

Research Article

Spatiotemporal Analysis of Residents in Shanghai by Utilizing Chinese Microblog Weibo Data

Li Hou ¹, Qi Liu,^{1,2} Mueen Uddin,³ Hizbullah Khattak ⁴, and Muhammad Asshad⁵

¹School of Information Engineering and Engineering Technology Research Center of Intelligent Microsystems of Anhui Province Huangshan University and Huangshan Ruixing Automotive Electronics Co., Ltd., Huangshan 245041, China

²School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

³Digital Science, Faculty of Science, Universiti Brunei Darussalam, Jalan Tungku Link, Gadong BE1410, Negara, Brunei Darussalam

⁴Department of Information Technology, Hazara University Mansehra K-P, Mansehra, Pakistan

⁵Department of Information Technology, The University of Haripur, Haripur, Pakistan

Correspondence should be addressed to Li Hou; houli_shu@126.com

Received 17 June 2021; Revised 25 July 2021; Accepted 11 August 2021; Published 13 September 2021

Academic Editor: Sadaqat ur Rehman

Copyright © 2021 Li Hou et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mobile applications are really important nowadays due to providing the accurate check-in data for research. The primary goal of the study is to look into the impact of several forms of entertainment activities on the density dispersal of occupants in Shanghai, China, as well as prototypical check-in data from a location-based social network using a combination of temporal, spatial, and visualization techniques and categories of visitors' check-ins. This article explores Weibo for big data assessment and its reliability in a variety of categories rather than physically obtained information by examining the link between time, frequency, place, class, and place of check-in based on geographic attributes and related implications. The data for this study came from Weibo, a popular Chinese microblog. It was preprocessed to extract the most important and associated results elements, then converted to geographical information systems format, appraised, and finally displayed using graphs, tables, and heat maps. For data significance, a linear regression model was used, and, for spatial analysis, kernel density estimation was utilized. As per results of hours-to-day usage patterns, enjoyment activities and frequency distribution are produced. Our findings are based on the check-in behaviour of users at amusement locations, the density of check-ins, rush periods for visiting amusement locations, and gender differences. Our data provide light on different elements of human behaviour patterns, the importance of entertainment venues, and their impact in Shanghai. So it can be used in pattern recognition, endorsement structures, and additional multimedia content for these collections.

1. Introduction

In recent decades, academics have been more interested in mining location-based social network (LBSN) data for relevant outlines and awareness. A considerable quantity of data is created as a result of the vast number of LBSN-based apps available today, which is then analyzed to extract useable information, primarily from a practical standpoint, such as public transportation flows, disaster management, and collision avoidance [1]. These programmers' online capabilities enable users to contribute and share their opinions, habits, photographs, video files, and other content

with their network acquaintances, resulting in a large quantity of data that researchers may utilize to pinpoint visitors' actions and likings only over study. Internet services acquire and retain users' personal information as well as their current location.

The data are frequently augmented with information, multimedia, geolocations, and text, which may be utilized to do further study on various human behaviour features. Previous researches datasets were either physically collected data for populations in various classifications, such as leisure, or LBSNs data for the whole population without predefined implementations. If appropriately categorized,

the extensive elements of the LBSNs facts might demonstrate to be a valuable data source for analyzing human behaviour in various areas such as amusement, education, tourism, restaurants, and travel. As a consequence, we will cover the research gap of using LBSN data in article or substance in this study by determining which entertainment venues Shanghai residents want to visit. A variety of studies have been conducted to evaluate and recreate human activities using geodata. Recent studies, for example, use check-in data from internationally popular LBSNs like Twitter, Facebook, and Foursquare to uncover correlations and patterns across users like female and male, skilled or less skilled classes, and age groups [2, 3].

Unlike the widespread usage of Facebook, Twitter, and other LBSNs throughout the world, the majority of these LBSNs are outlawed or limited in China. As a result, Chinese residents prefer to utilize national microblog such as Weibo (Sina Weibo) [4, 5]. So, check-in data from Weibo may be acceptable for LBSNs statistical study in China. These outlines duplicate functional features inside the city and between cities and include activity behaviour, mobility, and density [6, 7]. The term “check-in” refers to a user using an LBSN application to confirm her/his location by participating in an activity or sharing her/his site with someone in a note [8, 9]. Weibo is well known not only among users but also among researchers, who execute a variety of topics to extract useful information from its geodata, such as the analysis of traffic accidents in Shanghai [10, 11], the analysis of tourism hotspot appeal features, the evaluations of the growth of Beijing’s urban boundaries [12], and spatiotemporal analysis by sex [13]. The majority of this research focuses on check-in data study for specific users or implementation details such as tourism, road accidents, determining urban borders, spring-festival rush, or sex [14]. There is a need to connect these spatiotemporal features to the locations where guests check in to get more details.

To the best of our knowledge, other researchers have not included this in the past. As a result, we concentrated on three distinct elements of analysis on Weibo’s check-in records from Shanghai city for two years, from July 2015 to June 2017, to discover spatiotemporal patterns and population estimates using entertainment venue categorization and density estimation.

As a result, the latest research highlights three main features of the analysis, our contribution to the field’s available research:

- (i) Time scales of an hour, a week, and a day.
- (ii) Data grouping and entertainment venue research.
- (iii) Modeling and abundance estimate using spatial analysis.

Tableau, a famous statistical tool among academics, was used to do temporal analysis to find many patterns from user data in “time,” verifying the data appropriateness for some further evaluation by presenting typical individual contacts. We deliver thorough clarifications of the findings in both statistical tables and graphical graphics. The geographical analysis was approved using kernel density estimation

(KDE) in ArcMap of ArcGIS Desktop 10.7.1 and map data from OpenStreetMap (OSM) to prove the variability of the check-ins.

Section 2 of this research includes pertinent work on big data, LBSNs, and their critical implications in a variety of fields, as well as pieces on Weibo, Shanghai, and China. Section 3 gives a summary of the dataset as well as the analytical technique. The findings and a summary are included in Section 4, and the study’s conclusion and recommendations are included in Section 5.

