IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 11, 2020, accepted March 30, 2020, date of publication April 2, 2020, date of current version April 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2985329

Revealing Spatio-Temporal Patterns and Influencing Factors of Dockless Bike Sharing Demand

PENGFEI LIN[1], JIANCHENG WENG[1], SONG HU[1], DIMITRIOS ALIVANISTOS[2], XIN LI[3], AND BAOCAI YIN[1]
[1]Key Laboratory of Transportation Engineering, Beijing University of Technology, Beijing 100124, China
[2]Elsevier B.V., 1043 NX Amsterdam, The Netherlands
[3]Research Institute of Highway, Ministry of Transport of the People's Republic of China, Beijing 100088, China

Corresponding author: Jiancheng Weng (youthweng@bjut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51578028 and Grant U1811463, in part by the Beijing Nova Program by the Beijing Science and Technology Commission under Grant Z171100001117100, and in part by the Program of China Scholarships Council under Grant 201906540004.

ABSTRACT Dockless bike sharing plays an important role in complementing urban transportation systems and promoting the sustainable development of cities worldwide. To improve system operational efficiency, it is critical to study the spatiotemporal patterns of dockless bike sharing demand as well as factors influencing these patterns. Based on bicycle trip data from Mobike, Point of Interest (POI) data and smart card data in Beijing, we built a spatially embedded network and implemented the Infomap algorithm, a community detection method to uncover the usage patterns. Then, the Gradient Boosting Decision Tree (GBDT) model was adopted to investigate the effect of the built environment and public transit services by controlling the temporal variables. The spatiotemporal distribution shows imbalanced characteristics. About half of the total trips occur in the morning/evening rush hours and at noon. The community detection results further reveal a polycentric pattern of trip demand distribution and 120 sub-regions with a significant difference in connection strength and scale. The result of the GBDT model indicates that factors including subway ridership, bus ridership, hour, residence density, office density have considerable impacts on trip demand, contributing about 62.6% of the total influence. Factors also represent complex nonlinear relationships with dockless bike sharing usage. The effect ranges of each factor were identified, it indicates rebalancing schemes could be changed according to spatial location. These findings may help planners and policymakers to determine the reasonable scale of bike deployment and improve the efficiency of redistribution in local regions while reducing rebalance costs.

INDEX TERMS Dockless bike sharing system, spatiotemporal patterns, built environment, community detection, gradient boosting decision tree.

I. INTRODUCTION

In the last decade, the bike sharing system, an environmentally friendly urban transportation mode has been deployed and become popular in many cities around the world. The public bike sharing system provides people with convenient public bicycle access at numerous unattended stations, primarily serving daily mobility [1]. The potential benefits of public bike sharing include increasing use of public transit, relieving traffic congestion, avoiding the problems of maintaining and parking of private bicycles, and also reducing energy consumption and emissions [2], [3]. Therefore, many cities have identified public bike sharing as an effective way to complement the urban transportation systems and contribute to urban sustainability [4], [5].

In early 2017, dockless bike sharing, an innovative shared-transportation service, emerged in several cities in China and soon expanded. This new dockless bike sharing system has a significant difference compared with the traditional public bike sharing system, dockless bike sharing significantly enhances enjoyment and convenience for riders. For example, bike accessibility is improved through a built-in satellite positioning device and an intelligent lock, riders can rent the bike by using a smartphone to scan the QR code printed on

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

the bicycle body, the bike can be parked in any designated spot [6]–[8].

According to Mobike [9], one of the most famous dockless bike sharing companies in China, with increasing popularity in many cities, these innovative dockless bike sharing services reduce car usage by 3.2% in China's urban transportation systems. Furthermore, the usage of bicycles (especially bike sharing) had twice the contribution rate, at 11.6% of the volume of urban transportation. Meanwhile, like the traditional public bike sharing system, the problem of fluctuating spatiotemporal demand requires periodic redistribution of bicycles. Inefficient bike rebalancing can discourage users from choosing dockless bike sharing and increase the operating costs for service providers [8], [10]. Moreover, dockless bike sharing faces several unique problems, such as users parking illegally, which in turn encroaches road resources, and creates challenges in managing parking areas.

Unbalanced spatiotemporal demand distribution of bike sharing results in ineffective redistribution and higher operating costs. Improving the bike sharing system efficiency and encouraging more people to use bike sharing are challenges that have drawn the attention of researchers, leading to studies regarding the spatiotemporal characteristics and how determining factors affect the usage of conventional public bike sharing usage at the station level [11]–[15]. However, only a few studies have analyzed the spatiotemporal characteristics of dockless bike sharing usage based on large-scale, real-world trip data [6]–[8], [16], [17].

This study seeks to identify the spatiotemporal patterns of dockless bike sharing demand and clarify the correlations with the built environment, existing public transit systems, and the temporal factors based on a large-scale citywide trip dataset in Beijing provided by Mobike. This study implements a community detection method to reconstruct the spatiotemporal usage patterns of bike sharing. A machine learning approach, Gradient Boosting Decision Tree (GBDT), is then employed to uncover the relative importance and marginal effects of factors contributing to spatiotemporal fluctuations of dockless bike sharing demand.

This paper is structured as follows. Section II provides a comprehensive review of relevant studies. Sections III and IV represent the multi-source dataset and detailed descriptions of methodologies, respectively. Section V presents the results and analyses of models. Finally, in Section VI, we conclude our research findings and discuss future work.

II. LITERATURE REVIEW

A. SPATIOTEMPORAL CHARACTERISTICS OF BIKE SHARING TRIP DEMAND

Analyzing the spatiotemporal patterns of bike sharing trip demand can support operation scheduling. Gebhart and Noland [14] found that bicycle usage can vary considerably from month to month. Bergström and Magnusson [18] found that bike usage declined by 47 percent from summer to winter in Sweden. Significant variations have also been found between workday and non-workday usage. People prefer to use bike sharing during peak hours on workdays, which indicates bike sharing is dominantly used for commuting [4].

Many studies employed clustering methods to classify stations with similar spatiotemporal patterns and to better reveal the dynamics of the public bike sharing system [19]–[23]. Munoz-Mendez *et al.* [20] used a novel clustering method to explore spatiotemporal patterns based on bike sharing trip data in London, the results revealed self-contained and interconnected community structures. Jia *et al.* [22] applied a two-level affinity propagation clustering algorithm to divide bike sharing stations into some groups based on the trip distribution among stations and geographical locations. Shi *et al.* [23] employed the Latent Dirichlet Allocation model to investigate the use patterns of bike sharing systems in New York and Hangzhou, the findings highlighted the decisive role of bike sharing in commuting to work in the morning peak hour.

