IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 28, 2019, accepted April 14, 2019, date of publication April 17, 2019, date of current version April 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2911664

Fine-Grained Subjective Partitioning of Urban Space Using Human Interactions From Social Media Data

MENGLING QIAO[1], YANDONG WANG[1,2,3], SHANMEI WU[1], AN LUO[4], SHISI RUAN[1], AND YANYAN GU[1]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China
[2]Collaborative Innovation Center for Geospatial Information Technology, Wuhan University, Wuhan 430072, China
[3]Faculty of Geomatics, East China University of Technology, Nanchang 200237, China
[4]Chinese Academy of Surveying and Mapping, Beijing 100830, China

Corresponding author: Yandong Wang (ydwang@whu.edu.cn)

This work was supported in part by the National Key Research Program of China under Grant 2016YFB0501403, in part by the National Natural Science Foundation of China under Grant 41271399, and in part by the China Special Fund for Surveying, Mapping and Geoinformation Research in the Public Interest under Grant 201512015.

ABSTRACT Fine-grained subjective partitioning of urban space using human activity flows reveals actual human activity spaces with high resolution, which has great implications for the development and validation of planning strategies. This paper presents a new method for fine-grained subjective partitioning of urban space based on the combination of network analysis and human interactions from social media. Three main procedures are involved in this method: 1) a cut-off point for hierarchical partitioning is determined by fitting the probability distribution function of human activity patterns; 2) based on this cut-off point, improved hierarchical weighted-directed spatial networks are constructed by incorporating a gravity model into conventional spatial networks to take into account the importance of the attraction of nodes in shaping urban space; and 3) the hierarchical and fine-grained partitioning results, which reveal the actual human activity spaces with high resolution at multiscale are obtained by implementing a spatial community detection algorithm in these networks. A case study, using a real-world dataset from the capital of China validates the effectiveness of the proposed method. By analyzing the results from Beijing, we concluded that the social media, a gravity model, and the hierarchical subjective communities detected from the hierarchical human activity networks are all outstanding contributors to the fine-grained subjective partitioning of urban spaces.

INDEX TERMS Gravity model, hierarchy, network analysis, social media, subjective partitioning.

I. INTRODUCTION

Subjective partitioning of urban space refers to a conceptual categorization of the spatial variance resulting from cognitive taxonomic processes [1]. Different from official administrative partitioning in an objective way, subjective partitioning of urban space depends more on human activities. These subjective spaces, emphasizing the more natural ways that people interact, have great implications for the official management and configuration of cities.

The rise of big geo-data and network analysis science brings new opportunities for the subjective partitioning of urban space. Accompanied by the subjectivity of human beings, various unprecedented types of activity data have been produced. These activity data, such as taxi GPS data, mobile phone data, and social media data, are regarded as inherent driving forces for the subjective partitioning of urban space [2]–[4]. Combined with human activity data, network analysis methods have been expanded as a support for the subjective partitioning of urban space. Based on this support, the spatial community has been developed as an intuitive representation of the subjective partitioning of urban space [5]–[8]. The theory of a spatial community is to divide a whole region into several sub-regions so that the connections are dense within each sub-region but sparse between different sub-regions [9], [10]. These spatial communities

The associate editor coordinating the review of this manuscript and approving it for publication was Vijay Mago.

are shaped by both the urban form and inherent interactions among people, materials, freight and information [7], [11]. Therefore, these spatial communities put more emphasis on the actual activity space of human beings, which is valuable for the management and configuration of cities [12].

The resolutions of the partitioning results obtained in most existing studies vary in sizes, such as close to the size of a country, a city, a district or a town. Resolutions at the country level, city level and district level are determined by the coverage of experimental data. Specifically, results with a country-level resolution can be obtained using experimental data on a worldwide scale [3]. Similarly, by using experimental data covering a country, results with resolution at the city level can be achieved [13]. Moreover, results with a district-level resolution are determined by using experimental data covering a city [8]. Resolution at the town level in the suburbs is recommended to be obtained by hierarchical community detection [6]. However, some limitations still exist in the fine-grained subjective partitioning of urban space. Town-level resolution has not yet been achieved in urban areas, let alone higher resolution than that of a town, such as a scenic area or a commercial center.

In our study, a new method for fine-grained subjective partitioning of urban space has been proposed to reveal actual human activity spaces with high resolution based on a combination of a network analysis and human interactions obtained from social media. An improved way of constructing a spatial network has been developed by taking into account the importance of the attraction of nodes in shaping an urban space. In addition, hierarchical human activity patterns are derived and form the basis of detecting the hierarchical spatial community to refine the results of subjective partitioning. The effectiveness of our method has been validated in the case of Beijing. Interestingly, we have found that social media can refine the results of subjective partitioning, which will be demonstrated and discussed in the following sections.

The remainder of this paper is organized as follows. In Section 2, we review related studies with regard to spatial community detection with a network analysis method. Section 3 describes the details of our method. Section 4 introduces the case study area, the datasets and our partitioning results and findings. Section 5 discusses the functions of social media data and the gravity model in refining the subjective partitioning results. Additionally, the applicability of our method in other case studies is discussed in Section 5. Finally, the conclusions are stated in Section 6.

II. LITERATURE REVIEW

This study focuses on fine-grained subjective partitioning of urban space, which can be revealed by spatial community detection. Combined with human activity data, network analysis methods have been widely adopted in previous studies to depict the spatial community. This section will review the related studies on spatial community detection via network analysis methods.

Network analysis relies on mathematical graph theory, which can effectively investigate network structure [14]. A structure derived from a human activity network is a comprehensive consideration of resources such as people, material, freight and information. Network nodes near each other in geographic space are more related and strongly connected to form a local cluster structure [15], [16]. Inspired by these structures, many studies have used network-analysis-based approaches to investigate spatial community structure [17]–[20]. The construction of a network is based on the flows of human activity between different spatial units, where each node refers to a spatial unit and each edge is modeled by human activity flows between two nodes. Various aspects of the spatial community structure derived from the network, such as the division of human activity space, the dynamics of urban space, the hierarchy of the spatial community, and the framework to improve spatial community detection, have been investigated in previous studies. Typical studies on these aspects are reviewed below.

