

Article

Big Data Analysis to Observe Check-in Behavior Using Location-Based Social Media Data

Muhammad Rizwan ^{1,2,*} and Wanggen Wan ^{1,2}

¹ School of Communication & Information Engineering, Shanghai University, Shanghai, 200444, China; wanwg@staff.shu.edu.cn

² Institute of Smart City, Shanghai University, Shanghai, 200444, China

* Correspondence: rizwan@shu.edu.cn; Tel.: +86-131-220-98748

Received: 12 September 2018; Accepted: 11 October 2018; Published: 20 October 2018



Abstract: With rapid advancement in location-based services (LBS), their acquisition has become a powerful tool to link people with similar interests across long distances, as well as connecting family and friends. To observe human behavior towards using social media, it is essential to understand and measure the check-in behavior towards a location-based social network (LBSN). This check-in phenomenon of sharing location, activities, and time by users has encouraged this research on the frequency of using an LBSN. In this paper, we investigate the check-in behavior of several million individuals, for whom we observe the gender and their frequency of using Chinese microblog Sina Weibo (referred as “Weibo”) over a period in Shanghai, China. To produce a smooth density surface of check-ins, we analyze the overall spatial patterns by using the kernel density estimation (KDE) by using ArcGIS. Furthermore, our results reveal that female users are more inclined towards using social media, and a difference in check-in behavior during weekday and weekend is also observed. From the results, LBSN data seems to be a complement to traditional methods (i.e., survey, census) and is used to study gender-based check-in behavior.

Keywords: social media; LBSN; check-in; gender; time; behavior; geolocation

1. Introduction

Human mobility and human behavior towards services are closely intertwined with personal behavior and characteristics. In recent research [1–5], human mobility and population density are observed by using data collected through traditional methods (i.e., survey, census), which are considered expensive and require more processing time that results in very sparse data and is not very helpful in policymaking [6]. Social network sites [7], location-based social network (LBSN) services [8], and virtual social networks [9] (e.g., Facebook [10], Twitter [11], and Weibo [12]), has enabled users to share their location (hereby referred as a “check-in” [13]). As part of a social interaction, sharing a check-in allows users to announce the places they visit (e.g., restaurants, shopping malls, and popular scenic areas). This check-in phenomenon generates an enormous amount of user data (also referred “Big Data” [14]) and has attracted more than 222 million subscribers; statistics showed there were 500 million users with more than 100 million daily users on Weibo by the third quarter of 2015 [15–17]. Regardless of some limitations on representing check-in behavior, e.g., the bias of gender, a low sampling frequency, and the bias of location category, check-in data can uncover check-in behavior within a city. Compared to the aforementioned traditional methods, LBSN data are highly available at low cost. Moreover, this data contains rich information about geolocation [18], which can be used to study check-in behavior. Thus, geo-location data offers new dimensions towards studying check-in behaviors and can help to create new techniques and approaches to analyze LBSN data. Moreover, it seems that LBSN data can be a supplement to, rather than a substitute for, traditional data sources

for policy making [6]. Therefore, LBSN data can be considered as a supplement while taking policy decisions related to urban planning and public services by identifying the sentiment about a topic or community detection and user analysis for identification of the actors involved [19–25].

In the current study, we have the reasonable prospect of using LBSN data as a novel perspective to observe individual level check-in behavior and intensity of check-ins during the period within a city. The current study is an empirical exploration using a dataset from a dominant social media site in China referred as “Weibo” (launched by Sina Corporation on August 14, 2009), where 50.10% of Weibo users are male and 49.90% are female [26]. We consider LBSN data be helpful for observing check-in behavior by males and females during the weekday and weekend and over a period.

The rest of the paper is organized as follows. Section 2 overviews related works. Section 3 describes the study area and dataset used in the current study and presents the methodology. Section 4 presents the results and discussion of the experimental results performed on the dataset. Finally, Section 5 concludes the paper and proposes some further research issues.

2. Related Work

Research on gender studies [27] has long been limited to analyze traditional datasets (i.e., survey, census), but with the enhanced capabilities to capture and process geo-location information, the field of spatial analysis has blossomed [28]. Alongside social media data, mobile phone datasets have been used to understand human activity behaviors and individual mobility patterns [29,30]. However, mobile phone data sets are not considered to be a feasible choice to study human mobility pattern analysis. To avoid the limitation of mobile phone data, social media datasets are collected and used. The social media data sources are so diverse that it includes log files from smart devices and websites, social media data, and geotagged audio, video, and graphics data [31].

Various studies have been conducted to study check-in behavior under different perspectives such as privacy [32,33], gender differences [34], and geographical distances [7]. However, Benevenuto et al. [35] presented an analysis of user workloads in online social networks to study opportunities for better interface design, richer studies of social interactions, and improved design of content distribution systems. Results reveals vital features of the social network workloads, frequency, and for how long, as well as the types and sequences of activities that users conduct on the online social network. Scellato et al. [36] explored socio-spatial properties among different LBSN platforms to study the check-in behavior and the type of places users visit. Noulas et al. [37] analyzed check-in patterns of Foursquare users to study check-in behavior and the mobility patterns within a city. Ruggles [38] explored LBSN data to study the place to health relationships to expand opportunities for public health. Chorley et al. [39] developed a web-based participatory application that examines the personality characteristics and check-in behavior of Foursquare users and stated that personality traits help to explain individual differences in LBSN usage and the type of places visited. Maia et al. [40] proposed a methodology for characterizing and identifying user behaviors in online social networks and introduced a clustering algorithm to group users that share the similar behavioral pattern in the social network. Pucci et al. [41] presented an analysis of the large-scale event and aimed to observe inconsistency of urban spaces and to formulate policies in keeping with the molecular daily practices, and emerging demands can be made by diverse populations using the city and its services at varying rhythms and intensities. While Hong [42] explored the user participation in order to provide insight into Seoul city and analyzed social media data from Foursquare. Moreover, observed venues were based on user participation and the city’s characteristics. Jin et al. [43] presented a survey to give a comprehensive review of the state-of-the-art research related to user behavior in the social network from several perspectives, i.e., social connectivity and interaction, social behaviors, and malicious behaviors of social network users. Gyarmati and Trinh [44] presented a large-scale measurement analysis of user behavior in social network and created a measurement framework in order to observe user activity, characterization of user activities, and usage patterns in the social network.

LBSN datasets have been used in many studies of development and prediction, for example, Preoȋuc-Pietro and Cohn [45] studied the whereabouts of users with an emphasis on the type of places and their evolution over time, and uncovered patterns across different temporal scales for venue category usage. Shen and Karimi [46] proposed a framework to characterize urban streets by conceptualizing visual paths and stated that the usage of ubiquitous big social media data can enrich the current description of the urban network system and enhance the predictability of network accessibility on socioeconomic performance. Wu et al. [47] highlighted the use of LBSN data to observe the willingness of buyers to pay for various factors. The opinions and geographical preferences of individuals for places can be represented by visit frequencies and are given different motivations. Luarn et al. [48] developed and refined a conceptual framework to provide a theoretical understanding of the motivations that induce consumers to engage in check-in behavior. The study indicates that the social conditions (e.g., tie strength, subjective norms, expressiveness, social support, and information sharing) play the most critical role in motivating people to engage in check-in behavior. While Wang and Stefanone [49] studied how personality traits influence self-disclosure and, in turn, impacts the intensity of check-ins on Facebook and highlighted the physical and informational mobility of the users relating individual activities into spaces.

A substantial amount of studies [50–55] have been carried out over recent years to study user demographics of social network site (SNS) users and identified a few differentiating factors that lead a male or a female user to use a social media network. These studies suggest that there are different motivations for using social media networks among men and women. Smith [50] showed that women are more likely to consume social media to connect with families than men. While Muscanell and Guadagno [56] found that male social media network users reported using it for making new relationships while women reported more for relationship maintenance. The usage patterns and motivation for using social media network by gender seems to be slightly different. Research on social media network indicates that motivations for information seeking, entertainment, and self-expression have a positive effect on male users' active involvement; however, the motivations for socializing and entertainment have a positive effect on female users' effective involvement [57,58].

On the other hand, many studies have used LBSN data to study differences in check-in behavior by gender. For example, Blumenstock et al. [59] analyzed data from Rwanda to observe population density and mobile phone use behavior by different genders. Rizwan et al. [60] investigated the check-in behavior of Chinese microblog Sina Weibo and observed the gender difference and their frequency of use over a period. While in another study Rizwan et al. [61] examined how check-in behavior varies in the same weeks but in different years. Moreover, regarding mobility patterns and practices in terms of time and space in Shanghai, Lei et al. [62] used location data from Weibo to study the human dynamics of the spatial-temporal characteristics of gender differences and check-in behavior in Beijing's Olympic Village and stated that female users outnumbered male users in social media activity. Zheng et al. [63] designed an approach to mine the correlation between locations from a large amount of people's location histories by using social media data, and Comito et al. [64] presented a novel methodology to extract and analyze the time- and geo-references associated with social data to mine information about human dynamics and behaviors within the urban context. Also, previous research [36,65–67] on LBSN studied users' check-in data to predict users' location and mobility patterns. LBSN datasets have now been used in many studies for urbanization and its environmental effects [68], development and prediction [69–71], travel and activity patterns [72,73], emergency response [74–76], and urban sustainability [77]. This line of research is helpful in understanding differences in check-in behavior by gender, but the current study did not consider the connection with other indicators (i.e., equal access to education, equal access to economic resources, and end of violence) of gender equality [78,79] in a society which is out of the scope of the current study.

3. Material and Methods

3.1. Dataset and Study Area

The Shanghai, China dataset used in the current study comes from Chinese microblog Weibo during April–May 2016. Shanghai, China (lying between $30^{\circ}40'–31^{\circ}53'N$ and $120^{\circ}52'–122^{\circ}12'E$ [80]) is located on the eastern edge of the Yangtze River Delta [81]. According to Gu, X., S. Tao, and B. Dai [82], in 2015, Shanghai had a total area of 8359 km² with a population of around 24.15 million people. In 2016, Shanghai was divided into 16 county-level divisions: 15 districts (Baoshan, Changning, Fengxian, Hongkou, Huangpu, Jiading, Jingan, Jinshan, Minhang, Pudong New Area, Putuo, Qingpu, Songjiang, Xuhui, and Yangpu) and 1 county (Chongming) [83]. Seven of the districts (Changning, Hongkou, Huangpu, Jingan, Putuo, Xuhui, and Yangpu) are located in Puxi (literally Huangpu West). These seven districts are referred to as the city center [84,85] as shown in Figure 1. For the current study, ten districts of Shanghai (Baoshan, Changning, Hongkou, Huangpu, Jingan, Minhang, Pudong New Area, Putuo, Xuhui, and Yangpu) are analyzed, which are connected to the city center.

