Using Location-Based Social Media Data to Observe Check-In Behavior and Gender Difference: Bringing Weibo Data into Play

Muhammad Rizwan [1,2,*]**, Wanggen Wan** [1,2]**, Ofelia Cervantes** [3] **and Luc Gwiazdzinski** [4]

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
3   Computing, Electronics and Mechatronics Department, Universidad de las Américas Puebla, Puebla 72810,
    Mexico; ofelia.cervantes@udlap.mx
4   Institut de Géographie Alpine (IGA), Université Grenoble Alpes, 38100 Grenoble, France;
    lucmarcg@gmail.com
*   Correspondence: rizwan@shu.edu.cn; Tel.: +86-131-220-98748

check for
updates

Abstract: Population density and distribution of services represents the growth and demographic shift of the cities. For urban planners, population density and check-in behavior in space and time are vital factors for planning and development of sustainable cities. Location-based social network (LBSN) data seems to be a complement to many traditional methods (i.e., survey, census) and is used to study check-in behavior, human mobility, activity analysis, and social issues within a city. This check-in phenomenon of sharing location, activities, and time by users has encouraged this research on gender difference and frequency of using LBSN. Therefore, in this study, we investigate the check-in behavior of Chinese microblog Sina Weibo (referred as "Weibo") in 10 districts of Shanghai, China, for which we observe the gender difference and their frequency of use over a period. The mentioned districts were spatially analyzed for check-in spots by kernel density estimation (KDE) using ArcGIS. Furthermore, our results reveal that female users have a high rate of social media use, and significant difference is observed in check-in behavior during weekdays and weekends in the studied districts of Shanghai. Increase in check-ins is observed during the night as compared to the morning. From the results, it can be assumed that LBSN data can be helpful to observe gender difference.

Keywords: big data; social network; lbsn; check-in; gender difference

1. Introduction

Personal behavior and characteristics are intimately intertwined with city planning and human mobility [1] although, in past, many traditional methods (i.e., survey, census) are used to collect data about human mobility and population density, but these traditional methods are expensive and require more processing time, produce sparse data and not that effective in policymaking.

With the introduction of LBSN's (i.e., Weibo [2], Facebook [3], Twitter [4]), users can share their location as well as the activity (referred as "check-in" [5]). Sharing check-ins allows users to announce and discuss places they visit (e.g., eating at local restaurants, shopping, visiting popular area) as part of their social interaction online. This check-in phenomenon and fast sharing of information have 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 [6]. These activities generate an enormous amount of users data (also referred "Big Data" [7]) based on human mobility. Despite some limitations on representing check-in behavior, e.g., bias of gender, a low sampling frequency, and bias

of location category, check-in data has the ability to uncover check-in behavior within a city. Compared to the aforementioned traditional methods, LBSN data are highly available and low cost. Moreover, this data contains rich information about geolocation [8], which can be used to study check-in behavior. Thus, geo-location data offers new dimensions toward studying check-in behaviors and helps to create new techniques and approaches to analyze LBSN data. Moreover, it seems that LBSN data can be a supplement to than a substitute of traditional data sources for policy making [9]. Therefore, LBSN data can be considered as a supplement while taking policy decision 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 [10–16].

In this research, we reconnoiter 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. We will explore check-in behavior in 10 districts (Baoshan, Changning, Hongkou, Huangpu, Jingan, Minhang, Pudong New Area, Putuo, Xuhui, and Yangpu) of Shanghai, China, which are interconnected to the boundaries of the city center. We discuss an empirical exploration using Weibo (launched by Sina Corporation on 14 August 2009) dataset, which is a dominant social media site in China. Since each Weibo account carries information about the gender of the user, we can differentiate between LBSN usage behavior by males and females. Furthermore, we consider LBSN data can be helpful to observe check-in frequencies during weekday and weekend.

The rest of the paper is organized as follows. Section 2 overviews related works. Section 3 describes the study area and data set used in the current study. Section 4 presents the methodology. Section 5 presents the results and discussion for the experimental results performed on dataset. Finally, Section 6 concludes the paper and proposes some further research issues.

2. Related Work

Studying people's behavior toward services has long been constrained to analyze traditional datasets due to enhanced capabilities of capturing, analyzing, and processing geo-location data, and the field of spatial analysis has blossomed [17]. The origin of social networks lies in the early 1990s with simple communication mechanism to meet people over the internet, where people could exchange ideas. The term "social network site" (SNS) refers to web-based services. It gives people three significant capabilities: (1) to construct a public or semi-public profile, (2) to identify a list of other users with whom a connection shared, and (3) to view and track individual connections and those made by others within the system [18].