A. THE DIVISION OF HUMAN ACTIVITY SPACE

Researchers have focused on the division of human activity space to discover more actual ways that people interact across space. For example, Adam, Delvenne, and Thomas [2] investigated the subjective partitioning of urban space in Brussels and confirmed the sprawling of the metropolitan territory. Kallus et al. [3] redrew the geo-political borders from the worldwide Twitter communication networks, which showed the effects of local conflicts and globalization on entire continents. Taking taxi GPS data as a proxy of human activities, Tang et al. [21] successfully divided the urban space of Harbin into a series of traffic zones.

B. THE DYNAMICS OF SPATIAL COMMUNITY

Understanding the dynamics of a spatial community can improve our knowledge of both the development of urban areas in different periods and the evolution of human activities over time. Zhong et al. [7] detected the dynamic development of spatial community borders from weighted directed networks constructed from smart card data in different years and identified emerging and disappearing communities. Zhou et al. [12]; Jiang et al. [22]; and Wang et al. [23] all revealed the diurnal patterns of spatial communities from time-evolving networks.

C. THE HIERARCHY OF SPATIAL COMMUNITY

Hierarchy is also a valuable aspect of a spatial community in recognizing space at different levels. Two ways are useful to realize the hierarchy of spatial communities. One way is to perform hierarchical clustering on a network, which defines the hierarchy based on the characteristics and topology of a network. For example, Rinzivillo et al. [24] implemented a hierarchical clustering to a human mobility network to portray hierarchical human activity borders and evaluated to what extent the administrative borders represent the real basin of human movement. Another way is to perform flat clustering on hierarchical networks, which defines the hierarchy based on the hierarchical relations for forming hierarchical networks. For example, Sobolevsky et al. [25] implemented an additional iterative partitioning of each of the first-level communities into sub-communities, revealing that the cohesiveness and matching of official regions can also be observed on a second level. Liu et al. [6] revealed a two-level hierarchical city community from taxi datasets composed of different trip distances, which showed the district-level and town-level human activity spaces. Similarly, Yin et al. [13] delineated the hierarchical anthropography boundaries based on the different physical movement ranges of Twitter users in Great Britain, which revealed the city-level and district-level human activity spaces. Fusco and Caglioni [26] applied both the modularity optimization and significant dominant flows recursively to partition space hierarchically, which highlighted the inadequacy of the official methods in integrating administrative boundaries in the functional area definition.

D. OPTIMIZING THE FRAMEWORK OF SPATIAL COMMUNITY DETECTION

Many studies have put efforts into this optimization to consider the characteristic of geographic space. Gao et al. [5] improved the community detection algorithm of modularity [27] by integrating a gravity model to remove the influence of distance decay. Similarly, Chen et al. [15] proposed a geo-distance-based method to detect communities in spatially constrained networks. Guo et al. [10] developed a new community detection algorithm named STOCS (Spatial Tabu Optimization for Community Structure), which successfully guaranteed the stability of the results and filtered out noise with a spatial contiguity constraint. Zhu et al. [8] proposed using streets as analytical units (which are usually referred to as nodes in networks) instead of regular grids, which can effectively minimize the modifiable area unit problem.

Previous studies provide a good foundation for the current study to achieve the fine-grained subjective partitioning. Network-analysis-based spatial community detection is adopted to partition actual human activity space. In addition, hierarchical community detection and a gravity model are two important tools for refining the partitioning, as will be demonstrated and discussed below. Finally, noise removal is also considered in our method. The details and the effectiveness of our method are described below.

III. METHODOLOGY

The presented approach includes three main parts to achieve fine-grained subjective partitioning of urban spaces by using human interactions obtained from social media data. As Figure 1 shows, first, the valid and invalid activity flows are determined, and a cut-off point is detected as the basis for hierarchical partitioning. Second, improved hierarchical spatial networks are constructed by incorporating a gravity model. Third, the hierarchical spatial community is detected to obtain hierarchical and fine-grained partitioning results. It is worth highlighting that the hierarchical spatial community is obtained by performing flat clustering on multilevel hierarchical networks because the hierarchical relations between human activities are more able to adjust to the characteristic of subjectivity and contribute to more actual results. The details are provided below.

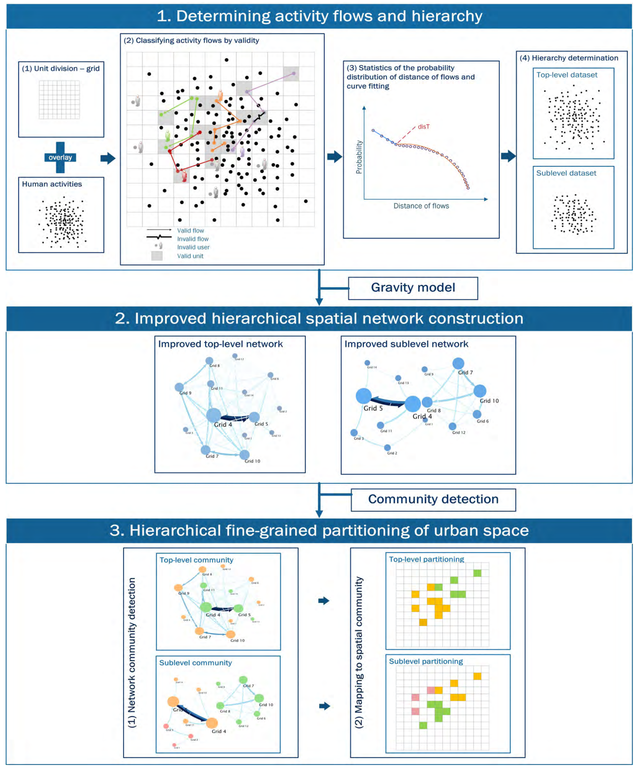

FIGURE 1. A schematic diagram of the fine-grained subjective partitioning of urban spaces.

A. DETERMINING ACTIVITY FLOWS AND HIERARCHY

Liu et al. [6] proposed an outstanding method for hierarchical spatial community detection, though this method as the drawback of requiring repeated attempts. Yin et al. [13] proposed an insightful method to overcome this drawback by defining a cut-off point in the probability distribution of the human activity radii, while they found no cut-off point at the city level. In our study, the idea of defining a cut-off point for hierarchical partitioning is applied. Instead of calculating human activity radii that may smooth the probability distribution curve, the valid human activity flows are used as a replacement. The detailed steps are as follows:

1) RESEARCHING UNIT DIVISION

A set of units is first obtained by region division, which is implemented on the study area to solve the problem that one address may be located by different coordinates of social media data. In our study, regular grids are suggested as research units, defined as $U = \{u_i | i \in \{1, 2, \ldots, s\}\}$.

