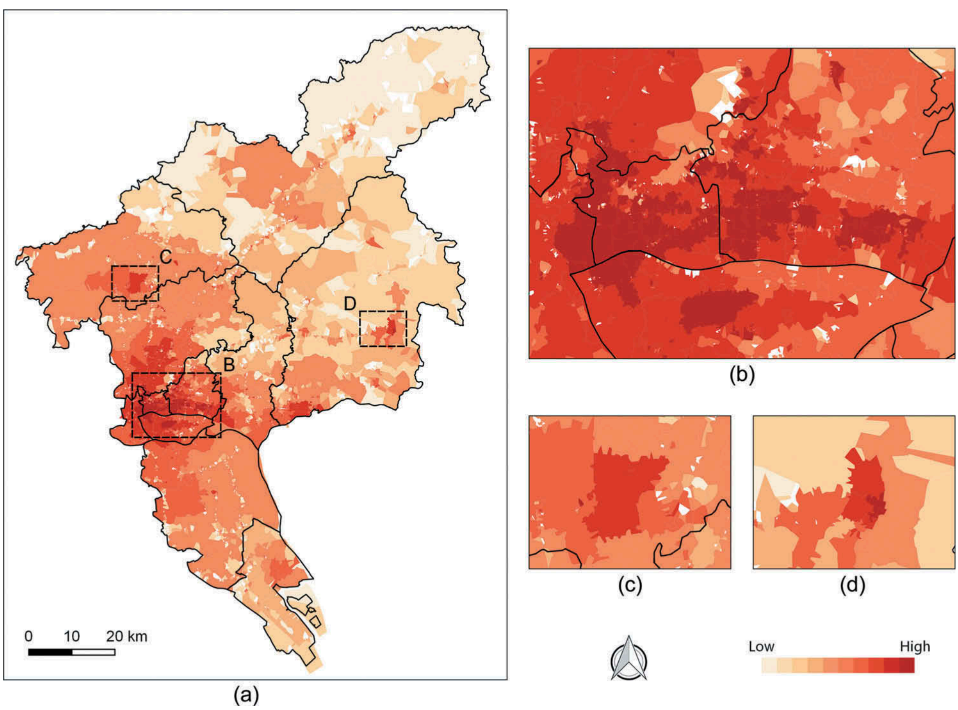

Figure 7. The POIs density of (a) Guangzhou, (b) Central districts of Guangzhou, (c) Center of Huadu District, (d) Center of Zengcheng District.

political and economic center, the diverse functional degree of the majority communities in Yuexiu and Tianhe Districts are higher (the average mixed index is 1.097 and 1.035, respectively), while that of Liwan and Haizhu Districts, known as the residential and industrial zone, are relatively low (with an average of 0.940 and 0.923, respectively). It is worth noting that Nansha District ranked the highest among all the districts in land-use mix level, with an average mixed index reaching 1.144, which is because Nansha has been the satellite town of Guangzhou and possesses complete and diverse urban function types.

Moreover, we introduce the POIs enrichment factor to determine the dominance degree of each functional category in each community. Figure 9 shows the enrichment factor of the seven POI categories, except for class OTH, in the entire study area and urban center. An enrichment index greater than one indicates that the enrichment degree of the current function in the community is higher than that of the entire city. Various functional area classes are clustered in the communities, with a high POIs mixed index in Yuexiu and Tianhe Districts. Among them, OBS and FIN are enriched in the Tianhe Central Business District (CBD) (average enrichment index > 1.41), while RES and