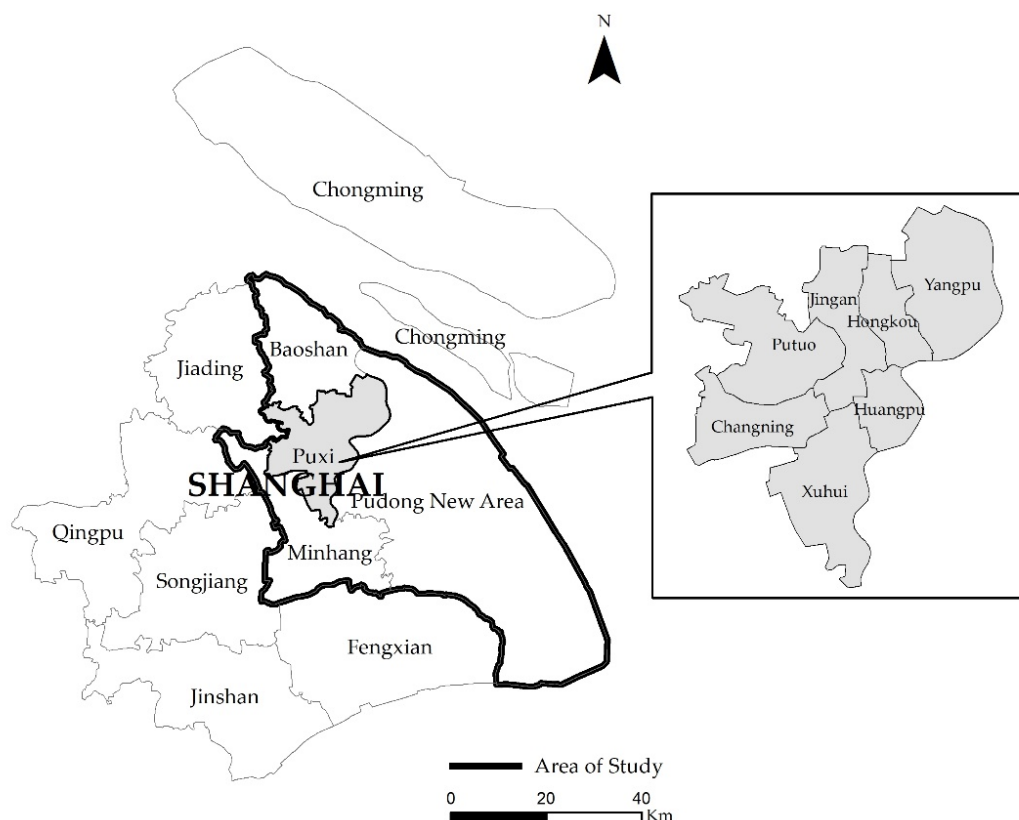


Figure 1. Map of Shanghai.

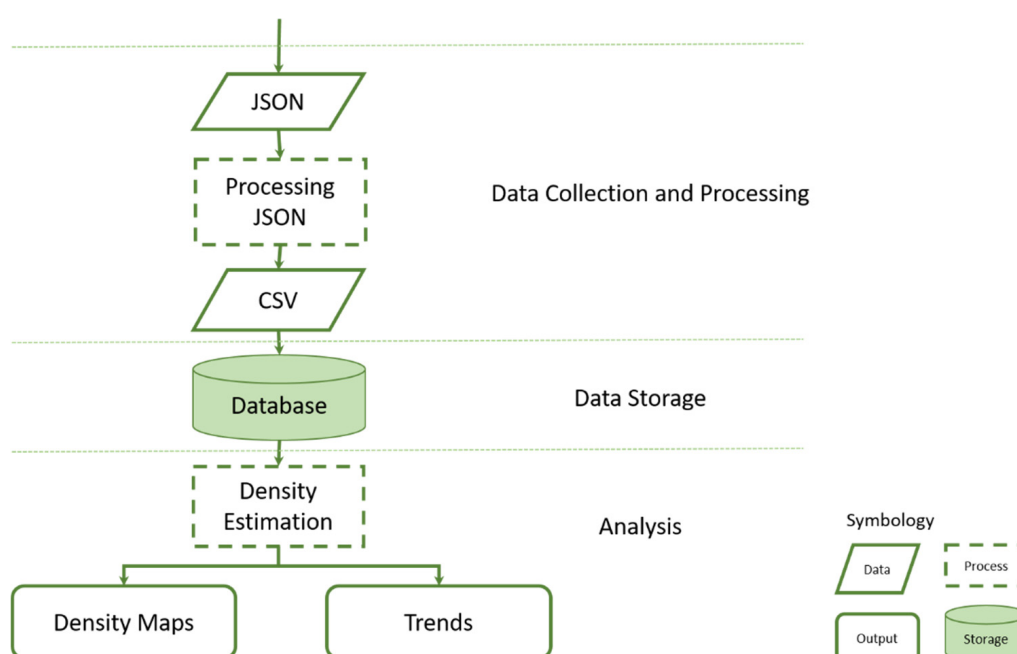
Table 1 presents the details of the Shanghai dataset used in the current study, which has information like user ID, gender, date, time, and geo-location (longitude and latitude), but no personal information like name and address is available. Therefore, check-in data records the daily life patterns and users' behaviors towards the services, and it helps to reflect the average person's daily activities. An example of a "check-in" looks like: check-in (3943172612597320) = {2537813697, ####, 3943172612597320, Sun Mar 13 23:02:12 +0800 2016, m, 121.5012107, 31.334979}. where 3943172612597320 is "status_id", 2537813697 is "user_id", #### is the "user_name" for privacy it is represented as "#", Sun Mar 13 23:02:12 +0800 2016 is "day, month, date, time and year), m is "gender" and 121.5012107, 31.334979 represents the geo-location of check-in.

Table 1. Shanghai dataset details.

Study Sample	
Total number of check-ins	503,521
Total number of processed check-ins	474,442
Total number of users	14,872
Total number of male users	7270
Total number of female users	7602
Range	March–April 2016
City	Shanghai, China

3.2. Methodology

Figure 2 presents the overall process flow for data collection and analysis. The location-based social media data analysis methodology is divided into two stages: data collection and storage, and analysis. The primary task of data collection and storage stage is to download a large number of Weibo data. During the data collection stage, the results came in separate files in JSON (JavaScript Object Notation) format by using a python-based [86] Weibo Application Programming Interface (API) [87]. JSON is considered one of the most widely used standard data formats and most major programming languages already have public reader and writer modules for it [88,89]. In order to be analyzed properly and stored in the database with the selected software, the dataset was transformed into one single file in CSV (Comma-Separated Values) format so all the check-ins could be listed regarding their publishing time. However, in the data analysis stage, the critical task was to extract and analyze the features (i.e., user ID, gender, date, time, and geo-location (longitude and latitude)) of data. The analysis phase used statistical analysis (probabilities of check-ins made), and data visualization by using ArcGIS (www.arcgis.com) to produce density maps and trends.

**Figure 2.** The process flow for data collection and analysis.

Weibo data is pre-processed to avoid noises, invalid records, and fake users. Data is pre-processed and filtered using the following criteria: (i) the location of check-in is in Shanghai based on the geographical location; (ii) users have at least checked-in twice during a month, and users who have only one check-in record are considered invalid; (iii) each check-in must have the following

information available: user ID, date, time, gender, and geo-location (longitude and latitude). After the pre-processing (noises, invalid records, and fake users) of 503,521 anonymized check-in records, 474,442 check-in records related to the study area were acquired for April–May 2016.

Figure 3 presents a general framework for check-in behavior analysis. Before detecting the hot-spot status of gender differences and check-in behavior, we analyzed the overall spatial patterns by using the kernel density estimation (KDE) and ArcGIS for estimating the density function.

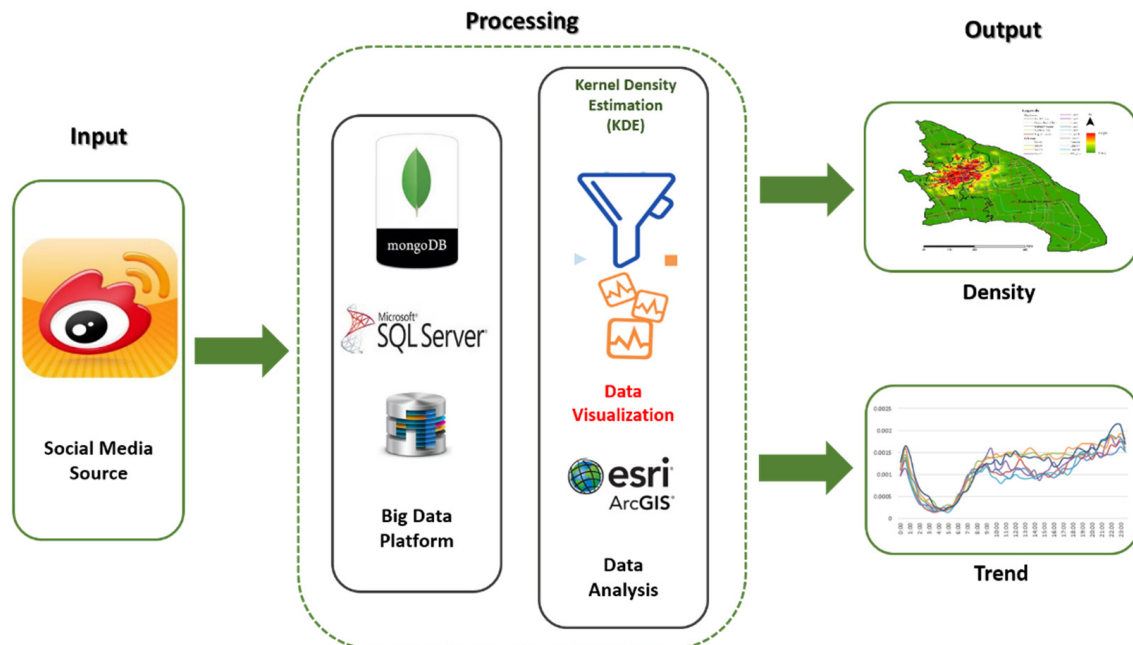


Figure 3. A general framework for check-in behavior analysis.