2. Related Work

Concerning other areas of computer science, research in big data analysis has advanced at an accelerating rate during the period, attracting enormous attention from numerous academic communities. Big data is becoming near, but potentially too near [15] for the greater advantage or the infringement of privacy. “Big data” is a phrase that should be used. Reference [16] provides the idea that only volumes are necessary; nevertheless, when defining “big data,” more factors must be considered, such as behaviour, difficulty, construction, tools, procedures, and technologies utilized to evaluate and handle the massive amounts of data [17]. Dumbill established the renowned three “V’s,” denoting dimensions of “big data,” meaning volume, variation, and velocity of its contents [18]. Mayer-Schönberger and Cukier [19] emphasized critical “big data” challenges such as association rather than interconnection, chaos rather than order, and populations rather than sample data.

Ovadia, on the other hand, showed and underlined the relevance of “big data” for academics and social scientists, stating that “big data” is too crucial to ignore because most social scientific research requires a large quantity of data and large datasets [20]. Many study areas, such as time and space geography, human mobility, user activity, and urban functions, began with statistical data derived from trip diaries, interviews, surveys, questioners, and other by-hand composed datasets [21]. These approaches, however, may not be sufficient to detect patterns in data since mobile phones, the Global Navigation Satellite System (GNSS), smart cards, and location-based Internet apps, including geoinformation, have lately disseminated efficient data sources for such scientific articles [22]. With the rapid advancement of mobile technology and the widespread usage of portable devices, tracking users’ whereabouts and activities has become easy. Gonzalez et al. [23]. Despite providing the estimated position near the mobile phone’s base transmitter where the calls are connected, the dataset exhibited success in predicting the placements of users with limited time, which was subsequently employed in predicting user movements [24]. Zhu discussed several components of GIS (geographic information system) and its role in pattern pulling out and urban studies by proving how to analyze and show the spatiotemporal components of recyclable waste, gathering, and restoration [25].

Researchers can organize quantitative analyses of human behaviours, trends, and relevant characteristics such as social contacts, personal preferences, and dwelling areas in the

current digitized environment [26, 27]. Fan et al divided their user behaviour research into three sections: location suggestion, trajectory mining, and location prediction [28]. Stressing the significance of understanding user patterns of behaviour and how they may help with applications such as traffic regulation, mobile advertising, disaster assistance, human health, and urbanism, “big data” can be processed, stored, and analyzed at several levels. Business intelligence (BI), which initially arose in the late 1990s [29], is an essential field of “big data” at the organizational level. BI helps in decision making by lowering doubt through forecasting, ad-hoc enquiries, and aggregation-based reporting, as well as handling structured and unstructured data and integrating “big data”-based systems [30, 31].

BI systems incorporate data warehouses for storing clean, accurate, and extensive data from various sources, as well as online analytical processing (OLAP) for real-time multidimensional assessment with operations such as grouping, filtering, roll-up, rotation, and drilling-down for specifics [32]. OLAP is one of the most well-known and renowned methodologies for “big data” research in BI systems [33]. However, BI, data warehouse, and OLAP are potent tools for dealing with large amounts of data and a wide range of activities, additionally posing a difficulty due to the high cost, storage, and processing resources required [34]. As it has been utilized and is developing rapidly in practically every area globally, online social networks have proven to be the most momentous source of “big data” to study individual’s behaviour. The LBSNs’ online services allow users to post and share their activities, interests, and whereabouts, resulting in vast volumes of data to perform various studies on various subjects. Papers [35, 36] go into great detail about the research on human behaviour analysis approaches.

Lindqvist [37] investigated the usage of LBSNs, which was led by other research publications, including empirical investigations and sociospatial features employing LBSNs [38] and a personalized geosocial suggestion based on a dataset from two independent LBSNs, namely Gowalla and Foursquare, by Zhang and Chow [39]. Colombo et al. [40] used similar check-in data from LBSN to improve recommendation algorithms in two UK cities by collecting frequent users at diverse venues. Li et al. [41] conducted a rigorous analysis that analyzed foursquare data from 2.4 million sites in 14 states to discover the factors influencing place reputation. The study’s outcomes revealed three primary criteria influencing a venue’s fame: venue profile, venue age, and venue nature. One other analysis on user behaviour at several venue classifications fixated on “food” in Riyadh, Saudi Arabia, discovered that customers are more open to sharing their practices when visiting meal venues. The check-ins of approximately 19,000 swarm (Foursquare) participants from three urban centers, namely San Francisco, New York, and Hong Kong, were used to debate relationships among distinct venues at varying periods of the day [42].

Several studies have been undertaken throughout the world to explore various characteristics of users and check-ins using LBSN data sources such as Foursquare and

Twitter. These characteristics have been applied in a diversity of domains, including mobility patterns, venue classification, and urban planning and expansion [43]. Weibo, a renowned Chinese LBSN, was employed and shown to be effective for this study. In research for Shenzhen, Gu and colleagues [11] used Weibo data to examine the appeal powers of visitor attractions. Another study on people’s movements and activity patterns which was undertaken in order to examine urban borders in Beijing [12] also used Weibo check-in. In a similar vein, Shi et al. [43], stemming from different details given by the LBSN, Weibo data were used to explore features of tourism destinations and took the analysis in conjunction with sentiments from user comments. Wu et al. [44] investigated spatiotemporal analysis based on the time of day and the variation in check-in patterns among weekdays and weekends. Wu et al. [45] also investigated check-in studies at 21 of Wuhan’s most renowned lakes.

3. Dataset and Methodology

3.1. Study Area and Data Source. The earlier study was based on data attained for Shanghai, a well-known Chinese metropolis located on the Yangtze River between 3040J–3153JN and 12052J–12212JE and spanning an area of 83592 km. It was allocated into a county and sixteen districts, which were named “Baoshan,” “Chongming,” “Fengxian,” “Hongkou,” “Area,” “Minhang,” “Putuo,” “Qingpu” and “Xuhui” “Huangpu,” “Songjiang,” “Changning,” “Jingan,” “Jiading,” “Jinshan,” “Pudong New”, and “Yangpu”. Shanghai is recognized as a financial powerhouse that connects China to the global economy, with a GDP of over 2.7 trillion Yuan and an average growth rate of 7.4 per cent during the past five years. Shanghai has a population density of 3854 persons per square kilometre in metropolitan areas. With an increase of 0.66 million people per year, it has exceeded Beijing as China’s most populated city. It has surpassed New York, the world’s fifth most populous metropolis. The massive influx of migrants, which accounted for nearly 39% of Shanghai’s total population in 2010 [46], is the major driver of population expansion. Figure 1 depicts the research area, which is based on Shanghai’s ten districts.

