

A Two-Phase Clustering Approach for Urban Hotspot Detection With Spatiotemporal and Network Constraints

Feng Li , Wenzhong Shi , and Hua Zhang 

Abstract—Urban hotspots are regions with intensive passenger flow, sound infrastructure, and thriving business during a certain period of time, which mirror the travel behavior of residents. Taxi trajectory is one of the important data sources for urban hotspot detection. Unfortunately, it should be pointed out that quite a few of the relevant studies have ignored the temporal dynamics or network-constrained characteristics of urban hotspots, making the detecting results less reasonable and reliable. In this study, a two-phase clustering approach is proposed to detect urban hotspot with taxi trajectory. Concretely, in the first phase, spatiotemporal hierarchical density-based spatial clustering of applications with noise is utilized to cluster the trajectory points with spatial and temporal attributes, which is essential for understanding the evolution of urban hotspots over time. In the second phase, the idea of region growing is introduced to further filter noise, in which the spatial similarity between data points is measured by the route distance, considering that the trajectory data are constrained by the road network. A case study is carried out by the proposed method. Meanwhile, in combination with the LuoJia1-01 night-time light remote sensing data and POI data, the reliability of the clustering results is verified and the semantic meaning of the discovered clusters is enriched. Furthermore, not only the spatiotemporal distribution but also the trip lengths and directions of the detected hotspots are explored. These findings can serve as a scientific basis for policymakers in traffic control, public facilities planning, as well as location-based service.

Index Terms—Night-time light, route distance, spatiotemporal clustering, taxi trajectory, urban hotspots.

I. INTRODUCTION

URBAN hotspots refer to the regions with developed transport and commercial prosperity, which can reflect human mobility patterns to a great extent [1]–[3]. Thus, effectively analyzing the spatiotemporal distributions of urban hotspots is of great significance not only in trip recommendation [4], but also in traffic prediction [5] and location-based service [6]. Equipped

Manuscript received January 13, 2021; revised March 8, 2021; accepted March 19, 2021. Date of publication March 23, 2021; date of current version April 12, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB2103102, and in part by the National Natural Science Foundation of China under Grant 41971400. (Corresponding authors: Wenzhong Shi and Hua Zhang.)

Feng Li and Hua Zhang are with the School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou 221116, China (e-mail: 07152665@cumt.edu.cn; zhhua_79@163.com).

Wenzhong Shi is with the Department of Land Surveying and Geo-Informatics, Smart Cities Research Institute, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (e-mail: lswzshi@polyu.edu.hk).

Digital Object Identifier 10.1109/JSTARS.2021.3068308

with GPS devices, taxis can easily capture sequences of locations and time stamps at a certain frequency, and record a wealth of information about human mobility [7]. Hence, taxi trajectory data have been widely used in urban planning [8], social sensing [9], and urban hotspot detection [10]–[12].

On the basis of taxi trajectory data, recent research works have devised a set of methods to identify urban hotspots, including but not limited to kernel density estimation (KDE), data field, and DBSCAN [13]. Hu *et al.* utilized KDE to generate a smooth and continuous surface network that summarized the spatial distribution of trajectory points, and locally high areas on the surface were defined as hotspots [14]. This method has formed a nearly rigorous statistical basis and effectively reduced the subjectivity of analysis that makes it suitable for detecting hotspots with trajectory data. However, it is important to note that the bandwidth selection was a vital issue and summarizing large collections of points was a challenge. Qin *et al.* introduced a data field clustering algorithm which was applied to simulate the mutual interaction among points in data space, then urban hotspots were shaped by the aggregation of trajectory points and visualized on the map [15]. The clustering method can identify hotspots with any shape and be available for uneven distribution datasets. Unfortunately, the choice of parameter is a challenging problem that will affect the clustering results significantly. Tang *et al.* used the DBSCAN algorithm to cluster the pick-up and drop-off points of taxis and further explored the spatial distribution of urban hotspots [16]. The benefit of this method lies in that it can group data in arbitrary shape and size with two important parameters, namely the neighborhood of a point and the minimum number of points within the neighborhood to form a cluster. Besides, data with a sparse distribution will be identified as noise points and filtered out [17]. Nevertheless, much attention needs to be paid on selecting the appropriate parameters.

Although spatial clustering has received increasing attention in urban hotspot detection [18], the lack of temporal attribute limits the ability in discovering the points that are spatially adjacent but distant in time [19]. It should be emphasized that taxi trajectory is a typical kind of spatiotemporal data, including location and time attributes [20]. And the status and size of urban hotspot change over time due to the movement of human [21]. Thus, spatiotemporal clustering method is desired on the study of urban hotspots detection so as to provide more accurate results. ST-DBSCAN extends DBSCAN to cater for spatiotemporal

data, making it possible to cluster in both space and time [22]. Liu *et al.* applied this method to several simulated datasets that were composed of spatiotemporal events, and the ability of ST-DBSCAN to discover clusters with different shapes and sizes was demonstrated in part by the results [23]. However, similar to DBSCAN clustering algorithm, this method is sensitive to the selection of parameters and performs less well on datasets with uneven density distribution. Qin *et al.* improved the data field-based cluster technique by introducing the time influence and applied it to extract city hotspots from taxi trajectory data [15]. The experiment results indicated that the sustaining and inconstant hotspots area could be found with the help of the improved method, but the choice of proper parameters needs further consideration.

Moreover, quite a few of clustering methods adopt Euclidean distance to measure the similarity between points. In fact, taxi trajectory data are considered as a particular type of spatiotemporal data that overlaid on the street network. Accordingly, in order to enhance the reliability of the results, route distance is more suitable for discovering urban hotspots from taxi trajectory data. Zhao *et al.* represented drop-off events of taxi trips as linear features and then used them to detect urban hotspots in network space [18]. For the purpose of delimitating the centeredness surfaces of the urban hotspot, Xia *et al.* calculated a spatiotemporal potential value for each pick-up point and then assigned the resulting values onto the basic linear units derived from a network segmentation algorithm [24]. Deng *et al.* utilized network-constrained Delaunay triangulation to calculate the spatial proximity between points in the process of forming clusters [25]. From the above researches, it can be inferred that the route distance plays a significant role in improving the accuracy of urban hotspot detection.

Night-time light (NTL) remote sensing can capture the intensity, distribution, and changing patterns of artificial light at night, which is closely related to human activities [26]. Consequently, it is popular in a range of research fields represented by socio-economic monitoring and population estimation. Xin *et al.* used the NTL data to monitor the urban expansion in Wuhan, and a conclusion was drawn that socio-economic had a strong correlation with the NTL data [27]. By fusing geo-tagged tweets along with night-time light data, Devkota *et al.* proposed an original approach to discover tourism areas of interest in remote places [28]. In order to improve the precision of population estimation, Yu *et al.* attempted to integrate NTL data together with taxi trajectory data to estimate population at fine scales in Shanghai, and the results demonstrated that the newly proposed method could cope with the overestimation problem caused by using NTL data alone [29]. Enlightened by the previous research works on NTL data, we attempt to integrate taxi trajectory and Luojia1-01 data for urban hotspot detection, so as to enrich the semantic meaning of clustering results in this article.

In light of the spatiotemporal-constrained and network-constrained characteristics of taxi trajectory data, a two-phase clustering method is proposed for urban hotspot detection in this article. First, enlightened by ST-DBSCAN and HDBSCAN clustering algorithm, we propose an improved method ST-HDBSCAN to discover uneven clusters from spatiotemporal data, which is essential for understanding the evolution of urban

hotspots. Second, we introduce the idea of region growing to further filter noise, and use the route distance to measure the spatial similarity between data points, which is more in line with the actual situation. The optimal clustering parameters for hotspot detection are determined on the basis of an internal validation indicator called the density-based clustering validation index (DBCVI) and prior knowledge interpreted from the OpenStreetMap, so as to minimize the impact of subjectivity in parameter selection. The effectiveness of our proposed method is illustrated with a case study in Shanghai, China. Besides, Luojia1-01 NTL data, which have a finer spatial resolution of 130 m and can serve as a useful source in the reflection of human activities, are integrated along with geographic data to verify and analyze the spatiotemporal distribution of urban hotspots. According to the detection results, not only the spatial and temporal distributions of urban hotspots are analyzed, but also the trip length and direction of the hotspots are explored.