KDE is a spatial analysis technique that accounts for the location of features (i.e., destination, time) relative to each other and is an emerging spatial tool that has previously been applied [47,60,61,90–92] to the examination of various aspects of social media data analysis such as human activity and mobility patterns [93], check-in behavior [94], defining city boundaries [95,96], and point of interest recommendation [97]. Moreover, it examines the distribution of destinations in neighborhoods, enables researchers to see where destinations are sparsely distributed, and where they are more intensely distributed. Finally, it attempts to produce a smooth density surface of spatial point events in geographic space [98], the aim is to construct a smooth surface that represents the density of the point group. The algorithm is operated by setting the search scope (window). The weight of each grid unit is given from the central grid of the window to an outward grid according to the principle of an anti-distance weight. Then the kernel density value of the central grid is the sum of the product of all kernel density values and weights in the window and is defined as:

$$f(x) = \sum_{i=1}^n \frac{1}{\pi h^2} k\left(\frac{D_{is}}{h}\right) \quad (1)$$

Let $f(x)$ be the KDE function at location “ x ”, where “ h ” is the bandwidth. “ D ” is the distance from the point “ i ” to a specific location “ s ,” and “ k ” represents the Gaussian kernel function. In the KDE, bandwidth is considered an important parameter. If the bandwidth is too large, then the point density surface will become too smooth, while being too small will change point density distribution abruptly [99]. Therefore, the optimal bandwidth is determined by repeatedly setting the bandwidth and comparing the smoothness of the point density surface. In this analysis, we partitioned the output cells as 200×200 m areas. The cell was partitioned to provide a precise way to measure and compare urban intensity; this method is also applied in many spatial analysis studies [100]. Since the focus

of this study is to observe the check-in behavior in Shanghai towards using Weibo and analyze the check-in behavior for males and females to observe the number of males, females, and total Weibo users in Shanghai, we also calculated the frequency of check-ins over the period (time slices) and during weekdays and the weekend.

4. Results

For the analysis, we used Weibo geo-location check-in data set, used KDE to analyze the spatial distribution of check-in data, and used ArcGIS for visualization. As shown in Figure 3, the areas shaded in red indicate a higher density of people, higher activity frequency, and higher concentration of social media use. It is no surprise that the city center has large clusters of activity and Figure 4 shows the overall check-in density in Shanghai over a period.

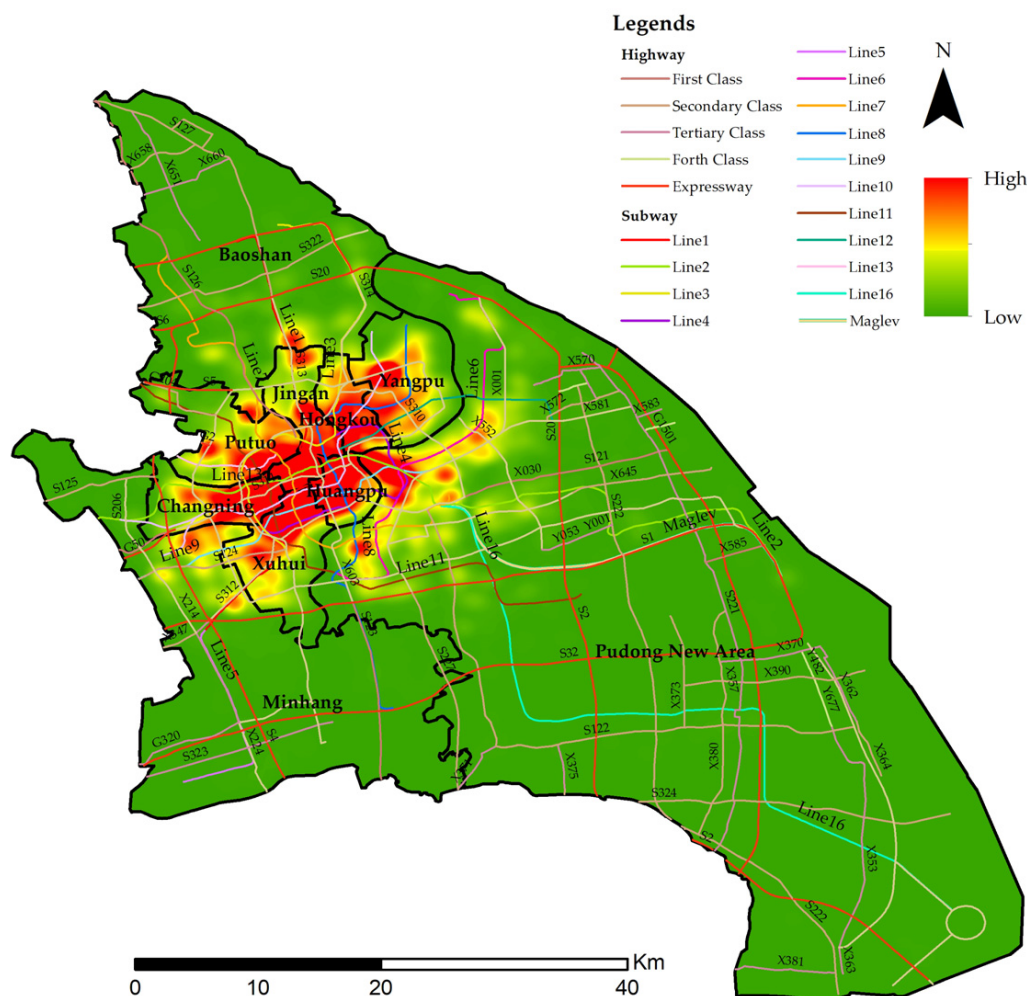


Figure 4. Check-in density in Shanghai.

Figure 5 shows the difference of users' check-in density during weekdays and the weekend in Shanghai; it is important to remember that the ratio between the total number of days for weekdays and the weekend is 5:2.

For analysis purposes, we sliced the time duration into three intervals 00:00–08:59, 09:00–16:59, and 17:00–23:59 to observe the users' check-in behavior during different hours of the day. In Figure 6, it can be observed that the check-in activity increased during the time interval 09:00–16:59 and 17:00–23:59.

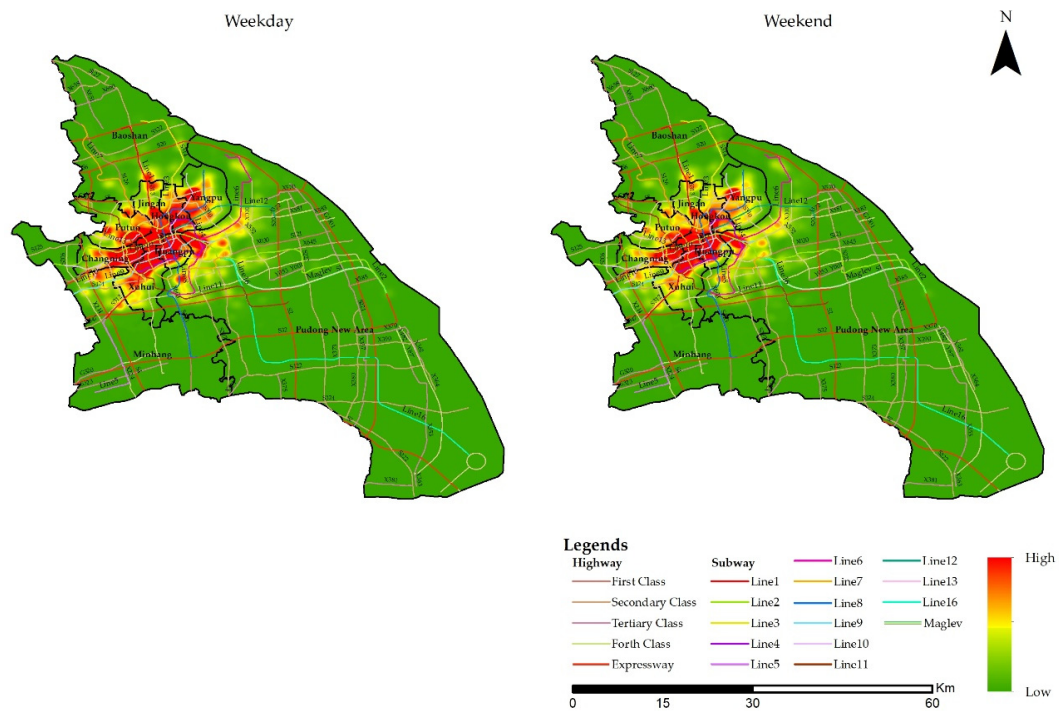


Figure 5. Check-in density during weekdays and the weekend.