The information was collected from one of China’s most famous LBSNs, Weibo, which is a hybrid of Twitter and Facebook (the world’s two most popular LBSNs) and was founded on August 14, 2009. It is a popular micro blog where users can express themselves by writing articles, check-ins, and communication with friends and family by sharing their thoughts, favorites, activities, shots, audio/video messages, and locations.

Weibo provides a variety of geographic data, with real-time location sharing, sites referenced in posts, and user profile location being three of the most useful. By the end of 2018, the number of visitors had climbed to 500 million, with 462 million monthly active users and 200 million daily active users. This research is based on two years of social basis gathered spatiotemporal check-ins from Weibo in Shanghai, from July 2015 to June 2017.

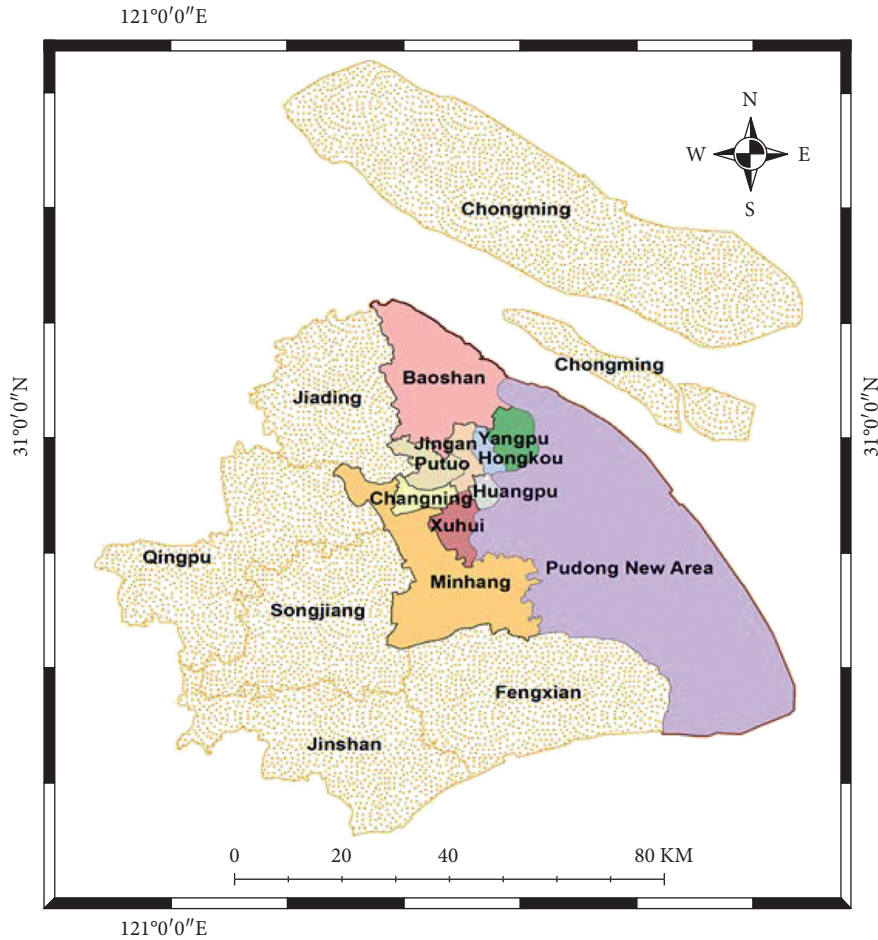


FIGURE 1: Study area.

The primary motivation for using LBSN is to share activities and remarks, which leads to the formation of a new intimate social companionship circle. This enables professionals to extrapolate a wide range of human behaviour and satisfaction from the geodata gathered by these LBSNs. The data for this article came from LBSN’s Weibo account. To gather information from check-ins in Shanghai, we used the Weibo API (application programming interface) based on Python. It was gathered in 2017, with around 3.5 million cumulative check-ins from around two million individuals. The gathered data were translated from JavaScript Object Notation (JSON), the standard API Java programming language, to comma-separated values (CSV) using MongoDB for the present study. Figure 2 depicts the data processing path.

Anomalies, missing values, and irrelevant characteristics were removed from the dataset after being filtered for valuable data. Finally, the researcher evaluated the entertainment places with multiple check-ins for a more thorough study. This study employed a filtered dataset with 87480 check-ins from 20152 regular users. Table 1 shows a sample of the records found in the final dataset.

The data inquiry is divided into three parts: temporal analysis, statistical analysis using spatiotemporal analysis of leisure sites, and density estimate. One of the most

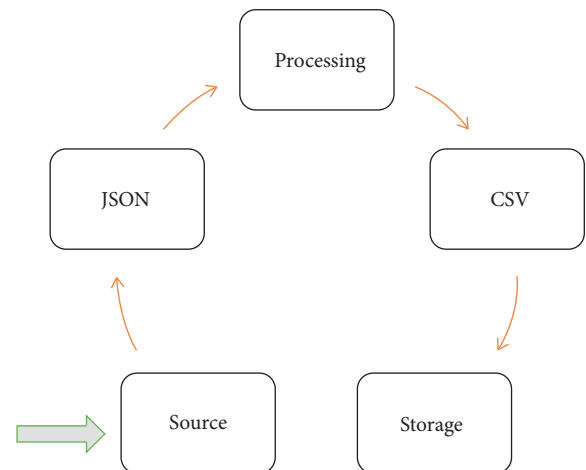


FIGURE 2: Process flow of data.

important goals of this type of research is to demonstrate the dataset’s legitimacy by demonstrating some observable human behaviour, such as less LBSN utilize after midnight until early morning due to sleeping habits and much more check-ins after working hours and on weekends due to more social events during relaxation time. In addition, because both genders are represented in distinct colours in all

TABLE 1: The sample of attributes in the dataset.

| User_id | Activity | Date | District | Lon | Lat | Gender | Time | Day |
|-------------|-----------|-----------|----------|----------|----------|--------|----------|-----|
| 1.78 E + 09 | Amusement | 5/8/2017 | Xuhui | 121.4487 | 31.19813 | M | 2:20:12 | Mon |
| 3.06 E + 09 | Amusement | 6/8/2016 | Yangpu | 121.5208 | 31.29679 | F | 18:20:23 | Thu |
| 2.46 E + 09 | Amusement | 12/5/2015 | Huangpu | 121.4634 | 31.21838 | M | 7:25:40 | Wed |
| 1.78 E + 09 | Amusement | 8/24/207 | Pudong | 121.5592 | 31.24204 | M | 3:30:45 | Mon |

descriptive data, several fascinating patterns emerge for males and females. The check-ins collection criteria have been implemented, as shown in Figure 3.