The remainder of this article is structured as follows: Section II introduces the dataset used and the preprocessing methods. In Section III, the two-phase clustering method is described in detail. A case study in Shanghai is carried out to validate the effectiveness of the proposed method in Section IV and the characteristics of the detected hotspots are further explored from different perspectives in Section V. Finally, Section VI summarizes the article and outlines the potential future work.

II. DATA DESCRIPTION AND PREPROCESSING

A. Data Description

The trajectory data for this study is acquired from the GPS devices of more than 12 000 taxicabs in Shanghai, one of the most populous cities in the world and located in east China. Over 85 million trajectory records are sampled with a time interval of approximately 10 s on April 1, 2018. Each record contains detailed taxi information, including not only the collecting time and location but also the vehicle number, status, and velocity. The street network of Shanghai, consisting of 46 470 nodes and 80 544 edges, is selected from OpenStreetMap by importing OSMnx library [30]. The POI data are collected from the Amap. After converting to the WGS84 coordinate system, a total of 493 675 POIs in 2018 are available in this work, including business area, residential zones, railway stations and so on. And the night-time light data are originally obtained from Luojia1-01 satellite launched in 2018. It has a finer spatial resolution of 130 m and greater light sensitivity in comparison with the commonly used NPP-VIIRS data, which make it possible to clearly discover the urban extent and even the distribution of roads or large-scale houses. The trajectory data of one taxi in Shanghai are presented in Fig. 1(a), and Fig. 1(b) displays the night-time light data of Luojia1-01.

B. Data Preprocessing

In order to reduce the impact of incorrect trajectory data and better achieve the targets of the work [31], several processing steps are performed as follows:

Duplicate data cleansing: Sliding window is used to delete duplicated adjacent points with the same coordinate position or

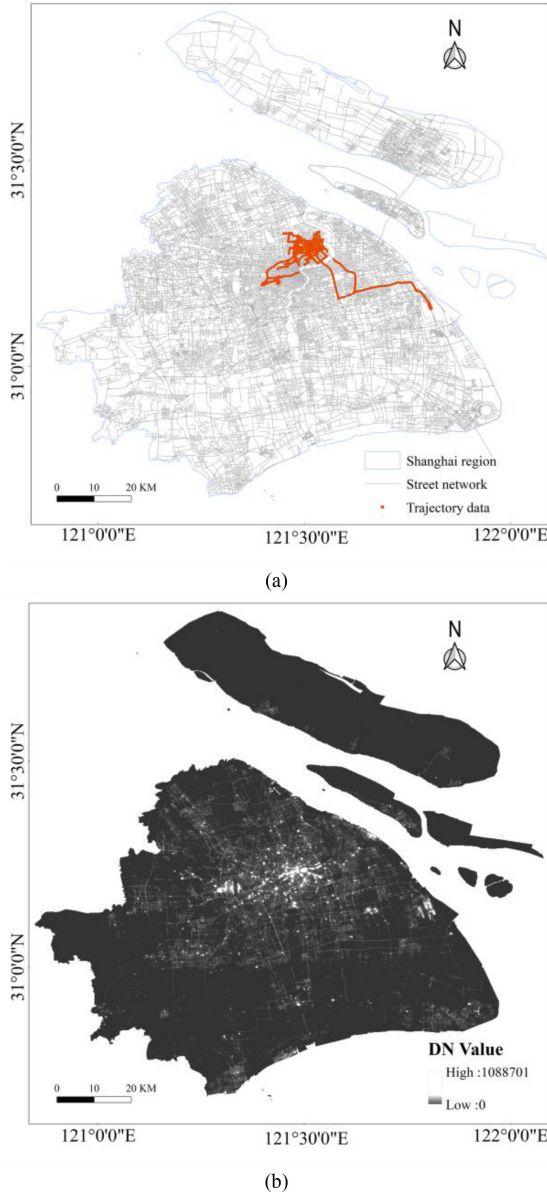


Fig. 1. (a) Study area with trajectory data of one taxi and (b) Luojia1-01 night-time light data.

collected time. Meanwhile, irrelevant attributes are removed to decrease the dimension of the data.

Anomalous data cleansing: The records whose attribute value exceeds the valid range are regarded as abnormal data. Besides, few taxis may keep the same status through all day owing to invalid operation or equipment failure.

Drift data cleansing: The first is to remove the sampling points with poor signal quality, that is, to delete the points with less than three positioning satellites; the second is to smooth the trajectory with multiple speed thresholds.

In view of the fact that trajectory points are constrained by the street network, a weight-based map matching method is introduced to project these points onto the corresponding road sections [32], [33]. Map matching plays a significant role in reducing the positioning errors of trajectory points and facilitating the subsequent calculation of route distance between two points.

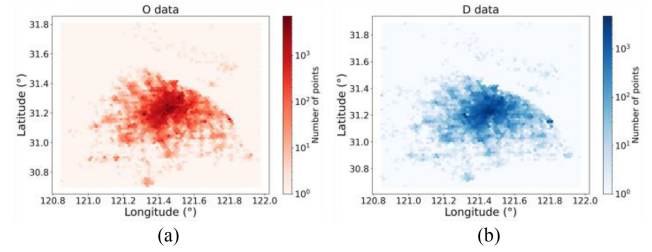


Fig. 2. Density scatter plot: (a) O points and (b) D points.

The quantization formula is put forward by involving two weight factors: the driving direction of the taxi and the distance to the adjacent road section

$$\Omega = \varepsilon_{\theta} \times \cos(\Delta\theta) + \varepsilon_d \times \frac{50 - d}{50} \quad (1)$$

in which θ represents the difference between the instant driving direction of taxi and the azimuth of road section; d indicates the vertical distance from the point to the road section; ε_{θ} and ε_d denote direction weight coefficient and distance weight coefficient respectively, and the sum of them equals 1. And referring to the width of the main road in Shanghai, 50 m was adopted in the formula, so as to ensure that the candidate road sections are within this limit. The matching point can be obtained by projecting the trajectory point vertically onto the center line of the best matching road section with the maximum value of Ω .

OD points, as the pick-up and drop-off events, are extracted from the processed data in terms of the status information of each trajectory and play a crucial role in urban hotspot detection. When the operation status change from vacant to occupied, it indicates a pick-up event marked as the origination of the ride, and a shift from occupied to vacant suggests a drop-off event marked as the destination of the ride. As shown in Fig. 2, the extracted OD points are displayed in the density scatter plot. Since the pick-up and drop-off events have a similar distribution, only the latter are selected to detect urban hotspots for the convenience of calculation.

In the process of acquiring and transmitting the night-time light data, there is an issue of radiation distortion that leads to high DN values and large bright area, so the original image of Luojia1-01 is not directly usable. Hence, the DN values are converted to the absolute radiance in accordance to the following conversion formula. Considering that the absolute radiance is extremely small floating-point data, a further step is needed to map it to a specified grey scale level like 10 bits

$$L = DN^{3/2} \cdot 10^{-10} \quad (2)$$

in which L denotes the calibrated absolute radiance and DN denotes the gray value of original image.

Then, the night-time light remote sensing data are clipped in terms of the city boundary shapefile originated from OSM data under the same coordinate reference system. Next, natural breaks method that can identify logical break points for each class is utilized to classify the grid cells with radiance intensity to different groups. Finally, in order to intuitively display the location of urban hotspots, the classification results are shown in pseudo-color and superimposed on the basic map of Shanghai.