B. DETERMINING FACTORS OF THE USAGE OF BIKE SHARING

Built environment factors have an appreciable effect on the choice of travel mode as well as travel behavior. People living in areas with higher density and more diverse land use might prefer to use non-motorized travel modes [8], [15], [24]–[27]. Xu *et al.* [8] found that the bike sharing demand was associated with residential density, commercial density and the number of intersections in Singapore. Noland *et al.* [15] applied Bayesian regression to quantify the association of bike sharing usage at the station level with factors such as population and employment density, land use and public transit accessibility. Wang and Chen [24] found bike sharing stations of the Citi Bike system around cafes and restaurants would generate more ridership in New York City. Evidence indicated that more mixed land use attributes could generate more trips than a single land use attribute [25]. Built environment significantly influenced the reallocation count of bike sharing in Nanjing, for example, the docked stations with higher densities of restaurants and employment in the service areas required more bicycle removal in the morning and evening bicycle refilling in the afternoon [26]. Bao *et al.* [27] classified the bike sharing stations into different categories based on the distribution of POI within the service areas, then found the influences of the factors such as bicycle infrastructures, station capacity, and socioeconomic variables were varied across different station categories.

Bike infrastructures such as bike lanes also played a key role in increasing bike sharing demand, creating a bicycle-friendly environment and protecting riders from collisions with vehicles [13], [28]. Street lights, station connectivity, density might be positively correlated with bike sharing trip demand [4]. Improving the bike sharing accessibility would generate more ridership, however, the effects were varied with the different built environment, especially in areas with higher bike sharing service [29].

Considerable studies have investigated the correlation between bike sharing and the public transport system. Bike sharing plays an important role in combining the bike with bus and rail transit and helps to solve the first mile/last mile problem [11], [30]–[33]. People living near the bike sharing stations would like to transfer to public transportation to go to work as well as use bike sharing to ride home after work [11]. Bike sharing might provide a faster and more convenient alternative for point-to-point travel rather than transferring rail lines to get to a destination, and more limited rail coverage may contribute to use bike sharing as a first mile/last mile connection [33]. The above research shows that the built environment and public transportation facilities have an important impact on the demand for bike sharing.

In addition, since cyclists are in direct contact with the natural environment, some studies have investigated the impacts of weather on the bike sharing usage considering variables such as temperature, rainfall, snowfall, wind speed, air pollution, and relative humidity [12], [14], [34]–[36]. In summary, the literature suggests that extreme weather conditions reduce both bike usage and trip duration. Furthermore, temperature and precipitation are the most predominant factors for bicycle usage with complex non-linear relationships.

C. DISTRIBUTION PATTERNS OF DOCKLESS BIKE SHARING DEMAND

Emerging dockless bike sharing programs allow riders to start and end their trips wherever they want. Therefore, the usefulness of previous studies concentrating on docked bike sharing usage may be limited because the spatial distribution of the dockless bike sharing trip is no longer confined to fixed stations. Few studies have revealed the distribution patterns of dockless bike sharing demand [6]–[8], [10], [16]. Xu *et al.* [8] developed LSTM neural networks to forecast the temporal and spatial distribution of dockless bike sharing usage in the center area of Nanjing, China, considering exogenous factors including weather conditions, air quality, and the built environment. Shen *et al.* [16] implemented a spatial autoregressive model to investigate the associations between dockless bike sharing demand and built environment, weather variables and in Singapore.

Numerous previous research studies have provided valuable insights into understanding spatiotemporal patterns and factors such as built environment and public transport contributing to spatiotemporal fluctuations of conventional docked bike-sharing demand. Emerging dockless bike sharing programs also suffer from the problem of fluctuating spatiotemporal demand, so it is critical to uncover how factors determine dockless bike sharing usage at the regional level. To fill this gap, this study intends to investigate spatiotemporal attributes of dockless bike sharing demand and their associations with built environments and public transportation services.

III. Data

A. DOCKLESS BIKE SHARING TRIP DATA

The trip dataset for ten consecutive weekdays starting from May 10th, 2017, was provided by Mobike which is one of the most famous dockless bike sharing operators in China. Our study area is within the Fifth Ring Road of Beijing, the trips within the Fifth Ring Road account for more than 90 percent of the total trips of Mobike in Beijing. Each raw of data includes start point and end point location and time, bike ID, and anonymized user ID. The location information was provided in Geohash geocoding by Mobike, which is a hierarchical data structure and can convert longitude and latitude into strings and divides the space into grids [37]. The trip dataset uses 7-bit Geohash geocoding, which can accurately represent the grid with an area of about 0.018 square kilometers. We removed the records of the dataset of trips that occurred outside of the study area and those that occurred from midnight to 5:00 a.m., and about 3 million records were finally included in the study. Our study focuses on the impact of the built environment and public transport service, other external factors, for example, weather conditions should be consistent as much as possible during the study period to avoid the biases. It is worth noting that the weather was good in Beijing during the study period.

B. BUILT ENVIRONMENT DATA

Many existing studies suggested using density, diversity, and design to describe the built environment of the region [38], [39]. To explore the effect of the built environment on the distribution of dockless bike sharing demand, Point of Interest (POI) data and transport infrastructure data were used to reflect the built environment conditions of regions, POI data and road network vector data are both collected from AMAP, one of the most popular digital map providers in China (https://www.amap.com/). POI data includes the name, latitude, and longitude as well as the category of each facility. AMAP divides the POI into 20 categories, and we chose 5 categories, which are considered to significantly influence factors according to the existing studies, such as residence, office, entertainment, leisure, and education. We measured the densities of each category of POI. Bike lane length, as an indicator of design, was used to measure the regional convenience of cycling. We also used the Shannon entropy to measure the land-use diversity in each region considering the densities of residence, office, entertainment, leisure, and education. The detailed built environment factors of each region will be described in Section V.

C. SMART CARD DATA

Previous studies have shown that bike sharing effectively facilitates the first mile/last mile connection to public transportation [4], [31], [32]. The larger usage of public bike sharing stations adjacent to public transport systems could be ascribed to the public transit ridership [4]. To deepen understanding about associations between dockless bike sharing and public transport systems, smart card data of the same period was collected from Beijing Municipal Transportation Operations Coordination Center. Smart card data records the information of passengers traveling by bus or subway, including card ID, boarding and alighting time, and boarding and alighting line and station, etc. About 13 million bus trip records and 5 million subway trip records were obtained every day.

IV. METHODOLOGY

A. WEIGHTED DIRECTED NETWORK CONSTRUCTION AND COMMUNITY DETECTION

People riding in the city can integrate discrete areas within a network. The dockless bike sharing trip dataset includes the information of origins and destinations, spatial interactions between these areas can be further understood by aggregating trip flow. A spatially embedded network will be constructed by regarding the origin and destination grids as nodes and trip flow as edges, and complex network methods such as community detection can be employed to reveal the properties and structures of the network [19], [20].