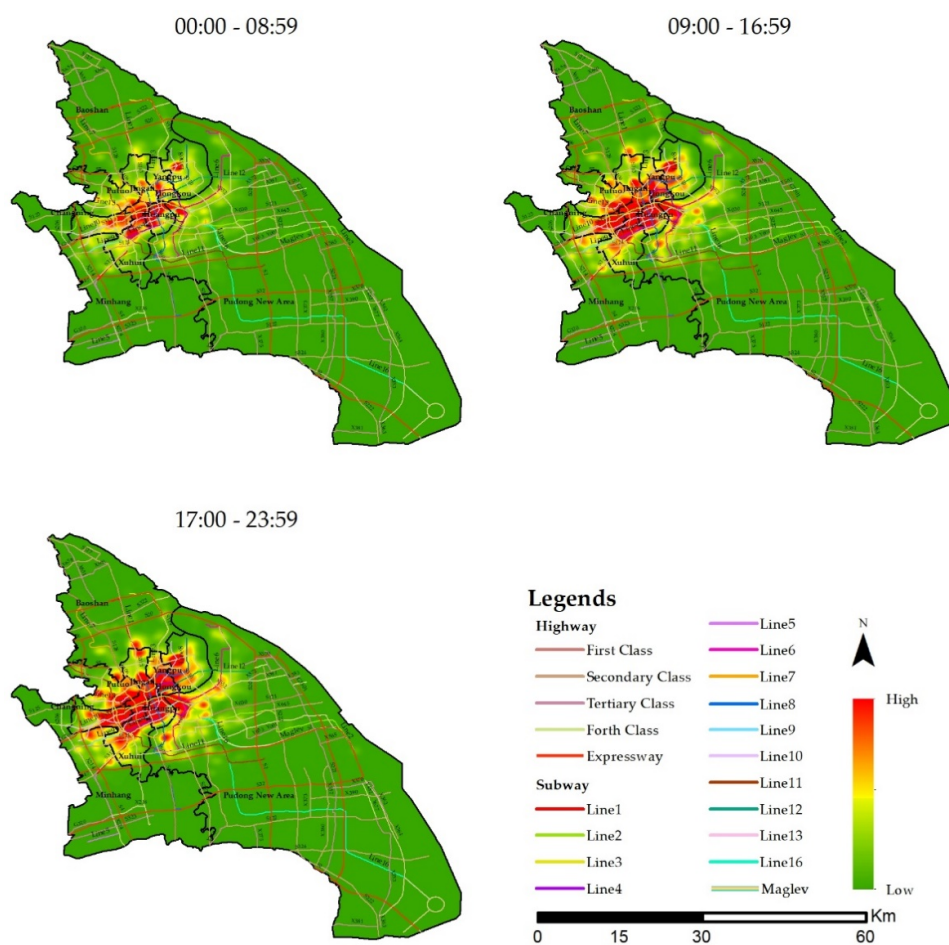


Figure 6. Check-in density for different hours of the day.

To investigate the check-in frequency and behavior, we analyzed the data regarding gender (male and female). In Figure 7a, it can be observed that females were more inclined toward using Weibo in Shanghai as compared to males during a week-long period. Moreover, Figure 7b also shows females were more inclined toward using Weibo as compared to male during weekdays and during the weekend.

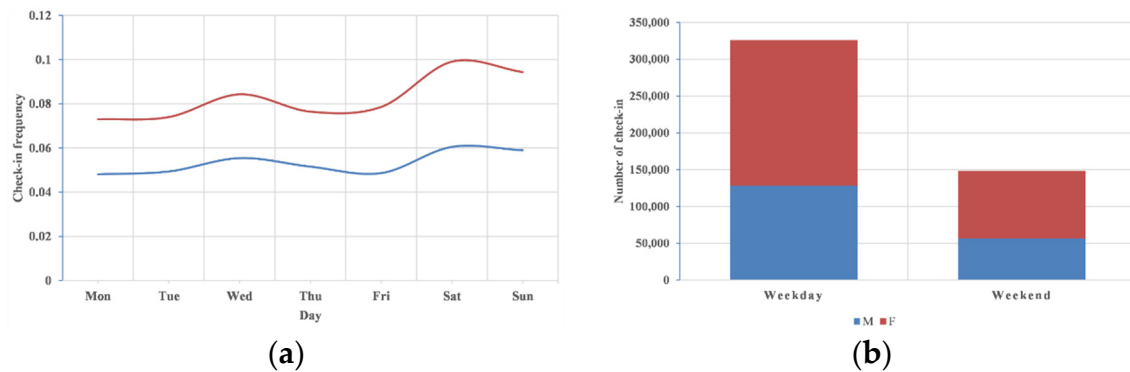


Figure 7. Check-in frequency of males and females during (a) a week, and (b) weekday and weekend.

Figure 8 represents the daily check-in behavior of males and females. It can be observed from Figure 8a that frequency for a female was comparatively high relative to males. From Figure 8a,b, it can also be observed that the frequency of females was also high as compared to males during weekdays and the weekend. Moreover, high frequency of use was observed during the time interval 18:00–22:30 by males and females during weekdays and the weekend.

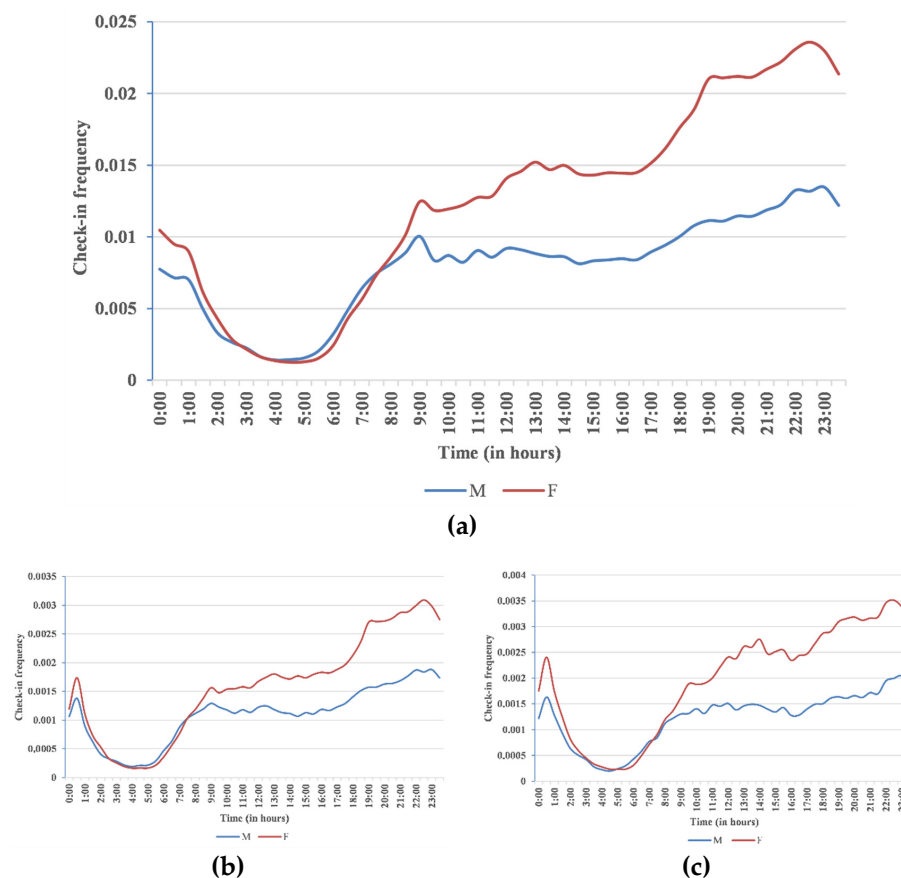


Figure 8. Check-in frequency of male and female during (a) a day, (b) a weekday, and (c) a weekend.

To further investigate and study the check-in frequency and behavior, we analyzed the data from three different research angles: weekday and weekend, time intervals, and days to observe the trends of males and females. Figure 9 shows the check-in density of males during April–May 2016.

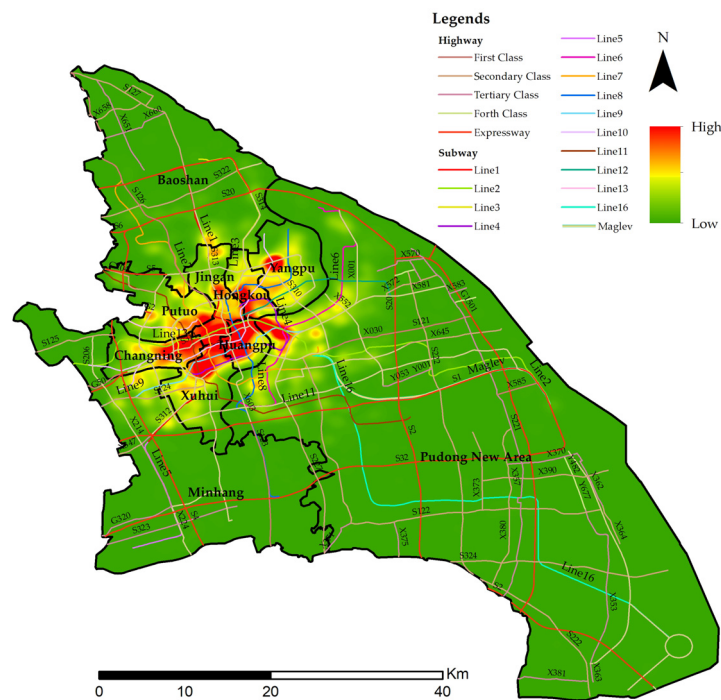


Figure 9. Check-in density of males.

Figure 10 shows the difference of male check-in density during the weekday and weekend in Shanghai. An increasing trend of check-in density can be observed during weekday as compared to the weekend.

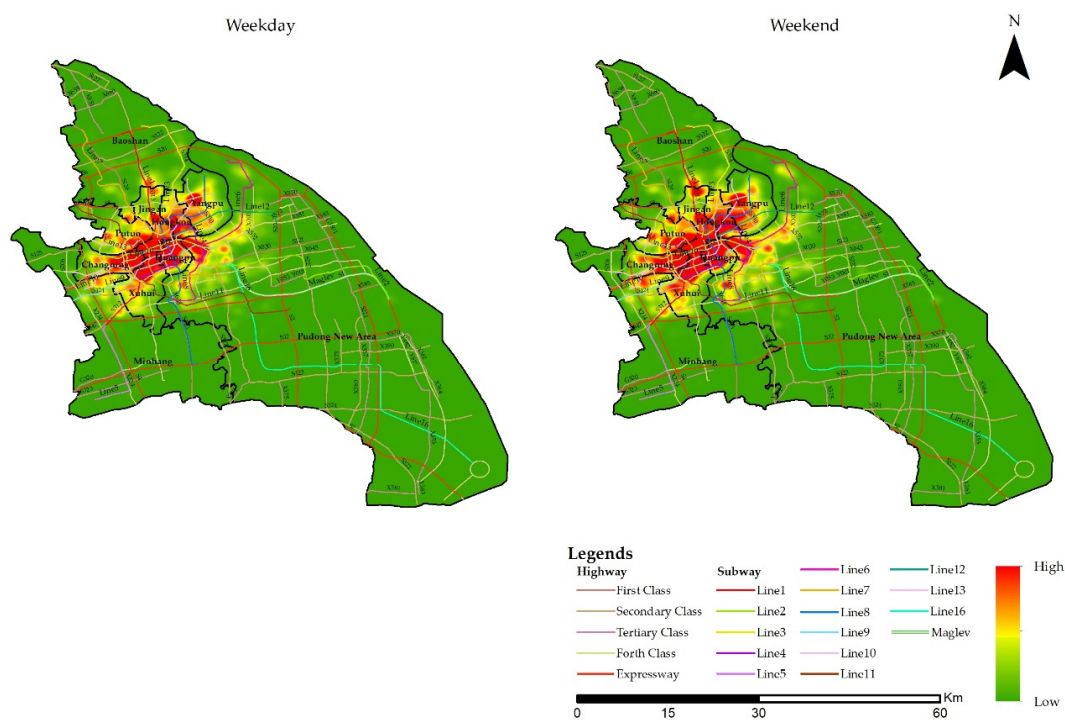


Figure 10. Check-in density of males during weekday and the weekend.