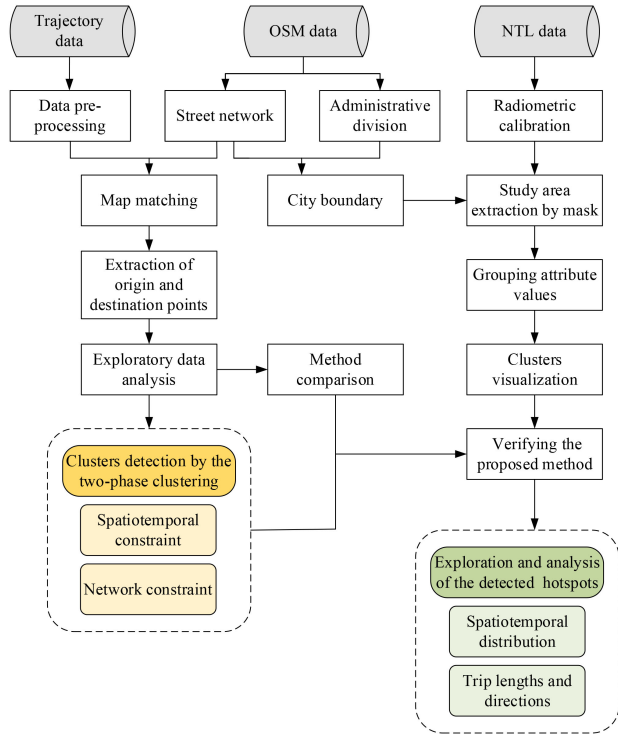


Fig. 3. Workflow for urban hotspot detection.

III. METHODOLOGY

The workflow devised for urban hotspots detection is shown in Fig. 3. Concretely, drop-off events are first extracted from the processed trajectory data in accordance with the variations of status information. Second, exploratory data analysis is applied to intuitively understand the distribution of drop-off events, ensuring that the dataset is suitable for performing spatiotemporal clustering. Third, the two-phase clustering approach is utilized to detect urban hotspots with spatiotemporal and network constraints. Fourth, the effectiveness and rationality of the proposed method are illustrated by the visual inspection of high-density POI regions and experimental comparison with other methods. Finally, the detailed characteristics of the detected hotspots are further explored, including spatiotemporal distribution, trip lengths, and trip directions. Besides, Luojial-01 NTL data are utilized as a complementary data to enrich the static features of urban hotspots. The improved clustering approach mentioned in the third step is elaborated below.

A. Phase One: Spatiotemporal Clustering With ST-HDBSCAN

Inheriting the merits of DBSCAN and hierarchical clustering, HDBSCAN has the ability to efficiently discover spatial clusters with arbitrary shapes, filter out noise with fast velocity, and handle clusters with varying densities [34], which make it well suitable for using large quantities of GPS trajectory points to detect urban hotspots in different shapes, sizes, and numbers. However, this method is inappropriate for clustering spatiotemporal data and more specifically it fails to address the dynamic change of hotspots through time [35]. As shown in Fig. 4, the spatially adjacent points are not necessarily close in the

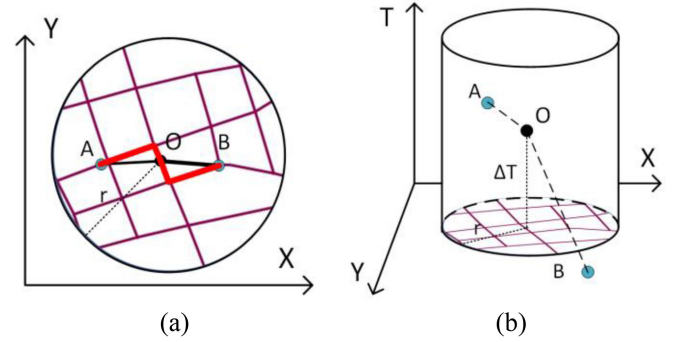


Fig. 4. Spatiotemporal neighborhood: (a) spatial similarity, high similarity between A and B; spatial similarity measurement, red line indicates route distance and black line indicates Euclidean distance, and (b) spatiotemporal similarity, low similarity between A and B.

time dimension, illustrating that the interaction between points is closely associated with their temporal attribute. Therefore, ST-HDBSCAN, which is an extension of HDBSCAN along the time axis, is utilized to discover the urban hotspots in light of the spatiotemporal similarity.

The clustering result of HDBSCAN is affected by one main input parameter m_{pts} , which describes the smallest size grouping to form a cluster, while a new initialization parameter ε_t needs to be introduced in ST-HDBSCAN to describe temporal threshold. For an accurate representation of the extended clustering approach, several concepts and components are given as follows.

Definition 1 (spatiotemporal core distance): The spatiotemporal core distance of a point x is denoted as $st-d_{core}(x)$ and defined as the spatial distance to the k th nearest neighbour within temporal threshold ε_t , where the value of k is the same as m_{pts} .

Definition 2 (Spatiotemporal mutual reachability distance): The spatiotemporal mutual reachability distance between two points a and b is formulated as $st-dmreach(a, b)$, which equals to $\max\{st-d_{core}(a), st-d_{core}(b), d(a, b)\}$ and $d(a, b)$ is the origin metric distance between them.

At the first stage, the spatiotemporal mutual reachability distance between points is calculated to push away sparser noise and then set as the edge weight of minimum spanning tree. With the help of minimum cluster size, the next step is to condense the cluster hierarchy, which is converted from the minimum spanning tree. In the final step, the stable clusters are extracted from the condensed tree. The cluster should not be considered as urban hotspot if the points within it are generated throughout the day. Therefore, the temporal threshold can further control the value of spatiotemporal mutual reachability distance and has a dramatic effect on hotspot detection.

B. Phase Two: Density Filtering Within Road Network Space

The significant advantage of ST-HDBSCAN is that it can handle variable spatiotemporal cluster. However, owing to the fact that fewer drop-off events are distributed in suburbs, some of the detected clusters are sparse in density and large in size, which may lead to the reduction of homogeneity in a cluster [36]. Moreover, the taxi trajectory is constrained by the street network in urban space. As the red and black lines in Fig. 4(a) illustrate, the route distance between the origin and the destination is

slightly longer than the Euclidean distance, which means the number of points in a cluster may be overestimated. The spatial similarity between points should be measured by the shortest route distance, so as to improve the accuracy of clustering. Therefore, a filter method based on route network distance is applied for postprocessing. Several of the definitions in the method are described below.

Definition 3 (Network distance threshold): The network distance threshold is utilized to measure the similarity between points and denoted as ε_s .

Definition 4 (Core point): The core point denotes such an object that there exists at least m_{pts} many points within the network distance.

The method starts with grouping the clustering results derived from ST-HDBSCAN according to their labels, and then the filter method based on the route distance is applied for each valid cluster. The noise points which are far from the core points are eliminated and the large sparse clusters are broken up into several compact clusters or directly removed, so as to ensure the homogeneity of each cluster.

In the process of filtering, all drop-off points in the spatiotemporal cluster are first projected in 2-D plane, and then an arbitrary point unmarked is selected as a seed point. The second is to judge whether the seed is a core point according to the number of points contained within the network distance threshold. If so, a new cluster is created and the idea of region growing is adopted to iteratively assign relevant points into this cluster [37], [38]. Otherwise, this point is temporarily marked as noise point. The above steps are repeated with a new point, until all of the points have been processed.

Besides, the origin cluster is converted into a graph form to facilitate the calculation of network distance. Considering that the street network consists of nodes and edges, where each node has a spatial coordinate and each edge has a spatial length, it is appropriate to transform it into a graph format for convenient calculation. After projecting the drop-off events onto the street network, the projection points and corresponding road segments are added to the graph to act as new nodes and edges, and then the shortest route distance between any two points is obtained by means of Dijkstra's algorithm.

C. Cluster Validation

Internal validation and external validation are two principal ways for measuring the effectiveness of clustering. However, urban hotspot detection is an unsupervised learning with no known class labels, so it is rather difficult to evaluate the clustering results eternally. In this article, with the high-density POI data and the geographical environment interpreted from the distribution of infrastructure, visual inspection is used to verify whether the cluster can be better explained from outside. Meanwhile, the DBCV index, is applied to the clustering results. In comparison with the commonly used validation indicator CDbw in density-based clustering algorithm, DBCV index performs better because it can properly find clusters with arbitrary shapes [39]. On the basis of the minimum spanning tree formed by entities, the validity index DBCV adopt the density sparseness of a cluster (DSC) and the density separation of a pair of clusters

(DPSC) to measure the density within a cluster and between clusters. Ranging from -1 to 1, the greater the value of DBCV, the better the clustering assignments

$$V(C_i) = \frac{\min_{1 \leq i, j \leq n; i \neq j} (DSPC(C_i, C_j)) - DSC(C_i)}{\max(\min_{1 \leq i, j \leq n; i \neq j} (DSPC(C_i, C_j)), DSC(C_i))} \quad (3)$$

$$DBCV = \sum_{i=1}^n \frac{|C_i|}{|O|} V(C_i) \quad (4)$$

in which n represents the quantity of clusters, $|C_i|$ and $|O|$, respectively, indicate the size of each cluster and the total number of entities.