To build a weighted and directed network $G = (V, E, W)$ from dockless bike sharing trip data, in this study each Geohash grid corresponds to a node v_i, and $V = \{v_i\,|1, 2, \cdots N\}$is the set of vertices. Trips between two grids represent a directed edge e_{ij} from node v_i to node v_j, $E = \left\{e_{ij}\,|i, j = 1, 2 \cdots N,\ i \neq j\right\}$is the set of edges, and $W = \left\{w_{ij}\,|i, j = 1, 2 \cdots N,\ i \neq j\right\}$is the set of weights, with w_{ij} representing the weight of e_{ij} and equals the trip volume passing through e_{ij}, the network construction steps are shown in Fig. 1(a), (b) and (c). In the dockless bike sharing trip network, some nodes are intensely connected while others are sparsely connected, reflecting the spatial interactions between the areas. Community detection can reduce the complexity of the network, and it is naturally suited for bike sharing trip networks [20]. The trip network will be divided into closely connected sub-networks based on community detection (Fig. 1(d)), called communities (or modules), in turn revealing the cluster characteristics of the network. The city will also be divided into some intensely connected subregions, as shown in Fig.1(e). The spatial distribution of dockless bike sharing trips can then be better understood, which provides technical support for the scheduling of shared bicycles.

In numerous community detection algorithms, the Infomap algorithm can more efficiently, steadily, and accurately deal with the large-scale weighted and directed network [41]. Infomap algorithm seeks to minimize the expected description length of the information flow of the random walk path in the network. More details of the Infomap algorithm can be found in the study constructed by Rosvall and Bergstrom [42]. Therefore, we employed the Infomap algorithm in the igraph Python package (https://igraph.org/python/) to handle the dockless bike sharing trip network in this study. We applied the modularity to measure the accuracy of community division, which ranges from 0 to 1. A good division with higher modularity indicates that there are many edges within communities and only a few between them.

B. INFLUENCE MODEL CONSTRUCTION

1) GRADIENT BOOSTING DECISION TREE

Some research studies have adopted the Gradient Boosting Decision Tree (GBDT) model, a machine learning approach

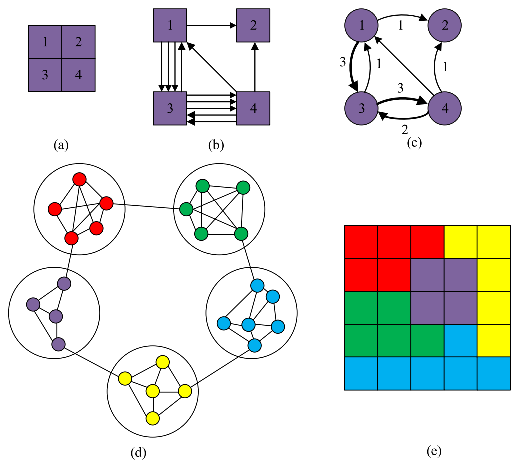

FIGURE 1. Weighted directed network construction and community detection [40].

to investigate how factors influence the crash occurrence [43] and travel behavior [44], [45], which can better illustrate the nonlinear and interactive impacts of independent variables. Therefore, this study employs the GBDT model to explore the impact of the built environment and public transit factors on average hourly trips of dockless bike sharing in each subregion, after controlling the temporal variable. Based on a robust tree-based structure generated by the GBDT model, we can obtain the relative importance and rank of each of the built environment and public transit factors contributing to regional dockless bike sharing trip demand, and identify the complex nonlinear relationship between them and the influence extent using partial dependence plots. We measured the average hourly trips in each subregion based on the same period from Monday to Friday.

The GBDT model is constructed based on an ensemble of many base decision trees. Supposing that $x = \left\{x^1, x^2, \cdots x^k\right\}$ is a set of independent variables (i.e., built environment and public transit ridership variables in this study), y is the (i.e., average hourly trips of dockless bike sharing) and $T = \{(x_1, y_1), (x_2, y_2), \cdots (x_n, y_n)\}$is the training data set. The GBDT model generates M base decision trees $h(x; a_1), \cdots h(x; a_m)$, and the feature space will be divided into J non-overlapping regions $R_{1m}, \cdots, R_{jm}$ with the corresponding predicted value r_{jm}. The approximation function $F(x)$ of the independent variableyis estimated by summation of the basis functions $h(x; a_m)$ [46]

$$F(x) = \sum_{m=1}^{M} f_m(x) = \sum_{m=1}^{M} \beta_m h(x; a_m) \tag{1}$$

$$h(x; a_m) = \sum_{j=1}^{J} r_{jm} I(x \in R_{jm}), \tag{2}$$

$$\text{where } I = 1, \quad \text{if } x \in R_{jm};\ I = 0,\ \text{otherwise}$$

where a_m represents a set of parameters of each tree $h(x; a_m)$ regarding splitting variable, splitting locations and predicted values, β_m represents the expansion coefficients. The gradient boosting procedure is employed to estimate parameters a_m and β_m, it generates the model in a stagewise way and updates the model by gradually reducing the expected value of a specific loss function $L(y, f(x))$.

To improve accuracy and avoid overfitting, some hyper-parameters such as the number of trees M, learning rate ξ, and max depth D need to be fine-tuned in the GBDT model. A larger number of trees can make the GBDT model fit the data well, but can also lead to an over-fitting problem. The learning rate ξ controls the contribution of the individual base decision tree as follows:

$$f_m(x) = f_{m-1}(x) + \xi\beta_m(x; a_m), \textit{ where } 0 < \xi < 1 \quad (3)$$

A smaller learning rate leads to better minimization of the loss function, but it also needs to increase the number of trees. There is a trade-off between the aforementioned two hyper-parameters. Max depth, referring to the maximum depth of each base decision tree, also influences the performance of the GBDT model. In general, the best performance of the GBDT model relies on the optimal combination of the number of trees, learning rate, and max depth. A five-fold cross-validation was employed to achieve the optimal performance of the model with different combinations using the Scikit-learn package in Python (https://scikit-learn.org/stable/). With a learning rate of 0.15, a number of the trees of 110, and a max depth of 5, the GBDT model achieves lowest predictive deviance.

2) RELATIVE IMPORTANCE OF INFLUENTIAL FACTORS

The GBDT model can identify relative importance or contribution of each factor on dockless bike sharing trip demand, according to the number of times that variables are selected for splitting and the degree of improvement of the model in the splitting process [46]. For an individual base decision tree T, the relative importance of variables x^k on the dependent variable can be measured as follows:

$$R_k^2(T) = \sum_{t=1}^{J-1} E_t^2 I(v_t = k) \quad (4)$$

where the sum is over $J-1$internal nodes of the tree, v_t is the splitting variable related to the node t, E_t^2 is an improvement in squared error after splitting.

