Journal of Transport Geography 92 (2021) 103018

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Using mobile phone data to determine spatial correlations between tourism facilities

Chen Qian [a], Weifeng Li [a,*], Zhengyu Duan [a], Dongyuan Yang [a], Bin Ran [b]

[a] The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China
[b] Department of Civil and Environmental Engineering, University of Wisconsin, Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Mobile phone data
Spatial correlation
Tourism transportation
Tour bus
Transportation accessibility

ABSTRACT

Mobile phone data provide a more complete and accurate description of tourism transportation demand than traditional and Internet data sources. In this paper, a framework is proposed to determine spatial correlations between tourism destinations, rest places, and transportation hubs based on mobile phone data. Firstly, nine rules for identifying visitors based on four spatial and temporal features are established. Then, the spatial correlations are analyzed from three aspects. A case study of Shanghai is carried out to verify the proposed methodology, and the addition of a tour bus network based on the evaluation of transportation accessibility is discussed. It is concluded that tourists tend to rest near next-day destinations and choose transportation hubs in the city center. The rest places that are sightseeing destinations, amusement parks, and convention centers exhibit polycentric characteristics. The research framework and results of this study are useful for tourism transportation planning.

1. Introduction

With the development of metropolitan areas and urban agglomeration, travel between cities has increased, and the purposes of intercity travel have become more diversified. The boundaries between tourism, business, and intercity commuting are getting increasingly blurred. For example, business tourism, a limited and focused subset of regular tourism, has a significant market share, especially in primary MICE (meetings, incentive travel, conventions, and exhibitions) destinations. The diversified purposes of visitors may lead to diversified patterns of tourism-related activities, as well as diversified requirements of tourism facilities. Most tourism facilities, such as airports, restaurants, and hotels, are frequented by visitors for different purposes. Moreover, the attractions of scenic spots may vary for different travel purposes. Tourism transportation typically connects famous scenic spots in the city center. However, due to the increasing demand for tourism facilities, such as hotels and transportation hubs, there is a need to connect these tourism-related activity sites in addition to tourist attractions. Therefore, it is necessary to obtain a better understanding of the spatial correlations between tourism-related facilities to satisfy the diversified demand of tourism transportation. The spatial correlation data can be used to evaluate transportation accessibility, the establishment of tour bus lines, and the supply of catering and accommodation.

Knowledge of the tourism market and tourist behavior is usually obtained from traditional data sources, such as statistical yearbooks (Shanghai Bureau of Statistics, 2020) and surveys (Rudjanakanoknad and Rattanasuwan, 2011; Huang et al., 2011; Coban, 2012; Bieland et al., 2016) and Internet data sources, including online trip diaries (Ma et al., 2018) and geotagged social media data (Kotus et al., 2015; Sun et al., 2018; Wu et al., 2018; Khan et al., 2020; Mou et al., 2020). Traditional methods include analyses of tourism facility statistics, market trends, and tourist preferences. Internet-based methods provide more detailed information on tourists. However, deficiencies still exist regarding the accuracy and real-time analysis of tourism transportation. Due to today's diversified demands of tourism transportation, existing methods cannot identify tourism-related activities and determine the relationship between activities and facilities at the individual level.

With the development of mobile communication technology, the use of mobile phone data has become a novel technical approach for the pervasive tracking of the movements of individuals. Compared with existing data sources, mobile phone data provide a more complete and accurate depiction of tourist flows. First, in terms of the sample rate and

* Corresponding author.
E-mail addresses: 969549096@qq.com (C. Qian), liweifeng@tongji.edu.cn (W. Li), d_zy@163.com (Z. Duan), yangdyk@tongji.edu.cn (D. Yang), bran@engr.wisc.edu (B. Ran).

https://doi.org/10.1016/j.jtrangeo.2021.103018
Received 30 September 2020; Received in revised form 28 February 2021; Accepted 4 March 2021
Available online 19 March 2021
0966-6923/

the sample size, mobile phone data meet the analysis requirements to track individual visitors from other cities. Second, mobile phone data describe the trajectories of individuals in the city comprehensively because of regular communications between individual mobile devices and local base stations. Third, in addition to tourist hotspots (De Cantis et al., 2016; Wu et al., 2018; Mou et al., 2020), sleeping facilities (Kotus et al., 2015) and transportation hubs (Zhong et al., 2016, 2018, 2019), which are important facilities that tourists frequently visit, can be detected accurately and economically by mobile phone data rather than geotagged social media data, such as Weibo (Sun et al., 2018; Khan et al., 2020) and Flickr (Wu et al., 2018; Mou et al., 2020).

The main focus of this study is to determine the diversified demands of transportation between tourism facilities using mobile phone signaling data. The following steps are used to achieve this objective: 1) to identify visitors from unlabeled mobile subscribers at the city level; 2) to extract tourism-related activities from the diversified activities of visitors; 3) to describe the spatial correlation between tourism facilities. Subsequently, the transportation accessibility between tourism facilities is evaluated considering demand and supply. Our study provides three contributions to the existing literature. First, we divide the tourism facilities into three types, including tourist destinations, rest places, and transportation hubs, for a comprehensive description of the spatial correlations. Second, we use the spatial and temporal behavior of individuals to determine which of the unlabeled mobile subscribers are visitors. Third, we provide constructive suggestions for potential application scenarios of the spatial correlations between tourism facilities, such as the evaluation of public transit accessibility and the addition of a tour bus network.

The remainder of the paper is organized as follows. Section 2 reviews the publications on data sources used in tourism studies, mobile phone data applications, and the use of mobile phone data in tourist investigations. Section 3 introduces the datasets applied in this study. Section 4 describes the proposed framework, including data preprocessing, visitor identification, and the correlation analysis of the tourism facilities. Section 5 presents the results of the identified tourists, the spatial correlations between tourist destinations, rest places, and transportation hubs, and the evaluation of transportation accessibility. The applications of the proposed framework to tourism transportation, such as the design of a tour bus network, are discussed briefly in Section 6. The last section provides the conclusions.

2. Literature review

2.1. Data sources used in tourism studies

Statistical yearbooks (Shanghai Bureau of Statistics, 2020) contain information on the annual number of local tourists and tourists from other cities and foreign countries, the tourism income, and the number of tourist destinations, star hotels, and travel agents. However, statistical yearbooks do not describe the characteristics of tourists in detail, including the spatial distribution and correlation. Surveys were conducted to understand tourist demands and preferences. Due to limitations of the sample size and cost, the respondents were typically limited to a small research area, such as tour bus lines (Rudjanakanoknad and Rattanasuwan, 2011), transportation hubs (Huang et al., 2011), hotels (Coban, 2012), and destinations (Bieland et al., 2016; De Cantis et al., 2016).

With the development of the Internet, tourists have shared their travel experiences through website diaries and social media, which cover broader areas in the city than surveys. According to the travelogues of the Ctrip website, individual movement patterns in Taiwan were explored by extracting trajectories (Ma et al., 2018). Social media data provide accurate geolocation information, such as Weibo (Chinese microblog App) and Flickr (International photo-sharing website). The characteristics of tourism hotspots, such as their popularity (Sun et al., 2018), the network attributes of tourist movements (Wu et al., 2018; Mou et al., 2020), and the tourists' spatiotemporal patterns (Khan et al., 2020), were described and modeled using machine learning algorithms, including a long short-term memory model (Sun et al., 2018) and complex network models (Wu et al., 2018; Mou et al., 2020). Although these Internet data sources provided detailed information on tourists, the studies suffered from two types of bias. First, the records were strongly influenced by the tourists' subjective selection when they recalled their travel experiences, took photos, or wrote microblogs. Second, the samples lacked representativeness of the tourist population of the target city because Ctrip/Flickr/Weibo users might possess more professional cameras and might be financially more stable than the majority of tourists.

2.2. Applications of mobile phone data

Mobile phone data cover the population of a city comprehensively and contain information on the space-time trajectories of individuals. Therefore, mobile phone data provide information on the mobility patterns of individuals (Gonzalez et al., 2008; Calabrese et al., 2013) and the spatial distribution (Kang et al., 2012; Deville et al., 2014; Louail et al., 2014). Spatial correlations (Gao et al., 2013; Iqbal et al., 2014; Li et al., 2014; Yu et al., 2018) were discovered using this dataset. It was verified that mobile phone data represented a reasonable alternative for individual mobility studies compared with the results of common data sources, such as vehicle safety inspection data (Calabrese et al., 2013), the national census, and remotely sensed data (Deville et al., 2014). The relationship between travel distance and location frequency was discussed adequately by Gonzalez et al. (2008) and Kang et al. (2012). However, these features were limited by the city's size and shape in the spatial analysis conducted by Kang et al. (2012) and Louail et al. (2014). Mobile phone data have proven to be more economical than traditional approaches of origin-destination matrix estimation in spatial correlation analysis (Iqbal et al., 2014). In addition, daily spatial interactions (Li et al., 2014; Gao et al., 2013) and job-housing relationships (Yu et al., 2018) were investigated using frequent itemset mining (Li et al., 2014) or community detection algorithms (Gao et al., 2013; Yu et al., 2018). In these studies, the home location had to be extracted from mobile phone data because the residents were the primary research object.

2.3. Mobile phone data in tourist investigations

Mobile phone data have been used since 2007 to analyze tourist/visitor mobility. The first problem addressed was the differentiation of tourists/visitors from residents. Based on roaming records in Estonia, the evaluation of mobile phone data (Ahas et al., 2008), the seasonality of foreign tourists' space (Ahas et al., 2007), and the comparison of travel distances of event visitors and regular visitors (Nilbe et al., 2014) were discussed. However, given that the registration location of anonymous mobile subscribers was not included in the data due to privacy protection, tourists/visitors were identified based on deterministic rules based on domain knowledge rather than only machine learning methods (Furletti et al., 2012; Mamei and Colonna, 2018; Wu et al., 2020). Thus, the definition of tourists was at the core of these studies. In the study by Furletti et al. (2012), commuters, people in transit, and tourists were placed in different categories from residents. We consider these groups of people visitors. The presence of tourists whose homes and workplaces were outside of study area A was limited to a threshold that allowed them to participate in activities in area A. Mamei and Colonna (2018) studied commuters, people in transit, tourists, and excursionists. We consider these groups of people visitors. Excursionists were defined as visitors who stayed for more than four hours but not overnight. Tourists were defined as overnight visitors who stayed in the city for more than one and less than four days. Wu et al. (2020) considered people who came to a city for no more than ten days as visitors and visitors who visited the top 50 attractions no less than two times as tourists. However, travelers who came to the city for 11 to 20 days and did not register their