The venues were classified by comparing physical locations inside the city to location names and location. The place names were decoded and retrieved from the Weibo dataset as the “location” property, as shown in Table 1, and categorized according to the functionalities of these various venues, such as those that were regularly frequented and well known in Shanghai. Every check-in was assigned to the class that most suited the type of the entertainment and amusement movement done at that place, such as “cinema, KTV entertainment hall, theatre, and Disney Park.” Figure 4 illustrates the flowchart of our technique.

3.2. Statistical Analysis. The study was conducted using SPSS (Version 17.0) and linear regression to see if the dependent variables are predictors of the independent variable. Table 2 shows the summary of the model.

$$\begin{aligned} \sum \text{Check in}_t = & \beta_0_t + \beta_1_t \sum \text{District} + \beta_2_t \sum \text{Days} \\ & + \beta_3_t \sum \text{Gender} + \mu_t. \end{aligned} \quad (1)$$

The models display R and R square values. The R -value (0.128) and the R square reflect a straightforward correlation (0.016). The value of R demonstrates that check-in can be predicted by 16 percent using characteristics such as days, gender, and districts.

Table 3 indicates that significance value of F-statistic in the model showed that the whole model is significant as $F(3, 87476) = 482.358$, $p = 0.000$.

As shown in Table 4, the independent variables are significant predictors for the dependent variable. The results show a statistically significant ($P = 0.000$) and negative (-6325.933) connection between gender and check-in. All other predictors, including days and districts, have a significant and positive relationship with the check-in having the P values 0.002 and 0.000, respectively.

3.3. Spatial Analysis. The KDE technique is a nonparametric way to estimate density from randomly selected evidence. KDE generates smooth distributions by removing some local noise, regarded as a nonprobability distribution with maximum bandwidth and lower error. KDE is a bulk study technique used to recognize numerous location-based features such as time and destination and is an imperative density approximation technique that has been widely studied for the analysis of unlike aspects of location-based social media data such as vital city limits, visitor’s mobility and activity patterns, recommendation for a position of

interest, and check-in manners. The KDE method has been used in a variety of applications, including medicine, marketing, and ecology for modeling spatial densities.

Let E be a set of historical data where $e^j = \langle x, y \rangle$ is the geocoordinates of a location, and $1 \leq j \leq n$, for an individual i . h_j is the Euclidean distance to k th nearest neighbor e^j in the training data. The KDE is expressed as follows:

$$\begin{aligned} f_{\text{KD}}(e|E) &= \frac{1}{n} \sum_{j=1}^n K_{h_j}(e, e^j), \\ K_h &= \frac{1}{2\pi h} \exp\left(-\frac{1}{2}\left(\frac{e - e^i}{h}\right)^t \sum_h^{-1}(e, e^i)\right). \end{aligned} \quad (2)$$

ArcGIS software has been used for spatial analysis and all the maps have been created through this software which is believed to be a significant geographic information systems software nowadays.

3.4. Temporal Analysis. Tableau software is used to do a temporal analysis for daily, weekly, and district-based check-ins.

4. Results and Discussion

Mobile technology, wireless connectivity, the Internet, and location-based services have evolved dramatically during the previous two decades. As a result, services based on these features, such as LBSN like Twitter, Weibo, and Facebook, are attracting an increasing number of academics to evaluate the vast amounts of data generated by these services. The study was incredibly helpful in identifying basic patterns concerning essential tasks like crisis and catastrophe management, urban planning, innovative city development, and other significant data-related sectors. This section looks at two different kinds of check-in results: temporal and spatial analysis.

We gathered check-ins based on entertainment purposes, and all of the check-ins from entertainment places are shown in Figure 5. The study area encompasses ten Shanghai districts.

The density hotspots were identified using kernel density estimation, and as shown in Figure 6, the enormous black dots represent a high density of visits. It is worth mentioning that Shanghai’s seven districts, together known as the city center, have the highest check-in density.

Figure 7 can see the temporal result of entertainment check-ins in districts, and all the district’s check-ins activity can be seen. Putuo and Pudong are two districts that have a higher number of amusement check-ins compared with others.

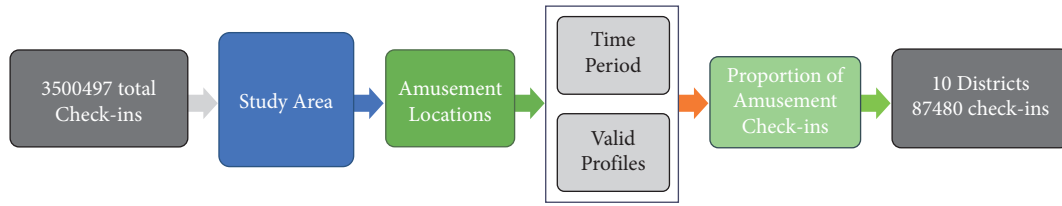


FIGURE 3: Criteria.