For a collection of decision trees $\{T_m\}_1^M$ in the GBDT model, the importance measure is generalized by averaging all trees as follows:

$$R_k^2 = \frac{1}{M}\sum_{m=1}^{M} R_k^2(T_m) \quad (5)$$

3) PARTIAL DEPENDENCE PLOTS

After identifying the most important factors, we investigate the marginal effect of a factor on dockless bike sharing trip demand with the help of partial dependence function [46]. We suppose that S is a subset of independent variables x, C is the complement subset and $S \cup C = x$. Generally, $f(x)$ is determined by all the independent variables: $f(x) = f(x_S, x_C)$, the partial dependence of x_S on $f(x)$ can be defined as follows:

$$f_S(x_s) = E_{x_C} f(x_S, x_C) \quad (6)$$

It can be estimated by

$$\bar{f}_S(x_s) = \frac{1}{N}\sum_{i=1}^{N} [f(x_S, x_{iC})] \quad (7)$$

where $\{x_{1c}, x_{2c}, \cdots x_{Nc}\}$ is the value of $x_C$occurring in the training data set, it means that the entire training data set should be used to calculate the partial dependence of a specific independent variable. Therefore, the partial dependence function defined in (6) represents the effect of x_S on $f(x)$ accounting for the average effects of the other variables x_C on $f(x)$.

V. CASE STUDY

A. SPATIOTEMPORAL PATTERNS AND COMMUNITY DETECTION RESULTS

The results indicate the imbalanced spatiotemporal distribution of dockless bike sharing demand in Beijing. The spatial distribution of bike sharing trip demand in the grid level is shown in Fig. 2 (a). Hot spots are mainly distributed in areas within the Fourth Ring Road, with over 75 percent of grids having more than 91 trips. Each grid includes 74 trips on average, and the standard deviation is 121 trips, it indicates a high variable spatial distribution of demand across different grids. The trips in some grids are close to zero. One reason is that bike sharing is prohibited in some areas, such as Tiananmen Square and some park areas.

The community detection result is shown in Fig. 2(b). The best division generated 120 subregions with the modularity value of 0.77, it means the result of the community detection is reasonable. It is worth noting that we removed the modules in the peripheral area including less than 10 grids with low trip flows, avoiding generating unstable structures. The subregions have 240 grids and 15 thousand trips on average, respectively. On average, 76.7% of the trips start and end in the same modules. The results reveal the polycentric structure of the city and identify the borders between them. The borders of modules generated by the collective travel patterns are often affected by major roads and natural barriers such as rivers, mountains, and parks. The polycentric structure can help to improve local rebalance efficiency in each sub-region. Fig. 2(c) represents the temporal distribution of demand in each subregion; the distributions have obvious peak agglomeration characteristics. In addition to the morning peak and evening peak hours, a considerable number of trips occur at noon, about 48 percent of total trips occur during these three peak periods.

Four indicators were selected to reflect the spatial interaction characteristics among grids in each subregion, including total trip, the strength of node, average clustering coefficient, and betweenness [40], [47]. The total trip indicator depicts

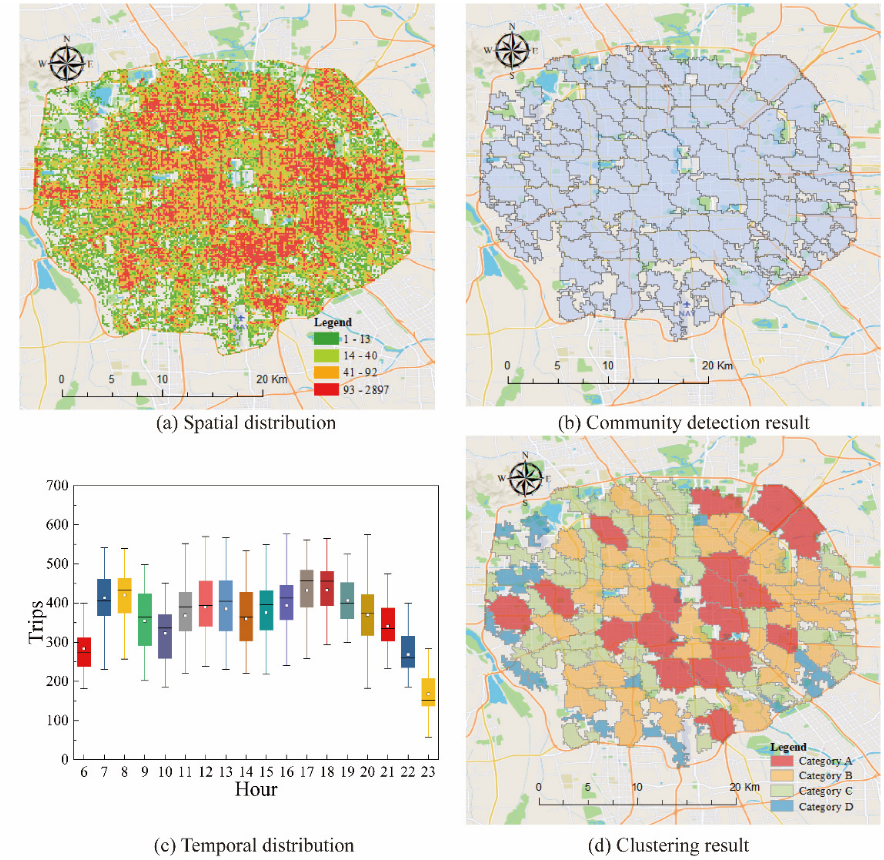

FIGURE 2. Spatiotemporal patterns and community detection results.

the total amount of trip productions and attractions in a subregion. The strength of a node is defined as the total weight of edges connecting to it, which is characterized by both inflow and outflow trip volume of each node. The average clustering coefficient measures the possibility of nodes in a network tending to cluster together. The betweenness is a measure of the importance of nodes in organizing flows along the shortest paths in the network.

We employed the K-means clustering method [48] to cluster subregions into some categories based on the above four indicators, we also applied Calinski-Harabasz score [49] to determine the best number of clusters. When the number of clusters is 4, the Calinski-Harabasz score reaches the maximum, indicating the clustering result is reasonable. The clustering result is shown in Fig.2 (d). The size of modules gradually decreases from Category A to Category D. Category A and Category B generally contain the largest number of trips, mainly distributed in the central urban area within the Third Ring Road, and the outer ones including the Wangjing area in the north of the city and the Yizhuang and Xinfadi areas in the south, where many residential neighborhoods are located. Category C mainly lies in the east and north areas of the city. Category D is mainly located in the western and southern periphery area, and the community size is relatively small.

The four indicators of each type of module are shown in Fig. 3. Compared with Category D, Categories A and B are higher in total trip and strength, indicating that these areas have border interactions between nodes within a subregion and riders may travel in various ways. However, Category D, with fewer trips and strength, has a higher value in average clustering coefficient and betweenness, meaning that human dynamics are more intensified than in Categories A and B. This finding is similar to the results of a case study developed by taxi short length trip data [40]; the different intensity of connections between nodes within a subregion may be caused by built environment conditions, population density, and other geographical heterogeneity factors.