IV. COMPARISON AND ANALYSIS

A. Exploratory Data Analysis

Through exploratory data analysis, the distribution pattern of OD points will be initially discovered, which is useful for determining whether the dataset is suitable for spatiotemporal clustering [40].

Specifically, the purpose of spatial autocorrelation statistics is to describe the degree of similarity between adjacent observations. The areas with high density of drop-off events are generally considered as urban hotspots, so we utilize the global autocorrelation index Moran's I to measure the correlation between the density of drop-off events and their spatial location, where the density is quantified as the number of points within their spatiotemporal neighborhood. As depicted in Fig. 2(b), the spatiotemporal neighborhood is well defined as a cylinder with a base radius as r and a height as $2\Delta T$. And these two parameters are the same as the spatial radius $Eps1$ and temporal radius $Eps2$ in ST-DBSCAN. The Moran scatter plot is shown in Fig. 5(a), where the horizontal axis indicates the standardized density value and the vertical axis indicates their spatially lagged counterparts. Most of the data lie in the upper right and lower left quadrants, and the value of Moran's I is up to 0.901, indicating that the density of drop-off events exhibits an obvious positive spatial autocorrelation. The p -values and z -scores are 0.001 and 304.244, respectively, which means that the drop-off events present a remarkable clustering pattern in space with a confidence level of 99%.

Besides, the target of space-time stationary analysis is to intuitively understand the variations in the data. In this article, we select second-order polynomials to fit the trends of drop-off events. A 3-D perspective of the data is shown in Fig. 5(b), where the X-axis and Y-axis represent the spatial coordinate of drop-off events, and the Z-axis indicates the collecting time expressed in minutes. As the red and green curves shown, the projected trends in both directions tend to be stable. Therefore, it can be inferred that the drop-off events are spatiotemporal stationary without the necessity of eliminating the trends from the data in advance.

In general, the drop-off events exhibit strong clustering pattern and are stationary in space and time, which are suitable for spatiotemporal clustering.

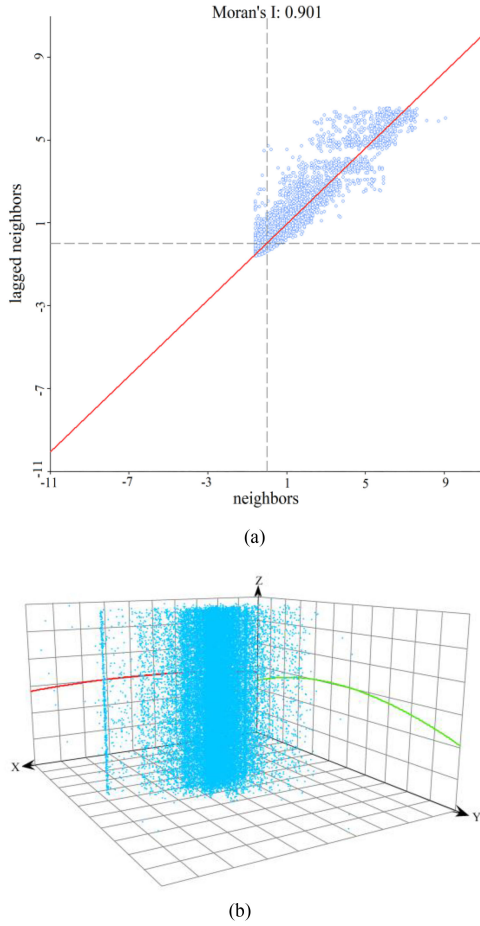


Fig. 5. Exploratory data analysis. (a) Spatial autocorrelation analysis, and (b) space-time stationary analysis.

B. Validation of the Proposed Method

In order to verify the effectiveness of the two-phase clustering base on spatiotemporal and network constraints, the drop-off events extracted from the taxi trajectory dataset at 10:00–12:00 are selected as the experimental data. Besides, given that the distribution of POI has a strong positive correlation with human activities [41], the POI data are used as an external reference. The accuracy of urban hotspots detection can be partly illustrated if there exists a well matching degree between the detected hotspots and the high-density POI areas. At first, the study area has been divided into different blocks in terms of the network structure rather than the commonly used grids, so as to ensure the independence and integrity of each block. Then, we project the POI data onto the block and further calculate its density Db with the following equation:

$$Db = \frac{Qp}{Ba} \quad (5)$$

in which Qp represents the quantity of POIs in each block, and Ba denotes the block area in hectares. Next, the natural breaks method is utilized to classify POI hotspots into different groups, and the blocks with top 3% high-density values are selected as POI hotspots. As shown in Fig. 6(b), from light red to dark red, the deeper the color, the more active the hotspot is.

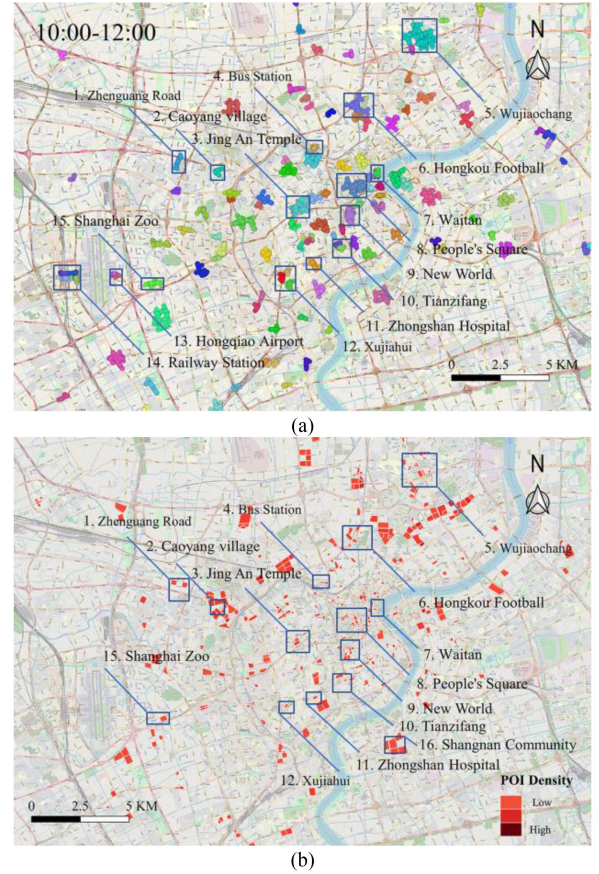


Fig. 6. Validating the effectiveness of the proposed method. (a) Urban hotspots detected and (b) high-density POI areas.

It is noteworthy that the selection of three initial parameters is a critical issue that will heavily affect the quality of clustering results. Nevertheless, without knowing the intrinsic characteristic of trajectory points, it is hard to provide meaningful values for these parameters. Hence, the optimal parameters are chosen from multiple sets of values by using the combination of DBCV index and visual inspection, so as to reduce the impact of subjectivity. In phase one, in order to select the optimal parameters for minimum cluster size m_{pts} and temporal threshold ε_t , a total of 20 sets for the two parameters are examined. The value of DBCV index is calculated with spatiotemporal constraint and the evaluation information for different combinations is listed in Table I. It indicates that the DBCV index goes up with the increase of temporal threshold and tend to be stable at 20 min. Thus, minimum cluster size and temporal threshold are set to 10 and 20 min, respectively, for the experimental dataset. In phase two, several trials are conducted with fixed values for parameter m_{pts} and ε_t , while changing spatial threshold ε_s from 50 to 300 m with the interval of 50 m. By using DBCV index, 200 m that has a relative maximum is determined as the value for parameter ε_s .