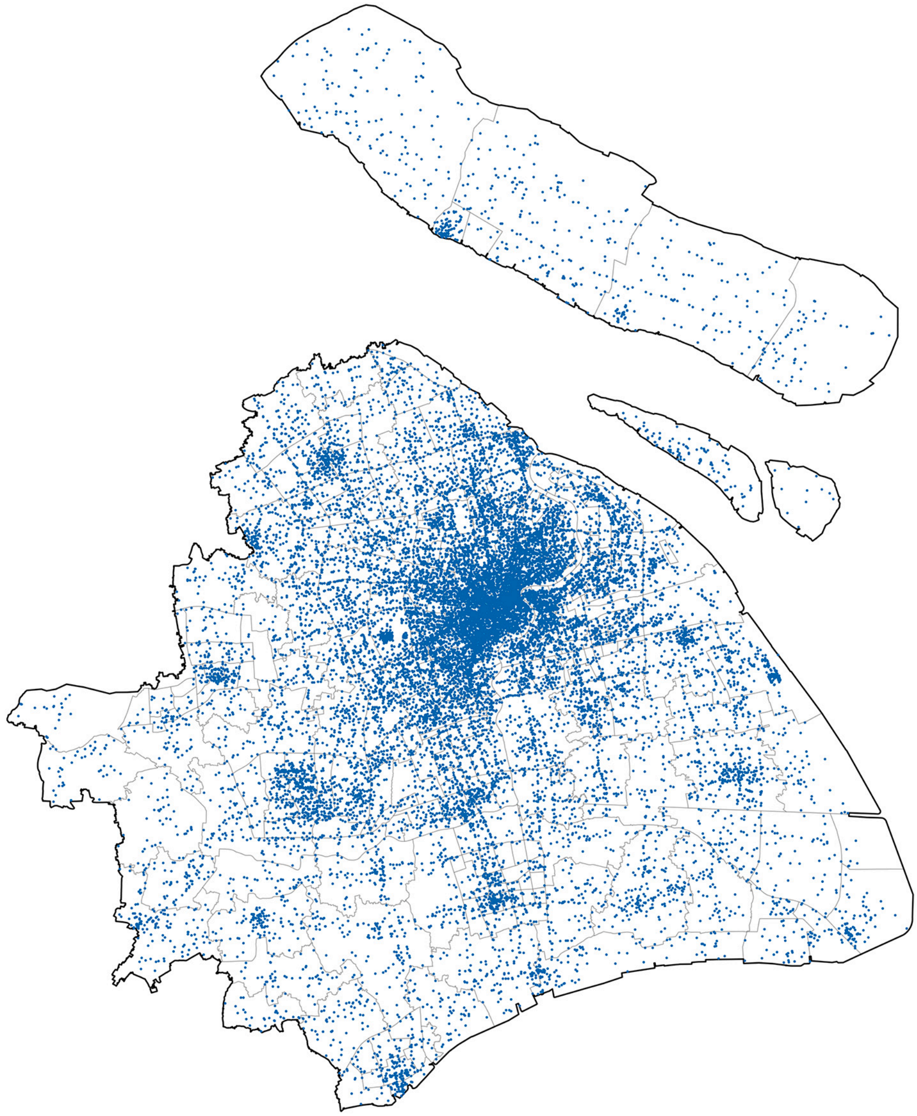


Fig. 1. The distribution of base stations in Shanghai.

phones in the city were not included in visitors. These three definitions indicate that different researchers use different definitions of the visit duration of tourists. Visitors/tourists who visited for an intermediate number of days were not clearly identified. Moreover, the difference between visitors and tourists was unclear. In fact, visitors are the opposite of residents, whereas non-local tourists are a subset of visitors. According to the tourist definition, the indicators of the spatial patterns have to be considered as well.

In the 2010s, the tracking technologies for tourism research were used in the issues from space-time discovery to tourism and consumer behavior (Shoval and Ahas, 2016), including the spatial correlation between tourism facilities. Jia et al. (2013) analyzed the distribution of the areas of the Expo tourist activity over time. Zhong et al. (2016, 2018, 2019) examined the distribution of the origins and destinations of passengers entering or exiting the Hongqiao transportation hub in Shanghai using kernel density estimation (KDE), raster analysis, and evaluation of traffic analysis zones (TAZs). The authors also mined the relationship between the core area and the surrounding area of the hub, the city area excluding the hub, and the area outside the city using association rules (Zhong et al., 2019). Graph theory has been commonly used to determine the spatial structure (Wu et al., 2018; Qian et al., 2019; Mou et al., 2020). Wu et al. (2018) developed a spatially-embedded tourism hotspot network of Beijing and conducted complex network analysis to determine the network characteristics. Mou et al. (2020) examined the changes in the network structure of the tourism areas of Shanghai before/during/after the Expo. The degree of the vertices, the clustering coefficient, and the shortest path length are the three most common indicators used in network analysis. Qian et al. (2019) used three S-dimension measurements, i.e., the strength, symmetry, and structure of the adjacency matrix of tourist flows, to evaluate the spatial correlation between activity areas of tourists in TAZs. Transportation hubs and rest places, which had strong correlations with tourism hotspots, were not separated from tourist activity locations in the reviewed studies. Moreover, the applications of the correlation results were not discussed adequately in these studies. Accessibility is an indicator of the serviceability of a city's transportation system. Since the tour bus system is designed to transport tourists, it is necessary to evaluate and modulate the system dynamically based on novel tracking data. Generally, accessibility measures include two parts: (1) the potential opportunities, such as the number of attraction points in a given area; (2) the spatial distance or the travel time between the origin and the destination (Guy, 1983; Liu and Kwan, 2020; Zuo et al., 2020). Thus, before the accessibility of tour bus lines is determined, the transportation demand of tourists should be considered.

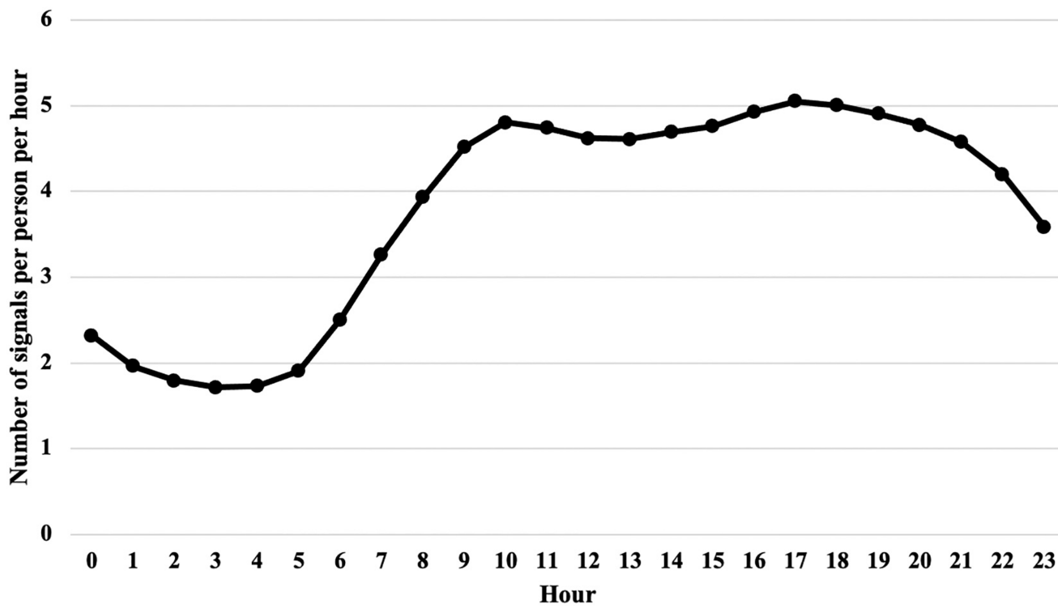


Fig. 2. The number of signals per person per hour.

Table 1
Datasets.

(a) mobile phone data

Field	Explanation	Type	Example
MSID	Encrypted ID of the mobile subscriber	Varchar(32)	0EC935E7D4192B8929E605ABD1EA7FE8
DATETIME	Time(yyyyMMddhhmmss)	Varchar(14)	20111016120337
LAC	Location Area Code	Varchar(5)	06164
CI	Cell ID	Varchar(5)	16977
EVENTID	Type of the signal transition event	Varchar(3)	002
LON	Longitude of the base station	Decimal(9,6)	121.367633
LAT	Latitude of the base station	Decimal(9,6)	31.275217

(b) POI data

Field	Explanation	Type	Example
NAME	Name of the POI	Varchar(100)	Sheshan State Resort
CLASS	Category of the POI	Varchar(20)	View spot
LON	Longitude of the POI	Decimal(9,6)	121.199296
LAT	Latitude of the POI	Decimal(9,6)	31.088591

Notes: The POI data were obtained from http://lbsyun.baidu.com/index.php?title=webapi/guide/webservice-placeapi.

3. Data

Mobile phone data, point of interest (POI) data, and Direction API data are used in this study. The mobile phone data used in this study were confidential and provided by a mobile operator in China through a governmental fund. Mobile phone data depict the spatiotemporal characteristics of the activities of individuals. POI data and Direction API data, which were obtained from the Baidu Map Open Platform, reflect the built environment and the connectivity and accessibility of locations where the tourism-related activities occur.

Mobile phone data store spatiotemporal information on mobile subscribers, i.e., the location and time of the signal transition events of mobile devices of individuals. The mobile phone data used in this study were collected during October 2011. There were 21,813 base stations in the Shanghai region, as shown in Fig. 1. The coverage radius of the base stations in the center of Shanghai ranges from 500 to 800 m. The location area code (LAC) and the Cell ID (CI) determine the position of the base station whose longitude and latitude are indicated as LON and LAT. In this study, 16 million mobile subscribers produced 2–5 records of cellular signal information on average per person per hour, as shown in Fig. 2. The frequency met the analysis requirements of the trajectories of individual movements. The data record of each signal transition event is stored in the dataset. The signal transition events activating the connection between the base stations and mobile devices include device on/off, sending/receiving messages, making/receiving phone calls, periodic location updates, and handoffs in cellular telecommunications. The data record generated by one subscriber is presented in Table 1(a), showing an example of the mobile phone signaling data.

From the POI data, 775,846 POIs in Shanghai city were classified into 19 categories, such as view spots, hotels, transportation facilities, companies, communities, and government agencies. An example of POI data is listed in Table 1(b).

Direction API data were used to analyze the transportation accessibility between tourism facilities. Thus, it was necessary to access the trip duration of public transit and taxi. An example of Direction API data is listed in Table 1(c), showing a 68-min metro journey from Pudong

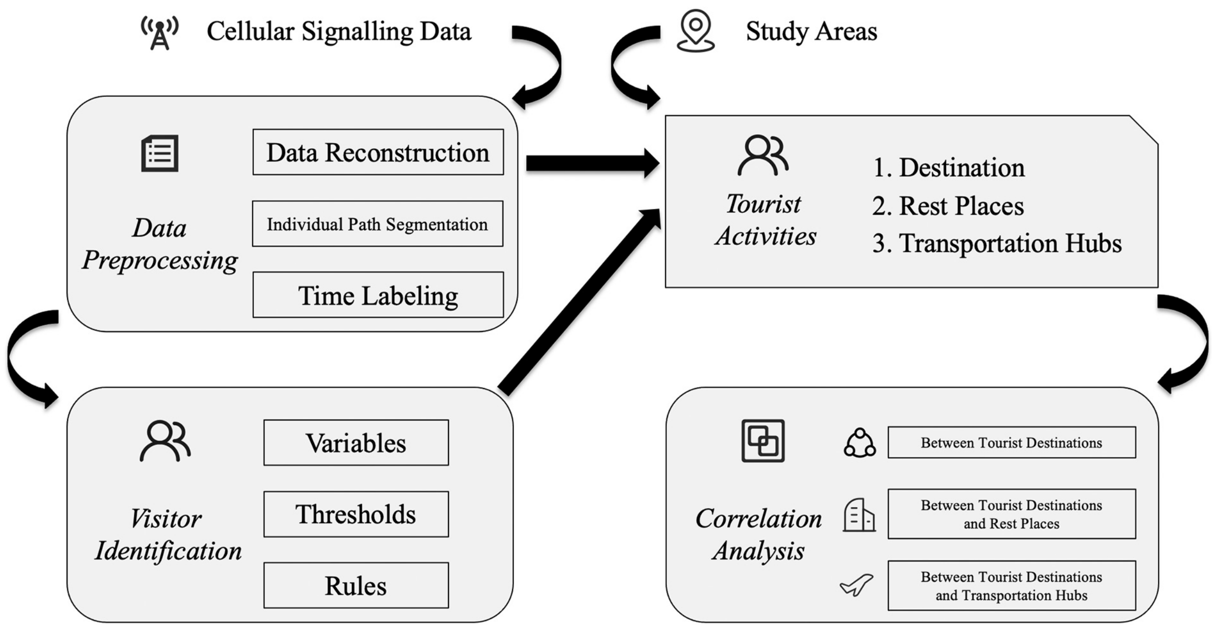


Fig. 3. The research framework for the analysis of the spatial correlations between tourism facilities.