2) CLASSIFYING ACTIVITY FLOWS BY VALIDITY

In this step, directed OD (Origin-Destination) flows of users are built from social media data in chronological order. As the first part of Figure 1 shows, if the origin and destination of a flow are located in the same unit u_i, then the flow is regarded as an invalid flow. Otherwise, the flow is regarded as a valid flow.

3) CURVE FITTING

Fitting the probability distribution of the distances of valid flows to find a cut-off point $disT$ can make sense. The probability distribution function should be different at the both sides of $disT$. Different probability distribution functions represent different human activity patterns, which is the basis for hierarchical partitioning.

4) DETERMINING HIERARCHY

After finding $disT$, the hierarchical datasets can be obtained to define the hierarchy. All valid flows are gathered into a top-level dataset D_1, which is used for the top-level partitioning of urban space. Flows with distances less than $disT$ are gathered into a sublevel dataset D_2, which is used for the sublevel partitioning of urban space. It is worth mentioning that the number of valid flows at different hierarchy levels should be reasonable.

B. IMPROVED HIERARCHICAL SPATIAL NETWORK CONSTRUCTION WITH THE GRAVITY MODEL INCORPORATED

Based on hierarchical activity datasets, a gravity model is added to improve hierarchical weighted-directed spatial network construction. Let $G = (V, E, W)$ be an improved weighted-directed spatial network, where V, E *and* W refer to the vertices, directed edges and weights of a network respectively, which are obtained as follows:

1) GENERATING VERTICES SET *V* AND DIRECTED EDGES SET *E*

A subset consisting of research units with valid flows is regarded as the set of vertices $V = \{v_i | i \in \{1, 2, \ldots, n\}\}$ of the network G. Then, the directed edges set $E = \{e_i = <v_o, v_d> | i \in \{1, 2, \ldots, m,\}\{v_o, v_d\} \in V\}$, is generated between vertices corresponding to the units with at least one valid flow.

2) GENERATING WEIGHTS SET *W*

In addition to the valid flows, there are many invalid flows attached to each vertex. The number of invalid flows indicates the strength of the attraction of each vertex, which will be taken into account in generating weights. Based on the relationship of distance decay and spatial community [5], [13], a gravity model is incorporated into our method to take into account the importance of the attraction of nodes in refining the partitioning of urban space. Weights incorporating the gravity model are defined as $W = \{w_i = g_i * c_i | i \in \{1, 2, \ldots, m\}\}$, where w_i refers to the weight of e_i, g_i refers to the gravitation of e_i calculated using equation (1), and c_i refers to the accumulated flows of e_i.

$$g_i = \frac{P_o * P_d}{dis^{\beta}}, \quad i \in \{1, 2, \ldots, m\} \text{ and } P_o,\ P_d \in V \tag{1}$$

where P_o and P_d refer to the numbers of invalid flows in unit O and unit D, respectively. dis refers to the Euclidean distance between unit O and unit D. β refers to the distance-decay parameter, which is calculated by fitting a power law [5].

Finally, by using a method that incorporates a gravity model, the hierarchical spatial networks G_1 and G_2 are constructed from the hierarchical datasets D_1 and D_2, respectively.

C. HIERARCHICAL FINE-GRAINED PARTITIONING OF URBAN SPACE

Based on the improved hierarchical spatial networks, a spatial community is used as an intuitive representation of the subjective partitioning of urban space. As the third part of Figure 1 shows, detection of the spatial community consists of two steps:

1) DETECTING THE NETWORK COMMUNITY

Infomap, a common community detection algorithms developed by Rosvall, Axelsson, and Bergstorm [28], is advantageous in handling a weighted directed network, which can identify the most important components of a network by using random walk as a proxy for information flows. Moreover, *Infomap* can identify small community structures in networks, which is meaningful for the detection of fine-grained human activity spaces. Therefore, *Infomap* is totally appropriate and used in our hierarchical spatial networks G_1 and G_2 to detect the hierarchical network communities.

2) MAPPING TO THE SPATIAL COMMUNITY

After the detection of hierarchical network communities, the results are mapped to geographical units to achieve hierarchical partitioning of urban space. Based on the corresponding relationships of geographical units and network nodes, it is easy to determine to which spatial community each unit belongs.

In addition, because of the absence of the spatial contiguity constraint, several units apparently disperse away from the gathering areas of the communities to which they belong. DBSCAN is suggested to be implemented in every spatial community to filter these units.

Finally, based on our method described above, the hierarchical fine-grained partitioning of urban space will be obtained with human interactions obtained from social media data. To improve the readability of our method, the detailed algorithm steps are shown in Table 1. To verify the effectiveness of this method, a real-world dataset from the capital of China is used as a case study.

IV. APPLICATION TO BEIJING

A. STUDY AREA AND DATA PREPROCESSING

As the capital of China, Beijing attracts many people to carry out a variety of activities, which provides valuable data for our study. Of the 16 districts of Beijing, 10 districts

TABLE 1. **Algorithm table of the proposed method for fine-grained subjective partitioning of urban space.**

Input

Weibo data with GPS embedded

Researching units (such as regular grids)

A. DETERMINING ACTIVITY FLOWS AND HIERARCHY

1. Embedding the unit id $U = \{u_i | \, i \in \{1, 2, \dots, s\}\}$ into each data
2. Building the directed O/D of each user's data in chronological order
3. Classifying activity flows by validity:
   invalid flow set *IVF*: the O/D of a flow located in the same unit u_i,
   valid flow set *VF*: the O/D of a flow not located in the same unit,
4. Calculating the distances of *VF*, saved in set *DIST*
5. Generating the probability distribution according to the O/D distances of *VF*
6. Perform curve fitting to find a cut-off point $disT$ representing hierarchical human activity patterns
7. Generating hierarchical datasets:
   Top-level dataset D_1: *VF*
   Sub-level dataset D_2: *VF* with distance < $disT$

B. IMPROVED HIERARCHICAL SPATIAL NETWORK CONSTRUCTION WITH THE GRAVITY MODEL INCORPORATED

1. For D_1 and D_2, performing steps 2-7, and output hierarchical networks G_1 and G_2
2. Obtain units U with valid flows as vertices V of the network G.
3. Generate directed edge set *E* according to *VF* between units.
4. Calculate the amount of *IFV* at each vertex, referred to as P_o, P_d
5. Calculate the amount of *VF* at each edge, referred to as c_i
6. Calculate the gravitation g_i of each edge: $g_i = \frac{P_o * P_d}{dis^{\beta}}$
7. Calculate the weight of each edge: $w_i = g_i * c_i$

C. HIERARCHICAL FINE-GRAINED PARTITIONING OF URBAN SPACE

1. Performing *Infomap* algorithm on hierarchical G_1 and G_2
2. Mapping communities to spatial units
3. Performing *DBSCAN* to remove noise units

Output

Hierarchical subjective partitioning results

are selected as the study area: the Changping, Shijingshan, Haidian, Dongcheng, Xicheng, Shunyi, Tongzhou, Fengtai, Daxing, and Chaoyang Districts. The urban and suburban areas of Beijing are included.