Figure 11 represents the check-in frequency of males during weekday and weekend, and it can be observed that a slightly increasing trend of check-ins occurred during 07:00–16:00. Which indicates that during breakfast and lunch hours males preferred to use Weibo.

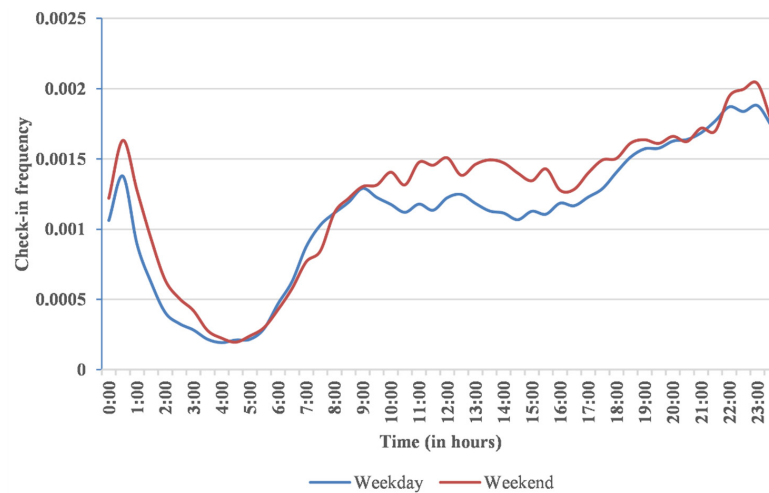


Figure 11. Check-in frequency of males during weekdays and the weekend.

For analysis purposes, we sliced the time duration into three intervals: 00:00–08:59, 09:00–16:59, and 17:00–23:59 to observed the male check-in behavior during different hours of the day. In Figure 12, it can be observed that check-in density by males is almost consistent but during the time interval 17:00–23:59 and 00:00–08:59, we can observe a slight increase in check-in density.

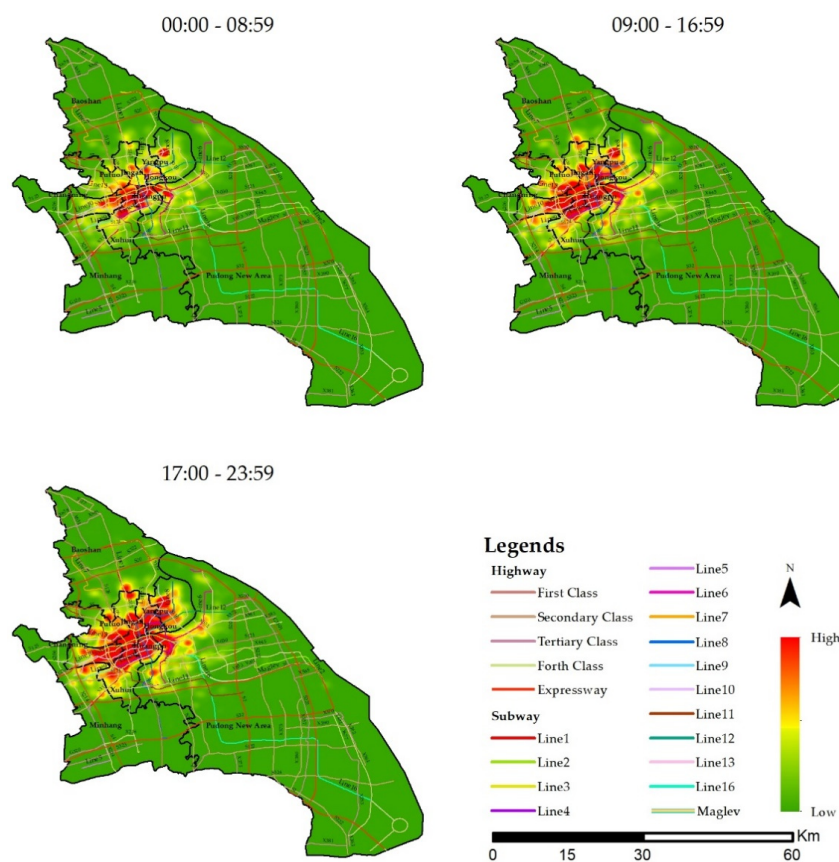


Figure 12. Check-in density of male for different hours of the day.

However, for further and in-depth analysis we observed the male check-in behavior over the course of a day. Figure 13 shows the detailed analysis during the whole week and we observe a notable increase in check-in behavior by males during the weekend. Moreover, it can be observed that during the weekdays (Monday–Friday), the frequency was consistent, but check-in behavior changed dramatically during Wednesday, which raises a new research question to explore the factors that influence this dramatic change in check-in behavior during Wednesday. This question will be addressed in the future by exploring the activities behavior analysis.

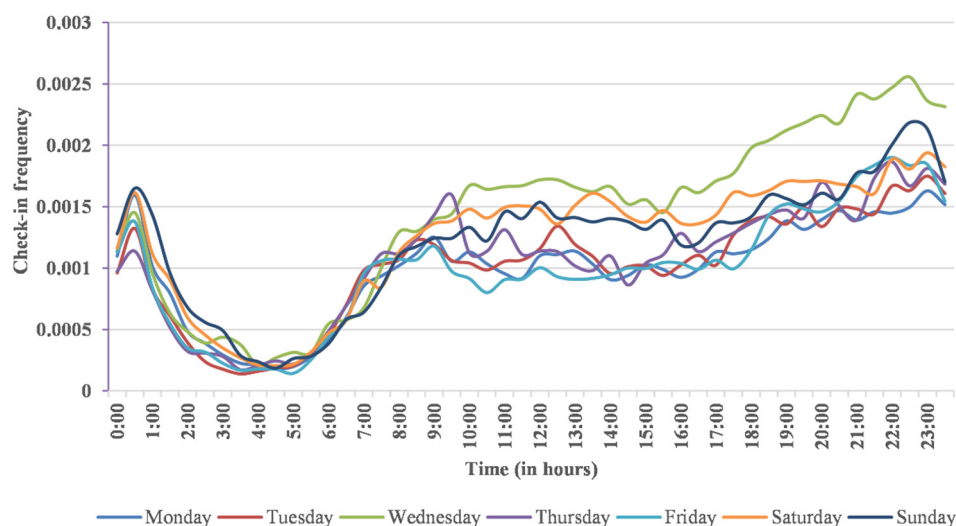


Figure 13. Check-in frequency of males during the course of a week.

Figure 14 shows the overall check-in density of females during April–May 2016.

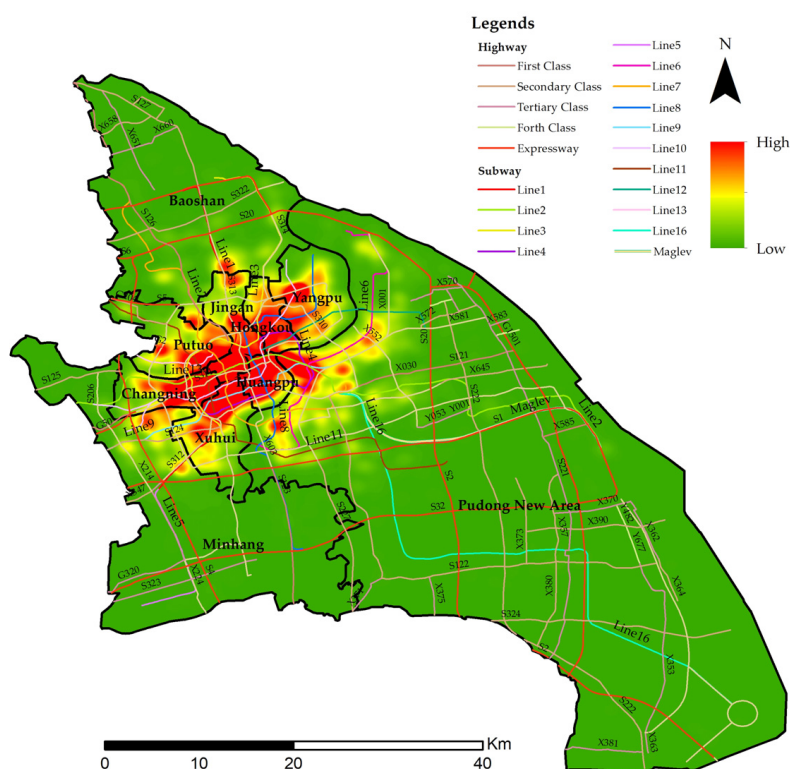


Figure 14. Check-in density of females.

Figure 15 shows the difference of female check-in density during weekdays and the weekend in Shanghai. It indicates the increasing trend of activity during weekdays as compared to the weekend, but on the other hand, Figure 16 shows an evident increasing trend of check-ins during noon and lunch hours by females as compared to males.