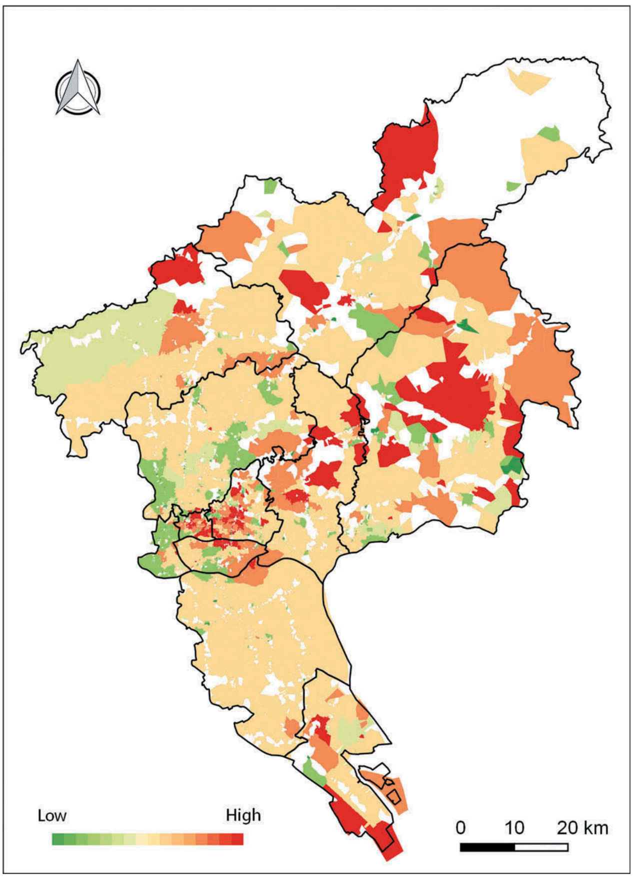

Figure 8. The distribution of POIs mixed index in Guangzhou.

Table 4. The POIs mixed index indicators of the communities in each district, ranked according to the average mixed index from large to small.

District	Community number	Average mixed index	Standard Deviation of mixed index
Nansha	27	1.144	0.245
Yuexiu	108	1.097	0.341
Huangpu	67	1.082	0.245
Tianhe	221	1.035	0.335
Conghua	58	1.026	0.300
Zengcheng	117	1.005	0.273
Huadu	45	0.991	0.250
Panyu	61	0.959	0.232
Haizhu	85	0.940	0.298
Liwan	54	0.923	0.274
Baiyun	196	0.878	0.273

MED are gathered on the periphery of the Tianhe CBD area (average enrichment index > 2.02). The ENT functional area is less enriched in the CBD zone (average enrichment index < 0.90) and more concentrated in the near-central area distant to the CBD (i.e. the higher education mega center of Guangzhou). The distribution of these functional areas reflects the typical structure of the 'work – residence – leisure' in Guangzhou (Yao *et al.* 2019). Meanwhile, LIS is sparsely distributed in Yuexiu and Tianhe Districts, and more concentrated in Liwan and Haizhu. The result of GOV reflects the distribution of government functions in Guangzhou.

Table 5 illustrates the Moran's I index value and z-score and p-value of different functional areas. Given the z-score and p-value, except for the pattern of GOV, which does not appear to be significantly different than random, the spatial distribution of all