We have built a data crawler frame to collect human activities from the Sina Weibo platform (a famous social media platform in China) within the study area. Data with GPS information embedded are collected through the application programming interfaces (APIs) provided by Sina Weibo. With a long period of collection (from 2014.02.01 to 2014.09.31), noise removal (removing advertisements, marketing accounts and virtual person accounts) and preprocessing (removing users with fewer than 3 messages posted during our collection period), we finally obtained 4,111,364 valid flows and 2,854,369 invalid flows for the fine-grained subjective partitioning of urban space.

The cut-off point $disT$ for hierarchical division is identified as $5km$ by fitting the probability distribution of the distances of valid flows. Flow distances within $5km$ follow a power-law distribution $P(d) \sim x^{-\beta}$, accounting for more than 40% of all flows, while the probability density function of flow distances greater than $5km$ can be approximated by $P(d) \sim k_1 \left(1 - e^{-x/t1}\right) + k_2 \left(1 - e^{-x/t2}\right)$. The probability distribution function is different on each side of $5km$ so $disT$ is determined to be $5km$. In addition, to determine the distance decay parameter of the gravity model, the probability distribution of the valid flows within $[0km, 30km]$ (accounting for about approximately 90% of all flows) are fitted. Finally, we obtain the distance decay parameter $\beta = 1.06$. Details are shown in Figure 2.

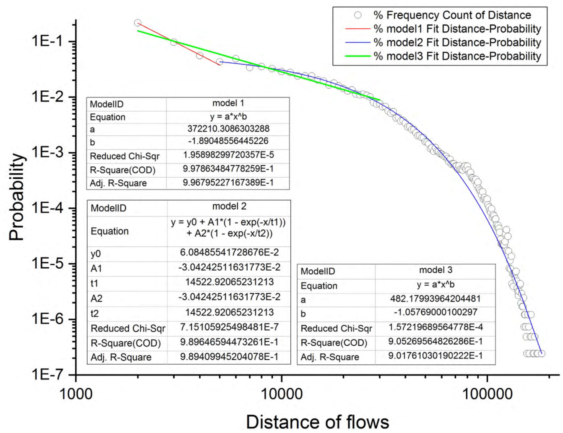

FIGURE 2. **The probability distribution of distances of valid flows. Flows with a distance less than $5km$ are fitted by model1, shown by the red line, which follows a power-law distribution $P(d) \sim x^{-\beta}$; flows with a distance greater than $5km$ are fitted by model2, shown by the blue line, which can be approximated by $P(d) \sim k_1(1 - e^{-x/t1}) + k_2(1 - e^{-x/t2})$; flows with a distance within $[0km, 30km]$ are fitted by model3, shown by the green line, which follows a power-law distribution $P(d) \sim x^{-\beta}$. Detailed parameter values are shown in these tables embedded in this figure.**

B. CHARACTERISTICS OF IMPROVED HIERARCHICAL NETWORKS

Hierarchical spatial networks are obtained by using the gravity-model-based network construction method described in Section 3.2. In this process, the study area is partitioned into $1km \times 1km$ regular cells as the nodes of networks. The distance-decay parameter β is set as 1.06. We define N_1 and N_2 as networks constructed by using all valid flows and valid flows of distances less than $5km$, respectively. The characteristics of these two networks are provided in Table 2.

TABLE 2. **Characteristics of networks.**

ID	Nodes	Edges	Flows	Average flows of edges	Average weights of edges
N_1	11,408 (100%)	1,029,936 (100%)	4,111,364 (100%)	3.99 (100%)	37,206.8 (100%)
N_2	9,581 (84.0%	131,866 (12.8%)	1,755,684 (42.7%)	13.31 (333%)	229,788.9 (617%)

The differences between the characteristics of these two networks in Table 2 further confirmed the existence of two different human activity patterns. The numbers of nodes in these two networks are similar, which means the ranges of the study area covered by the two networks are almost equal. Nevertheless, the difference between the edges is much larger than the difference between flows among these two networks, which indicates that N_2 is less accessible to the global scale than N_1 but has a stronger connectivity to the local scale. This finding is further confirmed by the characteristics of the average flows of edges and the average weights of edges. Especially in regard to the average weights of edges, the gap between N_1 and N_2 is further highlighted. This pattern is also consistent with the results of Liu et al. [6], in which short trips are more spatially stable. Networks N_1 and N_2 are then used for the fine-grained partitioning of urban space.

C. RESULTS OF FINE-GRAINED SUBJECTIVE PARTITIONING OF URBAN SPACE IN BEIJING

The results of the fine-grained subjective partitioning of urban space achieved by this method are shown in Section 3.3. Figure 3(1) shows the top-level partitioning results obtained from N_1, and Figure 3(2) shows the sublevel results obtained from N_2.