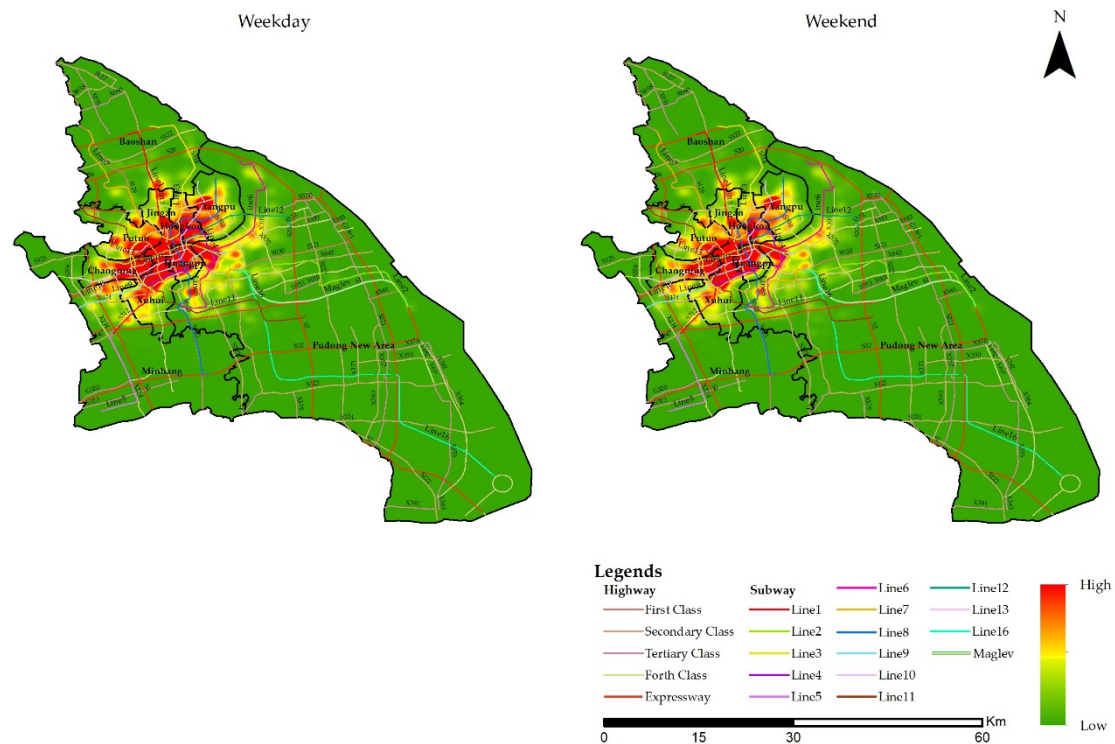


Figure 15. Check-in density of females during weekdays and the weekend.

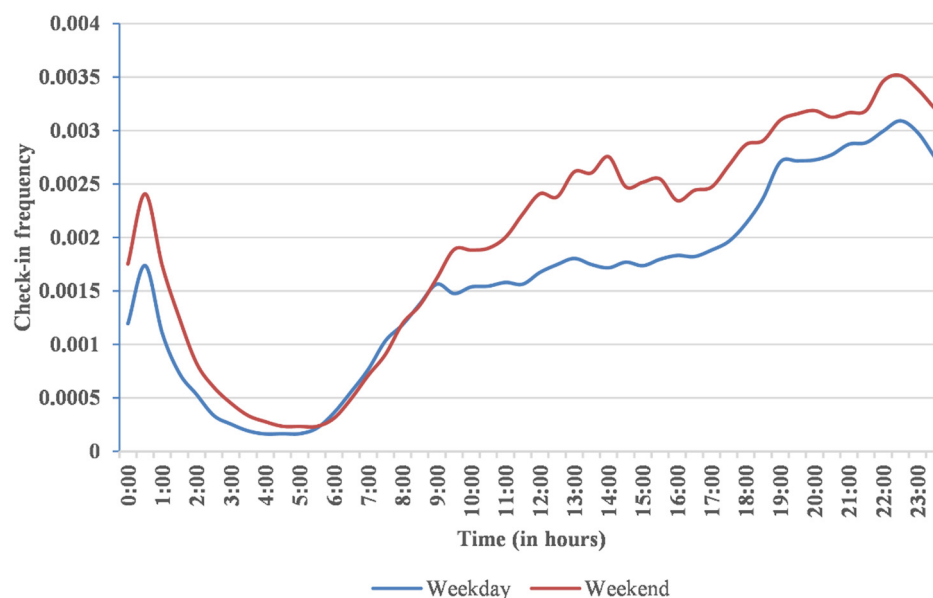


Figure 16. Check-in frequency of females during weekdays and the weekend.

Figure 17 represents check-in density during the time intervals 00:00–08:59, 09:00–16:59, and 17:00–23:59. From Figure 17b,c, increasing check-in density can be observed during the time intervals 09:00–16:00 and 17:00–23:59.

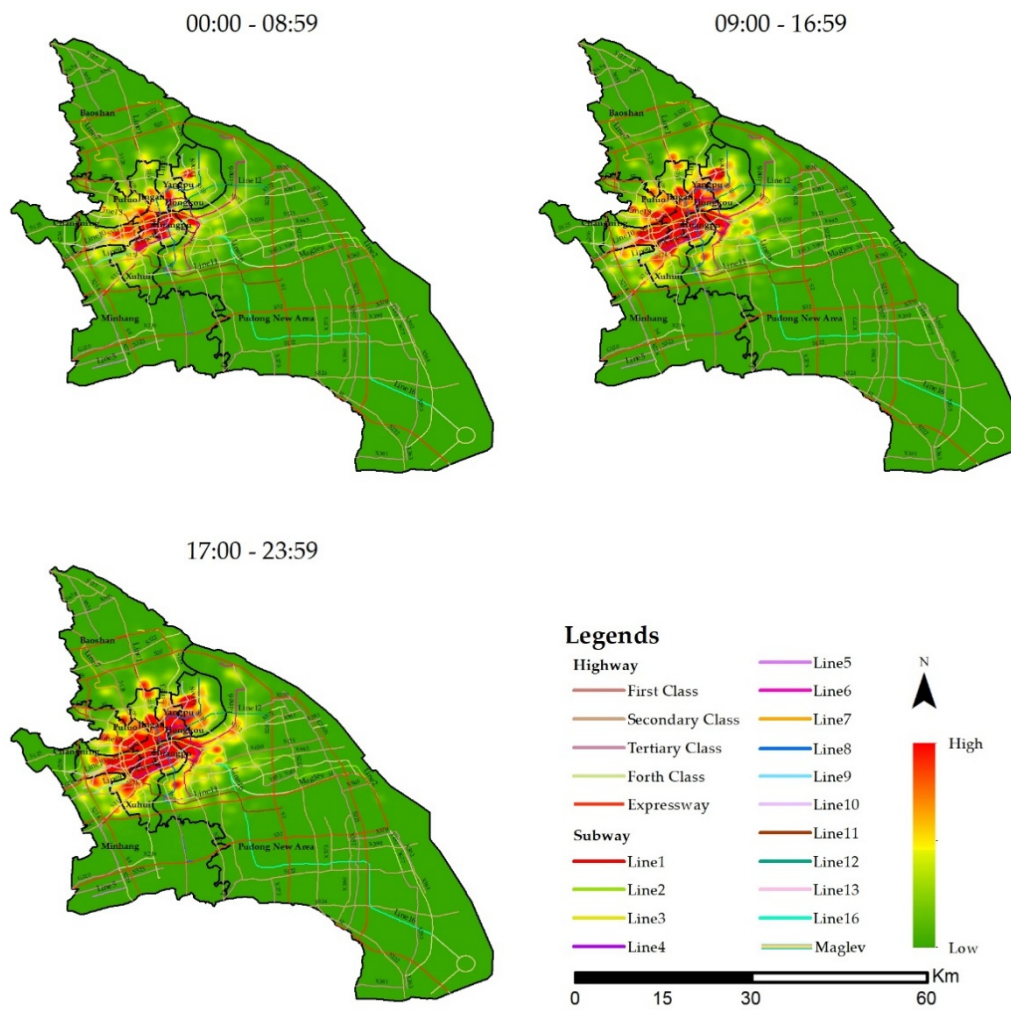


Figure 17. Check-in density of females for different hours of the day.

Moreover, in Figure 18, we observe consistent behavior during weekday excluding Wednesday, but as a result on the weekend, we observed an unusual increase in check-in behavior by females, as well as an increase in the trend after the time interval 17:00–23:59 on Friday.

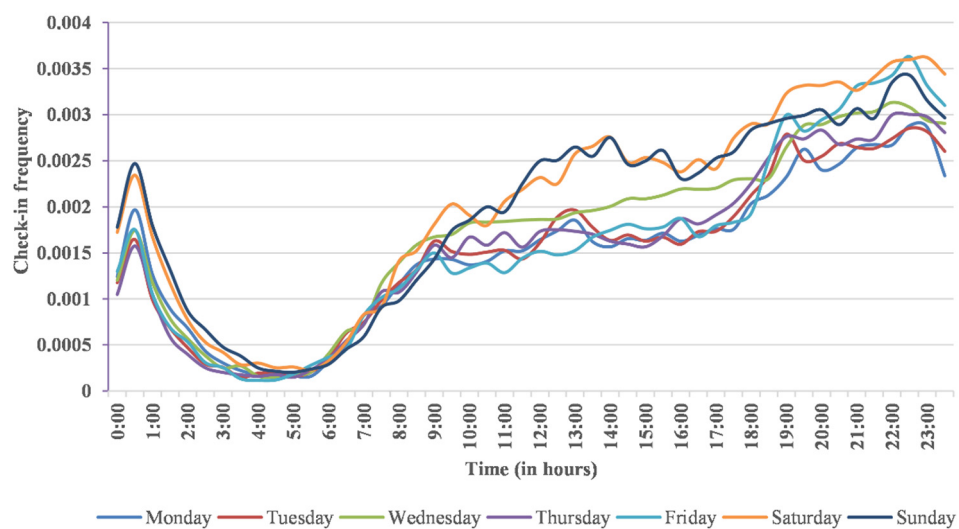


Figure 18. The check-in frequency of females during the course of a week.

From the above-reported results, it can be observed that the city center of Shanghai shows high check-in density during April–May 2016. Additionally, it can be observed that a higher check-in density is observed near the subway and highways, which is considered to be due to easy access of transportation services. Therefore, by observing the check-in data, it is likely that females prefer to use Weibo more for recreational, business, and personal purposes rather than males.

In the current study, the LBSN data from Weibo was used to draw inferences about male and female behaviors, which can be used in future studies as a valuable indicator of gender differences in Shanghai, China.

5. Conclusions

The current study presented the in-depth empirical investigation of check-in behavior by males and females using density maps and trends by using ArcGIS. We investigated the check-in behavior from different angles: the difference in gender, during weekday and weekend, daily and hourly patterns. In the results, we observed high rates of social media usage from female users and a difference in check-in behavior during weekdays and the weekend in Shanghai by both males and females.

On one hand, LBSN check-in data have some advantages, such as its low cost and high spatial precision. However, check-in data also has some limitations, such as bias of gender, a low sampling frequency, and bias of location category. In summary, LBSN data is more likely to be a supplement to, than a substitute for, traditional data sources.