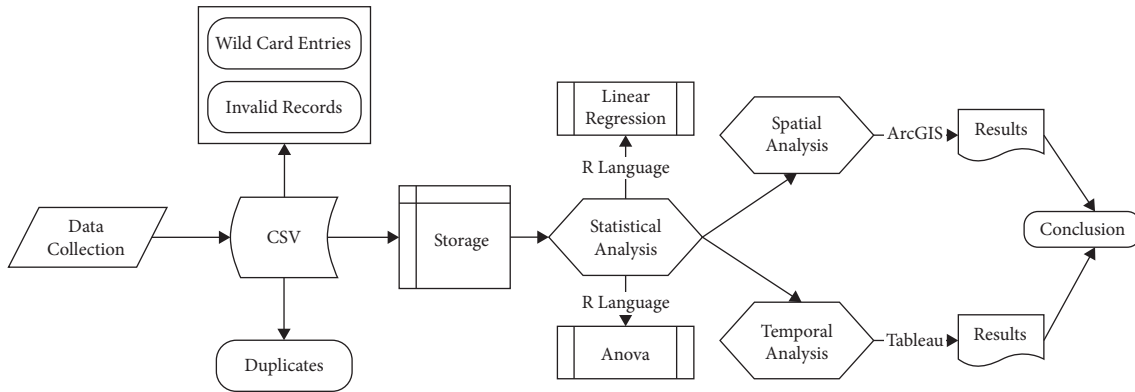


FIGURE 4: Methodology.

TABLE 2: Model summary.

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|--------------------|----------|-------------------|----------------------------|
| 1 | 0.128 ^a | 0.016 | 0.016 | 25047.553 |

a. Predictors: (constant), gender, districts, and days.

TABLE 3: ANOVA ^a.

| Model | | Sum of Squares | df | Mean Square | F | Sig. (b) |
|-------|------------|--------------------|-------|------------------|---------|----------|
| 1 | Regression | 907864706534.297 | 3 | 302621568844.766 | 482.358 | 0.000 |
| | Residual | 54880685702175.710 | 87476 | 627379917.945 | | |
| | Total | 55788550408710.010 | 87479 | | | |

a. Check in is a dependent variable. b. Predictors: gender, districts, (constant), and days.

TABLE 4: Coefficients.

| Model | Unstandardized coefficients | | | Standardized coefficients | | t | Sig. |
|-------|-----------------------------|------------|---------|---------------------------|--|---------|-------|
| | B | Std. Error | Beta | | | | |
| 1 | (Constant) | 51366.226 | 380.167 | | | 135.115 | 0.000 |
| | Districts | 333.889 | 32.151 | 0.035 | | 10.385 | 0.000 |
| | Days | 127.114 | 40.212 | 0.011 | | 3.161 | 0.002 |
| | Gender | -6325.933 | 172.959 | -0.123 | | -36.575 | 0.000 |

The weekly data indicate an odd phenomenon even after Saturday and Sunday being holidays. Saturday had more check-ins than Sundays. One reason is that Saturday is followed by another holiday (Sunday), allowing people to avoid getting up early and enjoy and amuse themselves until beyond late night every day. The weekly data indicate an odd phenomenon: while both Saturday and Sunday being holidays, Saturday had more check-ins than Sundays. Aside from that, check-in frequency was substantially lower and relatively consistent on all working days from Monday

through Thursday, increased dramatically on Fridays, and was followed by a significant number of check-ins on weekends. Figure 8 depicts the daily trend.

Figure 9 shows the hour-based check-ins of Weibo users, with the peak period of entertainment venue check-ins occurring between 8 and 10 p.m., and this pattern continued until midnight, with numbers declining after midnight.

In Shanghai, we looked at category (male and female) data to see how check-in frequency and mindset differed. Figure 10 depicts the gender check-in density in Shanghai. It

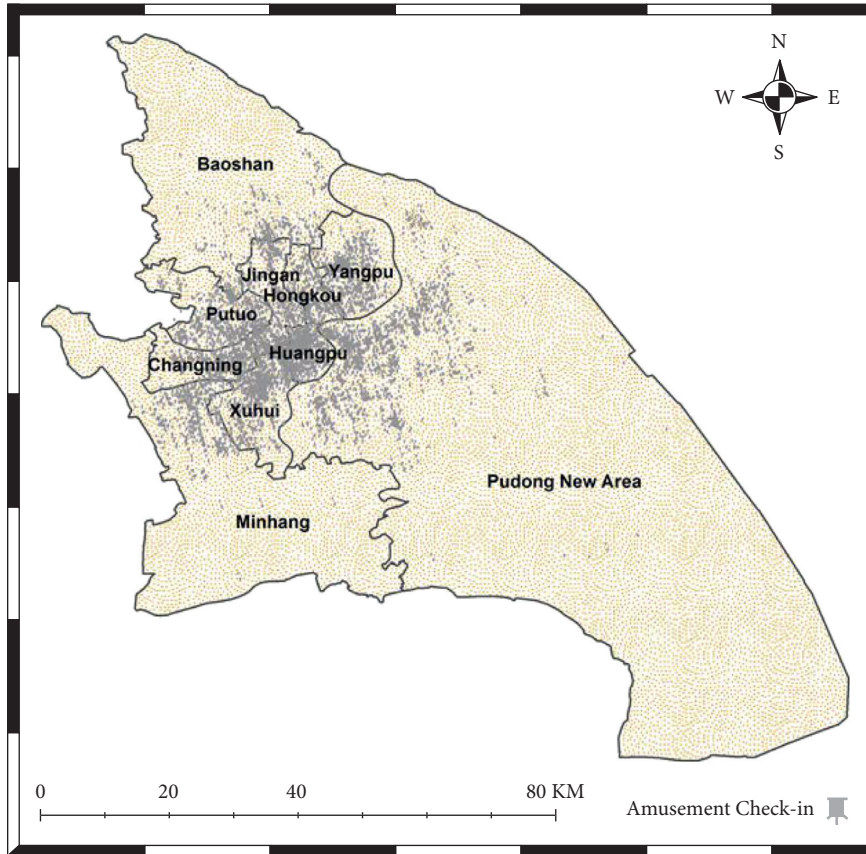


FIGURE 5: Amusement check-ins.