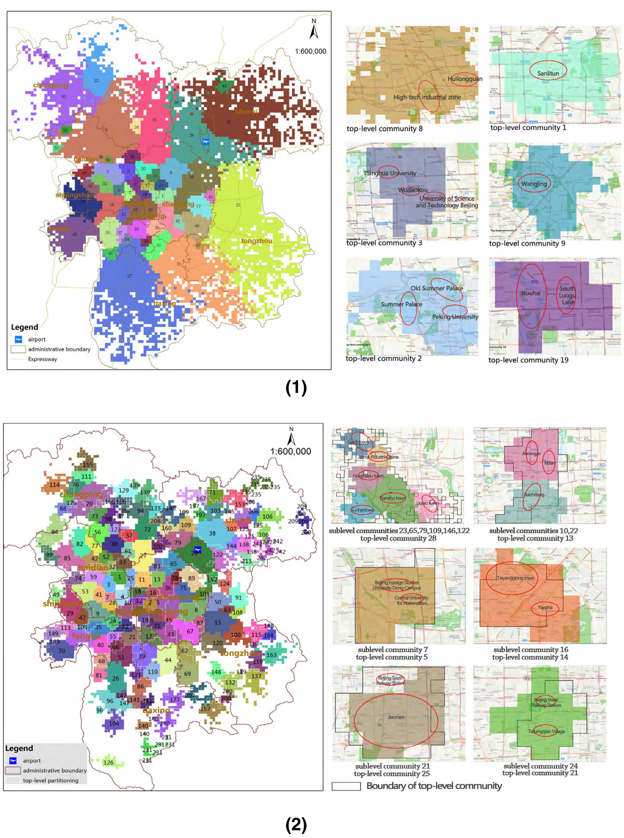

FIGURE 3. **Fine-grained subjective partitioning of urban space in Beijing. (1) Top-level partitioning of urban space in Beijing. (2) Sublevel partitioning of urban space in Beijing. Some famous areas such as famous blocks, attractions and commercial centers are labeled by red ellipses.**

To evaluate the top-level results, the district-level administrative boundaries and the urban expressway of Beijing are overlapped in Figure 3(1). Three characteristics can be revealed in Figure 3(1).

First, the top-level results have greater consistency with the administrative boundaries and expressways of Beijing. In detail, there is greater consistency with the combination of administrative boundaries and expressways in suburban areas such as *communities* 8, 15, 23, 27, 28, 31 *and* 39, while there is higher consistency with the expressways of Beijing in urban areas because of the similar mosaic texture. This characteristic indicates that the administrative boundaries and expressways determined in a top-down fashion have preconceived influences on actual human activity spaces.

Second, some mismatches within the results are visible and reasonable. The actual human activity spaces are not only constrained by administrative boundaries and expressways but also influenced by preference, religion, economic level and so on. For example, many people live in Huilongguan in the eastern part of *community* 8 and work in the high-tech industrial zone of the Shangdi sub-district in the southwest part of *community* 8. Strong interactions between these two regions combine the regions into a community.

Third, on the whole, the top-level partitioning results are more refined than those of the official district-level boundaries, especially in urban areas. For example, Sanlitun is situated in *community* 1, where the famous commercial mansion Taikoo Li and bar street are locatedt. *Community* 3 corresponds to the Wudaokou subdistrict, where many famous universities are located at, such as Tsinghua University and the University of Science and Technology Beijing. The Wangjing sub-district is clearly delineated in *community* 9. Famous tourist attractions, such as Peking University, the Summer Palace and the Old Summer Palace are located in *community* 2. *Community* 19 includes the tourist attractions of Houhai and South Luogu Lane.

In short, our top-level results have revealed the fine-grained actual human activity spaces, especially in urban areas. The resolution of the top-level results is higher than that of the district boundary, which confirms the importance of integrating of social media data and the gravity model to refine the subjective partitioning of urban space. Nevertheless, suburban areas yield poor partitioning results that are not sufficiently fine-grained. The advantage of hierarchical partitioning in refining is then considered to optimize the results. Sublevel results are shown in Figure 3(2).

To analyze the sublevel results, we have overlapped the top-level and sublevel results, as shown in Figure 3(2). Two characteristics can be revealed.

First, the results from human activities with short flows show an obvious refining in the subjective partitioning of urban space. For example, $top - level\ community$ 28 (a space mainly formed by the influence of Beijing Capital International Airport) becomes the collection of *sublevel communities* 23, 65, 79, 109, 146 *and* 122, which respectively correspond to Tianzhu Town with Beijing Capital International Airport, Sunhe Town, Houshayu Town, Jinma industrial zone, Gaoli Town and Liqiao Town. $Top - level\ community$ 13 (a space that mainly corresponds to the boundary of Xicheng District) becomes the collection of *sublevel communities* 10 *and* 22, which respectively correspond to the combination of Jinrongjie and the Xidan commercial center, and the Baizhifang sub-district containing the famous ethnic culture street Niujie.

Second, the degree of refinement in urban areas is less obvious than that in suburban areas. At the extreme, some communities remain almost unchanged such as the *top − level communities* 5, 14, 25, *and* 21, which respectively correspond to *sublevel communities* 7, 16, 21, *and* 24, which are formed by a university town (including the Beijing Foreign Studies University Dong Campus and the Central University for Nationalities), the combination of Taiyanggong Town and Yansha, the combination of the Beijing South Railway Station and Jiaomen sub-district, and the combination of the Beijing West Railway Station and Taipinqiao Village, respectively. This finding indicates that people create stable and mostly short flows in these areas to generate more stable activity spaces.

In short, the comparison of Figure 3(1) and Figure 3(2) confirms that hierarchical partitioning contributes substantially to the fine-grained subjective partitioning of urban spaces. The integration of social media data, a gravity model and hierarchical community is valuable for achieving fine-grained subjective partitioning of urban space.

In the case of Beijing, the effectiveness of our method has been validated and the characteristics of the fine-grained subjective partitioning of urban space in Beijing have been explored. Through the analysis, the top-level results show the advantage of combining social media data and a gravity model in refining the partitioning. With this advantage, the sublevel results further prove the function of the hierarchical community in refining the partitioning. In addition, the respective functions of social media data and the gravity model will be discussed in Section 5.

V. DISCUSSIONS

The subjective partitioning of urban spaces by using human interactions can effectively reveal the natural ways that people interact and the actual human activity spaces, which has great implications for the official management and configuration of cities. Meanwhile, the development of research on the fine-grained subjective partitioning of urban space is still in its infancy. Our study contributes to this emerging body of research by developing a new method that combines the advantages of hierarchical community obtained from hierarchical human activity networks, a gravity model and human interactions obtained from social media in refining subjective partitioning results. The function of the hierarchical community in refining partitioning has been revealed above. In this section, the functions of social media data and the gravity model in refining partitioning will be discussed in the first part. In addition, the weakness of our method with regard to features of noise and noise removal will be discussed in the second part. The applicability of our method in other case studies will be discussed in the third part.