Based on the results of the current study, LBSN data has the potential to provide a new outlook as a supplement to observe gender differences and intensity of check-ins during weekdays and the weekend, and can help policymakers to define policies regarding the supply of services within a city. In the future, we plan to use LBSN data as a means to investigate check-in behavior related to activities (i.e., food and drink, residence, shopping, work, and travel) within the city in space and time.

Author Contributions: M.R. and W.W. conceived the research; M.R. designed the research and performed the simulations. M.R. and W.W. wrote the article and proofread the article for language editing. All authors read and approved the final manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (61711530245) and the key project of Shanghai Science and Technology Commission (17511106802).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Glaeser, E.L.; Kolko, J.; Saiz, A. Consumer city. *J. Econ. Geogr.* **2001**, *1*, 27–50. [[CrossRef](#)]
2. Carlino, G.A.; Mills, E.S. The determinants of county growth. *J. Reg. Sci.* **1987**, *27*, 39–54. [[CrossRef](#)] [[PubMed](#)]
3. Chen, C.; Gong, H.; Paaswell, R. Role of the built environment on mode choice decisions: Additional evidence on the impact of density. *Transportation* **2008**, *35*, 285–299. [[CrossRef](#)]
4. Batty, M. The size, scale, and shape of cities. *Science* **2008**, *319*, 769–771. [[CrossRef](#)] [[PubMed](#)]
5. Fraser, D.J.; Coon, T.; Prince, M.R.; Dion, R.; Bernatchez, L. Integrating traditional and evolutionary knowledge in biodiversity conservation: A population level case study. *Ecol. Soc.* **2006**, *11*, 4. [[CrossRef](#)]
6. Charalabidis, Y.; Loukis, E. Participative public policy making through multiple social media platforms utilization. *Int. J. Electron. Government Res.* **2012**, *8*, 78–97. [[CrossRef](#)]
7. Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. *J. Comput. Commun.* **2007**, *13*, 210–230. [[CrossRef](#)]
8. Symeonidis, P.; Ntempos, D.; Manolopoulos, Y. Location-based social networks. In *Recommender Systems for Location-Based Social Networks*; Springer: New York, NY, USA, 2014; pp. 35–48.
9. Gao, H.; Liu, H. Synthesis Lectures on Data Mining and Knowledge Discovery. In *Mining Human Mobility in Location-Based Social Networks*; Han, J., Getoor, L., Wang, W., Gehrke, J., Grossman, R., Eds.; Morgan & Claypool Publishers: San Rafael, CA, USA, 2015; Volume 7, pp. 1–115.
10. Facebook. Available online: <https://www.facebook.com/> (accessed on 30 August 2018).

11. Twitter. Available online: <https://twitter.com/> (accessed on 30 August 2018).
12. Weibo. Available online: <http://www.weibo.com> (accessed on 30 August 2018).
13. Lu, E.H.-C.; Chen, C.-Y.; Tseng, V.S. Personalized trip recommendation with multiple constraints by mining user check-in behaviors. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 6–9 November 2012; pp. 209–218.
14. De Mauro, A.; Greco, M.; Grimaldi, M. A formal definition of Big Data based on its essential features. *Library Review* **2016**, *65*, 122–135. [[CrossRef](#)]
15. Lin, X.; Lachlan, K.A.; Spence, P.R. Exploring extreme events on social media: A comparison of user reposting/retweeting behaviors on Twitter and Weibo. *Comput. Hum. Behav.* **2016**, *65*, 576–581. [[CrossRef](#)]
16. Zhao, J.; Wu, W.; Zhang, X.; Qiang, Y.; Liu, T.; Wu, L. A short-term trend prediction model of topic over Sina Weibo dataset. *J. Comb. Optim.* **2014**, *28*, 613–625. [[CrossRef](#)]
17. Chen, Z.; Liu, P.; Wang, X.; Gu, Y. Follow whom? Chinese users have different choice. *arXiv* **2012**, arXiv:1212.0167.
18. Miller, H.J.; Goodchild, M.F. Data-driven geography. *GeoJournal* **2015**, *80*, 449–461. [[CrossRef](#)]
19. López-Ornelas, E.; Abascal-Mena, R.; Zepeda-Hernández, S. Social Media Participation in Urban Planning: A New way to Interact and Take Decisions. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2017**, *42*, 59. [[CrossRef](#)]
20. Criado, J.I.; Sandoval-Almazan, R.; Gil-Garcia, J.R. Government innovation through social media. *Government Inf. Q.* **2013**, *30*, 319–326. [[CrossRef](#)]
21. Zheng, L.; Zheng, T. Innovation through social media in the public sector: Information and interactions. *Government Inf. Q.* **2014**, *31*, S106–S117. [[CrossRef](#)]
22. Sobaci, M.Z.; Karkin, N. The use of twitter by mayors in Turkey: Tweets for better public services? *Government Inf. Q.* **2013**, *30*, 417–425. [[CrossRef](#)]
23. Agostino, D. Using social media to engage citizens: A study of Italian municipalities. *Public Relat. Rev.* **2013**, *39*, 232–234. [[CrossRef](#)]
24. Graham, M.W.; Avery, E.J.; Park, S. The role of social media in local government crisis communications. *Public Relat. Rev.* **2015**, *41*, 386–394. [[CrossRef](#)]
25. Tursunbayeva, A.; Franco, M.; Pagliari, C. Use of social media for e-Government in the public health sector: A systematic review of published studies. *Government Inf. Q.* **2017**, *34*, 270–282. [[CrossRef](#)]
26. Sabrina. Sina Weibo User Demographics Analysis in 2013. Available online: <https://www.chinainternetwatch.com/5568/what-weibo-can-tell-you-about-chinese-netizens-part-1/> (accessed on 22 August 2018).
27. Thelwall, M. Social networks, gender, and friending: An analysis of MySpace member profiles. *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 1321–1330. [[CrossRef](#)]
28. Reed, P.J.; Khan, M.R.; Blumenstock, J. Observing gender dynamics and disparities with mobile phone metadata. In Proceedings of the Eighth International Conference on Information and Communication Technologies and Development, Ann Arbor, MI, USA, 3–6 June 2016.
29. Kung, K.S.; Greco, K.; Sobolevsky, S.; Ratti, C. Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS ONE* **2014**, *9*, e96180. [[CrossRef](#)] [[PubMed](#)]
30. Hoteit, S.; Secci, S.; Sobolevsky, S.; Ratti, C.; Pujolle, G. Estimating human trajectories and hotspots through mobile phone data. *Comput. Netw.* **2014**, *64*, 296–307. [[CrossRef](#)]
31. Hesse, B.W.; Moser, R.P.; Riley, W.T. From big data to knowledge in the social sciences. *Ann. Am. Acad. Political Soc. Sci.* **2015**, *659*, 16–32. [[CrossRef](#)] [[PubMed](#)]
32. Benson, V.; Saridakis, G.; Tennakoon, H. Information disclosure of social media users: Does control over personal information, user awareness and security notices matter? *Inf. Technol. People* **2015**, *28*, 426–441. [[CrossRef](#)]
33. Strater, K.; Richter, H. Examining privacy and disclosure in a social networking community. In Proceedings of the 3rd Symposium on Usable Privacy and Security, Pittsburgh, PA, USA, 18–20 July 2007; pp. 157–158.
34. Stefanone, M.A.; Huang, Y.C.; Lackaff, D. Negotiating Social Belonging: Online, Offline, and In-Between. In Proceedings of the 44th Hawaii International Conference on System Science, Kauai, HI, USA, 4–7 January 2011; pp. 1–10.

35. Benevenuto, F.; Rodrigues, T.; Cha, M.; Almeida, V. Characterizing user behavior in online social networks. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, Chicago, IL, USA, 4–6 November 2009; pp. 49–62.
36. Scellato, S.; Noulas, A.; Lambiotte, R.; Mascolo, C. Socio-spatial properties of online location-based social networks. In Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, 17–21 July 2011; pp. 329–336.
37. Noulas, A.; Scellato, S.; Mascolo, C.; Pontil, M. An empirical study of geographic user activity patterns in foursquare. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
38. Ruggles, S. Big microdata for population research. *Demography* **2014**, *51*, 287–297. [[CrossRef](#)] [[PubMed](#)]
39. Chorley, M.J.; Whitaker, R.M.; Allen, S.M. Personality and location-based social networks. *Comput. Hum. Behav.* **2015**, *46*, 45–56. [[CrossRef](#)]
40. Maia, M.; Almeida, J.; Almeida, V. Identifying user behavior in online social networks. In Proceedings of the 1st Workshop on Social Network Systems, Glasgow, Scotland, 1 April 2008; pp. 1–6.
41. Pucci, P.; Manfredini, F.; Tagliolato, P. *Mapping Urban Practices through Mobile Phone Data*; Springer: New York, NY, USA, 2015.
42. Hong, I. Spatial analysis of location-based social networks in seoul, korea. *J. Geogr. Inf. Syst.* **2015**, *7*, 259. [[CrossRef](#)]
43. Jin, L.; Chen, Y.; Wang, T.; Hui, P.; Vasilakos, A.V. Understanding user behavior in online social networks: A survey. *IEEE Commun. Mag.* **2013**, *51*, 144–150.
44. Gyarmati, L.; Trinh, T.A. Measuring user behavior in online social networks. *IEEE netw.* **2010**, *24*, 26–31. [[CrossRef](#)]
45. Preoțiuc-Pietro, D.; Cohn, T. Mining user behaviours: A study of check-in patterns in location based social networks. In Proceedings of the 5th Annual ACM Web Science Conference, Paris, France, 2–4 May 2013; pp. 306–315.
46. Shen, Y.; Karimi, K. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities* **2016**, *55*, 9–21. [[CrossRef](#)]
47. Wu, C.; Ye, X.; Ren, F.; Wan, Y.; Ning, P.; Du, Q. Spatial and social media data analytics of housing prices in Shenzhen, China. *PLoS ONE* **2016**, *11*, e0164553. [[CrossRef](#)] [[PubMed](#)]
48. Luarn, P.; Yang, J.-C.; Chiu, Y.-P. Why people check in to social network sites. *Int. J. Electron. Commer.* **2015**, *19*, 21–46. [[CrossRef](#)]
49. Wang, S.S.; Stefanone, M.A. Showing Off? Human Mobility and the Interplay of Traits, Self-Disclosure, and Facebook Check-Ins. *Soc. Sci. Comput. Rev.* **2013**, *31*, 437–457. [[CrossRef](#)]
50. Smith, A. Why Americans Use Social Media. Available online: <http://www.pewinternet.org/2011/11/15/why-americans-use-social-media/> (accessed on 30 September 2018).
51. Zhang, L.; Pentina, I. Motivations and usage patterns of Weibo. *Cyberpsychol. Behav. Soc. Netw.* **2012**, *15*, 312–317. [[CrossRef](#)] [[PubMed](#)]
52. Li-Barber, K.T. Self-disclosure and student satisfaction with Facebook. *Comput. Hum. Behav.* **2012**, *28*, 624–630.
53. Pentina, I.; Basmanova, O.; Zhang, L. A cross-national study of Twitter users' motivations and continuance intentions. *J. Marketing Commun.* **2016**, *22*, 36–55. [[CrossRef](#)]
54. Shao, W.; Ross, M.; Grace, D. Developing a motivation-based segmentation typology of Facebook users. *Marketing Intell. Plann.* **2015**, *33*, 1071–1086. [[CrossRef](#)]
55. Kim, H.-S. A Study on Use Motivation of SNS and Communication Behavior. *J. Korea Acad.-Ind. Cooper. Soc.* **2012**, *13*, 548–553. [[CrossRef](#)]
56. Muscanell, N.L.; Guadagno, R.E. Make new friends or keep the old: Gender and personality differences in social networking use. *Comput. Hum. Behav.* **2012**, *28*, 107–112. [[CrossRef](#)]
57. Chun, M.-h. The affective/cognitive involvement and satisfaction according to the usage motivations of social network services. *Manage. Inf. Syst. Rev.* **2012**, *31*.
58. Hwang, H.S.; Choi, E.K. Exploring gender differences in motivations for using sina weibo. *KSII Trans. Internet Inf. Syst.* **2016**, *10*, 1429–1441.
59. Blumenstock, J.E.; Gillick, D.; Eagle, N. Who's calling? Demographics of mobile phone use in Rwanda. *Transportation* **2010**, *32*, 2–5.