International Airport to Lujiazui.

(c) Direction API data

Field	Explanation	Type	Example	
LON1	Longitude of the origin	Decimal (9,6)	121.813249	Pudong International Airport
LAT1	Latitude of the origin	Decimal (9,6)	31.157247	
LON2	Longitude of the destination	Decimal (9,6)	121.508984	Lujiazui
LAT2	Latitude of the destination	Decimal (9,6)	31.243812	
MODE	Traffic mode	Varchar (20)	Metro	
DURATION	Time spent during the trip (min)	Int(3)	68	

Notes: The Direction API data were obtained from http://lbsyun.baidu.com/index.php?title=webapi/direction-api-v2.

4. Methodology

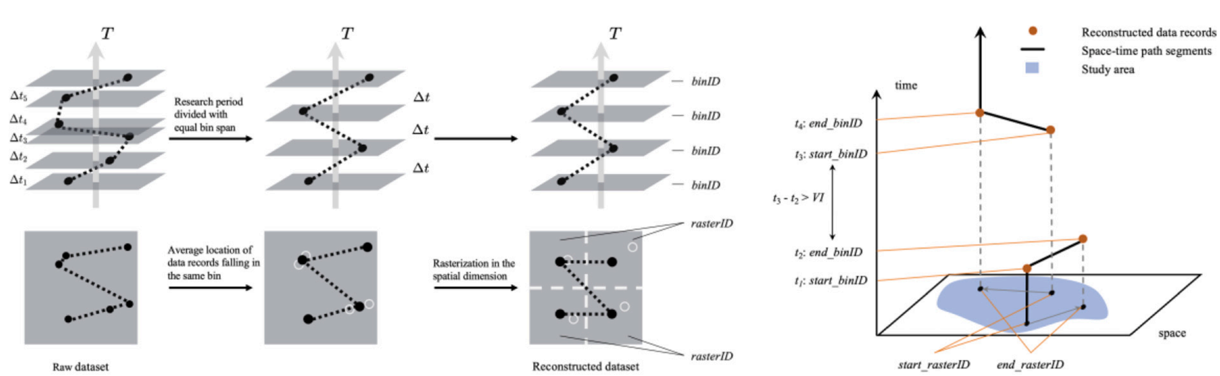
The mobile phone data of a city represent the population that consists of residents and visitors. Tourism-related activities most likely occur in areas such as the central business district (CBD), commercial circles, sightseeing locations, activity venues, and transportation hubs. Researchers selected these areas according to the background of the city. We propose a research framework (Fig. 3) based on mobile phone data to analyze the spatial correlations between tourism facilities. The framework consists of preprocessing of the mobile phone data, including data reconstruction, individual path segmentation, and time labeling, the identification of visitors based on rules defined by four spatiotemporal variables, and the analysis of the correlations between tourist destinations, rest places, and transportation hubs. Date table matching is performed based on the mobile subscriber identifier of the visitors who visited the tourism-related areas.

4.1. Preprocessing of mobile phone data

Due to the missing data and handovers in the raw dataset (Bekhor and Shem-Tov, 2015; Li et al., 2014; Mamei and Colonna, 2018; Zhong et al., 2018), the mobile phone data have to be preprocessed before identifying the visitors and analyzing tourism-related activities. The following data fields were selected from the raw dataset listed in Table 1 (a):

<MSID, DATETIME, LON, LAT>

Data preprocessing consisted of data reconstruction, individual path segmentation, and time labeling.



(a) Data reconstruction (b) Individual path segmentation

Fig. 4. Preprocessing of the mobile phone data.

Mobile phone data are spatiotemporal trajectory data that are not designed for tourism-related analysis. Irregular but frequent handovers between base stations lead to data noise and errors in the analysis of individual behavior. This study used the binning method and the raster data structure proposed by Li et al. (2014) to filter redundant data in the temporal and spatial domains. As shown in Fig. 4(a), a weighted location is generated using the records in the same bin. The weights of the longitudes and latitudes of the original records in the same bin are the duration from each timestamp to the next one in the same bin or from the timestamp to the end of the bin. The timestamp of the bin (*binID*) is the middle timestamp. In the raster data structure, the weighted locations represent the center of the cell, which is labeled as *rasterID*. In the first stage, the raw dataset was transformed into the following data structure:

<MSID, binID, rasterID>

where *binID* and *rasterID*, respectively, denote the timestamp and the location of the data records. In this stage, the bin span and the raster length, which represent the temporal and spatial resolutions, respectively, were determined based on the data quality, base station coverage, and individual activity patterns.

The reconstructed data records represent the space-time path of mobile subscribers. In the second stage, the space-time path was segmented to determine the mobility or activity of individuals during their visits to the city, as shown in Fig. 4(b). After the individual path segmentation, the reconstructed data were transformed into the following data structure:

<MSID, start_binID, start_rasterID, end_binID, end_rasterID, subgroupID, duration>

where *start_binID* and *end_binID* denote the start time and the end time of the path segment, respectively, *start_rasterID* and *end_rasterID* represent the change in individual's location in the path segment, and *subgroupID* denotes the visit to the city during which the path segment occurs. For each path segment, if *start_rasterID* = *end_rasterID*, it is assumed that the individual remains in the same raster cell *start_binID/end_binID*. If *start_rasterID* ≠ *end_rasterID*, it is assumed that the individual moves from *start_rasterID* to *end_rasterID*. The duration of the stay or the move is calculated as Eq. (1):

$$duration = (end\_binID - start\_binID) \times \Delta t \tag{1}$$

where Δt is the bin span. The *end_binID* of the first segment of two adjacent path segments is denoted as t_1, the *start_binID* of the latter segment is denoted as t_2, and the preset threshold of the visit interval is denoted as *VI*. If t_2-$t_1 > VI$, the two adjacent path segments are assumed to not belong to the same visit to the city, and two different values of *subgroupID* are assigned to the segments.

Since most tourism-related activities occur at a specific time of the day, we label the time information of the path segments as follows:

<MSID, start_binID, start_rasterID, end_binID, end_rasterID, subgroupID, duration, time>

If the *start_binID* or *end_binID* belongs to the active period of tourism-related activities, *time* is assigned to *daytime.*

4.2. Visitor identification

This study mainly focuses on domestic visitors due to the small proportion of foreign visitors. Since no data field shows the mobile subscriber is roaming, visitors were identified based on their intrinsic spatiotemporal activity patterns that were different from those of the residents. Four variables were defined as follows:

1. *D*: the number of total days that a mobile subscriber stayed in the city during the research period.
2. *V*: the number of a mobile subscriber's data subgroups, which were divided by large visit intervals (Fig. 4(b)) during the research period.

Table 2
Rules of visitor identification.

Rules	*D* (days)	*V* (times)	*NS*	*C*	Class
1	$1 \sim DLL$				Visitors
2	$DLL + 1 \sim DUL$	$1 \sim VL$	No		Visitors
3			Yes	$[0, CL]$	Unknown
4				$(CL, +\infty)$	Visitors
5		$> VL$			Visitors
6	$DUL + 1 \sim 31$	$1 \sim VL$			Residents
7		$> VL$	No		Visitors
8			Yes	$[0, CL]$	Unknown
9				$(CL, +\infty)$	Visitors

3. *NS*: whether a mobile subscriber stayed in rest places at night during the research period. A rest place was defined as the cell with the longest duration of stay at night.
4. *C*: the entropy of the rest places where a mobile subscriber stayed during the research period, which is defined as Eq. (2) (Song et al., 2010):

$$C = -\sum_{i=1}^{N} \frac{c_i}{D} log_2 \frac{c_i}{D} \tag{2}$$

where c_i is the number of days that a person stayed at the rest place i for the longest duration at night ($\sum_{i=1}^{N} c_i = D$), and N is the number of rest places.

Thresholds are used for the four variables to distinguish visitors from residents, including the lower limit of *D* (*DLL*), the upper limit of *D* (*DUL*), the limit of *V* (*VL*), and the limit of *C* (*CL*). As shown in Table 2, nine rules were proposed to identify visitors and residents based on the four thresholds. The rules differ for different cities because of the features of the city and the visitors' activity patterns. Thus, the distributions of the four variables should be considered when determining the values of the four thresholds.

4.3. Correlation analysis

In this study, tourists were the population group used to determine the correlations between tourism facilities. Thus, tourists represented visitors with tourism-related activities; they were extracted from the individual path segments by matching the raster cells to typical tourism-related activity areas. The duration in the same tourist destination (TD) had to be at least 30 min, regardless of whether the visitor moved or stopped, which meant that travel without stopping inside the TD was taken into consideration. The tourism-related activity information at the destinations during the daytime, the rest places at night, and the transportation hub locations where tourists entered or left the city were extracted.

The following spatial correlations of tourism-related activities were determined:

1) Correlations between tourist destinations:
These correlations were based on tourists who moved from one destination to another on the same day in the daytime. This table was created using <*VisitorID*, *day*, *start_TD*, *end_TD* > to sum the number of the two directed flows.
2) Correlations between tourist destinations and rest places:
These correlations were based on tourists who stayed at rest places and visited tourist destinations the next day. This table was created using <*VisitorID*, *day*, *TD*, *rest_gridID* > to calculate the distances between the two areas. KDE of the rest places around tourist destinations was performed.
3) Correlations between tourist destinations and transportation hub locations:

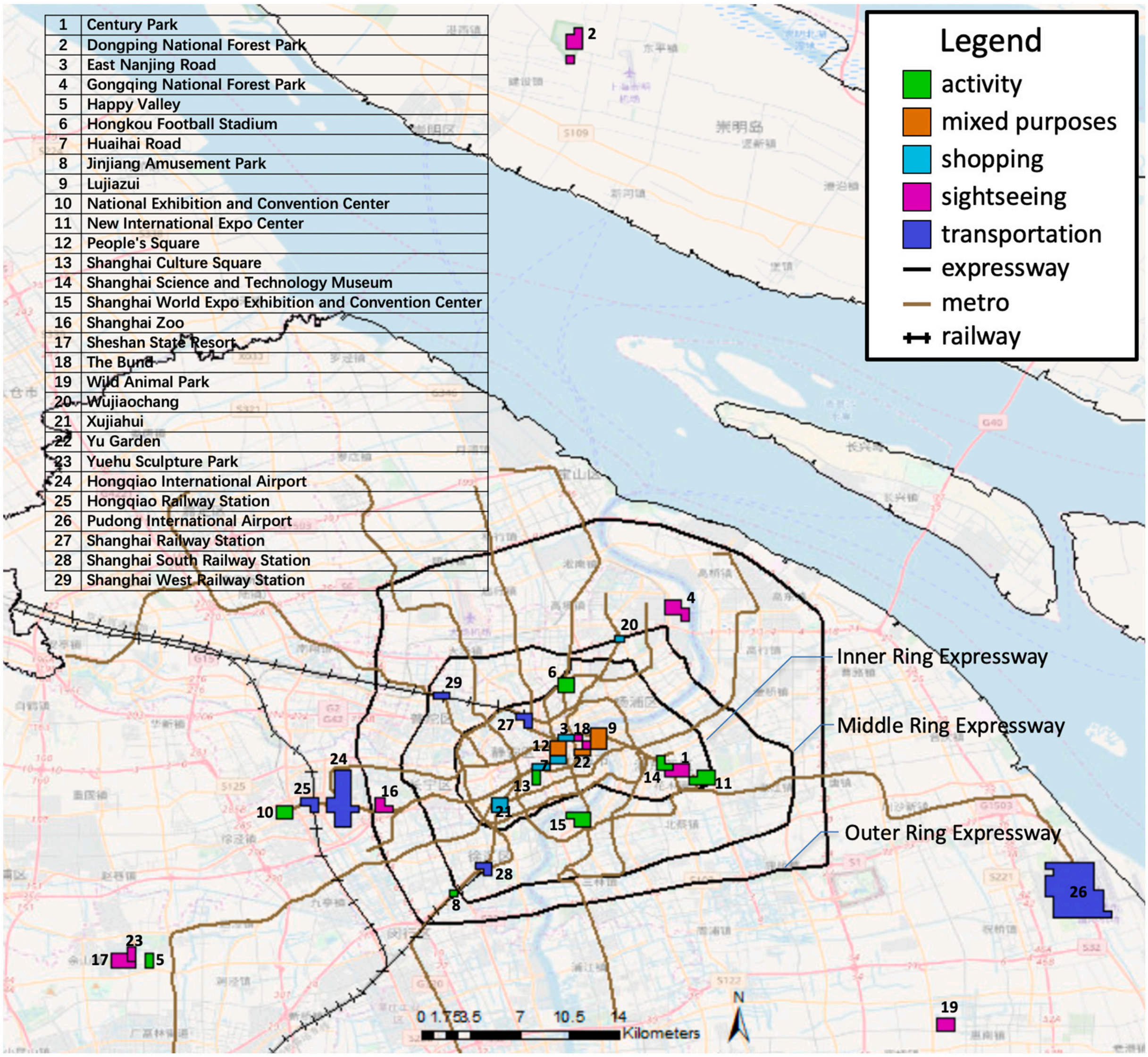


Fig. 5. Study area in Shanghai.

These correlations were based on tourists who entered or left the city through transportation hubs (airports and railway stations) and visited tourist destinations on the same day. In the first hour and last hour of each visit, records within the region of a given transportation hub were identified, which meant the tourist passed through this transportation hub. This table was created as < *VisitorID*, *day*, *transportation hub*, *TD* > to calculate the proportions of visitors passing through each transportation hub.

5. Case study

5.1. Study area

Shanghai is the core city in the Yangtze River Delta Urban Agglomeration, which is recognized as one of the six world-class urban agglomerations. The MICE market in China is massive, highly sustainable, and competitive globally. Shanghai, in particular, is positioned as a significant and sophisticated urban commercial hub and gateway. It has become a tremendously attractive benchmark as the leading destination in Asia and worldwide. About 181 million visitors from other cities in China and foreign countries visited Shanghai in 2019 (Shanghai Bureau of Statistics, 2020). Since MICE business plays an important role in the tourism industry of Shanghai, it is important to have an in-depth understanding of the visitors' requirements and their tourism-related activity patterns in the city.

In this study, 23 tourism destinations in Shanghai were selected to identify the tourism-related activities of visitors, including eight scenic spots, four shopping centers, eight recreation and sports facilities (e.g., exhibition halls, opera houses, and sports stadiums), and three tourist destinations with mixed functions. Additionally, six transportation hubs, including two airports and four railway stations, were selected to analyze the spatial correlations between them and tourist destinations. The locations of the selected areas are shown in Fig. 5 in a raster map with a 500 m by 500 m cell size.

5.2. Tourist identification results

Preprocessing of the mobile phone data was performed to improve the efficiency of data processing and noise elimination. Based on the coverage of the base stations in Shanghai, the activity patterns of individuals, and the resolution required in the subsequent analysis, the following values were used for the parameters in data preprocessing.

- The bin span was 10 min (Li et al., 2014).
- The raster size was 500 m by 500 m (Li et al., 2014).
- The visit interval was 1800 min (Qian et al., 2019).

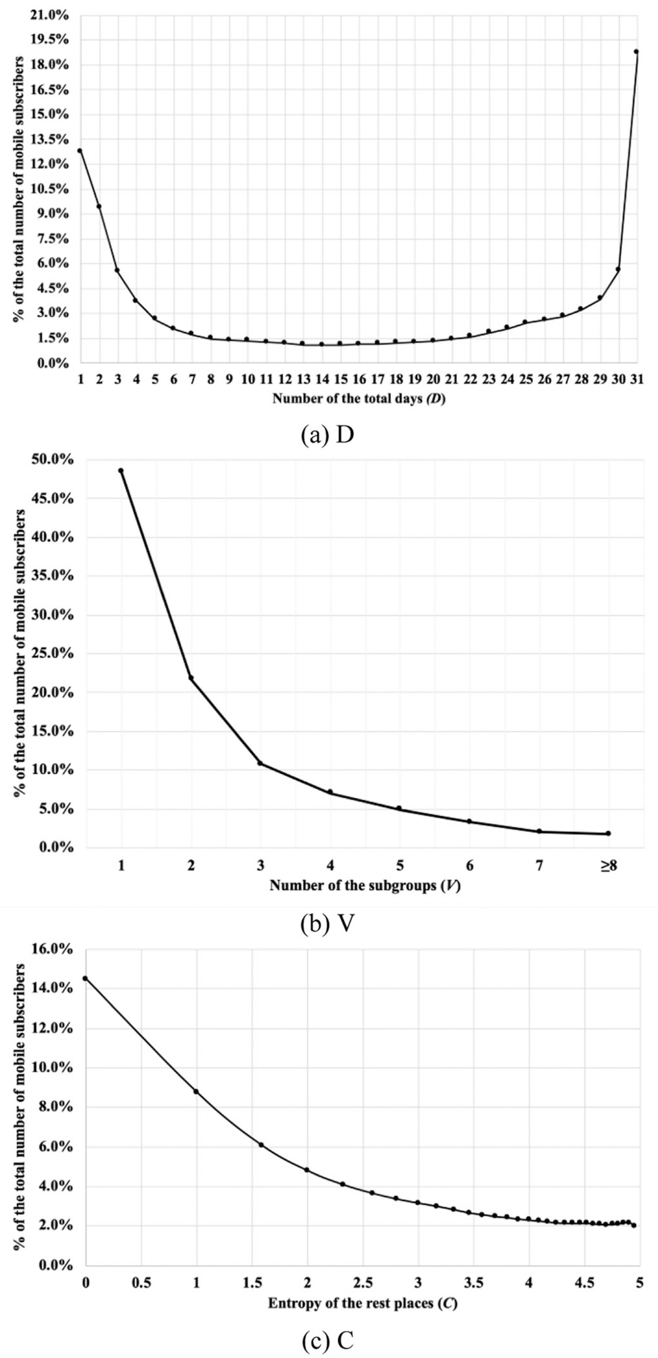


Fig. 6. Distributions of the variables used for the identification of visitors and residents.

- The active period of tourism-related activities or the daytime period was between 9 a.m. and 9 p.m., and the nighttime period was between 9 p.m. and 9 a.m. (Qian et al., 2019).

As shown in Fig. 6, according to the distributions of *D*/*V*/*C*, the reference values of the thresholds required in the visitor identification were set as follows.

- *DLL* was 7, *DUL* was 21; thus, *D* was divided into three sections (1–7 days, 8–21 days, and 22–31 days).
- *VL* was 4; thus, *V* was divided into two sections (1–4 times and > 4 times).
- *CL* was 1; thus, *C* was divided into two sections ([0,1] and (1, +∞)).

A sensitivity analysis was performed to evaluate the applicability of the values of the thresholds (Fig. 7). According to the Shanghai Statistical Yearbook in 2011, there were 23.47 million residents and 108.8 million visitors from other cities, indicating that the monthly visitors accounted for 27.9% of the total number of residents and visitors (Shanghai Bureau of Statistics, 2012). Thus, the average daily visitor proportion was calculated by changing one of the thresholds and keeping the other three thresholds the same. The results indicate that the change in the visitor proportion *DLL* and *DUL* is positively correlated with the visitor proportion, whereas *VL* and *CL* are negatively correlated with the visitor proportion. Moreover, the visitor proportion range shows that the visitor proportion is not sensitive to the *DLL* but is sensitive to the *DUL*, *VL*, and *CL*. Since visitors who stayed for more than one month in Shanghai were not counted in this dataset, it is reasonable that the proportion of visitors obtained from the reference values of the four thresholds (26.8%) is slightly smaller than the values obtained from the yearbook (27.9%).

Based on the nine deterministic rules developed by using the four thresholds, 16,301,266 visitors (accounting for 56.2%), 11,915,596 residents (accounting for 41.1%), and 766,682 unclassified individuals (accounting for 2.7%) were identified using the mobile phone data, as listed in Table 3. The proportion of identified visitors in the mobile subscribers using the proposed nine rules was 55.5% larger than that of the visitors identified using rule 1 by Wu et al. (2020). The visitors, which were not identified in the existing literature, fell into three groups: 1) visitors with medium stays, low visit frequency, and non-fixed accommodation (rule 4); 2) visitors with medium stays and high frequency (rule 5); 3) visitors with long stays and non-fixed accommodation (rule 9).

The results show a good performance for the ability to distinguish visitors from residents. Fig. 8 illustrates the temporal variation of the number of visitors and residents and the proportion of visitors in the population. The proportion of visitors staying in Shanghai was 29.3%, on average, during the Chinese National Day holidays (from October 1 to October 7), whereas the proportion fell to 26.1% on the other days. This result showed that tourism in Shanghai has a large potential market. Fig. 8 shows the number of tourists and the proportion of tourists among visitors after matching the visitor path segments with the tourism facilities.

5.3. Spatial correlations between tourism facilities

5.3.1. Correlations between tourist destinations

Fig. 9 shows the spatial correlation between two tourist destinations. The width and the color of the curves indicate the number of tourist movements. The radius of the vertex represents the number of visits to tourist destinations in the daytime. The figure shows that six areas in the center of the graph had strong spatial correlations. These areas are adjacent geographically in the CBD of Shanghai. The metro stations adjacent to these areas have large passenger volumes because of their strong spatial correlations.

The color of the vertex shows the potential purpose of tourists who visited the site. Except for the central locations (The Bund, Lujiazui, Yu Garden, and People's Square) in the figure, sightseeing flows accounted for a small proportion. The purposes of the tourist were variable. Tourists intended to watch a football match, go to an opera, visit a museum, attend a meeting or exhibition, or play in an amusement park. Wild Animal Zoo and the two national forest parks are far from the metro stations; thus, few tourists added them to their travel lists. In 2015, Shanghai Metro Line 16 was built and started to operate, and Wild Animal Park was more easily accessible. However, Happy Valley, Sheshan State Resort, and Yuehu Sculpture Park could have a considerable tourist flow if their accessibility were improved.