B. DESCRIPTION OF THE INDEPENDENT VARIABLES

To comprehend the formation of the imbalanced spatiotemporal distribution of dockless bike sharing demand and the polycentric structure, we explored how built environment conditions and public transit services impact dockless bike sharing ridership in subregions. The built environment features ware quantified within the subregion polygons and linked to each subregion based on the Spatial Join Toolbox in ArcGIS software. Subway and bus trip records were aggregated by the hour for each subregion according to the boarding and

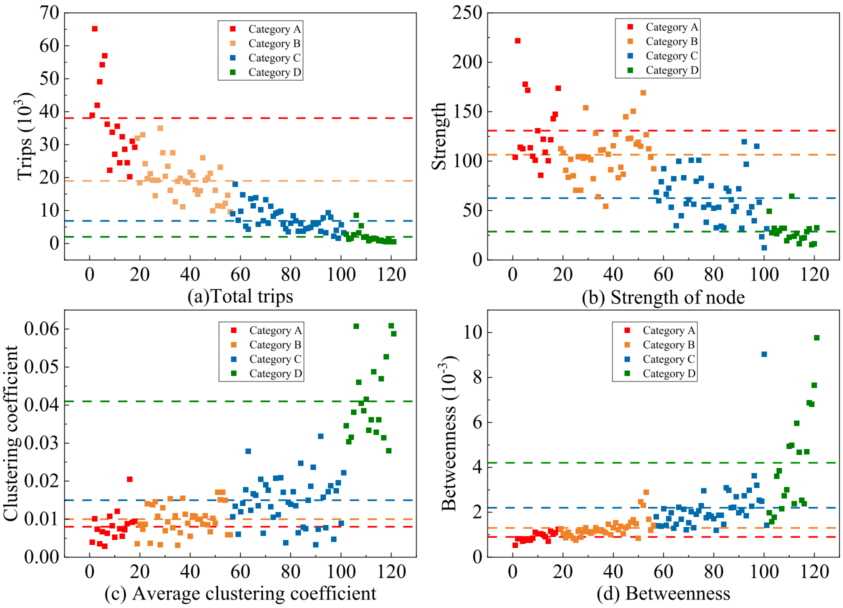

FIGURE 3. Network properties of 4 type subregions.

TABLE 1. Descriptive statistics for independent variables.

Variable	Description	Average	S.D.
Density			
Residence	The number of residential buildings per km^2 in each subregion.	31.7	22.5
Office	The number of companies, government agencies, enterprises per km^2 in each subregion.	27.9	38.0
Entertainment	The number of shopping malls, bars, cinemas, stadiums per km^2 in each subregion.	2.6	3.0
Education	The number of universities, research institutions, libraries per km^2 in each subregion.	6.4	17.9
Leisure	The number of parks, zoos, botanical gardens, scenic spots per km^2 in each subregion.	1.6	1.5
Diversity			
Landusemix	The mixture of residence, office, entertainment, leisure, and education land-use types in each subregion.	0.5	0.2
Design			
Bikelength	Length of bike lanes in per km^2 in each subregion (km / km^2).	7.4	2.1
Public transit ridership			
Subflow	The hourly average subway boarding and alighting ridership per km^2 in each subregion (10^4 trips/ km^2).	0.06	1.4
Busflow	The hourly average bus boarding and alighting ridership per km^2 in each subregion (10^4 trips/ km^2).	0.05	0.32

alighting time and location. The descriptive statistics for independent variables of each subregion are shown in Table 1.

C. EFFECTS OF INDEPENDENT VARIABLES

To explore the effects of various variables on dockless bike sharing demand, the relative importance of each influential factor was calculated. A higher value of relative importance a stronger effect; it is worth noting that the effects of all influential factors add up to 100%.

The subway ridership and bus ridership are two of the most important variables that contribute to dockless bike sharing demand with a value of 16.8% and 14.1%, respectively, indicating that public transit plays an important role in promoting dockless bike sharing use. This finding is in accord with previous studies [2], [11], [33]. Hour of the day is the third most crucial influential factor with a contribution of 11.5%. Collectively, the residence and office variables contribute to 20.3% of the total impact on dockless bike sharing demand. Previous studies have shown that dockless bike sharing demand is closely related to residential density and business density [8], [15]. Subway ridership, bus ridership, hour, residence and office attributes collectively carry out approximately 62.6% of the total contribution to dockless bike sharing demand, highlighting the significant role of bike sharing in facilitating the first mile/last mile trips.

The entertainment variable has about an 8.8% contribution to dockless bike sharing demand. The bike lane length factor has an 8.7% contribution, it indicates that bike lane length also acts as a pivotal part in correlation to dockless bike sharing demand. It is largely consistent with previous studies [13]. The leisure variable carries out about a 7.6% contribution. Land use mix contributes 6.9% to dockless bike sharing demand. It highlights the effects of mixed land use on promoting to use bike sharing [25]. The education variable has a trivial effect on the fluctuation of dockless bike sharing in Beijing with a contribution of 5.4%. The aforementioned findings can help develop seasonable rebalance strategies to increase system efficiency.

To shed further light upon how the built environment and public transit variables influence dockless bike sharing

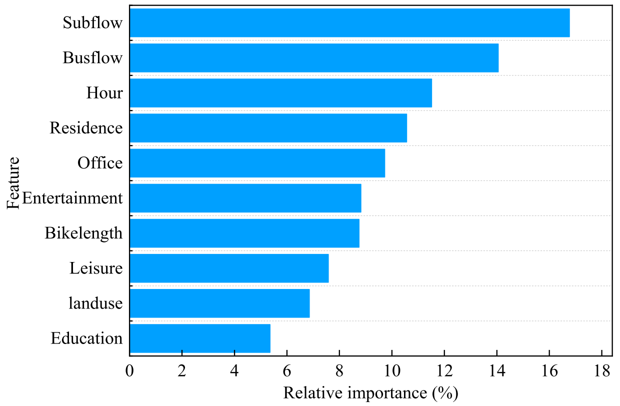

FIGURE 4. The relative importance of variables.