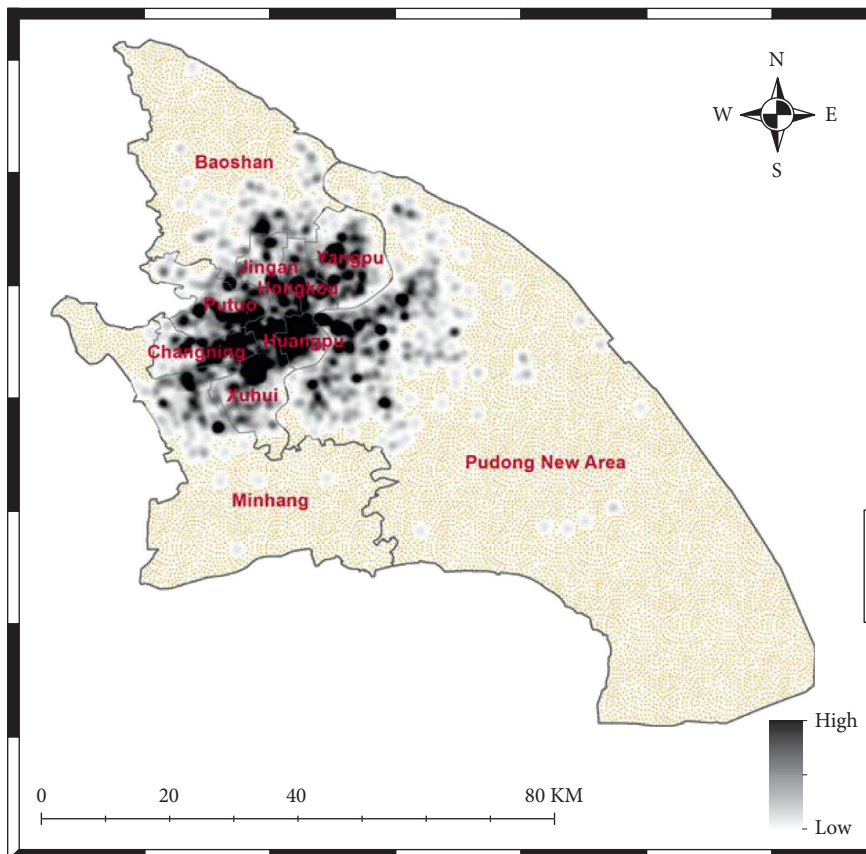


FIGURE 6: Check-ins density.

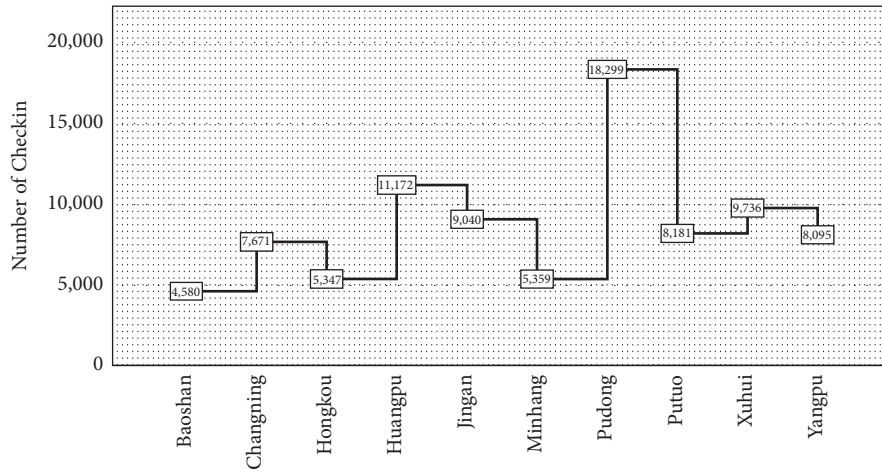


FIGURE 7: Amusement check-ins in districts.

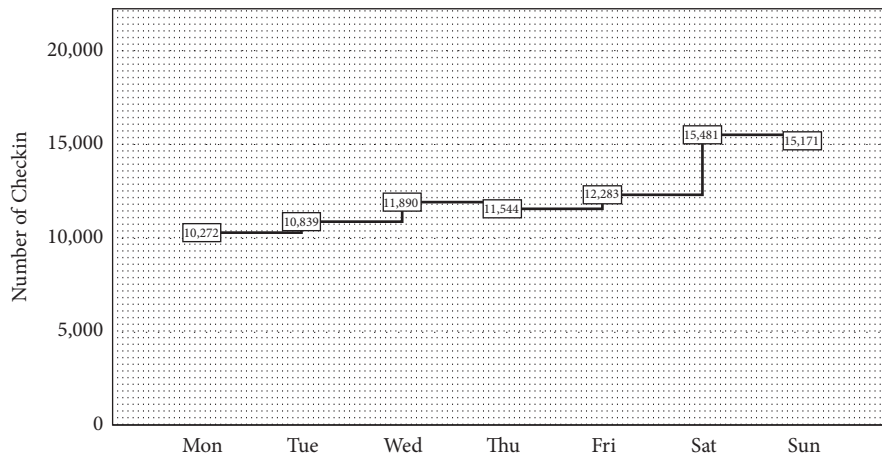


FIGURE 8: Check-ins pattern for days.

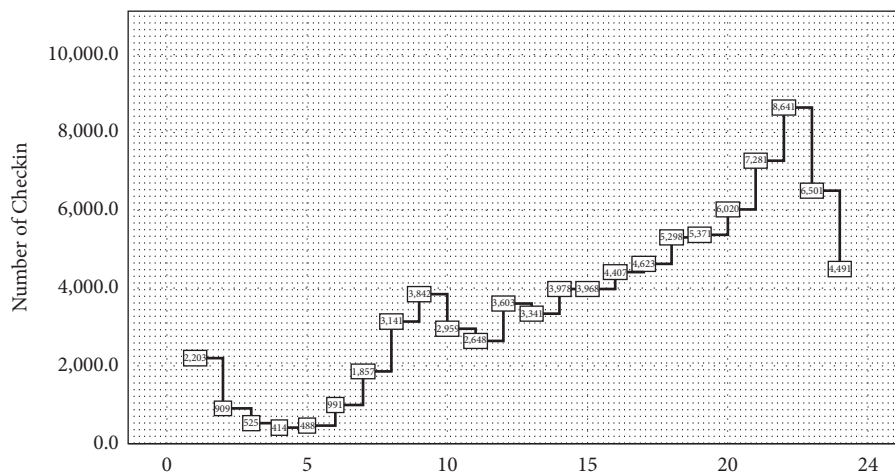


FIGURE 9: Hour-based check-ins.