5.3.2. Correlations between tourist destinations and rest places

Did tourists intend to choose rest places near their tourist

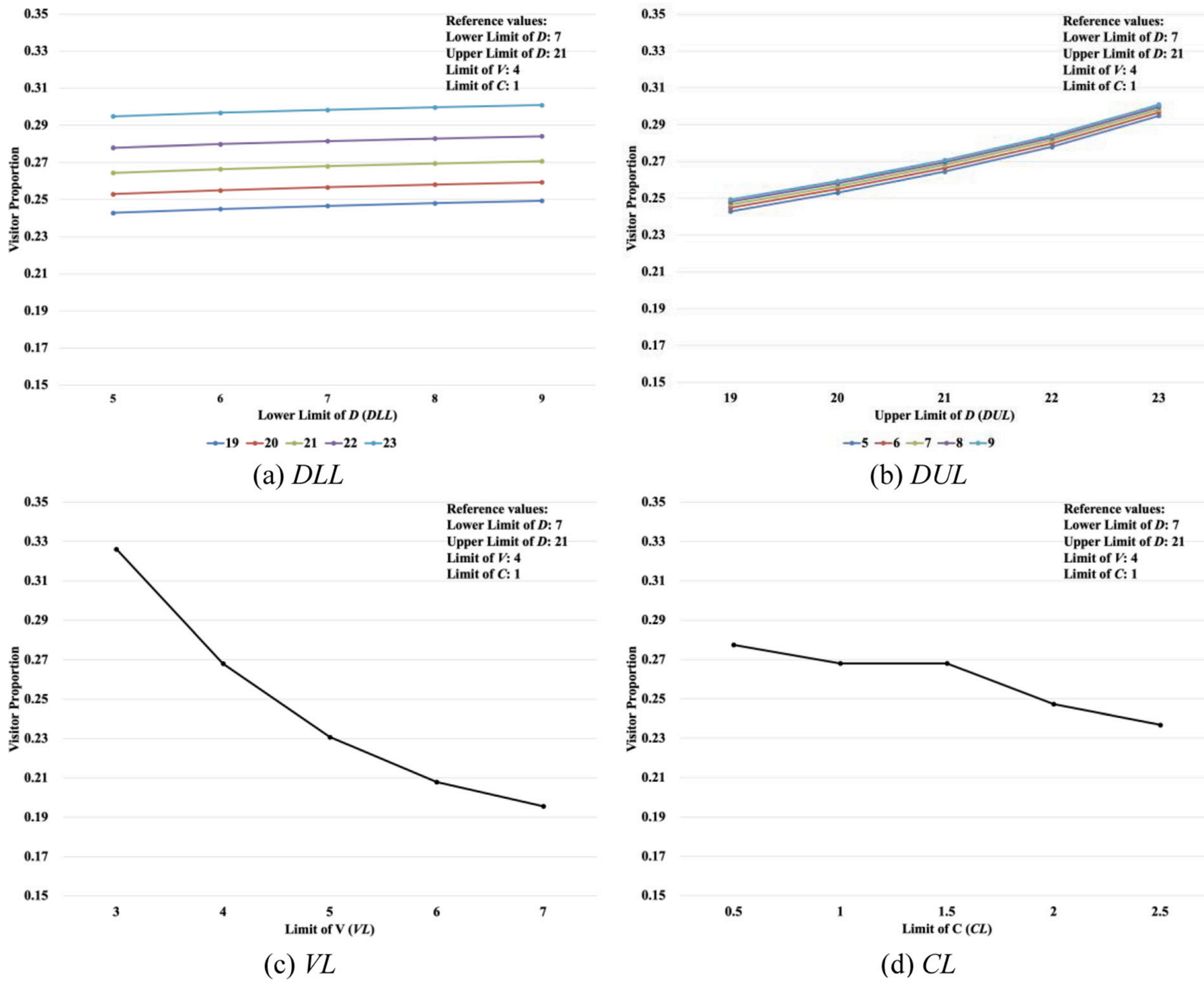


Fig. 7. Sensitivity analysis of visitor proportion.

Table 3
Results of identifying visitors and residents.

Rule	Class	Number	Proportion
1	Visitors	10,486,823	36.2%
2	Visitors	222,405	0.8%
3	Unknown	567,606	2.0%
4	Visitors	2,482,374	8.6%
5	Visitors	1,928,606	6.7%
6	Residents	11,915,596	41.1%
7	Visitors	40,599	0.1%
8	Unknown	199,076	0.7%
9	Visitors	1,140,459	3.9%

destinations the next day? Fig. 10 illustrates the spatial patterns of the rest places for each destination. The data were generated in ArcGIS software using KDE of the center points in the raster data. The KDEs were categorized into nine classes. The brighter color represents more people gathered at the locations. The area of each tourist destination is outlined in green. First, the results show the number of tourists that chose rest places near their next-day destination. Second, contrary to common belief, the rest places that were sightseeing destinations, amusement parks, and convention centers exhibited polycentric characteristics. The common area of these polycentric rest places was the city center because of its well-built environment. Unlike Gongqing National Forest Park, Dongping National Forest Park attracted tourists from a larger area because it provided unique experiences on farms in addition to sightseeing. Happy Valley and Jinjiang Amusement Park attracted people who preferred longer travel distances. Additionally, different types of exhibitions were held in the National Exhibition and Convention Center and New International Expo Center, resulting in visits by long-distance tourists.

5.3.3. *Correlations between tourist destinations and transportation hubs*

Did tourist destinations have stronger correlations with some transportation hubs than the others? In this study, the correlations between tourist destinations and transportation hubs included two conditions. First, tourists had to leave the transportation hub immediately to save time to visit tourist attractions or perform other scheduled activities. Second, tourists had to visit destinations before leaving the city late in the afternoon or at night. As shown in Fig. 11, the larger marker shows the transportation hub that was chosen preferentially by tourists.

Shanghai West Railway Station is not in the polar chart because the flow of each destination was less than 1%. Pudong International Airport was the least visited location because it is far from the city center, and tourists do not have the time to arrange other activities or travels. Hongqiao Railway Station was the most important transportation hub because of its large number of flows. The correlation between Shanghai Railway Station and tourist destinations, such as Happy Valley, Hongkou Football Stadium, Huaihai Road, Shanghai Culture Square, The Bund, Wild Animal Park, and Yu Garden, ranked second because of its location in the city center. Strong correlations existed between two locations, such as the Hongqiao Railway Station – National Exhibition and Convention Center/Shanghai Zoo and Shanghai South Railway Station and Jinjiang Amusement Park/Xujiahui. These locations are close to the

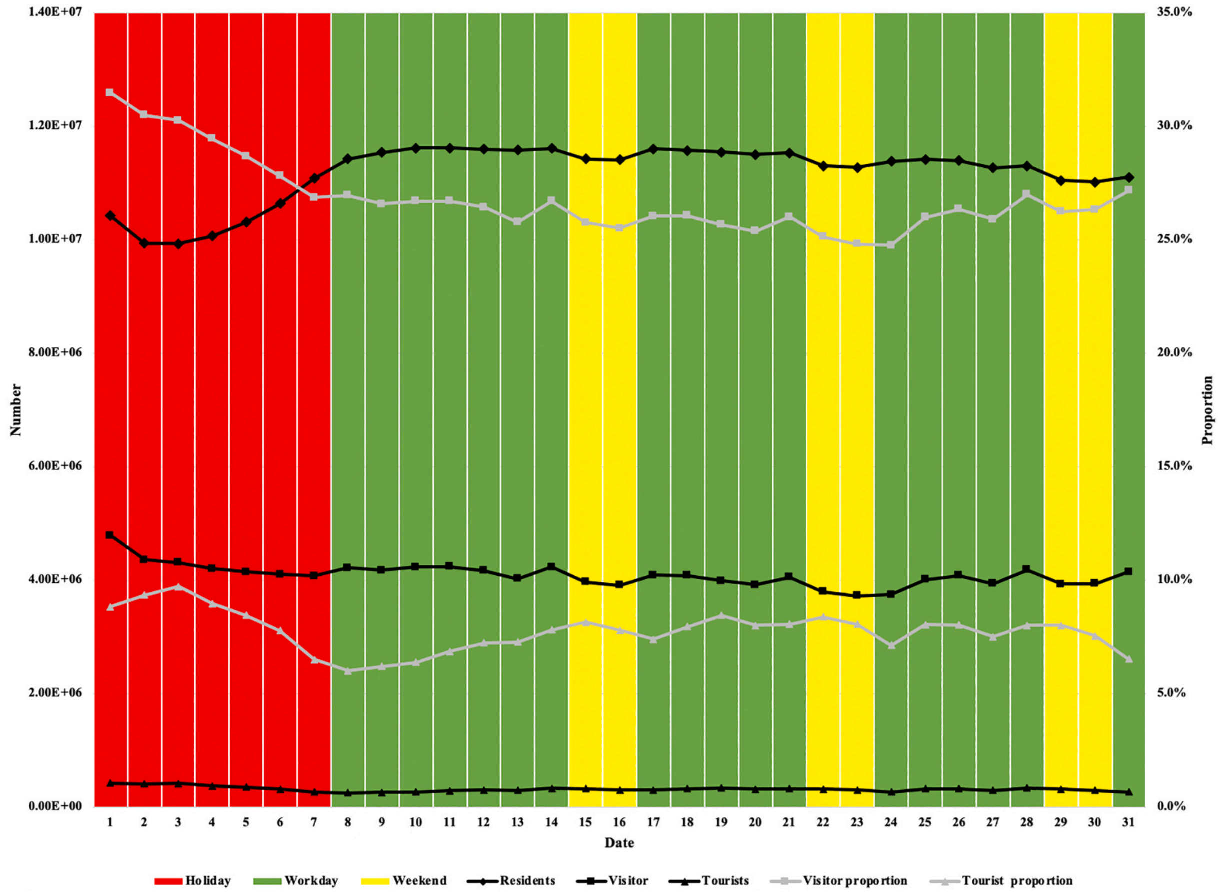


Fig. 8. The number of residents/visitors/tourists and the visitor/tourist proportion on different dates.

metro network. The long-distance locations, such as Hongqiao Railway Station – Sheshan State Resort/Yuehu Sculpture Park, Shanghai Railway Station – Gongqing National Forest Park, and Shanghai South Railway Station – Dongping National Forest Park, indicate that direct tour buses should be provided for the convenience of tourists. These correlations are of interest to tourism product managers.

5.4. Evaluation of transportation accessibility

The transportation accessibility was evaluated from the perspective of the supply. The indicator was the travel time by a specific travel mode. The improvement of transportation accessibility can be evaluated comprehensively by considering travel demand data. The transportation accessibility of 406 origin-destination (OD) pairs was evaluated using Direction API data and three indicators, namely, the duration difference, the number of tourists, and the economic effect.

1) The duration difference is the difference in duration between using public transit and automobiles, which means there are time savings if tour buses are considered automobiles because tour bus lines have fewer stops than the regular bus or metro.
2) The number of tourists refers to the number of tourists traveling between tourist destinations or between tourist destinations and transportation hubs.
3) The economic effect is measured by the product of the first two indicators and reflects the potential to increase destination flows and create a positive effect due to time-savings if a new tour bus line is operated.

Table 4 shows the ranking in descending order of the top 10 OD pairs in terms of the economic effect. The duration difference is larger than the threshold. The threshold is determined by the distribution of the duration difference. On the one hand, the number of tourists is much lower than that in the city center because transportation accessibility is low for public transit. On the other hand, if good access is provided, the number of tourists traveling to these destinations will increase, and a positive economic effect will be created. The data in the tables show that the tourism transportation accessibility between the tourism facilities has to be improved.