A. DISCUSSIONS ON THE RESPECTIVE FUNCTION OF SOCIAL MEDIA AND THE GRAVITY MODEL IN REFINING PARTITIONING

A comparative analysis of examples will be discussed. Figure 4 shows the partitioning results derived from different data sources and different network construction methods. Figure 4(1) shows the partitioning results obtained with human activities from taxi GPS data in Beijing (period from 2012.11.01 to 2012.11.30 with 6,281,618 valid flows, yielding a network with 8,383 nodes and 1,073,710 directed edges). Figure 4(2) uses social media data as a replacement data source (yielding a network with 11,408 nodes and 1,029,936 edges, and 4,111,364 valid flows; the size of this network is not much different from that of the taxi network) to partition urban space. Figure 4(1) and (2) are derived from networks without a gravity model added. The results obtained by incorporating a gravity model in the construction of the network are shown in Figure 4(3). The spatial networks of these three partitioning results with nodes corresponding to grids in *communities* 1, 2 *and* 3 are shown in Figure 4(1). For visualization purposes, only directed edges with weights greater than the 95th percentile are displayed. Figure 4(a), (b) and (c) show the sample networks corresponding to the respective partitioning results of Figure 4(1), (2) and (3), which will be used below to explain the reasons for the differences in the partitioning results.

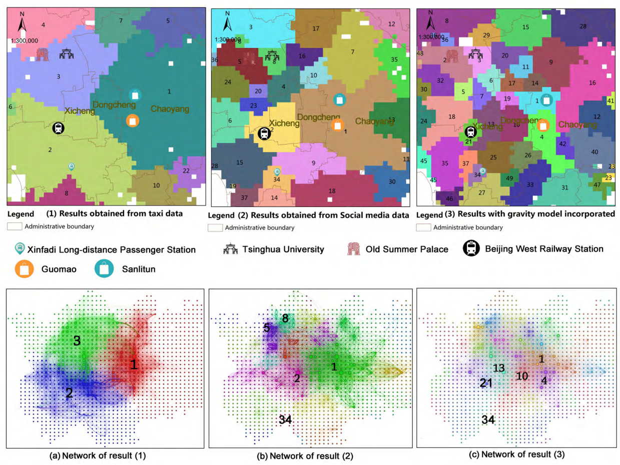

FIGURE 4. Samples illustrating the respective functions of social media data and the gravity model in refining the partitioning. (1) shows results obtained from taxi data, (2) shows results obtained from social media data, and (3) shows results obtained from social media data with a gravity model incorporated. some famous landmarks are labeled in (1)-(3). The corresponding networks of (1)-(3) are shown in (a)-(c).

1) FUNCTION OF SOCIAL MEDIA IN REFINING SUBJECTIVE PARTITIONING

Based on the same network construction and subjective partitioning method, using social media data can yield more fine-grained results than using taxi data. For example, the Xicheng District cannot be delineated in Figure 4(1). In addition, the mosaic texture similar to expressways vanishes in the results obtained with taxi data. Nevertheless, in Figure 4(2), not only is the mosaic texture similar to that of the expressways, but also more fine-grained spaces can be delineated, such as the Xicheng District boundary in *community* 2 and the Xinfadi Long-distance Passenger Station located in *community* 34 in Figure 4(2). In general, Figure 4(2) is a good demonstration of the role of social media data in refining subjective partitioning.

To explore the intrinsic reason for the function of social media data in refining subjective partitioning, the sample networks in Figure 4(a) and (b) are analyzed. Compared to the relatively uniform distribution of edges in the network constructed with taxi data, the edges in the network constructed with social media data in Figure 4(b) are shorter and concentrated more locally. As seen by tracing back to the source, these features in Figure 4(a) are contributed by the characteristics of social media data. Social media data are usually posted with a high variance of consciousness and purposiveness when people are aware of something [29], [30]. The high variance of awareness is effective in refining partitioning. For example, as Figures 4(2) and (b) show, *community* 5 is adjacent to *community* 8 in Figure 4(2); the former region is formed mainly by the awareness related to the Old Summer Palace, while the latter region is related mainly to a university town dominated by Tsinghua University. In short, the enormity and high variance of the information provided by social media data condenses the wisdom of a group of people and contributes to the short and locally concentrated edges in spatial networks, which make social media data effective in refining subjective partitioning.

However, some inaccuracies that exist in Figure 4(2) show the limitations of only using social media data in refining partitioning. For example, *community* 1 in Figure 4(2) has merged Dongcheng District and Chaoyang District into a single region, which is not sufficiently fine-grained. New techniques such as a gravity model are required to overcome these limitations. The function of a gravity model in refining subjective partitioning is discussed below.

2) FUNCTION OF GRAVITY MODEL IN REFINING SUBJECTIVE PARTITIONING

The incorporation of a gravity model in Figure 4(3) shows further fine-grained partitioning results compared to Figure 4(2). Obviously, Dongcheng District and Chaoyang District are successfully separated. More fine-grained spaces such as the Beijing West Railway Station (corresponding to *community* 21 in Figure 4(3)) and the Guomao commercial center (corresponding to *community* 4 in Figure 4(3)), are delineated accurately. The purpose of using a gravity model in our method is to take into consideration the importance of the attraction of nodes in shaping urban space. Taking the attraction of nodes into consideration is a reasonable and effective way to enhance the distinguishability between nodes, which plays an important role in refining subjective partitioning.

Similarly, the sample networks in Figure 4(b) and (c) are compared to explore the intrinsic reason of the function of the gravity model in refining subjective partitioning. In general, the network with a gravity model incorporated in Figure 4(c) highlights the domination of local nodes with strong connections and weakens the connections between nodes far away from each other. For example, *community* 1 in Figure 4(2) is dominated by both Sanlitun and Guomao. With the incorporation of the gravity model, the connections between Sanlitun and Guomao are weakened, and new communities dominated by these regions are formed (*community* 1 dominated by Sanlitun and *community* 4 dominated by Guomao in Figure 4(3)). Therefore, the gravity model is effective in refining subjective partitioning.

In conclusion, the use of social media data as a data source contributes to more refined results for the subjective partitioning of urban space, and the incorporation of a gravity model can also refine the results from the method level. The enormity and high variance of the information provided by social media data makes human activities more distinguishable to refine partitioning. The consideration of the attraction of nodes with the gravity model also increases the distinguishability between nodes, which results in more fine-grained spaces. Therefore, the combination of social media data and a gravity model in our method yields more fine-grained results than the methods used in previous studies.