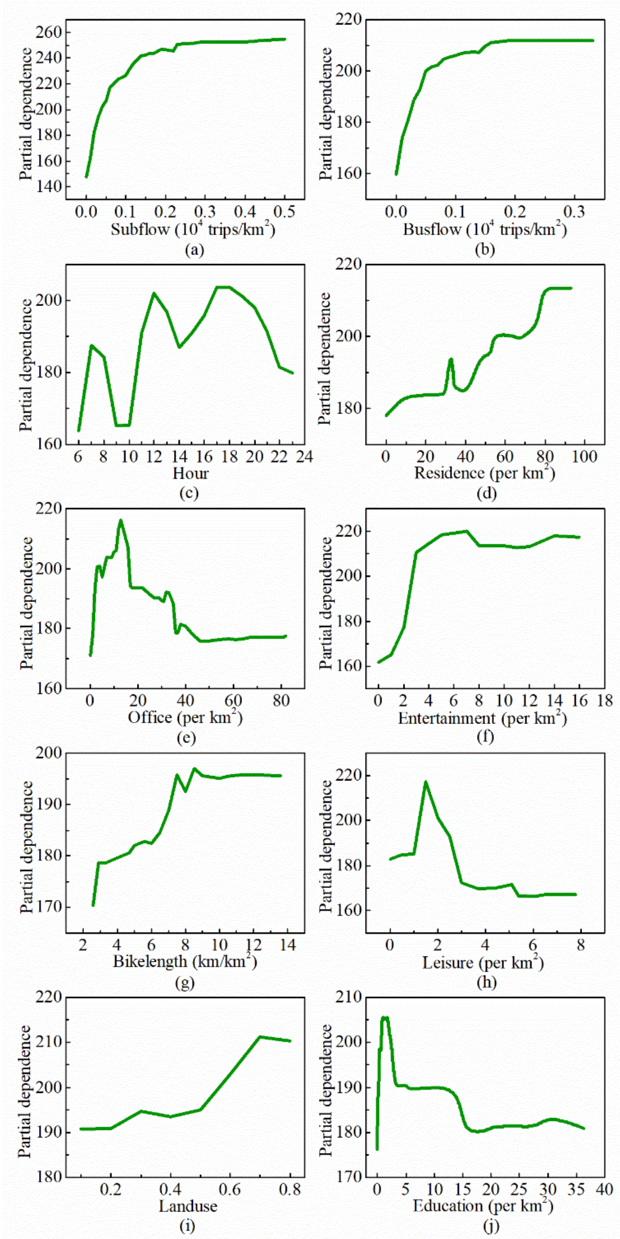

FIGURE 5. Marginal effects of variables on dockless bike sharing demand.

demand, partial dependence plots were applied to depict these relationships, as shown in Fig. 5. It is noted that the effects begin to even out after going through specific cut off points.

Fig. 5(a) indicates a nonlinearity relationship between subway ridership and dockless bike sharing demand. From 0 to 0.2, dockless bike sharing trips increased sharply as subway ridership increases; then the effect tends to reach a stable state. Similarly, the effect of the bus ridership variable also increases substantially from 0 to 0.15, and then its effect remains stable. Riders generally prefer to use dockless bike sharing in areas where more public transit lines are scheduled. Dockless bike sharing has also become a useful solution to the first mile/last mile problem in public transportation systems. These findings are also consistent with other empirical results regarding the association between public transit and bike sharing [30]–[33]. As for the hour factor, dockless bike sharing becomes more popular during rush hours in the morning and evening, it indicates commuting is a dominant use of dockless bike sharing in Beijing. In addition to that, there is also a strong positive effect at noon, possibly because some people choose to ride dockless shared bikes home for lunch or eat near their workplace during this period [17].

Fig. 5(d) indicates that residential land use has a strong positive effect on dockless bike sharing usage in general. Dockless bike sharing usage tends to increase with fluctuation as residence variable value increases. Areas with a higher residential population generally have more usage [15]. The office variable has a nonlinear impact on dockless bike sharing demand, influenced by a rapidly increasing rate when the office variable value is within 16 per square kilometer but then substantially decreasing rate for values between 16 and 87 per square kilometer. In Beijing, urban central business districts (CBD) are more pedestrian-friendly and with higher accessibility of public transport. Meanwhile, bike sharing is not superior to other travel modes considering travel time and speed.

When the entertainment density is within 7 per square kilometer, the usage of dockless bike sharing increases with a higher entertainment density. Beyond this range, its effect remains stable. This finding affirms that bike sharing usage has a positive correlation with the accessibility of entertainment facilities [21]. The bike lane length variable is positively correlated with the usage of dockless bike sharing. Dockless bike sharing trips sharply increase when the bike lane length is increased from 2.6 to 9.7 km per square kilometer. After that, the effect steadily increases and tends to remain stable. Supportive cycling facilities would encourage bike usage because bike lanes are generally constructed for areas with higher population density, meanwhile, users prefer a safer biking environment [13], [28].

A complex non-linear relationship between leisure land use and dockless bike sharing demand is identified as shown in Fig.5 (h). A leisure variable value below 1.2 per square kilometer seems to have a minor effect on dockless bike sharing usage. This effect increases substantially when the leisure variable value is between 1.2 and 1.5 per square kilometer. After that, its effect fluctuates and decreases. This may be due to the fact that some public parks offer a bicycle-friendly environment, while in other leisure places bike sharing is prohibited in Beijing. As shown in Fig. (i), land use mix has a small effect on when it is below 0.5, after that it has a significant positive impact on dockless bike-sharing trip demand. It indicates the area with a more heterogeneous land use mix appears to promote the usage of dockless bike sharing. The important role of mixed land use has also been certified in previous research [25]. The education variable has a negative effect on the dockless bike sharing demand. Beyond the range of 0 to 2, dockless bike sharing usage reduces as the education variable value increases. It may be ascribed to the majority of college students in Beijing living on campus and having personal bikes. In addition, some colleges prohibit using dockless bike sharing on campus, echoing the findings of case studies in Seattle [4] and in Beijing [36].

VI. CONCLUSION

This study utilized Mobike trip data, POI data and smart card data to reveal the temporal and spatial pattern and factor complexity of dockless bike sharing trip demand. The Infomap algorithm, a community detection method was implemented to reconstruct the spatiotemporal usage patterns of bike sharing. Gradient Boosting Decision Tree (GBDT), a machine learning method, was then employed to uncover the factor complexity of dockless bike sharing demand, considering correlations with the built environment, public transit ridership, and temporal factors. This may provide new and useful patterns for stakeholders attempting to determine a reasonable scale of the bike sharing system and improve the efficiency of redistribution in local regions.

We first employed spatiotemporal analysis and community detection to examine the mobility pattern of the dockless bike sharing, which indicates the imbalanced spatiotemporal distribution of bike sharing trips. Hot spots are mainly distributed in core areas within the Fourth Ring Road, characterized by more diverse land-use types. The temporal distribution has obvious peak agglomeration characteristics; a considerable number of trips occur in the morning and evening rush hours and at noon. The result of community detection uncovers a polycentric pattern of trip demand distribution, and 120 subregions with a significant difference in connection strength and scale are obtained. On average, 76.7% of the trips start and end in the same subregion, which indicates the subregions are self-contained and stronger local connectivity. These findings shed new light on local rebalancing schemes within subregions, operators should take full advantage of self-contained characteristic to develop rebalancing schemes for local areas within the subregions to make demand and supply balance. Moreover, the result reveals the locations (origins to destinations) are more frequently used by cyclists, it will help to set up parking areas.