The two-phase clustering method is implemented on the drop-off events with the parameters selected above. In Fig. 6(a), the urban hotspots identified are superimposed on the basic map of Shanghai. It can be found that most of the detected urban hotspots seem to be reasonable by the visual inspection and are

TABLE I
QUANTITATIVE EVALUATION OF CLUSTERING RESULTS WITH DIFFERENT
COMBINATIONS OF MINIMUM CLUSTER SIZE AND TEMPORAL THRESHOLD

Parameters	Cluster Numbers	DBCW Values
$m_{pts}=10, \epsilon_t=10$	140	0.0850
$m_{pts}=15, \epsilon_t=10$	74	0.0863
$m_{pts}=20, \epsilon_t=10$	44	0.0136
$m_{pts}=25, \epsilon_t=10$	18	-0.3031
$m_{pts}=30, \epsilon_t=10$	5	-0.5132
$m_{pts}=10, \epsilon_t=15$	163	0.0754
$m_{pts}=15, \epsilon_t=15$	91	0.0915
$m_{pts}=20, \epsilon_t=15$	47	-0.0089
$m_{pts}=25, \epsilon_t=15$	11	-0.6016
$m_{pts}=30, \epsilon_t=15$	9	-0.5942
$m_{pts}=10, \epsilon_t=20$	187	0.1386
$m_{pts}=15, \epsilon_t=20$	94	0.0975
$m_{pts}=20, \epsilon_t=20$	69	0.1026
$m_{pts}=25, \epsilon_t=20$	55	0.1029
$m_{pts}=30, \epsilon_t=20$	12	-0.6095
$m_{pts}=10, \epsilon_t=25$	190	0.1403
$m_{pts}=15, \epsilon_t=25$	108	0.1325
$m_{pts}=20, \epsilon_t=25$	82	0.1266
$m_{pts}=25, \epsilon_t=25$	60	0.1176
$m_{pts}=30, \epsilon_t=25$	42	-0.0023

in the vicinity of high-density POIs, which means the overlap degree between them is relatively high, and, thus demonstrating the reliability of the proposed method. Since the trajectory data are distributed along the road while the POI data are mainly located with the city blocks, the detected hotspots and high-density POI areas are not exactly matched. For example, Hongqiao Railway Station and Airport are successfully identified as urban hotspots in the bottom left corner of Fig. 6(a), owing to the large and variable passenger flow, but the density of POI data near them is relatively low. It may be due to the fact that the public transportation hubs are mainly designed for freight/passenger traffic and are far away the city center, resulting in less developed businesses. Besides, a few residential areas have high-density POI data, but without corresponding hotspots around, such as Shangnan Community in the lower right of Fig. 6(b). The reason for the abnormal case can be deduced from the origin map. Equipped with complete facilities and possess a large number of houses, this community is marked as a high-density POI area. But the roads within it are crowded and the taxis may even be restricted to enter, which makes it almost impossible to generate an urban hotspot within the community due to fewer drop-off and pick-up points.

C. Clustering Results and Comparison

By comparing with the clustering results of typical method DBSCAN and its extension in the time domain ST-DBSCAN, the advantages of the two-phase clustering algorithm constrained by space-time and road network are elaborated. With

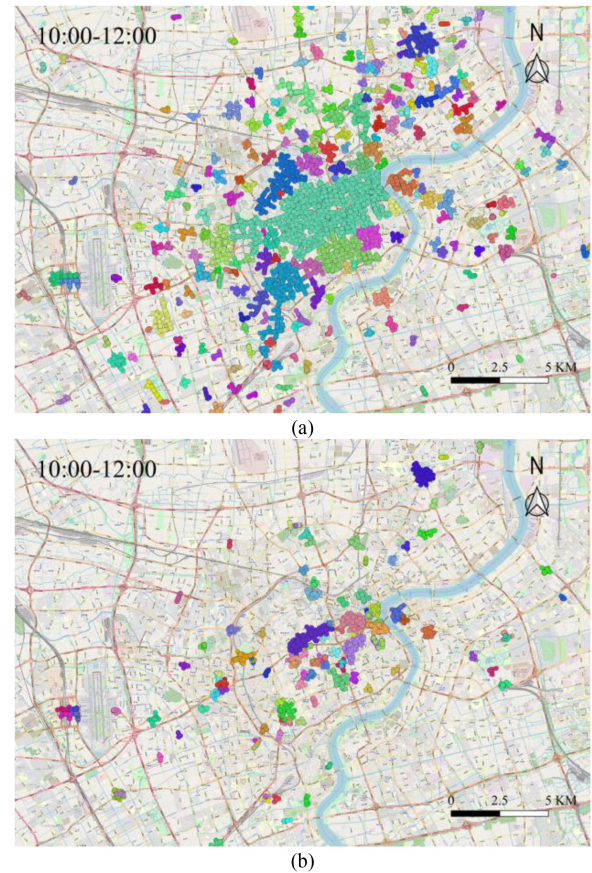


Fig. 7. Validating the advantages of the proposed method. (a) DBSCAN clustering and (b) ST-DBSCAN clustering.

the help of a simple heuristic, the input parameter $M_{in}P_{ts}$ of DBSCAN is roughly set to the natural logarithm of the number of datasets, while the value for parameter E_{ps} is picked depending on the k -distance graph. For ST-DBSCAN, the value of minimum points and spatial radius are inherited from DBSCAN and temporal radius E_{ps2} is selected from a range of 10 to 25 min via the clustering validation indicator DBCW. With the proper parameters, DBSCAN and ST-DBSCAN clustering method are adopted on the drop-off events and the detection results of urban hotspots in the selected period are presented in Fig. 7. It can be seen that DBSCAN could find clusters with arbitrary shape but lacks the ability to divide the drop-off events concentrated in the city center into independent clusters. The main reason why the clustering result of DBSCAN is undesirable is that it ignores the temporal attribute of drop-off events, which results in many spatial adjacent objects but with large time difference being grouped together. By introducing temporal similarity, both the proposed method and ST-DBSCAN can discover the dynamic patterns of urban hotspots, and thus providing preferable clustering results. As shown in Fig. 6(a) and 7(b), the two-phase clustering method and ST-DBSCAN are able to identify multiple urban hotspots that are at the same spatial location but differ greatly in time. However, compared with the proposed method, ST-DBSCAN shows a worse performance in handing clusters with different densities. Fig. 7(b) illustrates that quite a few of the clusters detected by the ST-DBSCAN are concentrated in downtown,

TABLE II
QUANTITATIVE EVALUATION OF DIFFERENT CLUSTERING ALGORITHMS

Clustering Method	Parameters	Cluster Numbers	DBC Values	Execution time
The proposed method	mpts=10, $\epsilon_t=20$ min, $\epsilon_s=200$ m	143	0.1591	124.9
DBSCAN	MinPts=10, Eps=200 m	264	-0.0347	42.4
ST-DBSCAN	MinPts=10, Eps1=200 m, Eps2=15 min	114	0.0204	60.5

leaving many urban hotspots in the outskirts undiscovered. In contrast, the proposed method works well with varying density clusters and the size of the detected hotspots is relatively smaller due to the constraint of the road network [see Fig. 6(a)].

Moreover, the quantitative evaluations of different clustering algorithms are provided to further validate the advantage of the proposed method. And the results of these comparisons are listed in Table II. It indicated that the two-phase clustering method performs better than DBSCAN and ST-DBSCAN in urban hotspot detection, and it has the potential to effectively discover spatiotemporal clusters with various shapes, sizes, and densities in the road network space. As for efficiency, with R-tree spatial index, both DBSCAN and ST-DBSCAN can reduce the time complexity to $O(n \log n)$, while the runtime complexity of our method is $O(n^2 \log n)$. We tested the time consuming by these methods on various size of dataset, and the result reveals that the proposed method performs better than the other two when the number of points is lower than 15 000, but the execution time increases rapidly for bigger size of dataset due to the calculation of large 2-D array and the introduction of temporal threshold ϵ_t . The size of the dataset we used in the comparison experiment is up to 21 558 and 10:00-12:00 is the time period with the largest number of drop-off points. The execution time of the three methods is 124.9, 42.4, and 60.5, respectively. Although the execution efficiency of our method in processing large dataset is unsatisfactory, the result of urban hotspot detection is much better than the other two method.