6. Discussion

The mobile phone data provided more valuable information on spatial correlations between tourism facilities than traditional and Internet data sources. The spatial structure between tourist destinations was reasonable compared to the network structure of foreign tourists in the post-World Expo period based on Flickr location data (Mou et al., 2020). Tourist flows between the destinations in the city center accounted for the major proportion. Shanghai World Expo Exhibition and Convention Center or Expo Park had medium connections with the central destinations. However, domestic tourists showed different preferences than foreign tourists. For example, domestic tourists preferentially visited People's Square and Yu Garden, whereas foreign tourists preferred visiting The Bund. In addition, domestic tourists preferred Hongqiao Railway Station, whereas foreign tourists preferred Pudong International Airport because of the distance between Shanghai and their origin cities. The travel chains obtained from the tourists' mobile phone data provided comprehensive spatial information on the rest places around the tourist destinations on the next day. The types of these tourist destinations and the structure of the urban built environment influenced the travel distances between the rest places and destinations.

The tour bus network in the city requires expansion due to the

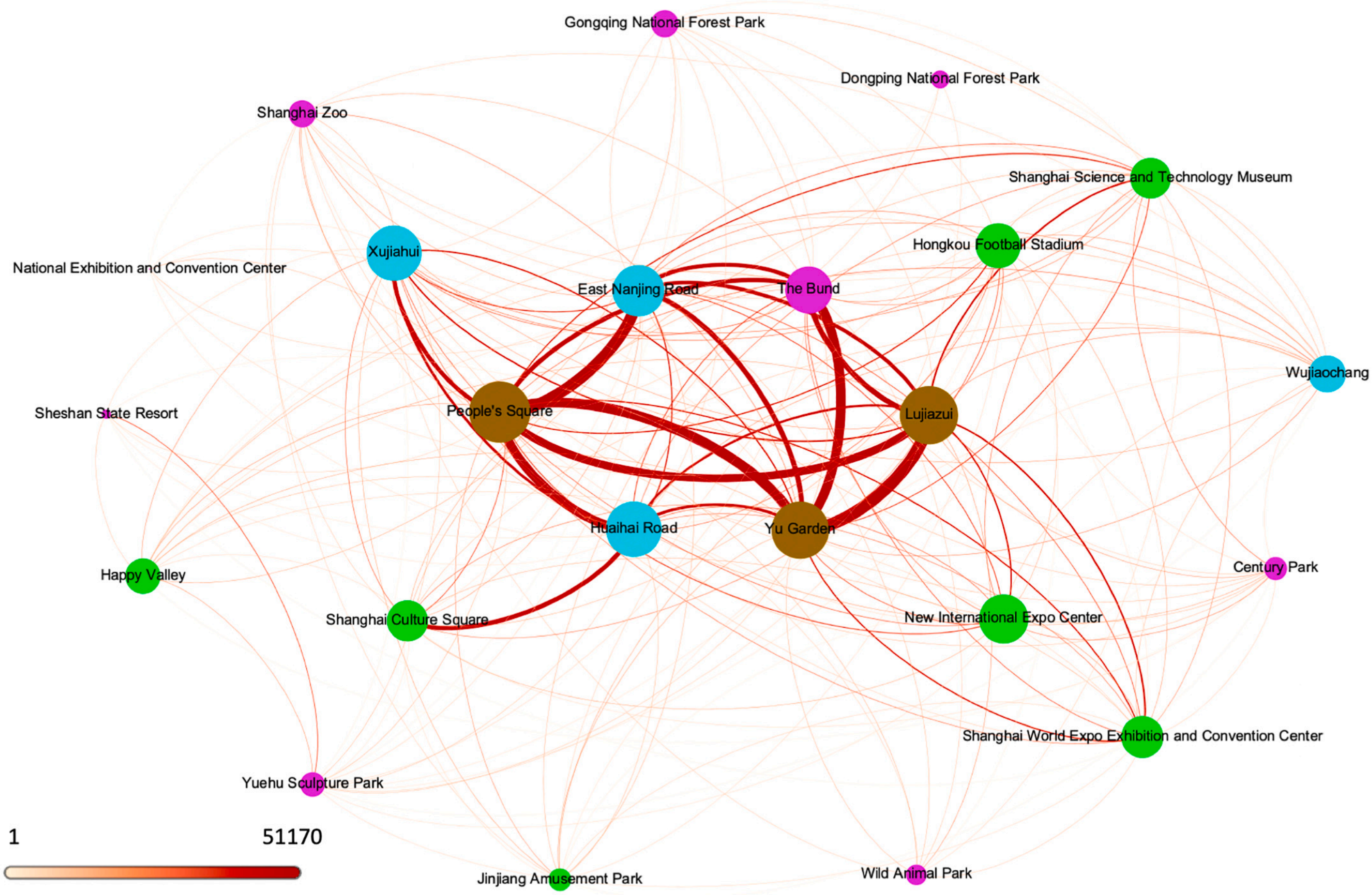


Fig. 9. The network of tourist flows between two tourist destinations on the same day.

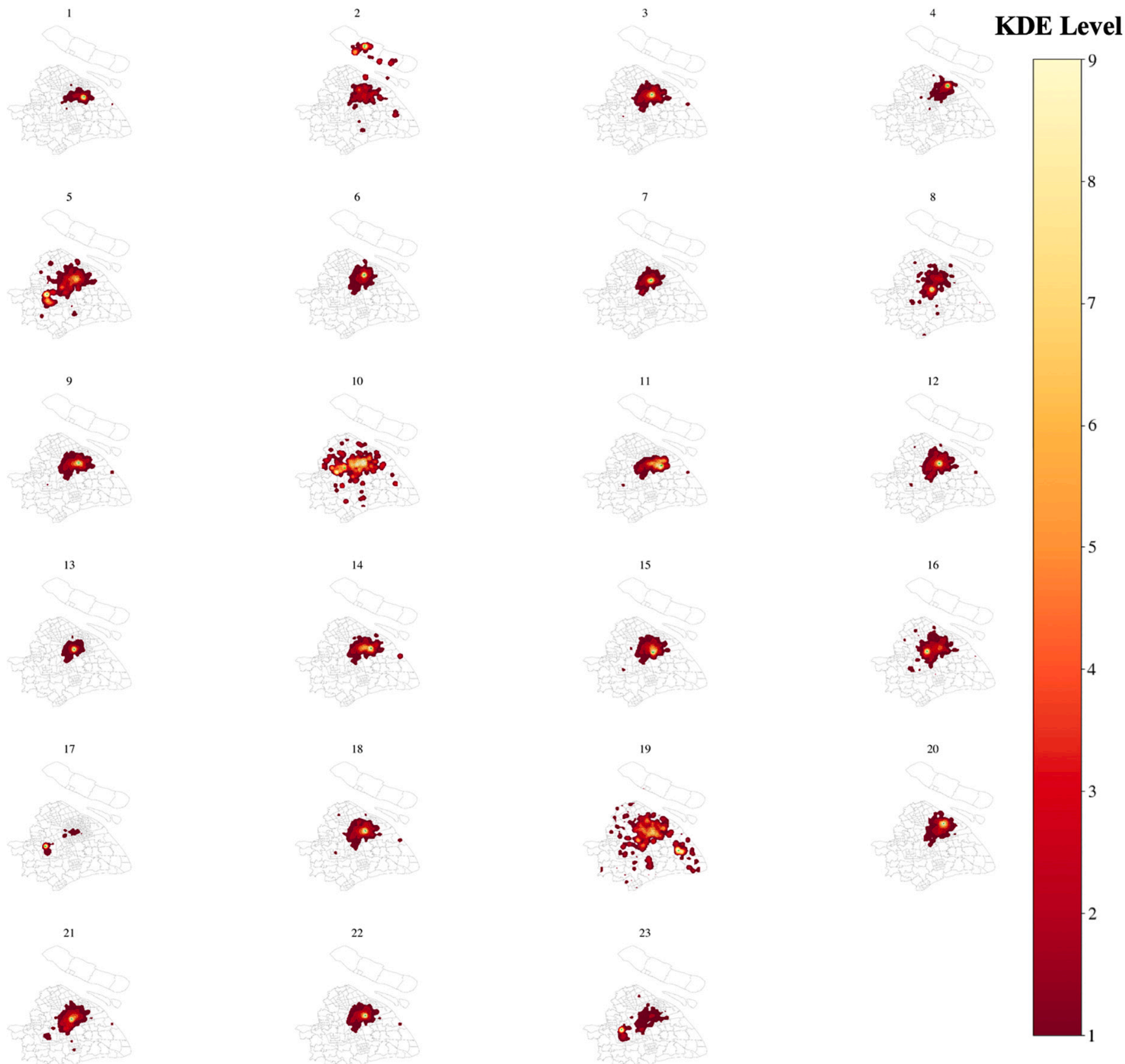
increasingly diversified demand for connectivity of tourism-related activity sites. The existing tour bus lines (Bus Tour and City Sightseeing) in Shanghai are circular and operate in the city center, as shown in Fig. 12. The tour buses meet the demand for visiting tourist destinations with high popularity (the Bund, East Nanjing Road, Lujiazui, Yu Garden, People's Square, and Huaihai Road) and a world-famous site (Shanghai World Expo Exhibition and Convention Center). In addition, these tour bus lines cover one hotel, two Buddhist temples, and several view spots in the North and South Bund along the Huangpu River. However, the tourist destinations inside the Middle Expressway, including one view spot (Century Park), two commercial circles (Xujiahui and Wujiaochang), and four venues for diversified activities (Shanghai Cultural Square, Hongkou Football Stadium, Shanghai Science and Technology Museum, and New International Expo Center), are not covered by the tour bus lines. It is suggested to add a bus station near Shanghai Cultural Square to the Bus Tour Green Line according to the strength of the correlation between Huaihai Road and Shanghai Cultural Square.

In addition to the existing tour bus lines in the city center, it is suggested to develop a hierarchical tour bus network based on the results of the evaluation of the transportation accessibility. Three principles should be followed. First, it is suggested to add a new loop route to cover destinations in close proximity, such as sightseeing sites (Sheshan State Resort, Yuehu Sculpture Park) and the amusement park (Happy Valley) in Songjiang District. Second, it is recommended to add new single routes between tourist destinations or between tourist destinations and transportation hubs. It is preferable to choose OD pairs whose economic effect is larger than 10 thousand, as suggested in Table 4. Third, tourists staying in rest places near other distant gathering centers of destinations can be served by new single tour bus lines that operate in the early morning or late afternoon.

In summary, the contributions of this study are twofold. First, in addition to the duration of visitors staying in the city, the visit counts, daytime visits, and the entropy of visitors' rest places were considered, improving the recognition rate of visitors who frequently visit, visit in the daytime, and stay overnight in many places. Their activities and interactions with tourism facilities, including tourist destinations, accommodation facilities, and transportation hubs, were objectively and comprehensively recorded by mobile phone data. Second, in addition to the duration difference between the two traffic modes, the traffic demand, i.e., the spatial correlations between tourism facilities, were considered in the evaluation of transportation accessibility to design the tour bus lines. This approach allows for science-based decisions in urban tourism traffic planning. Moreover, the spatial correlations have practical values for evaluating existing public transit and plan a hierarchical tour bus network in the next step. More tourists will be attracted to the tourist destinations and hotels along the tour bus lines.