B. DISCUSSIONS ON THE FEATURES OF NOISE AND NOISE REMOVAL

The absence of spatial constraints in the network-analysis-based method may make the grids that contain only several random data become noise. As Figure 5(1) shows, many grids apparently disperse away from the gathering areas of the communities to which they belong in the top-level examples. By filtering the flows by distance, less discrete grids were produced, while some discrete communities emerged in Figure 5(3). These discrete grids and discrete communities are all regarded as noise in our study. Due to the sparseness of data in the suburbs, one common feature of these two kinds of noise is that they are mainly distributed in the suburbs.

To remove noise from the discrete grid, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [31] is recommended in our method. Figure 5(2) displays the top-level examples with noise removal by DBSCAN. As seen by comparing Figure 5(1) and (2), most noise has been removed effectively by DBSCAN. However, some shortcomings exist when using DBSCAN to remove such noise. First, near the border of the two communities, there is the case where a grid still crosses into communities that it does not belong to. Another shortcoming of using DBSCAN to remove a discrete grid is that repeated implementation of DBSCAN is needed to ensure each community is processed. The consideration of a spatial contiguity constraint such as the STOCS, may overcome these problems, but new algorithms are required. Compared to designing new algorithms, DBSCAN with open source code is easy

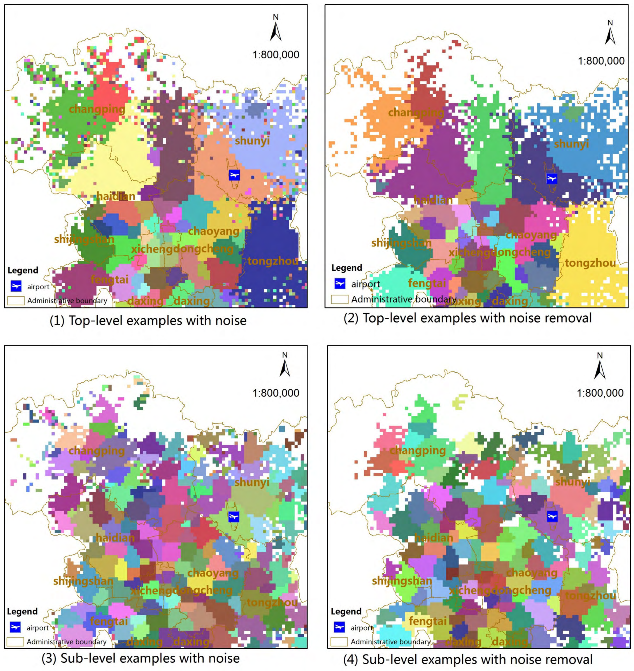

FIGURE 5. Examples of noise and noise removal. (1) shows top-level results with noise and (2) shows top-level results with noise removed by DBSCAN. (3) shows sub-level results with noise and (4) shows sub-level results with noise removed by DBSCAN.

to implement. In general, DBSCAN is a feasible but not perfect solution for removing discrete grids.

To remove the noise of discrete communities, an easy way is to remove the communities with fewer than 10 grids [6]. As Figure 5(4) shows, discrete communities in the suburbs were effectively removed. Nevertheless, the shortcoming of this solution was that several small communities in urban areas were also inevitably removed. In short, removing communities with fewer than 10 grids can improve the partitioning results in the suburbs. However, some mistakes emerge with regard to the urban areas.

In conclusion, applying DBSCAN and removing discrete communities are effective but not perfect ways to remove noise. In the future, spatial constraints, such as spatial contiguity, will be considered to solve this problem.

C. DISCUSSION ON THE APPLICABILITY OF OUR METHOD IN OTHER CASE STUDIES

There are three important inputs in our method: first, sufficient human interaction data from social media; second, the distance-decay parameter β of the gravity model for improving the construction of networks; and third, the hierarchy of human activity patterns, which can be simply thought of as short-distance flows and long-distance flows. Theoretically, if these three inputs are available, our method will be applicable to achieve fine-grained subjective partitioning of human activity spaces. For example, portioning of metropolises, such as Shanghai and Wuhan in China, is considered to be feasible because of massive population agglomeration and sufficient urban areas. Nevertheless, these are only theoretical speculations because of the lack of social media data. If data are available in the future, we will further prove the applicability of our method in other case studies.

VI. CONCLUSION

Revealing actual human activity spaces with high resolution has great implications for the development and validation of planning strategies. This paper has proposed a new method for the fine-grained subjective partitioning of urban space. In our method, key contributors such as human activity data obtained from social media, a gravity model and hierarchical subjective communities detected from hierarchical human activity networks are involved to refine the partitioning results. First, human activity data from social media, created with variety and strong subjective awareness, are separated into two datasets with hierarchical relationships by defining a cut-off point in the probability distribution of activity flows. Second, two improved networks are constructed from the hierarchical datasets by taking into account the importance of the attraction of nodes in shaping urban space using a gravity model. Finally, community detection is used in these hierarchical networks to obtain hierarchical and fine-grained partitioning results. The effectiveness of this method has been validated in the case of Beijing, and both the respective and integrated functions of these three contributors in fine-grained subjective partitioning of urban space have been revealed. The combination of these contributors realizes more high-resolution partitioning of urban space than the methods used in previous studies. In the case of Beijing, high-resolution spaces, such as railway stations, airports, subdistricts, tourist attractions, commercial centers, and university towns, are delineated accurately. These spaces represent the exact and actual human activity spaces, which have great implications for the official management and configuration of cities.

In future research, there are still many points in our study that need to be improved. The method for noise removal in our study is not scientific enough. Spatial constraints will be considered to improve this point in the future. In addition, only two different human activity patterns are determined based on the probability distribution of the distances of flows. If more factors are considered to sort out more human activity patterns, more accurate, more diverse and more fine-grained human activity spaces will be obtained in the future. Moreover, to expand our research, we will consider the use of these fine-grained and actual human activity spaces as basic research units in other urban studies. For example, these spaces obtained in a bottom-up fashion can be used as basic research units for land use investigation. If extended to remote sensing science, our results can be regarded as new segmentation results, which are valuable for image classification. Moreover, the interactions among people in the same space are stronger than those in two different spaces, and determining the formative reasons for these strong interactions or the stratum of these people also interests us. In short, these fine-grained and actual human activity spaces are useful and valuable for research studies.