V. EXPLORATION OF THE DETECTED URBAN HOTSPOTS

A. Spatiotemporal Distribution

For a better understanding of the detected hotspots, the spatiotemporal characteristics of them, including spatiotemporal distribution, travel distance, and direction distribution, are further explored in this section, which also illustrate the applicability of the proposed method [42]. At first, the drop-off events in three representative periods are selected to study the spatiotemporal distribution of urban hotspots, namely 10:00-12:00, 15:00-17:00, and 20:00-22:00. And with reference to the geographical environment interpreted from the OpenStreetMap, each of the detected clusters for different time spans is assigned with a particular semantic meaning such as school and mall. Besides, the extracted urban hotspots from the processed night-time light data are shown in Fig. 8(c). From blue to red, the deeper the color indicates the greater the value of radiance intensity, and the areas with distinct red color are deemed as urban hotspots.

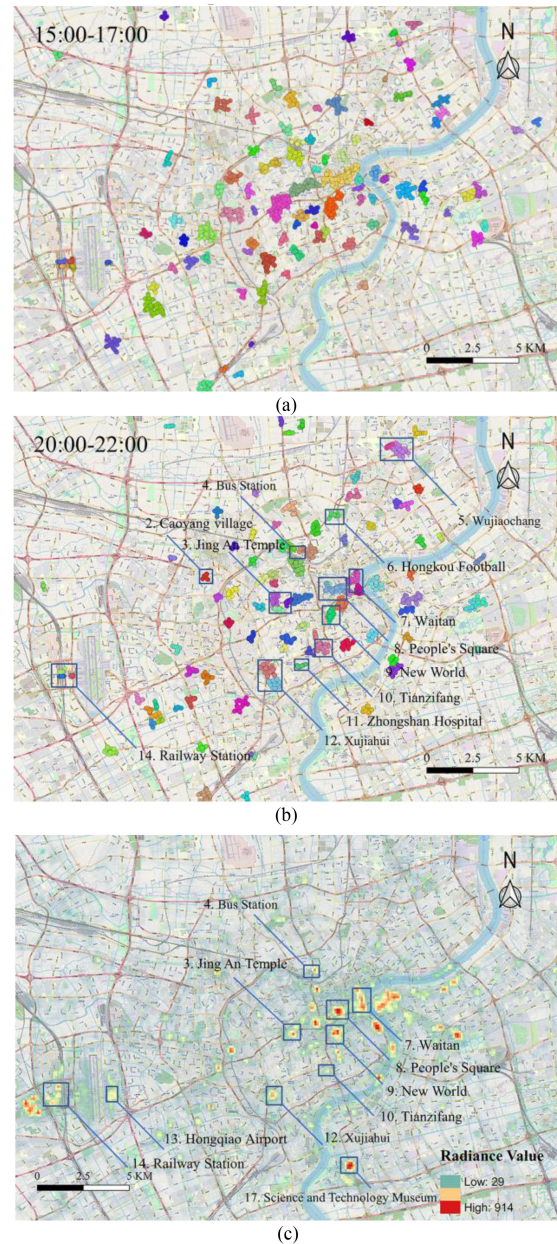


Fig. 8. Spatiotemporal distribution of the detected urban hotspots. (a) Using trajectory data in 15:00–17:00, (b) Using trajectory data in 20:00–22:00, and (c) Using LuoJia1-01 night-time light data.

As shown in Figs. 6(a) and 8, the overall spatial distribution of the identified hotspots is relatively similar in different periods and the majority of the hotspots are located within the outer ring of the city. Among them, the area around People's Square and Jing'an Temple are one of the largest urban hotspots and even tend to be connected and integrated at 15:00–17:00. In addition, public transportation hubs such as Hongqiao airport, railway stations, and bus stations have formed active urban hotspots due to the high volume of trips they maintain. Commercial blocks like Xujiahui, Tianzifang, and New World, which are acknowledged as preferred places for gathering, have been identified as urban hotspots with a long duration time but varying in scope at different time periods. Despite being far from the city center, Wujiaochang district is surrounded by numbers of universities

and zones for entertainment, and a majority of students prefer to go out for dining and shopping on a rest day, which makes it easier to generate urban hotspots within this area.

There is a certain spatiotemporal evolution of urban hotspots over the three periods, along with the change in the number of the urban hotspots detected. On the rest day, residents are more prone to do outdoor activities later than usual. The greatest number of clusters are identified during 10:00–12:00 with 143 clusters. In addition to some persistent urban hotspots such as commercial blocks and transportation hubs, residential areas like Caoyang Village, medical facilities like Zhongshan Hospital, and scenic spots like Hongkou Football stadium have formed different types of temporary urban hotspots due to the heavy travel demands of residents in the late morning and the diversity of outdoor activities distributed in various parts of the city. During 15:00–17:00, multiple previously independent clusters are gathered together in downtown and the quantity of the detected clusters decreases to 109. The more likely reason is that quite a few of people travel at this time and the volume of passenger flow around urban hotspots remains high. Only 103 cluster are discovered between 20:00–22:00, but the scope of some urban hotspots such as main business districts and large shopping malls have been expanded, of which the most notable change happens in Waitan. This is because it would be more appealing than ever with the lights on.

It is interesting to note that many striped urban hotspots are discovered along the major road, such as Zhenguang Road, which partly reflects the fact that the detection of clusters is constrained by the road network. This is mainly because these roads are covered with residential, educational, and commercial facilities that extremely appeal to residents and the drop-off points of taxi are usually determined in light of traffic conditions as well as the wishes of passengers, which are randomly distributed along with a certain range of the road with no fixed location. Moreover, it can be seen that there are two clusters around Waitan with different time span in 10:00–12:00, because the volume of passenger flow fluctuates over time and the state of urban hotspots may change dynamically between convergence and dispersion. Therefore, it is convenient to monitor the dynamic of the urban hotspots with the help of spatiotemporal constraints.

Both the trajectory data and the NTL data contain a wealth of information about human activities, which can directly affect the spatiotemporal distribution of urban hotspots. It can be seen from Fig. 8(b) and (c) that there is a strong positive correlation between them due to the similar distribution of the detected urban hotspots. Simultaneously, it also demonstrated the effectiveness of using taxi trajectory data to detect urban hotspots. Trajectory data have been extensively applied in detecting urban hotspots due to it tells a lot of stories about the area residents interested in during a day. But limited by the fact that trajectory data just account for a small part of human movement data and urban hotspots cannot be detected completely using trajectory data alone. For instance, in view of the travel cost, a significant number of residents tend to take the subway to Hongqiao Airport during the evening period, resulting in an underestimation of passenger volume and the disappearance of the hotspot detected by trajectory data at 20:00–22:00, but it can be discovered using

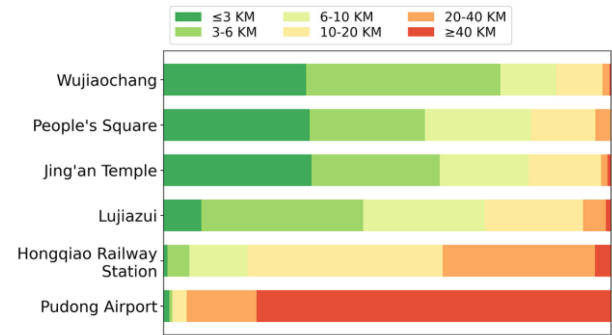


Fig. 9. Distribution of travel distance for six representative hotspots.

NTL data because of high light intensity maintained. With a finer spatial resolution of 130 m, LuoJial-01 night-time light remote sensing data create a new perspective to extract urban hotspots quickly and effectively. However, using only NTL data may cause identification errors in specific places and ignore the dynamics of urban hotspots. For example, owing to excessively high radiance intensity in surrounding roads, Science and Technology Museum was marked in deep red in Fig. 8(b), but it is unreasonable to regard it as a busy hotspot because fewer residents visit it at night. Hence, taking into consideration the advantages and limitations of trajectory and NTL data, the combination of the two types of data may provide considerable potential for urban hotspot detection to a certain extent.