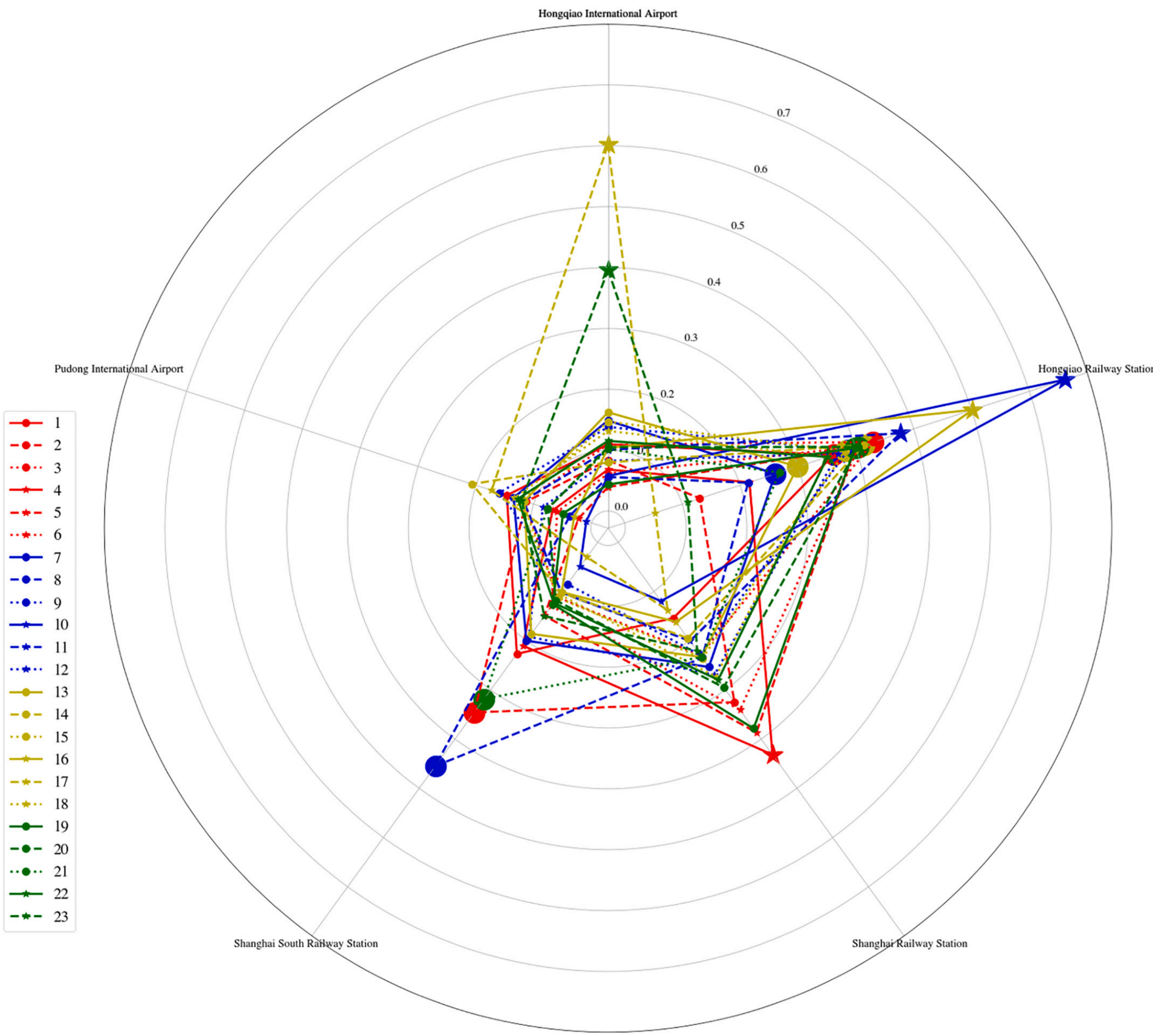
7. Conclusions

Spatial correlation between tourism facilities is an important indicator in tourism transportation. In this paper, a research framework was proposed to investigate spatial correlations between tourism facilities using mobile phone data, POI data, and Direction API data. Nine rules related to spatial and temporal features were defined to identify visitors using mobile phone data. The spatial correlations between tourist destinations, rest places, and transportation hubs were extracted and visualized. Subsequently, based on the results, potential application scenarios in tourism transportation were discussed, including the evaluation of transportation accessibility and the addition of a tour bus network. A case study in Shanghai was carried out to verify the



1	Century Park	2	Dongping National Forest Park	3	East Nanjing Road	4	Gongqing National Forest Park
5	Happy Valley	6	Hongkou Football Stadium	7	Huaihai Road	8	Jinjiang Amusement Park
9	Lujiazui	10	National Exhibition and Convention Center	11	New International Expo Center	12	People's Square
13	Shanghai Culture Square	14	Shanghai Science and Technology Museum	15	Shanghai World Expo Exhibition and Convention Center	16	Shanghai Zoo
17	Sheshan State Resort	18	The Bund	19	Wild Animal Park	20	Wujiaochang
21	Xujiahui	22	Yu Garden	23	Yuehu Sculpture Park		

Fig. 10. Distributions of the kernel density estimates of the number of tourists in the rest places that visited the tourist destinations the next day.



1	Century Park	2	Dongping National Forest Park	3	East Nanjing Road	4	Gongqing National Forest Park
5	Happy Valley	6	Hongkou Football Stadium	7	Huaihai Road	8	Jinjiang Amusement Park
9	Lujiazui	10	National Exhibition and Convention Center	11	New International Expo Center	12	People's Square
13	Shanghai Culture Square	14	Shanghai Science and Technology Museum	15	Shanghai World Expo Exhibition and Convention Center	16	Shanghai Zoo
17	Sheshan State Resort	18	The Bund	19	Wild Animal Park	20	Wujiaochang
21	Xujiahui	22	Yu Garden	23	Yuehu Sculpture Park		

Fig. 11. Polar chart of the proportion of tourists arriving at/departing from the transportation hub and visiting tourist destinations on the same day.

Table 4
Results of transportation accessibility comparison.

(a) Between tourist destinations (threshold = 26 min)

Rank	OD	Duration of public transit (min)	Duration of automobile travel (min)	Duration difference (min)	Number of tourists	Economic effect
1	New International Expo Center - The Bund	47.23	19.03	28.20	922	26,000.40
2	Happy Valley - Yuehu Sculpture Park	53.45	5.65	47.80	316	15,104.80
3	Gongqing National Forest Park - People's Square	64.62	37.50	27.12	136	3688.32
4	Wild Animal Park - Yu Garden	88.88	56.97	31.91	82	2616.62
5	Dongping National Forest Park - People's Square	224.77	152.83	71.94	28	2014.32
6	Gongqing National Forest Park - Yu Garden	66.93	36.95	29.98	65	1948.70
7	Dongping National Forest Park - Yu Garden	250.10	144.20	105.90	15	1588.50
8	Gongqing National Forest Park - Lujiazui	67.00	36.22	30.78	47	1446.66
9	The Bund - Wild Animal Park	88.37	51.35	37.02	35	1295.70
10	Gongqing National Forest Park - Huaihai Road	87.40	42.38	45.02	24	1080.48

(b) Between tourist destinations and transport hubs (threshold = 36 min)

Rank	OD	Duration of public transit (min)	Duration of automobile travel (min)	Duration difference (min)	Number of tourists	Economic effect
1	Pudong International Airport - Wujiaochang	84.55	43.44	41.11	679	27,913.69
2	Gongqing National Forest Park - Shanghai Railway Station	84.12	37.30	46.82	476	22,286.32
3	Happy Valley - Pudong International Airport	140.57	73.68	66.89	203	13,578.67
4	Hongqiao Railway Station - Wild Animal Park	101.17	49.68	51.49	248	12,769.52
5	Dongping National Forest Park - Shanghai South Railway Station	242.00	150.42	91.58	82	7509.56
6	Pudong International Airport - Yuehu Sculpture Park	132.18	75.42	56.76	125	7095.00
7	Pudong International Airport - Shanghai Zoo	94.37	55.49	38.88	167	6492.96
8	Gongqing National Forest Park - Pudong International Airport	122.81	38.60	84.21	74	6231.54
9	Dongping National Forest Park - Shanghai Railway Station	218.40	147.25	71.15	77	5478.55
10	Dongping National Forest Park - Pudong International Airport	261.18	111.34	149.84	27	4045.68

methodology and discuss the applications. The results indicated the following:

1) The spatiotemporal travel characteristics of visitors cannot be determined comprehensively and accurately by traditional and Internet data sources. In contrast, the use of nine rules based on mobile phone data significantly improved the recognition rate of visitors compared to using only the number of visit days in the observation period. The proportion of daily visitors to the total population of Shanghai was 26.1%.
2) Regular demands between tourism facilities should be considered in tourism transportation planning in a city. Tourist destinations in the city center were highly popular and strongly spatially correlated because of the good accessibility to public transit. Tourists intended to rest near their next-day destinations. A bicycle-sharing system and flat bicycle lanes are suggested to connect destinations and hotels. Steady passenger flows arriving at or leaving the tourist destinations for shopping or other purposes in the city center are needed to provide tour buses with fixed schedules.
3) The rest places that were sightseeing destinations, amusement parks, and convention centers exhibited polycentric characteristics. Arrival/departure peaks were observed at these destinations. Hongqiao Railway Station and Shanghai Railway Station ranked first in terms of the arrival/departure date, whereas Pudong International Airport ranked last due to its location far from the city center. Transportation accessibility and arrival/departure peaks of tourism facilities should be determined before adding new tour bus lines to the existing tour bus system. As a result, irregular transportation demands in tourism correlations should be considered in real-time.

This study provides an in-depth understanding of the travel demands and preferences of tourists and serves as a practical guide for administrators of urban tourism departments, transportation planners of city bus lines, and product managers of tour companies. The results can be used in comparative studies of cities. Other dense spatiotemporal data sources, such as mobile device GPS data, can be used to analyze and examine the spatial correlations using the proposed research framework. In a future study, the station selection and line direction of tour bus lines connecting tourist destinations, the gathering points of rest places, and transportation hubs will be investigated using the trip chains of tourists.

Funding

This work was supported by the National Key R&D Program of China (No. 2018YFB1601100) and Tsien Hsue-shen Urbanology Award of Hangzhou International Urbanology Research Center & Zhejiang Urban Governance Studies Center (No. 20QXS002).

Declaration of Competing Interest

None.