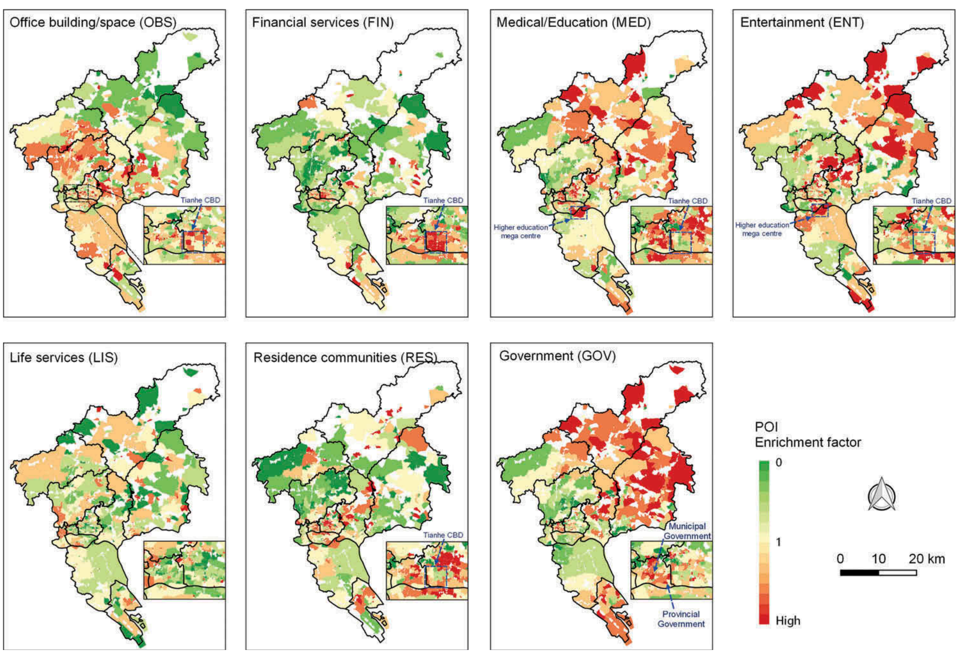

Figure 9. POI enrichment factors of functional areas in Guangzhou.

Table 5. The Moran's I index value, z-score and p-value of different functional areas.

Functional areas	Moran's I	z-score	p-value
OBS	0.30	5.56	0.00
FIN	0.51	5.35	0.00
MED	0.12	2.14	0.03
ENT	0.29	4.62	0.00
LIS	0.40	7.93	0.00
RES	0.24	3.63	0.00
GOV	0.04	0.50	0.62

other categories is very likely to be clustered. Moreover, all the functional areas except for GOV possess a positive spatial correlation, with FIN and LIS being the highest.

5. Discussion

Community detection in a road network provides a novel way to gather topologically densely connected road segments (Duan and Lu 2013). Moreover, hierarchical community detection can reveal modules generated by the network structure from macro to a finer scale. The detected community at different levels enables us to explore the partition of the urban space and could assist in understanding the urban functions and the resulting mobility behavior at different spatial scales. For example, the community detection results for the top- and second-level demonstrate the connection between urban areas. Since communities are internally well-connected but externally less so, areas that belong in the same community are more closely linked through roads. Based on this feature, it is of great interest to explore urban characteristics such as the urban functions within a single community or human mobility interaction among multiple communities.

We propose a way to link the urban roads with urban space and urban functional zones. The detected community can be used as a basic unit in other urban studies. Undoubtedly, there are many directions worth exploring. First, the study identified urban functional areas by directly calculating the indexes of POIs, which may not be able to adequately reflect complex urban internal functions (Yao *et al*. 2017). Using the semantic analysis model to mine deep semantic information implied in POIs can more effectively reveal function areas and land use (Zheng *et al*. 2014, Zhang and Du 2015). Second, the urban roads not only affect the distribution of urban space and functional areas but also play a crucial guiding role in the movement patterns of urban residents (Wang *et al*. 2012, Pan *et al*. 2013). Constructing road networks with additional traffic information such as road types and travel time or speed will help to understand the mobility behaviors in the community-based network. Third, as the POI data adopted in this study originated from a single source (i.e. Gaode map), the ability to generalize the proposed methods to other regions deserves further exploration (Tallman and Phene 2007). In the future, through spatial big data that reflects crowd dynamic information, such as taxi trajectory, mobile phone and social media data, we can explore the human mobility pattern within and among diverse communities (Peng *et al*. 2012), reveal the relationship between the current urban roads and residents' behaviors, and propose possible optimization directions for road planning.

6. Conclusions

As an essential infrastructure for economic development, urban roads carry extensive connectivity, transportation and transfer functions in cities. This study employed the Infomap method to derive the hierarchical structure of the Guangzhou road network and analyzed the urban space at different levels. Since community emerges only from the structural characteristics of the road network itself, the connection between urban road structure and functional area distribution was analyzed using geospatial big data.