B. Travel Distance Distribution

From a smaller spatial scale, we further adopt taxi trajectory points to study the travel distance and direction distribution of the detected hotspots. To simplify the experiments, only six typical hotspots with high travel demands are selected, namely People's Square, Jing'an Temple, Wujiaochang, Lujiazui, Hongqiao Railway Station, and Pudong Airport. First, the travel distance is roughly set to the cumulative distance between two adjacent track points, since it is more line with the actual situation than the shortest route distance. Then, it is divided into six levels and the proportion of travel distance in each level is shown in Fig. 9. Except for transportation hubs, the distribution of travel distance around other four hotspots are similar and the vast majority of travels are no more than 20 km, which reflects the service scope of these hotspots and also indicates that people are more willing to choose close destinations for activities. Compared to the most popular hotspots near People's Square and Jing'an Temple, there are more trips between 3 and 6 km around Wujiaochang since it is a subcenter of Shanghai and mainly caters to the travel needs of its surrounding areas, while there are fewer trips smaller than 3 km around Lujiazui because it is a central business district with limited residential building nearby. As for Hongqiao Railway Station, and Pudong Airport, they possess a higher percentage of long-distance trips than other hotspots as they are located in the suburbs and provide convenient services throughout the city.

C. Travel Direction Distribution

The travel direction is defined as the coordinate azimuth between the pick-up point and the drop-off point. First, the travel

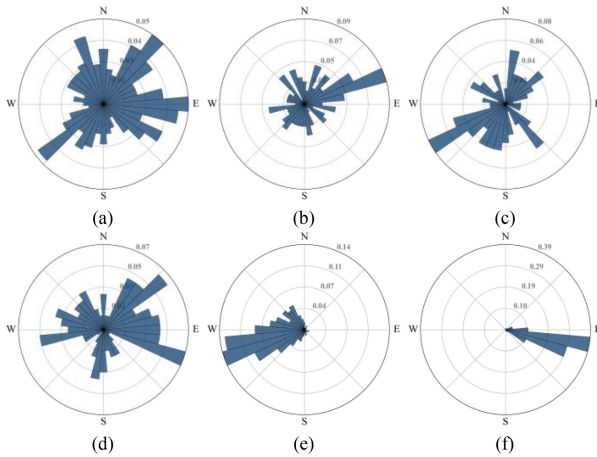


Fig. 10. Travel direction distribution and entropy value of six hotspots. (a) Wujiaochang ($H = 3.4977$), (b) People's Square ($H = 3.4324$), (c) Jing'an Temple ($H = 3.3834$), (d) Lujiazui ($H = 3.3976$), (e) Hongqiao Railway Station ($H = 2.9028$), and (f) Pudong Airport ($H = 1.6308$).

directions for all drop-off points around one urban hotspot are calculated, then we divide them into 36 equal-sized bins with an interval of 10° and take 0° as the due north. Besides, in order to get more detailed knowledge about the distribution of travel direction, Shannon entropy is employed to quantify it, where a lower value imply more regular distribution [43]. Based on the direction binned, the entropy can be written as

$$H = - \sum_{i=1}^n P(d_i) \log_e P(d_i) \quad (6)$$

where n denotes the numbers of bins and $P(d_i)$ represents the proportion of the directions in the i th bin. The distribution of travel direction for six urban hotspots is presented in Fig. 10. Of these hotspots, Wujiaochang has the most disordered distribution with travel directions across all bins and the entropy value of it is the greatest (3.4977), which means that it has a relatively high attraction to the surrounding areas, especially the direction toward People's Square. While Pudong Airport has the most regular distribution since the majority of passengers coming from the city center in the north-west, and the entropy value of it is the smallest (1.6308). Besides, an interesting phenomenon can be observed from the travel direction distribution of People's Square and Jing'an Temple, the two hotspots seem to have an active interaction due to the fact that there is a higher proportion of trips in the corresponding directions.

VI. CONCLUSION

In this article, using the taxi trajectory data, a two-phase clustering method is proposed for urban hotspots detection with spatiotemporal and network constraints. In the first phase, ST-HDBSCAN is applied so as to catch the dynamic patterns of urban hotspot over time and handle the uneven density. In the second phase, in order to make the clustering results more in line with the actual situation, the idea of region growing is utilized to filter out the noise points in the urban road network space. Then, the optimal clustering parameters for urban hotspot detection are determined on the basis of an internal validation DBCV

and prior knowledge interpreted from the OpenStreetMap. The effective of the proposed method can be partly demonstrated because there exists a well matching degree between the detected hotspots and the high-density POI areas. And compared with other conventional methods, the advantages of it are illustrated due to it could handle clusters with different sizes, shapes, and densities.

Furthermore, a case study is conducted to detect urban hotspots in Shanghai by the two-phased clustering. And for a better understanding of the detected hotspots, we further explore the spatiotemporal characteristics of them, including spatiotemporal distribution, travel distance, and direction distribution. Moreover, a new possibility is provided to combine the taxi trajectory data and NTL data, so as to detect the urban hotspots in a more comprehensive way. In essence, all these findings are determined by the travel regularities of residents and the layout of urban facilities, so exploring the characteristics behind the urban hotspots can serve as a scientific basis for policymakers in convenient travel, public facilities planning, as well as location-based service. In the future, attention could be paid to detecting the periodic patterns of urban hotspots using the trajectory data with a longer time span, various kinds of POI data and NTL data.

ACKNOWLEDGMENT

The authors would like to thank the editor, associate editor, and anonymous reviewers for their helpful comments and suggestions that improved this article greatly.

REFERENCES

- [1] T. Jia and Z. Ji, "Understanding the functionality of human activity hotspots from their scaling pattern using trajectory data," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 11, 2017, Art. no. 341.
- [2] S. Hoteit, S. Secci, S. Sobolevsky, C. Ratti, and G. Pujolle, "Estimating human trajectories and hotspots through mobile phone data," *Comput. Netw.*, vol. 64, pp. 296–307, 2014.
- [3] H. Jiang *et al.*, "A collective human mobility analysis method based on data usage detail records," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 12, pp. 2359–2381, 2017.
- [4] C. Kang, and K. Qin, "Understanding operation behaviors of taxicabs in cities by matrix factorization," *Comput., Environ. Urban Syst.*, vol. 60, pp. 79–88, Nov. 2016.
- [5] S. Zhao, P. Zhao, and Y. Cui, "A network centrality measure framework for analyzing urban traffic flow: A case study of Wuhan, China," *Physica A, Statist. Mechanics Appl.*, vol. 478, pp. 143–157, Jul. 2017.
- [6] L. Cai, F. Jiang, W. Zhou, and K. Li, "Design and application of an attractiveness index for urban hotspots based on GPS trajectory data," *IEEE Access*, vol. 6, pp. 55976–55985, 2018.
- [7] K. Siła-Nowicka, J. Vandrol, T. Oshan, J. A. Long, U. Demšar, and A. S. Fotheringham, "Analysis of human mobility patterns from GPS trajectories and contextual information," *Int. J. Geographical Inf. Sci.*, vol. 30, no. 5, pp. 881–906, 2016.
- [8] D. Zhang *et al.*, "Identifying region-wide functions using urban taxicab trajectories," *ACM Trans. Embedded Comput. Syst.*, vol. 15, no. 2, pp. 1–19, Mar. 2016.
- [9] Y. Liu *et al.*, "Social sensing: A new approach to understanding our socioeconomic environments," *Ann. Assoc. Amer. Geographers*, vol. 105, no. 3, pp. 512–530, May 2015.
- [10] Y. Dong, S. Qian, K. Zhang, and Y. Zhai, "A novel passenger hotspots searching algorithm for taxis in urban area," in *Proc. 18th IEEE/ACIS Int. Conf. Softw. Eng., Artif. Intell., Netw. Parallel/Distrib. Comput.*, 2017, pp. 175–180.
- [11] Y. Shen, L. Zhao, and J. Fan, "Analysis and visualization for hot spot based route recommendation using short dated taxi GPS traces," *Information*, vol. 6, no. 2, pp. 134–151, 2015.