The results of the GBDT model reveals the relative importance and marginal effects of factors contributing to spatiotemporal fluctuations of dockless bike sharing demand. Subway ridership and bus ridership contribute most to dockless bike sharing demand, collectively carrying out approximately 31% of the total impact. These findings affirm the relationship between bike sharing and public transport systems [11], [33]. The effect of the hour variable is also crucial in motivating to use dockless bike sharing, especially in the morning and evening peak hours and at noon. The aforementioned findings further support the role of dockless bike sharing in commuting and facilitating the first-mile/last-mile connections to public transport. Some built environment variables, such as residence, office, entertainment land use, and bike lane length, are strongly associated with the trip demand. Others, such as leisure and education factors, seem to have less impact. These findings echo with the existing studies on the relationship between built environment and bike sharing usage [4], [14], [15], [31], [36]. More mixed land use is also found to generate more trips [25]. Planners and operators can propose rebalancing schemes based upon the relative importance of these factors. Furthermore, the results show that all variables have non-linear relationships with dockless bike sharing ridership, the effect ranges of each variable have been identified, suggesting that dockless bike sharing providers could develop different scheduling strategies in areas with different levels of built environment conditions, the supply of public transport and periods. For example, the subway ridership has a significant non-linear relationship with the dockless bike sharing ridership, and the threshold value between them has been identified. Operators could dispatch different scales of the bike in areas with different levels of subway ridership so that bike sharing can better connect with the subway system and avoid piling up and blocking streets.

These findings contribute to system operators having a valuable basis to plan the best parking location and improve rebalance efficiency to enhance usage. However, our study has several limitations. First, we constructed the spatially embedded network without considering the temporal factor. We will uncover the spatial distribution pattern by constructing a dynamic graph and considering the temporal element to better understand the evolution process of community structure among different periods of the day. Second, it was found that more bike sharing trips were generated in areas with more population and employment in New York City [15], it has an important sense to examine the association of dockless bike sharing demand and the socio-economic variables, such as population and employment density. Third, we didn't explore the day-to-day variations at each subregion. Some external factors such as weather conditions and public holidays have an important influence on the usage of bike sharing. Understanding the effect of weather conditions and temporal variables can support the travel demand forecast and help optimize bike redistribution. The effects of socio-economic conditions, weather conditions and public holidays will be further studied when more data becomes available.

REFERENCES

[1] S. A. Shaheen, S. Guzman, and H. Zhang, "Bikesharing in Europe, the americas, and Asia: Past, present, and future," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2143, no. 1, pp. 159–167, Jan. 2010.
[2] H. Zhang, S. A. Shaheen, and X. Chen, "Bicycle evolution in China: From the 1900s to the present," *Int. J. Sustain. Transp.*, vol. 8, no. 5, pp. 317–335, Sep. 2014.
[3] Y. Zhang and Z. Mi, "Environmental benefits of bike sharing: A big data-based analysis," *Appl. Energy*, vol. 220, pp. 296–301, Jun. 2018.
[4] F. Sun, P. Chen, and J. Jiao, "Promoting public bike-sharing: A lesson from the unsuccessful pronto system," *Transp. Res. D, Transp. Environ.*, vol. 63, pp. 533–547, Aug. 2018.
[5] J.-J. Lin, P. Zhao, K. Takada, S. Li, T. Yai, and C.-H. Chen, "Built environment and public bike usage for metro access: A comparison of neighborhoods in beijing, taipei, and tokyo," *Transp. Res. D, Transp. Environ.*, vol. 63, pp. 209–221, Aug. 2018.
[6] Y. Jia and H. Fu, "Association between innovative dockless bicycle sharing programs and adopting cycling in commuting and non-commuting trips," *Transp. Res. Part A: Policy Pract.*, vol. 121, pp. 12–21, Mar. 2019.
[7] J. Yin, L. Qian, and J. Shen, "From value co-creation to value co-destruction? The case of dockless bike sharing in China," *Transp. Res. D, Transp. Environ.*, vol. 71, pp. 169–185, Jun. 2019.
[8] Y. Xu, D. Chen, X. Zhang, W. Tu, Y. Chen, Y. Shen, and C. Ratti, "Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system," *Comput., Environ. Urban Syst.*, vol. 75, pp. 184–203, May 2019.
[9] Beijing Mobike Technology Co., Ltd.. *Bike-sharing and the City: 2017 White Paper*. Accessed: Mar. 14, 2019. [Online]. Available: https://mobike.com/global/public/Mobike%20-%20White%20Paper%202017_EN.pdf
[10] A. Pal and Y. Zhang, "Free-floating bike sharing: Solving real-life large-scale static rebalancing problems," *Transp. Res. C, Emerg. Technol.*, vol. 80, pp. 92–116, Jul. 2017.
[11] J. Bao, C. Xu, P. Liu, and W. Wang, "Exploring bikesharing travel patterns and trip purposes using smart card data and online point of interests," *Netw. Spatial Econ.*, vol. 17, no. 4, pp. 1231–1253, Dec. 2017.
[12] A. A. Campbell, C. R. Cherry, M. S. Ryerson, and X. Yang, "Factors influencing the choice of shared bicycles and shared electric bikes in Beijing," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 399–414, Jun. 2016.
[13] E. Fishman, S. Washington, N. Haworth, and A. Watson, "Factors influencing bike share membership: An analysis of Melbourne and Brisbane," *Transp. Res. A, Policy Pract.*, vol. 71, pp. 17–30, Jan. 2015.
[14] K. Gebhart and R. B. Noland, "The impact of weather conditions on bikeshare trips in Washington, DC," *Transportation*, vol. 41, no. 6, pp. 1205–1225, Nov. 2014.
[15] R. B. Noland, M. J. Smart, and Z. Guo, "Bikeshare trip generation in New York City," *Transp. Res. A, Policy Pract.*, vol. 94, pp. 164–181, Dec. 2016.
[16] Y. Shen, X. Zhang, and J. Zhao, "Understanding the usage of dockless bike sharing in Singapore," *Int. J. Sustain. Transp.*, vol. 12, no. 9, pp. 686–700, Oct. 2018.
[17] C. Xu, J. Ji, and P. Liu, "The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 47–60, Oct. 2018.
[18] A. Bergström and R. Magnusson, "Potential of transferring car trips to bicycle during winter," *Transp. Res. A, Policy Pract.*, vol. 37, no. 8, pp. 649–666, Oct. 2003.
[19] X. Zhou, "Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in chicago," *PLoS ONE*, vol. 10, no. 10, 2015, Art. no. e0137922.
[20] F. Munoz-Mendez, K. Han, K. Klemmer, and S. Jarvis, "Community structures, interactions and dynamics in London's bicycle sharing network," in *Proc. ACM Int. Joint Conf. Int. Symp. Pervas. Ubiquitous Comput. Wearable Comput. (UbiComp)*, New York, NY, USA, 2018, pp. 1015–1023.
[21] H.-C. Liu and J.-J. Lin, "Associations of built environments with spatiotemporal patterns of public bicycle use," *J. Transp. Geography*, vol. 74, pp. 299–312, Jan. 2019.
[22] W. Jia, Y. Tan, and J. Li, "Hierarchical prediction based on two-level affinity propagation clustering for bike-sharing system," *IEEE Access*, vol. 6, pp. 45875–45885, 2018.
[23] X. Shi, F. Lv, D. Seng, B. Xing, and J. Chen, "Exploring the evolutionary patterns of urban activity areas based on origin-destination data," *IEEE Access*, vol. 7, pp. 20416–20431, 2019.
[24] K. Wang and Y.-J. Chen, "Joint analysis of the impacts of built environment on bikeshare station capacity and trip attractions," *J. Transp. Geography*, vol. 82, Jan. 2020, Art. no. 102603, doi: 10.1016/j.jtrangeo.2019.102603.
[25] Y. Zhang, T. Thomas, M. Brussel, and M. van Maarseveen, "Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China," *J. Transp. Geography*, vol. 58, pp. 59–70, Jan. 2017.
[26] D. Zhao, G. P. Ong, W. Wang, and X. J. Hu, "Effect of built environment on shared bicycle reallocation: A case study on nanjing, China," *Transp. Res. A, Policy Pract.*, vol. 128, pp. 73–88, Oct. 2019.
[27] J. Bao, X. Shi, and H. Zhang, "Spatial analysis of bikeshare ridership with smart card and POI data using geographically weighted regression method," *IEEE Access*, vol. 6, pp. 76049–76059, 2018.
[28] E. Martin, A. Cohen, J. L. Botha, and S. A. Shaheen. (2016). Bikesharing and Bicycle Safety. Mineta Transportation Institute Publications. Accessed: Mar. 14, 2019. [Online]. Available: https://scholarworks.sjsu.edu/mti_publications/207/
[29] J. Wang and G. Lindsey, "Do new bike share stations increase member use: A quasi-experimental study," *Transp. Res. A, Policy Pract.*, vol. 121, pp. 1–11, Mar. 2019.
[30] K. B. Campbell and C. Brakewood, "Sharing riders: How bikesharing impacts bus ridership in New York City," *Transp. Res. A, Policy Pract.*, vol. 100, pp. 264–282, Jun. 2017.
[31] C. Médard de Chardon, G. Caruso, and I. Thomas, "Bicycle sharing system 'success' determinants," *Transp. Res. A, Policy Pract.*, vol. 100, pp. 202–214, Jun. 2017.
[32] S. A. Shaheen, H. Zhang, E. Martin, and S. Guzman, "China's Hangzhou public bicycle: Understanding early adoption and behavioral response to bikesharing," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2247, no. 1, pp. 33–41, Jan. 2011.
[33] E. W. Martin and S. A. Shaheen, "Evaluating public transit modal shift dynamics in response to bikesharing: A tale of two U.S. Cities," *J. Transp. Geography*, vol. 41, pp. 315–324, Dec. 2014.
[34] T. Thomas, R. Jaarsma, and B. Tutert, "Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: The influence of weather on cycling," *Transportation*, vol. 40, no. 1, pp. 1–22, Jan. 2013.
[35] B. Caulfield, M. O'Mahony, W. Brazil, and P. Weldon, "Examining usage patterns of a bike-sharing scheme in a medium sized city," *Transp. Res. A, Policy Pract.*, vol. 100, pp. 152–161, Jun. 2017.
[36] P. Lin, J. Weng, Q. Liang, D. Alivanistos, and S. Ma, "Impact of weather conditions and built environment on public bikesharing trips in Beijing," *Netw. Spatial Econ.*, to be published, doi: 10.1007/s11067-019-09465-6.
[37] G. Neimeyer. *Geohash Tips & Tricks*. Accessed: Jun. 5, 2019. [Online]. Available: http://geohash.org/site/tips.html
[38] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: Density, diversity, and design," *Transp. Res. D, Transp. Environ.*, vol. 2, no. 3, pp. 199–219, Sep. 1997.
[39] R. Ewing and R. Cervero, "Travel and the built environment: A meta-analysis," *J. Amer. Planning Assoc.*, vol. 76, no. 3, pp. 265–294, Jun. 2010.
[40] 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.
[41] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 80, no. 5, Nov. 2009, Art. no. 056117.
[42] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 4, pp. 1118–1123, Jan. 2008.
[43] Y.-S. Chung, "Factor complexity of crash occurrence: An empirical demonstration using boosted regression trees," *Accident Anal. Prevention*, vol. 61, pp. 107–118, Dec. 2013.
[44] C. Ding, X. Cao, and Y. Wang, "Synergistic effects of the built environment and commuting programs on commute mode choice," *Transp. Res. A, Policy Pract.*, vol. 118, pp. 104–118, Dec. 2018.
[45] C. Ding and X. Cao, "How does the built environment at residential and work locations affect car ownership? An application of cross-classified multilevel model," *J. Transp. Geography*, vol. 75, pp. 37–45, Feb. 2019.