In this study, the identification results at different levels can reveal the partition of urban space at different scales; the top level expresses the connection between the city center and the surrounding area, the second level reflects the spatial relation within the core downtown area, and the results of the third level can assist in understanding the distribution of urban functional areas. Also, the study explored the functional area characteristics at the community scale and identified the distribution of various functional areas. The results show that the functional area of Guangzhou forms a distinct annular layer feature in space and presents a ternary feature of 'work – residence – leisure' from the inside to the outside.

The results of the community at all levels are more detailed than the traditional division with the administrative district, and owing to the road structure information contained in the identification process, we believe that community division is a reasonable and effective way to define the basic territorial units for other urban studies. Due to the accessibility of OSM road and POI data, the proposed method to detect the hierarchical community and identify functional zones can be readily generalized to other large cities in China. We believe studying from the community perspective will bring new insight into the understanding of urban space and urban functional zones.

Acknowledgments

The authors thank the editor and anonymous reviewers for their constructive comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Key R&D Program of China [2017YFB0503800]; the National Natural Science Foundation of China [Grant No. 41801306, 41671408];Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University [Grant No. 18S01] and Natural Science Foundation of HubeiProvince [2017CFA041].

Notes on contributors

Ye Hong is a master student at Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland. His research interest is geospatial big data analysis and applications.

Yao Yao is an Associate Professor at the School of Geography and Information Engineering at China University of Geosciences, Wuhan, China. At the same time, he worked as a visiting scholar and senior algorithm engineer at Alibaba Group's data center. His main research interests comprise multi-source geospatial big data mining, machine learning applications and fine-scale simulation of urban land-use dynamic changes.

ORCID

Ye Hong http://orcid.org/0000-0002-8996-3748
Yao Yao http://orcid.org/0000-0002-2830-0377

References

Barthélemy, M., 2011. Spatial networks. *Physics Reports*, 499 (1–3), 1–101. doi:10.1016/j.physrep.2010.11.002

Barthelemy, M., *et al*., 2013. Self-organization versus top-down planning in the evolution of a city. *Scientific Reports*, 3, 2153. doi:10.1038/srep02153

Batty, M., 2008. The size, scale, and shape of cities. *Science*, 319 (5864), 769–771. doi:10.1126/science.1151419

Chen, Y., *et al*., 2017. Delineating urban functional areas with building-level social media data: a dynamic time warping (DTW) distance based k -medoids method. *Landscape and Urban Planning*, 160, 48–60. doi:10.1016/j.landurbplan.2016.12.001

Cohen, R. and Havlin, S., 2010. *Complex networks: structure, robustness and function*. Cambridge, UK: Cambridge University Press.

Corcoran, P., Mooney, P., and Bertolotto, M., 2013. Analysing the growth of OpenStreetMap networks. *Spatial Statistics*, 3, 21–32. doi:10.1016/j.spasta.2013.01.002

Duan, Y. and Lu, F., 2013. Structural robustness of city road networks based on community. *Computers, Environment and Urban Systems*, 41, 75–87. doi:10.1016/j.compenvurbsys.2013.03.002

Duan, Y. and Lu, F., 2014. Robustness of city road networks at different granularities. *Physica A: Statistical Mechanics and Its Applications*, 411, 21–34. doi:10.1016/j.physa.2014.05.073

Estrada, E., 2012. *The structure of complex networks: theory and applications*. Oxford, UK: Oxford University Press.

Fan, H., *et al*., 2014. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28 (4), 700–719. doi:10.1080/13658816.2013.867495

Fortunato, S. and Hric, D., 2016. Community detection in networks: a user guide. *Physics Reports*, 659, 1–44. doi:10.1016/j.physrep.2016.09.002

Frank, L.D., Andresen, M.A., and Schmid, T.L., 2004. Obesity relationships with community design, physical activity, and time spent in cars. *American Journal of Preventive Medicine*, 27 (2), 87–96. doi:10.1016/j.amepre.2004.04.011

Gao, Y., *et al*., 2016. Efficient collective spatial keyword query processing on road networks. *IEEE Transactions on Intelligent Transportation Systems*, 17 (2), 469–480. doi:10.1109/TITS.2015.2477837

Garrison, W.L. and Marble, D.F., 1962. *The structure of transportation networks*. Northwestern Univ Evanston Il.