has been observed that female visitors favor Weibo over male visitors. As a result, Figure 10 illustrates the results. This study showed the influence of each check-in entertainment venue through temporal and geographic

outcomes. However, the truth remains that, in comparison to suburban districts, the city center has a more focused and higher density. Aside from these diverse spatial and temporal trends, the findings demonstrate Weibo’s potential as

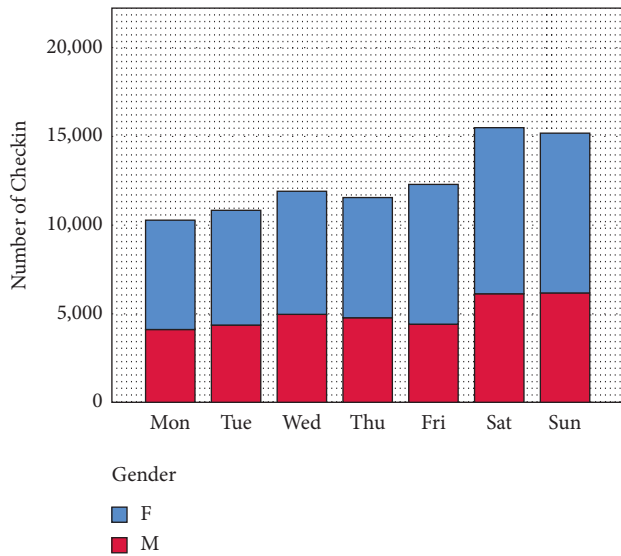


FIGURE 10: Gender difference.

an LBSN dataset by illustrating the critical involvement of every place type to the intensity and variety of Shanghai as a vibrant metropolitan center in terms of time and location. As long as the dataset is suitably classified into multiple groups, we may use LBSN data instead of human data collecting as a valuable source of big data analysis in numerous sectors.

5. Conclusion

We utilized the check-in data from Weibo for spatiotemporal analysis to investigate various trends in different activities linked to amusement in Shanghai during two years (July 2015 to June 2017). The study examines check-in behaviour over time (daily, weekly), as well as check-in data density predictions and the involvement of each check-in location to the strength. A temporal analysis revealed some noteworthy characteristics, including the low number of amusement events after midnight in a megacity like Shanghai and the most significant social activities on Saturday rather than Sunday. The findings of density estimates revealed that the city center performs the highest number of check-ins, as predicted, due to the greater availability of resources. In addition, females seem to be more likely to use social media when participating in entertainment. These conclusions are based on Weibo check-in data for Shanghai city, which has numerous qualities, portraying it as a chrono urban metropolis that is more reachable and friendly. The research might help create smart cities, recommendation systems, and LBSN studies in specific domains, including transportation, tourism, and entertainment. With the present research, there are various constraints to observe. Because Weibo was the sole source of data used, the collection was restricted and perhaps skewed. To acquire a more accurate sample, we may handle this problem by merging it with additional information sources such as WeChat, Trip Advisor, and other LBSNs.

Data Availability

The dataset can be downloaded from <http://www.weibo.com>.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the Anhui Natural Science Foundation (no. 1908085MF178), Anhui Key Research and Development Plan Project (no. 202104b11020031), Anhui Excellent Young Talents Support Program Project (no. gxyqZD201906), and Anhui Natural Science Research Key Project (no. KJ2019A0609).

References

- [1] R. Singh, Y. Zhang, and H. Wang, "Exploring human mobility patterns in Melbourne using social media data," in *Australian Database Conference* Springer, Berlin, Germany, 2018.
- [2] N. Alrumayyan, S. Bawazeer, R. AlJurayyad, and M. Al-Razgan, "Analyzing user behaviour s: a study of tips in Foursquare," in *Advances in Intelligent Systems and Computing* Springer, Berlin, Germany, 2018.
- [3] J. Preston and B. Stelter, *How government officials are using Twitter for Hurricane Sandy*, The New York Times, New York, NY, USA, 2012.
- [4] Weibo. [cited 2020 2 September]; Available from: <http://www.weibo.com>.
- [5] N. U. Khan, W. Wan, S. Yu, A. A. M. Muzahid, S. Khan, and L. Hou, "A study of user activity patterns and the effect of venue types on city dynamics using location-based social network data," *ISPRS International Journal of Geo-Information*, vol. 9, no. 12, p. 733, 2020.
- [6] B. Wang, F. Zhen, Z. Wei, S. Guo, and T. Chen, "A theoretical framework and methodology for urban activity spatial structure in e-society: empirical evidence for Nanjing City, China," *Chinese Geographical Science*, vol. 25, no. 6, pp. 672–683, 2015.
- [7] W. Bo, Z. Feng, and W. Zongcai, "The research on characteristics of urban activity space in Nanjing: an empirical analysis based on big data," *Human Geography*, vol. 29, no. 3, pp. 14–21, 2014.
- [8] A. W. Todd, A. L. Campbell, G. G. Meyer, and R. H. Horner, "The effects of a targeted intervention to reduce problem behaviors," *Journal of Positive Behaviour Interventions*, vol. 10, no. 1, pp. 46–55, 2008.
- [9] F. Zhen, Y. Cao, X. Qin, and B. Wang, "Delineation of an urban agglomeration boundary based on Sina Weibo microblog "check-in" data: a case study of the Yangtze River Delta," *Cities*, vol. 60, pp. 180–191, 2017.
- [10] B. P. Loo, S. Yao, and J. Wu, "Spatial point analysis of road crashes in Shanghai: a GIS-based network kernel density method," in *Proceedings of the 2011 19th International Conference on Geoinformatics*, June 2011.
- [11] Z. Gu, Y. Zhang, Y. Chen, and X. Chang, "Analysis of attraction features of tourism destinations in a mega-city based on check-in data mining-A case study of shenzhen, China," *ISPRS International Journal of Geo-Information*, vol. 5, no. 11, p. 210, 2016.