[46] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
[47] L. Sun and I. Rahwan, "Coauthorship network in transportation research," *Transp. Res. A, Policy Pract.*, vol. 100, pp. 135–151, Jun. 2017.
[48] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, Berkeley, CA, USA, Jun. 1967, vol. 1, no. 14, pp. 281–297.
[49] T. Caliński and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.*, vol. 3, no. 1, pp. 1–27, Jun. 2007.

PENGFEI LIN received the B.S. degree in transport engineering from Chongqing Jiaotong University, Chongqing, China, in 2015. He is currently pursuing the Ph.D. degree in transportation planning with the Beijing University of Technology, Beijing, China. His research interest includes spatiotemporal data modeling.

JIANCHENG WENG received the B.S. degree in transport engineering and the Ph.D. degree in transportation planning from the Beijing University of Technology, Beijing, China, in 2000 and 2007, respectively. He is currently an Associate Professor with the College of Metropolitan Transportation, Beijing University of Technology. His research interests include intelligent transportation systems, traffic information systems, and technology.

SONG HU received the B.S. degree in transport engineering from the Haibin College, Beijing Jiaotong University, Cangzhou, China, in 2015, and the M.S. degree in transportation planning from Chongqing Jiaotong University, Chongqing, China, in 2018. He is currently pursuing the Ph.D. degree in transportation engineering with the Beijing University of Technology, Beijing, China. His research interests include travel behavior and intelligent transportation systems.

DIMITRIOS ALIVANISTOS received the B.S. degree in informatics from the Athens University of Economics and Business, Athens, Greece, in 2015, and the M.S. degree in artificial intelligence from Vrije Universiteit Amsterdam, The Netherlands, in 2018. He is currently a Researcher with Elsevier, The Netherlands. His research interests include knowledge representation, graphs, and machine learning.

XIN LI received the B.S. degree in transport engineering from Chongqing Jiaotong University, Chongqing, China, in 2015, and the M.S. degree in transport planning and engineering from the University of Leeds, U.K., in 2016. He is currently a Researcher with the Research Institute of Highway, Ministry of Transport of the People's Republic of China. His research interest includes intelligent transportation systems.

BAOCAI YIN received the M.S. and Ph.D. degrees in computational mathematics from the Dalian University of Technology, Dalian, China, in 1988 and 1993, respectively. He is currently a Professor with the College of Metropolitan Transportation, Beijing University of Technology. He is also the Director of the Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Beijing, China.