Gulbahce, N. and Lehmann, S., 2008. The art of community detection. *BioEssays*, 30 (10), 934–938. doi:10.1002/bies.20820

Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. *Environment Planning B: Planning and Design*, 37 (4), 682–703. doi:10.1068/b35097

Hric, D., Darst, R.K., and Fortunato, S., 2014. Community detection in networks: structural communities versus ground truth. *Physical Review E*, 90 (6), 062805. doi:10.1103/PhysRevE.90.062805

Huffman, D., 1952. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40 (9), 1098–1101. doi:10.1109/JRPROC.1952.273898

Jiang, B., 2007. A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and Its Applications*, 384 (2), 647–655. doi:10.1016/j.physa.2007.05.064

Jiang, S., *et al*., 2015. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53, 36–46. doi:10.1016/j.compenvurbsys.2014.12.001

Lancichinetti, A. and Fortunato, S., 2009. Community detection algorithms: a comparative analysis. *Physical Review E*, 80 (5), 056117. doi:10.1103/PhysRevE.80.056117

Leskovec, J., Lang, K.J., and Mahoney, M., 2010. Empirical comparison of algorithms for network community detection. *Proceedings of the 19th international conference on World Wide Web*. 631–640, Raleigh, North Carolina, USA.

Levinson, D. and Moreno, Y., 2012. Network structure and city size. *PLoS One*, 7 (1), e29721. doi:10.1371/journal.pone.0029721

Li, H., Calder, C.A., and Cressie, N., 2007. Beyond Moran's I: testing for spatial dependence based on the spatial autoregressive model. *Geographical Analysis*, 39 (4), 357–375. doi:10.1111/gean.2007.39.issue-4

Li, Q., *et al*., 2011. Dynamic accessibility mapping using floating car data: a network-constrained density estimation approach. *Journal of Transport Geography*, 19 (3), 379–393. doi:10.1016/j.jtrangeo.2010.07.003

Liu, R., *et al*., 2011. The road network evolution of Guangzhou-Foshan metropolitan area based on kernel density estimation. *Scientia Geographica Sinica*, 1, 013.

Liu, X., *et al*., 2017. Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science*, 31 (8), 1675–1696. doi:10.1080/13658816.2017.1324976

Liu, X. and Long, Y., 2015. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design*, 43 (2), 341–360. doi:10.1177/0265813515604767

Manaugh, K. and Kreider, T., 2013. What is mixed use? Presenting an interaction method for measuring land use mix. *Journal of Transport and Land Use*, 6 (1), 63–72. doi:10.5198/jtlu.v6i1

Meng, Y., Hou, D.Y., and Xing, H., 2017. Rapid detection of land cover changes using crowdsourced geographic information: a case study of Beijing, China. *Sustainability*, 9 (9), 1547. doi:10.3390/su9091547

Pan, B., *et al*., 2013. Crowd sensing of traffic anomalies based on human mobility and social media. *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems*, 344–353.

Peng, C., *et al*., 2012. Collective human mobility pattern from taxi trips in urban area. *PLoS One*, 7 (4), e34487. doi:10.1371/journal.pone.0034487

Porta, S., Crucitti, P., and Latora, V., 2006. The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design*, 33 (5), 705–725. doi:10.1068/b32045

Porta, S., *et al*., 2009. Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and Design*, 36 (3), 450–465. doi:10.1068/b34098

Rodrigue, J.-P., Comtois, C., and Slack, B., 2009. *The geography of transport systems*. Routledge.