- [12] C. Chen, S. Jiao, S. Zhang, W. Liu, L. Feng, and Y. Wang, "TripImputor: Real-time imputing taxi trip purpose leveraging multi-sourced urban data," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3292–3304, Oct. 2018.
- [13] G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 123–144, Jan. 2017.
- [14] Y. Hu, H. J. Miller, and X. Li, "Detecting and analyzing mobility hotspots using surface networks," *Trans. GIS*, vol. 18, no. 6, pp. 911–935, 2014.
- [15] K. Qin, Q. Zhou, T. Wu, and Y. Q. Xu, "Hotspots detection from trajectory data based on spatiotemporal data field clustering," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, no. 2W7, pp. 1319–1325, 2017.
- [16] J. Tang, F. Liu, Y. Wang, and H. Wang, "Uncovering urban human mobility from large scale taxi GPS data," *Physica A, Statist. Mechanics Appl.*, vol. 438, pp. 140–153, 2015.
- [17] M. Pavlis, L. Dolega, and A. Singleton, "A modified DBSCAN clustering method to estimate retail center extent," *Geographical Anal.*, vol. 50, no. 2, pp. 141–161, 2018.
- [18] P. Zhao, X. Liu, J. Shen, and M. Chen, "A network distance and graph-partitioning-based clustering method for improving the accuracy of urban hotspot detection," *Geocarto Int.*, vol. 34, no. 3, pp. 293–315, 2019.
- [19] P. Zhao, K. Qin, X. Ye, Y. Wang, and Y. Chen, "A trajectory clustering approach based on decision graph and data field for detecting hotspots," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 6, pp. 1101–1127, 2017.
- [20] X. Wu, C. Cheng, R. Zurita-Milla, and C. Song, "An overview of clustering methods for geo-referenced time series: From one-way clustering to co- and tri-clustering," *Int. J. Geographical Inf. Sci.*, vol. 34, no. 9, pp. 1822–1848, 2020.
- [21] X. Yang, Z. Zhao, and S. Lu, "Exploring spatial-temporal patterns of urban human mobility hotspots," *Sustainability (Switzerland)*, vol. 8, no. 7, 2016, Art. no. 674.
- [22] D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial-temporal data," *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, 2007.
- [23] Q. Liu, M. Deng, J. Bi, and W. Yang, "A novel method for discovering spatio-temporal clusters of different sizes, shapes, and densities in the presence of noise," *Int. J. Digit. Earth*, vol. 7, no. 2, pp. 138–157, 2014.
- [24] Z. Xia, H. Li, Y. Chen, and W. Liao, "Identify and delimitate urban hotspot areas using a network-based spatiotemporal field clustering method," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 8, 2019, Art. no. 344.
- [25] M. Deng, X. Yang, Y. Shi, J. Gong, Y. Liu, and H. Liu, "A density-based approach for detecting network-constrained clusters in spatial point events," *Int. J. Geographical Inf. Sci.*, vol. 33, no. 3, pp. 466–488, 2019.
- [26] X. Li, C. Elvidge, Y. Zhou, C. Cao, and T. Warner, "Remote sensing of night-time light," *Int. J. Remote Sens.*, vol. 38, no. 21, pp. 5855–5859, 2017.
- [27] X. Xin *et al.*, "Monitoring urban expansion using time series of night-time light data: A case study in Wuhan, China," *Int. J. Remote Sens.*, vol. 38, no. 21, pp. 6110–6128, 2017.
- [28] B. Devkota, H. Miyazaki, A. Witayangkurn, and S. M. Kim, "Using volunteered geographic information and nighttime light remote sensing data to identify tourism areas of interest," *Sustainability (Switzerland)*, vol. 11, no. 17, 2019, Art. no. 4718.
- [29] B. Yu *et al.*, "Integration of nighttime light remote sensing images and taxi GPS tracking data for population surface enhancement," *Int. J. Geographical Inf. Sci.*, vol. 33, no. 4, pp. 687–706, Apr. 2019.
- [30] G. Boeing, "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Comput., Environ. Urban Syst.*, vol. 65, pp. 126–139, 2017.
- [31] X. Li *et al.*, "Prediction of urban human mobility using large-scale taxi traces and its applications," *Front. Comput. Sci. China*, vol. 6, no. 1, pp. 111–121, 2012.
- [32] N. R. Velaga, M. A. Qudus, and A. L. Bristow, "Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems," *Transp. Res. Part C, Emerg. Technol.*, vol. 17, no. 6, pp. 672–683, 2009.
- [33] M. N. Sharath, N. R. Velaga, and M. A. Qudus, "A dynamic two-dimensional (D2D) weight-based map-matching algorithm," *Transp. Res. Part C, Emerg. Technol.*, vol. 98, no. July 2018, pp. 409–432, 2019.
- [34] R. J. G. B. Campello, D. Moulavi, and J. Sander, "Density-based clustering based on hierarchical density estimates," in *Advances in Knowledge Discovery and Data Mining*, vol. 7819, Berlin, Germany: Springer, 2013, pp. 160–172.
- [35] R. W. Scholz and Y. Lu, "Detection of dynamic activity patterns at a collective level from large-volume trajectory data," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 5, pp. 946–963, 2014.
- [36] Y. Chen, Z. Huang, T. Pei, and Y. Liu, "HiSpatialCluster: A novel high-performance software tool for clustering massive spatial points," *Trans. GIS*, vol. 22, no. 5, pp. 1275–1298, 2018.
- [37] P. Xiao, X. Wang, X. Feng, X. Zhang, and Y. Yang, "Detecting China's urban expansion over the past three decades using nighttime light data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4095–4106, Oct. 2014.
- [38] P. N. Happ, G. A. O. P. Da Costa, C. Bentes, R. Q. Feitosa, R. D. S. Ferreira, and R. Farias, "A cloud computing strategy for region-growing segmentation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 12, pp. 5294–5303, Dec. 2016.
- [39] D. Moulavi, P. A. Jaskowiak, R. J. G. B. Campello, A. Zimek, and J. Sander, "Density-based clustering validation," in *Proc. SIAM Int. Conf. Data Mining*, 2014, vol. 2, pp. 839–847.
- [40] M. Deng, Q. L. Liu, J. Q. Wang, and Y. Shi, "A general method of spatio-temporal clustering analysis," *Sci. China Inf. Sci.*, vol. 56, no. 10, pp. 1–14, 2011.
- [41] X. Liu *et al.*, "Characterizing mixed-use buildings based on multi-source big data," *Int. J. Geographical Inf. Sci.*, vol. 32, no. 4, pp. 738–756, 2018.
- [42] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," *Cartogr. Geograph. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, 2016.
- [43] G. Boeing, "Urban spatial order: Street network orientation, configuration, and entropy," *Appl. Netw. Sci.*, vol. 4, no. 1, pp. 1–19, 2019.



Feng Li received the B.S. degree in surveying and mapping engineering from China University of Mining and Technology, Xuzhou, China, in 2019. He is currently working toward the Ph.D. degree with the School of Environmental Sciences and Spatial Informatics, China University of Mining and Technology.

His research interests include trajectory data mining, urban hotspot detection and human movement analysis.



Wenzhong Shi received the doctoral degree from University of Osnabrück in Vechta, Germany, in 1994.

He is Otto Poon Charitable Foundation Professor at Urban Informatics, Chair Professor in GISci and remote sensing, Director of Smart Cities Research Institute, The Hong Kong Polytechnic University. His current research interests include urban informatics and Smart Cities, GISci and remote sensing, intelligent analytics and quality control for spatial data, artificial-intelligence-based object extraction and change detection from satellite imagery, and mobile mapping and 3-D modeling based on LiDAR and remote sensing imagery. He has published more than 250 academic papers that are indexed by SCI and 15 books.



Hua Zhang received the Ph.D. degree from the China University of Mining and Technology, Xuzhou, China, in 2012.

He is currently an Associate Professor with GIS and Remote Sensing, Key Laboratory for Land Environment and Disaster Monitoring of SBSM, China University of Mining and Technology. He has authored or coauthored more than 20 peer-reviewed articles in international journals, such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and *ISPRS Journal of Photogrammetry and Remote*

Sensing. His current research interests include multi/hyperspectral and high-resolution remotely sensed images processing, uncertainty in classification, pattern recognition, and remote sensing applications.