- [12] Y. Long, H. Han, Y. Tu, and X. Shu, "Evaluating the effectiveness of urban growth boundaries using human mobility and activity records," *Cities*, vol. 46, pp. 76–84, 2015.
- [13] C. Lei, A. Zhang, Q. Qi, H. Su, and J. Wang, "Spatial-temporal analysis of human dynamics on urban land use patterns using social media data by gender," *ISPRS International Journal of Geo-Information*, vol. 7, no. 9, p. 358, 2018.
- [14] D. Preoŕtiuc-Pietro and T. Cohn, "Mining user behaviours: a study of check-in patterns in location based social networks," in *Proceedings of the 5th Annual ACM Web Science Conference*, New York, NY, USA, May 2013.
- [15] S. Lohr, *Big Data is Opening Doors, But Maybe Too Many*, New York Times, New York, NY, USA, 2013.
- [16] P. Chatterjee, "Big data: the greater good or invasion of privacy," *The Guardian*, vol. 12, 2013.
- [17] J. S. Ward and A. Barker, "Undefined by data: a survey of big data definitions," <https://arxiv.org/abs/1309.5821>.
- [18] E. Dumbill, "What is big data: an introduction to the big data landscape (article)," *Strata Oreilly*, vol. 11, 2012.
- [19] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution that Will Transform How We Live, Work, and Think*, Houghton Mifflin Harcourt, Boston, MA, USA, 2013.
- [20] S. Ovadia, "The role of big data in the social sciences," *Behavioural & Social Sciences Librarian*, vol. 32, no. 2, pp. 130–134, 2013.
- [21] Y. Chai, Y. Shen, Z. Xiao, Y. Zhang, Y. Zhao, and N. Ta, "Review for space-time behaviour research: theory frontiers and application in the future," *Progress in Geography*, vol. 31, no. 6, pp. 667–675, 2012.
- [22] M. Graham and T. Shelton, "Geography and the future of big data, big data and the future of geography," *Dialogues in Human geography*, vol. 3, no. 3, pp. 255–261, 2013.
- [23] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [24] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [25] X. Zhu, "GIS and urban mining," *Resources*, vol. 3, no. 1, pp. 235–247, 2014.
- [26] A. Wesolowski, T. Qureshi, M. F. Boni et al., "Impact of human mobility on the emergence of dengue epidemics in Pakistan," *Proceedings of the National Academy of Sciences*, vol. 112, no. 38, pp. 11887–11892, 2015.
- [27] L. Pappalardo, F. Simini, S. Rinzivillo, D. Pedreschi, F. Giannotti, and A. L. Barabási, "Returners and explorers dichotomy in human mobility," *Nature Communications*, vol. 6, no. 1, pp. 8166–8168, 2015.
- [28] C. Fan, Y. Liu, J. Huang, Z. Rong, and T. Zhou, "Correlation between social proximity and mobility similarity," *Scientific Reports*, vol. 7, no. 1, pp. 11975–11978, 2017.
- [29] B. Wixom and H. Watson, "The BI-based organization," *International Journal of Business Intelligence Research*, vol. 1, no. 1, pp. 13–28, 2010.
- [30] B. Wieder and M.-L. Ossimitz, "The impact of business intelligence on the quality of decision making - a mediation model," *Procedia Computer Science*, vol. 64, pp. 1163–1171, 2015.
- [31] T. Grublješić and J. Jaklič, "Conceptualization of the business intelligence extended use model," *Journal of Computer Information Systems*, vol. 55, no. 3, pp. 72–82, 2015.
- [32] S. Ahmad, S. Miskon, T. A. Alkanhal, and I. Tlili, "Modeling of business intelligence systems using the potential determinants and theories with the lens of individual, technological, organizational, and environmental contexts-a systematic literature review," *Applied Sciences*, vol. 10, no. 9, p. 3208, 2020.
- [33] Y. S. Singh, "Easy designing steps of a local data warehouse for possible analytical data processing," *ADB Journal of Engineering Technology*, vol. 8, 2019.
- [34] V. Santos, R. Silva, and O. Belo, "Towards a low cost ETL system," *International Journal of Database Management Systems*, vol. 6, no. 2, pp. 67–79, 2014.
- [35] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Diego, CA, USA, August 2011.
- [36] H. Gao, J. Tang, and H. Liu, "Exploring social-historical ties on location-based social networks," in *Proceedings of the International AAAI Conference on Web and Social Media*, Palo Alto, CA, USA, 2012.
- [37] J. Lindqvist, "I'm the mayor of my house: examining why people use foursquare-a social-driven location sharing application," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, Canada, May 2011.
- [38] S. Scellato, "Socio-spatial properties of online location-based social networks," in *Proceedings of the International AAAI Conference on Web and Social Media*, Barcelona, Spain, July 2011.
- [39] J.-D. Zhang and C.-Y. Chow, "iGSLR: personalized geo-social location recommendation: a kernel density estimation approach," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, NY, USA, November 2013.
- [40] G. B. Colombo, M. J. Chorley, M. J. Williams, S. M. Allen, and R. M. Whitaker, "You are where you eat: foursquare checkins as indicators of human mobility and behaviour," in *Proceedings of the 2012 IEEE International Conference on Pervasive Computing and Communications Workshops*, March 2012.
- [41] Y. Li, M. Steiner, L. Wang, Z.-L. Zhang, and J. Bao, "Exploring venue popularity in foursquare," in *Proceedings of the 2013 Proceeding IEEE INFOCOM*, April 2013.
- [42] S. Lin, R. Xie, Q. Xie, H. Zhao, and Y. Chen, "Understanding user activity patterns of the swarm app: a data-driven study," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, Maui, HI, USA, September 2017.
- [43] B. Shi, J. Zhao, and P.-J. Chen, "Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data," *Current Issues in Tourism*, vol. 20, no. 11, pp. 1186–1209, 2017.
- [44] C. Wu, X. Ye, F. Ren, and Q. Du, "Check-in behaviour and spatiotemporal vibrancy: an exploratory analysis in Shenzhen, China," *Cities*, vol. 77, pp. 104–116, 2018.
- [45] J. Wu, J. Li, and Y. Ma, "A comparative study of spatial and temporal preferences for waterfronts in Wuhan based on gender differences in check-in behaviour," *ISPRS International Journal of Geo-Information*, vol. 8, no. 9, p. 413, 2019.
- [46] C. Y. Liu, J. Chen, and H. Li, "Linking migrant enclave residence to employment in urban China: the case of Shanghai," *Journal of Urban Affairs*, vol. 41, no. 2, pp. 189–205, 2019.