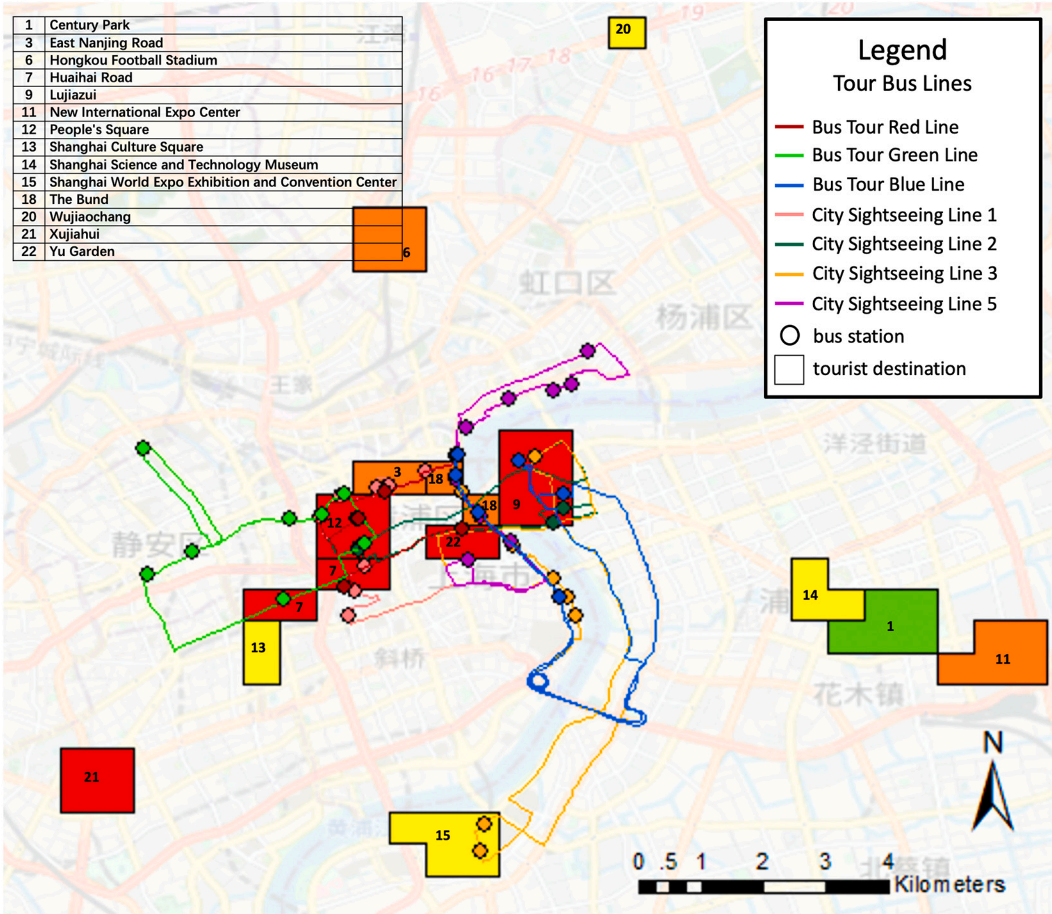


Fig. 12. The existing tour bus lines in Shanghai.
Notes: The bus lines of Bus Tour and City Sightseeing were obtained from http://www.bustourchina.com/bus_lines and http://map.baidu.com, respectively.

References

Ahas, R., Aasa, A., Mark, Ü., Pae, T., Kull, A., 2007. Seasonal tourism spaces in Estonia: case study with mobile positioning data. Tour. Manag. 28 (3), 898–910. https://doi.org/10.1016/j.tourman.2006.05.010.

Ahas, R., Aasa, A., Roose, A., Mark, Ü., Silm, S., 2008. Evaluating passive mobile positioning data for tourism surveys: an Estonian case study. Tour. Manag. 29 (3), 469–486. https://doi.org/10.1016/j.tourman.2007.05.014.

Bekhor, S., Shem-Tov, I.B., 2015. Investigation of travel patterns using passive cellular phone data. J. Loc. Based Serv. 9 (2), 93–112. https://doi.org/10.1080/17489725.2015.1066515.

Bieland, D., Sommer, C., Witte, C., 2016. A survey-based analysis of traffic behaviour of short vacationers and same-day visitors. Transp. Res. Procedia 14, 3228–3237. https://doi.org/10.1016/j.trpro.2016.05.266.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr., J., Ratti, C., 2013. Understanding individual mobility patterns from urban sensing data: a mobile phone trace example. Transport. Res. Part C Emerg. Technol. 26, 301–313. https://doi.org/10.1016/j.trc.2012.09.009.

Coban, S., 2012. The effects of the image of destination on tourist satisfaction and loyalty: the case of Cappadocia. Eur. J. Soc. Sci. 29 (2), 222–232.

De Cantis, S., Ferrante, M., Kahani, A., Shoval, N., 2016. Cruise passengers' behavior at the destination: investigation using GPS technology. Tour. Manag. 52, 133–150. https://doi.org/10.1016/j.tourman.2015.06.018.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Tatem, A.J., 2014. Dynamic population mapping using mobile phone data. Proc. Natl. Acad. Sci. 111 (45), 15888–15893. https://doi.org/10.1073/pnas.1408439111.

Furletti, B., Gabrielli, L., Renso, C., Rinzivillo, S., 2012. Identifying users profiles from mobile calls habits. In: Proceedings of the ACM SIGKDD International Workshop on Urban Computing, pp. 17–24. https://doi.org/10.1145/2346496.2346500.

Gao, S., Liu, Y., Wang, Y., Ma, X., 2013. Discovering spatial interaction communities from mobile phone data. Trans. GIS 17 (3), 463–481. https://doi.org/10.1111/TGIS.12042.

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L., 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782. https://doi.org/10.1038/nature06958.

Guy, C.M., 1983. The assessment of access to local shopping opportunities: a comparison of accessibility measures. Environm. Plann. B Plann. Design 10 (2), 219–237. https://doi.org/10.1068/b100219.

Huang, C., Yang, X., Gu, S., 2011. Airport passenger flow characteristics analyzing of Shanghai Hongqiao hub. Traffic Transport. 1, 133–136.

Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C., 2014. Development of origin–destination matrices using mobile phone call data. Transp. Res. Part C Emerg. Technol. 40, 63–74.

Jia, F., Cheng, X., Duan, Z., 2013. Analyzing the activity areas of non-resident tourists of shanghai expo using cellular phone data. Procedia Soc. Behav. Sci. 96, 1136–1145. https://doi.org/10.1016/j.sbspro.2013.08.130.

Kang, C., Ma, X., Tong, D., Liu, Y., 2012. Intra-urban human mobility patterns: an urban morphology perspective. Phys. A Statistic. Mech. Appl. 391 (4), 1702–1717. https://doi.org/10.1016/j.physa.2011.11.005.

Khan, N.U., Wan, W., Yu, S., 2020. Spatiotemporal analysis of tourists and residents in Shanghai based on location-based social Network's data from Weibo. ISPRS Int. J. Geo Inf. 9 (2), 70. https://doi.org/10.3390/ijgi9020070.

Kotus, J., Rzeszewski, M., Ewertowski, W., 2015. Tourists in the spatial structures of a big polish city: development of an uncontrolled patchwork or concentric spheres? Tour. Manag. 50, 98–110. https://doi.org/10.1016/j.tourman.2015.01.007.

Li, W., Cheng, X., Duan, Z., Yang, D., Guo, G., 2014. A framework for spatial interaction analysis based on large-scale mobile phone data. Comput. intellig. Neurosci. 2014. https://doi.org/10.1155/2014/363502.

Liu, D., Kwan, M.P., 2020. Measuring spatial mismatch and job access inequity based on transit-based job accessibility for poor job seekers. Travel Behav. Soc. 19, 184–193. https://doi.org/10.1016/j.tbs.2020.01.005.

Louail, T., Lenormand, M., Ros, O.G.C., Picornell, M., Herranz, R., Frias-Martinez, E., Barthelemy, M., 2014. From mobile phone data to the spatial structure of cities. Sci. Rep. 4, 5276. https://doi.org/10.1038/srep05276.

Ma, Y., Wang, Y., Xu, G., Tai, X., 2018. Multilevel visualization of travelogue trajectory data. ISPRS Int. J. Geo Inf. 7 (1), 12. https://doi.org/10.3390/ijgi7010012.

Mamei, M., Colonna, M., 2018. Analysis of tourist classification from cellular network data. J. Loc. Based Serv. 12 (1), 19–39. https://doi.org/10.1080/17489725.2018.1463466.

Mou, N., Yuan, R., Yang, T., Zhang, H., Tang, J.J., Makkonen, T., 2020. Exploring spatio-temporal changes of city inbound tourism flow: the case of Shanghai, China. Tour. Manag. 76, 103955. https://doi.org/10.1016/j.tourman.2019.103955.

Nilbe, K., Ahas, R., Silm, S., 2014. Evaluating the travel distances of events visitors and regular visitors using mobile positioning data: the case of Estonia. J. Urban Technol. 21 (2), 91–107. https://doi.org/10.1080/10630732.2014.888218.

Qian, C., Li, W., Yang, D., Ran, B., Li, F., 2019. Measuring spatial distribution of tourist flows based on cellular Signalling data: a case study of Shanghai. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2584–2590. https://doi.org/10.1109/ITSC.2019.8917262.

Rudjanakanoknad, J., Rattanasuwan, P., 2011. Study of sightseeing bus service for tourists on the Rattanakosin Island in Bangkok. In: Proceedings of the Eastern Asia Society for Transportation Studies Vol. 8 (The 9th International Conference of Eastern Asia Society for Transportation Studies, 2011). Eastern Asia Society for Transportation Studies, p. 165. https://doi.org/10.11175/eastpro.2011.0.165.0.

Shanghai Bureau of Statistics, 2012. 2011 Shanghai Economic and Social Development Statistics Bulletin. http://tjj.sh.gov.cn/tjgb/20120223/0014-239488.html.

Shanghai Bureau of Statistics, 2020. 2019 Shanghai Economic and Social Development Statistics Bulletin. http://tjj.sh.gov.cn/tjgb/20200329/05f0f4abb2d448a69e4517f6a6448819.html.

Shoval, N., Ahas, R., 2016. The use of tracking technologies in tourism research: the first decade. Tour. Geogr. 18 (5), 587–606. https://doi.org/10.1080/14616688.2016.1214977.

Song, C., Qu, Z., Blumm, N., Barabási, A.L., 2010. Limits of predictability in human mobility. Science 327 (5968), 1018–1021. https://doi.org/10.1126/science.1177170.

Sun, Y., Ma, H., Chan, E.H., 2018. A model to measure tourist preference toward scenic spots based on social media data: a case of Dapeng in China. Sustainability 10 (1), 43. https://doi.org/10.3390/su10010043.

Wu, X., Huang, Z., Peng, X., Chen, Y., Liu, Y., 2018. Building a spatially-embedded network of tourism hotspots from geotagged social media data. IEEE Access 6, 21945–21955. https://doi.org/10.1109/ACCESS.2018.2828032.

Wu, Y., Wang, L., Fan, L., Yang, M., Zhang, Y., Feng, Y., 2020. Comparison of the spatiotemporal mobility patterns among typical subgroups of the actual population with mobile phone data: a case study of Beijing. Cities 100, 102670. https://doi.org/10.1016/j.cities.2020.102670.

Yu, Q., Li, W., Duan, Z., Yang, D., 2018. Unveiling urban commuting structure from mobile phone data: a case study in Shanghai, China (No. 18-00784). In: Transportation Research Board 97th Annual Meeting.

Zhong, G., Wan, X., Zhang, J., Yin, T., Ran, B., 2016. Characterizing passenger flow for a transportation hub based on mobile phone data. IEEE Trans. Intell. Transp. Syst. 18 (6), 1507–1518. https://doi.org/10.1109/TITS.2016.2607760.

Zhong, G., Zhang, J., Li, L., Chen, X., Yang, F., Ran, B., 2018. Analyzing passenger travel demand related to the transportation hub inside a city area using mobile phone data. Transp. Res. Rec. 2672 (50), 23–34. https://doi.org/10.1177/0361198118774671.

Zhong, G., Yin, T., Zhang, J., He, S., Ran, B., 2019. Characteristics analysis for travel behavior of transportation hub passengers using mobile phone data. Transportation 46 (5), 1713–1736. https://doi.org/10.1007/s11116-018-9876-5.

Zuo, Y., Liu, Z., Fu, X., 2020. Measuring accessibility of bus system based on multi-source traffic data. Geo-Spatial Inform. Sci. 23 (3), 248–257. https://doi.org/10.1080/10095020.2020.1783189.