REFERENCES

[1] N. Gale and R. G. Golledge, "On the subjective partitioning of space," *Ann. Assoc. Amer. Geographers*, vol. 72, no. 1, pp. 60–67, Mar. 1982.
[2] A. Adam, J. C. Delvenne, and I. Thomas, "Cartography of interaction fields in and around brussels: Commuting, moves and telephone calls," *Brussels Stud.*, vol. 118, pp. 1–18, Dec. 2017. doi: 10.4000/brussels.1601.
[3] Z. Kallus, N. Barankai, J. Szüle, and G. Vattay, "Spatial fingerprints of community structure in human interaction network for an extensive set of large-scale regions," *PLoS ONE*, vol. 10, no. 5, May 2015, Art. no. e0126713.
[4] J. Tang, F. Liu, Y. Wang, and H. Wang, "Uncovering urban human mobility from large scale taxi GPS data," *Phys. A. Stat. Mech. Appl.*, vol. 438, pp. 140–153, Nov. 2015.
[5] S. Gao, Y. Liu, Y. Wang, and X. Ma, "Discovering spatial interaction communities from mobile phone data," *Trans. GIS*, vol. 17, no. 3, pp. 463–481, 2013.
[6] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *J. Transp. Geography*, vol. 43, pp. 78–90, Feb. 2015.
[7] C. Zhong, S. M. Arisona, X. Huang, M. Batty, and G. Schmitt, "Detecting the dynamics of urban structure through spatial network analysis," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 11, pp. 2178–2199, May 2014.
[8] D. Zhu, N. Wang, L. Wu, and Y. Liu, "Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data," *Appl. Geography*, vol. 86, pp. 152–164, Sep. 2017.
[9] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, no. 6, Jun. 2004, Art. no. 066133.
[10] D. Guo, H. Jin, P. Gao, and X. Zhu, "Detecting spatial community structure in movements," *Int. J. Geograph. Inf. Sci.*, vol. 32, no. 7, pp. 1326–1347, Feb. 2018.
[11] J.-P. Rodrigue, C. Comtois, and B. Slack, *The Geography of Transport Systems*. London, U.K.: Routledge, 2006.
[12] M. Zhou, Y. Yue, Q. Li, and D. Wang, "Portraying temporal dynamics of urban spatial divisions with mobile phone positioning data: A complex network approach," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 12, p. 240, Dec. 2016.
[13] J. Yin, A. Soliman, D. Yin, and S. Wang, "Depicting urban boundaries from a mobility network of spatial interactions: A case study of great britain with geo-located twitter data," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 7, pp. 1293–1313, Mar. 2017.
[14] D. O'Sullivan, "Spatial network analysis," in *Handbook of Regional Science*, M. Fischer P. Nijkamp, Eds. Heidelberg, Germany: Springer, 2014, pp. 1253–1273.
[15] Y. Chen, J. Xu, and M. Z. Xu, "Finding community structure in spatially constrained complex networks," *Int. J. Geograph. Inf. Sci.*, vol. 29, no. 6, pp. 889–911, Jun. 2015.
[16] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Econ. Geography*, vol. 46, pp. 234–240, Jun. 1970.
[17] A. J. Comber, C. F. Brunsdon, and C. J. Q. Farmer, "Community detection in spatial networks: Inferring land use from a planar graph of land cover objects," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 18, pp. 274–282, Aug. 2012.
[18] Z. X. Fang, L. H. Liu, S. L. Shaw, and L. Yin, "An extended community detection algorithm to compare human mobility flow based on urban polycentric cluster boundaries: A case study of Shenzhen city," in *Spatial Data Handling Big Data Era*, C. Zhou, F. Su, F. Harvey, J. Xu, Eds. Singapore: Springer, 2017, pp. 111–124.
[19] M. A. Garcia-López and I. Muñiz, "Urban spatial structure, agglomeration economies, and economic growth in barcelona: An intra-metropolitan perspective," *Papers Regional Sci.*, vol. 92, no. 3, pp. 515–534, Dec. 2011.
[20] Q. Huang and D. W. S. Wong, "Activity patterns, socioeconomic status and urban spatial structure: What can social media data tell us?" *Int. J. Geograph. Inf. Sci.*, vol. 30, no. 9, pp. 1873–1898, Feb. 2016.
[21] J. Tang, S. Zhang, W. Zhang, F. Liu, W. Zhang, and Y. Wang, "Statistical properties of urban mobility from location-based travel networks," *Phys. A, Stat. Mech. Appl.*, vol. 461, pp. 694–707, Nov. 2016.
[22] S. Jiang, J. Ferreira, Jr., and M. C. Gonzalez, "Discovering urban spatial-temporal structure from human activity patterns," presented at the 18th Int. Conf. KDD, Beijing, China, Aug. 2012, pp. 95–102.
[23] J. Y. Wang, F. Gao, P. Cui, C. Li, and Z. Xiong, "Discovering urban spatio-temporal structure from time-evolving traffic networks," in *Proc. APWeb*, in Lecture Notes in Computer Science, vol. 8709, 1st ed., Sep. 2014, pp. 93–104.
[24] S. Rinzivillo, S. Mainardi, F. Pezzoni, M. Coscia, D. Pedreschi, and F. Giannotti, "Discovering the Geographical Borders of Human Mobility," *KI-Künstliche Intelligenz*, vol. 26, no. 3, pp. 253–260, Aug. 2012.
[25] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti, "Delineating geographical regions with networks of human interactions in an extensive set of countries," *PLoS ONE*, vol. 8, no. 12, Dec. 2013, Art. no. e81707.
[26] G. Fusco and M. Caglioni, "Hierarchical clustering through spatial interaction data. the case of commuting flows in south-eastern france," in *Proc. ICCSA*, Santander, Spain, 2011, pp. 135–151.
[27] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech., Theory Exp.*, vol. 2008, no. 10, pp. 155–168, Jul. 2008.
[28] M. Rosvall, D. Axelsson, and C. T. Bergstrom, "The map equation," *Eur. Phys. J. Special Topics*, vol. 178, no. 1, pp. 13–23, Nov. 2009.
[29] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Toronto, ON, Canada, Aug./Sep. 2010, pp. 492–499.
[30] J. Cai, B. Huang, and Y. Song, "Using multi-source geospatial big data to identify the structure of polycentric cities," *Remote Sens. Environ.*, vol. 202, pp. 210–221, Dec. 2017. Accessed: Jul. 2017.
[31] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, Portland, OR, USA, Aug. 1996, pp. 226–231.

Authors' photographs and biographies not available at the time of publication.