Rosvall, M. and Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105 (4), 1118–1123. doi:10.1073/pnas.0706851105

Rosvall, M. and Bergstrom, C.T., 2011. Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS One*, 6 (4), e18209. doi:10.1371/journal.pone.0018209

Samaniego, H. and Moses, M.E., 2008. Cities as organisms: allometric scaling of urban road networks. *Journal of Transport and Land Use*, 1 (1), 21–39. doi:10.5198/jtlu.v1i1.29

Song, Q. and Wang, X., 2011. Efficient routing on large road networks using hierarchical communities. *IEEE Transactions on Intelligent Transportation Systems*, 12 (1), 132–140. doi:10.1109/TITS.2010.2072503

Strano, E., *et al.*, 2018. Mapping road network communities for guiding disease surveillance and control strategies. *Scientific Reports*, 8 (1), 4744. doi:10.1038/s41598-018-22969-4

Tallman, S. and Phene, A., 2007. Leveraging knowledge across geographic boundaries. *Organization Science*, 18 (2), 252–260. doi:10.1287/orsc.1060.0227

Tang, L. and Liu, H., 2010. Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2 (1), 1–137. doi:10.2200/S00298ED1V01Y201009DMK003

Verburg, P.H., *et al.*, 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems*, 28 (6), 667–690. doi:10.1016/j.compenvurbsys.2003.07.001

Wakita, K. and Tsurumi, T., 2007. Finding community structure in mega-scale social networks. *Proceedings of the 16th international conference on World Wide Web*, 1275–1276. doi:10.1096/fj.07-9420com

Wang, F., Antipova, A., and Porta, S., 2011. Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography*, 19 (2), 285–293. doi:10.1016/j.jtrangeo.2010.01.004

Wang, P., *et al.*, 2012. Understanding road usage patterns in urban areas. *Scientific Reports*, 2, 1001. doi:10.1038/srep00386

Xie, F. and Levinson, D., 2007. Measuring the structure of road networks. *Geographical Analysis*, 39 (3), 336–356. doi:10.1111/gean.2007.39.issue-3

Yao, Y. *et al.*, 2019. Sensing multi-levels urban funtional structures by using time series taxi trajectory data. *Geomatics and Information Science of Wuhan University*, In Press.

Yao, Y., *et al.*, 2018. Estimating the effects of “community opening” policy on alleviating traffic congestion in large Chinese cities by integrating ant colony optimization and complex network analyses. *Computers, Environment and Urban Systems*, 70, 163–174. doi:10.1016/j.compenvurbsys.2018.03.005

Yao, Y., *et al.*, 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *International Journal of Geographical Information Science*, 31 (4), 825–848. doi:10.1080/13658816.2016.1244608

Yuan, J., Zheng, Y., and Xie, X., 2012. Discovering regions of different functions in a city using human mobility and POIs. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 186–194. Beijing, China.

Yue, Y., *et al.*, 2016. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, 31 (4), 658–675. doi:10.1080/13658816.2016.1220561

Zhang, H. and Li, Z., 2012. Fractality and self-similarity in the structure of road networks. *Annals of the Association of American Geographers*, 102 (2), 350–365. doi:10.1080/00045608.2011.620505

Zhang, X. and Du, S., 2015. A linear dirichlet mixture model for decomposing scenes: application to analyzing urban functional zonings. *Remote Sensing of Environment*, 169, 37–49. doi:10.1016/j.rse.2015.07.017

Zhang, Y., *et al.*, 2015. Investigating the associations between road network structure and non-motorist accidents. *Journal of Transport Geography*, 42, 34–47. doi:10.1016/j.jtrangeo.2014.10.010

Zhao, S., Zhao, P., and Cui, Y., 2017. A network centrality measure framework for analyzing urban traffic flow: a case study of Wuhan, China. *Physica A: Statistical Mechanics and Its Applications*, 478, 143–157. doi:10.1016/j.physa.2017.02.069

Zheng, Y., *et al.*, 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology*, 5 (3), 1–55.