60. Rizwan, M.; Wanggen, W.; Cervantes, O.; Gwiazdzinski, L. Using Location-Based Social Media Data to Observe Check-In Behavior and Gender Difference: Bringing Weibo Data into Play. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 196. [\[CrossRef\]](#)
61. Rizwan, M.; Mahmood, S.; Wanggen, W.; Ali, S. Location based social media data analysis for observing check-in behavior and city rhythm in Shanghai. In Proceedings of the 4th International Conference on Smart and Sustainable City (ICSSC 2017), Shanghai, China, 5–6 June 2017.
62. Lei, C.; Zhang, A.; Qi, Q.; Su, H.; Wang, J. Spatial-Temporal Analysis of Human Dynamics on Urban Land Use Patterns Using Social Media Data by Gender. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 358. [\[CrossRef\]](#)
63. Zheng, Y.; Zhang, L.; Xie, X.; Ma, W.-Y. Mining correlation between locations using human location history. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 4–6 November 2009; pp. 472–475.
64. Comito, C.; Falcone, D.; Talia, D. Mining human mobility patterns from social geo-tagged data. *Pervasive Mob. Comput.* **2016**, *33*, 91–107. [\[CrossRef\]](#)
65. Humphreys, L. Mobile social networks and urban public space. *New Media Soc.* **2010**, *12*, 763–778. [\[CrossRef\]](#)
66. Roche, S. Geographic Information Science I: Why does a smart city need to be spatially enabled? *Prog. Hum. Geogr.* **2014**, *38*, 703–711. [\[CrossRef\]](#)
67. Anthopoulos, L.G.; Vakali, A. Urban Planning and Smart Cities: Interrelations and Reciprocities. In *The Future Internet. FIA 2012. Lecture Notes in Computer Science*; Álvarez, F., Cleary, F., Daras, P., Domingue, J., Galis, A., Garcia, A., Gavras, A., Karnourskos, S., Krco, S., Li, M.-S., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7281, pp. 178–189.
68. Cui, L.; Shi, J. Urbanization and its environmental effects in Shanghai, China. *Urban Clim.* **2012**, *2*, 1–15. [\[CrossRef\]](#)
69. Han, B.; Cook, P.; Baldwin, T. Geolocation prediction in social media data by finding location indicative words. In Proceedings of the COLING 2012, Mumbai, India, December 2012; pp. 1045–1062.
70. Schoen, H.; Gayo-Avello, D.; Takis Metaxas, P.; Mustafaraj, E.; Strohmaier, M.; Gloor, P. The power of prediction with social media. *Internet Res.* **2013**, *23*, 528–543. [\[CrossRef\]](#)
71. Backstrom, L.; Sun, E.; Marlow, C. Find me if you can: Improving geographical prediction with social and spatial proximity. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, CA, USA, 26–30 April 2010; pp. 61–70.
72. Sun, Y.; Li, M. Investigation of travel and activity patterns using location-based social network data: A case study of active mobile social media users. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 1512–1529. [\[CrossRef\]](#)
73. Gu, Z.; Zhang, Y.; Chen, Y.; Chang, X. Analysis of attraction features of tourism destinations in a mega-city based on check-in data mining—A case study of ShenZhen, China. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 210. [\[CrossRef\]](#)
74. Yin, J.; Lampert, A.; Cameron, M.; Robinson, B.; Power, R. Using social media to enhance emergency situation awareness. *IEEE Intell. Syst.* **2012**, *27*, 52–59. [\[CrossRef\]](#)
75. Yates, D.; Paquette, S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *Int. J. Inf. Manage.* **2011**, *31*, 6–13. [\[CrossRef\]](#)
76. Cervone, G.; Schnebele, E.; Waters, N.; Moccaldi, M.; Sicignano, R. Using Social Media and Satellite Data for Damage Assessment in Urban Areas During Emergencies. In *Seeing Cities through Big Data*; Thakuriah, P., Tilahun, N., Zellner, M., Eds.; Springer: Cham, Switzerland, 2017; pp. 443–457.
77. Wang, Y.; Wang, T.; Ye, X.; Zhu, J.; Lee, J. Using social media for emergency response and urban sustainability: A case study of the 2012 Beijing rainstorm. *Sustainability* **2015**, *8*, 25. [\[CrossRef\]](#)
78. Vikat, A.; Jones, C. *Indicators of Gender Equality*; UNECE: Geneva, Switzerland, 2014.
79. O'Dorchai, S.; Meulders, D.; Crippa, F.; Margherita, A. *She Figures 2009—Statistics and Indicators on Gender Equality in Science*; Publications Office of the European Union: Luxembourg, Luxembourg, 2009.
80. Li, J.; Fang, W.; Wang, T.; Qureshi, S.; Alatalo, J.M.; Bai, Y. Correlations between Socioeconomic Drivers and Indicators of Urban Expansion: Evidence from the Heavily Urbanised Shanghai Metropolitan Area, China. *Sustainability* **2017**, *9*, 1199. [\[CrossRef\]](#)
81. Guo, R. *Regional China: A Business and Economic Handbook by Rongxing Guo*; Palgrave Macmillan UK: New York, NY, USA, 2013.
82. Gu, X.; Tao, S.; Dai, B. Spatial accessibility of country parks in Shanghai, China. *Urban For. Urban Greening* **2017**, *27*, 373–382. [\[CrossRef\]](#)

83. Xiong, X.; Jin, C.; Chen, H.; Luo, L. Using the Fusion Proximal Area Method and Gravity Method to Identify Areas with Physician Shortages. *PLoS ONE* **2016**, *11*, e0163504. [CrossRef] [PubMed]
84. Shen, J.; Kee, G. Shanghai: Urban Development and Regional Integration Through Mega Projects. In *Development and Planning in Seven Major Coastal Cities in Southern and Eastern China*; Springer: Cham, Switzerland, 2017; pp. 119–151.
85. Shen, J.; Kee, G. *Development and Planning in Seven Major Coastal Cities in Southern and Eastern China*; Springer: Cham, Switzerland, 2016.
86. Python. Available online: <https://www.python.org/> (accessed on 30 July 2018).
87. Weibo API. Available online: <http://open.weibo.com/wiki/API> (accessed on 30 August 2018).
88. Fernandes, R.; D'Souza, R. Analysis of product Twitter data through opinion mining. In Proceedings of the 2016 IEEE Annual India Conference (INDICON), Bangalore, India, 16–18 December 2016; pp. 1–5.
89. Batrinca, B.; Treleaven, P.C. Social media analytics: A survey of techniques, tools and platforms. *AI Soc.* **2015**, *30*, 89–116. [CrossRef]
90. Lichman, M.; Smyth, P. Modeling human location data with mixtures of kernel densities. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 35–44.
91. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC press: Boca Raton, FL, USA, 1986.
92. King, T.L.; Bentley, R.J.; Thornton, L.E.; Kavanagh, A.M. Using kernel density estimation to understand the influence of neighbourhood destinations on BMI. *BMJ Open* **2016**, *6*, e008878. [CrossRef] [PubMed]
93. Hasan, S.; Zhan, X.; Ukkusuri, S.V. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago, IL, USA, 11 August 2013.
94. Li, L.; Goodchild, M.F.; Xu, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 61–77. [CrossRef]
95. Sun, Y.; Fan, H.; Li, M.; Zipf, A. Identifying the city center using human travel flows generated from location-based social networking data. *Environ. Plann. B Plann. Des.* **2016**, *43*, 480–498. [CrossRef]
96. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering urban functional zones using latent activity trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *27*, 712–725. [CrossRef]
97. Li, H.; Ge, Y.; Hong, R.; Zhu, H. Point-of-interest recommendations: Learning potential check-ins from friends. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 975–984.
98. Xie, Z.; Yan, J. Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst.* **2008**, *32*, 396–406. [CrossRef]
99. Zhang, Q.; Jin, X.; Zhang, Z.; Zhang, Z.; Liu, Z. An Evaluation Method for Spatial Distribution Uniformity of Plane Form Error for Precision Assembly. *Procedia CIRP* **2007**, *76*, 59–62. [CrossRef]
100. Yoon, H.; Srinivasan, S. Are they well situated? Spatial analysis of privately owned public space, Manhattan, New York City. *Urban Affairs Rev.* **2015**, *51*, 358–380. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).