When SNSs first emerged, they were only accessible through personal computers [19]. However, recent technological advancements of "smart" mobile devices have allowed users to access their social network accounts in fixed as well as mobile stations on the move. While users have the option to access, communicate, and exchange information on SNSs via their personal computer [20], the options to access SNSs on smartphones has allowed them to easily and conveniently communicate with their "friends" at any time, anywhere [21]. As mobile development continues to progress, users share information (text, audio, video) which contain location-specific information, i.e., geo-location. With rapid use of smartphones in the recent decade, the significant innovation is the geo-location capabilities, prompting the rise and commercialization of location-based services (LBSs) [22]. Sharing information is not only just about what users are doing; it is also about what, where, why and whom they are sharing. Integration of technologies drove the development of LBSNs. LBSNs are a type of social networking in which geographic services and capabilities such as geocoding and geo-tagging are used to enable additional social dynamics [23,24]. LBSNs allow users to share their current geo-location and see their friends' location, which opens the debate about user's privacy. Privacy in LBSN is not necessarily an individual issue but extends to organizational and institutional actors involved in data sharing [25]. Some of the private data are shared by the user unsuspectingly or voluntarily. Sometimes, information is intentionally shared by the users are extracted from them extrinsically by offering them some benefits. Through the location based social network Service (LBSNS) like FireEagle, Google

Latitude, Wechat, Nearby etc. are able to identify the location of a person. Some are even able to identify the location of his/her friends [26].

Various studies have been conducted to study check-in behavior under different perspectives like privacy [27,28], gender differences [29], and geographical distances [30]. Research [31,32] has found that the capacity of sharing information with millions of users is a simple method to meet with friends, make new friends, experience new things, and manage one's identity. Zheng, et al. [33] Designed an approach to mine the correlation between locations from a large amount of people's location histories. Beyond the geo-distance and the category relationship between locations, the correlation describes a more comprehensive relationship between locations in the space of human behavior and is a more nature way for human understanding. Comito, et al. [34] Presented a novel methodology to extract and analyze the time- and geo-references associated with social data so as to mine information about human dynamics and behaviors within urban context. In another study [35] presented a cloud-based software environment specifically designed for urban computing supporting smart city applications and described in detail the design and workflow for the implementation of the application and its execution by a workflow engine integrated in the environment. Brimicombe and Li [36] developed city intelligence idea that measures city ability to produce favorable conditions to get metropolitan operators (i.e., inhabitants, systems, and public/private groups) and Cheng, et al. [37] investigated the interrelation between the smart city and urban planning. Also, previous research [38–41] on LBSNs has also studied user's check-in data to predict user's location and mobility patterns. While [42–44] studied the uses and patterns of LBSN and examined the factors that predict the use of LBSNs regarding check-in.

For instance, mobile phone datasets have been used to understand the crowd and individual mobility patterns [45–47]. However, mobile phone data sets are not the only choice to study human mobility pattern analysis. Many other data sources of big data are collected and used, especially including geo-tagged data. This variety of new data sources is so diverse that it ranges log files from smart devices and websites, social media data and geo-tagged audio, video, and graphics data [48,49]. Ye, et al. [50] Proposed a novel definition of life pattern by presenting LP normal form to formalize the definition of individual life patterns and LP-Mine, an abstraction-and-mining framework to effectively retrieve life patterns from GPS data.

Many researchers [51–55] have concentrated on human mobility patterns, venue tagging, and check-in behavior toward using location-based social networks. Automatic venue tagging is one of the new concepts to observe spatial differences in many applications [56,57]. However, Gao and Liu [58] argued that when human mobility is integrated into an application that ranked locations based on a user's check-in history, temporal features were shown to be irrelevant. Ye, et al. [59] explored socio-spatial properties among different LBSN platforms, in another study Ye, et al. [60] analyzed check-in patterns of Foursquare users. A place to healthy relationships has been explored in [61] to expand opportunities for public health. Scellato, et al. [41] Presented a broad study of the spatial properties of the social networks arising among users in online location-based services and analyzed large dataset aimed to observe the inconsistency of urban spaces. Noulas, et al. [62] Explored user participation and provided insight of the city by analyzing social media data from foursquare in Seoul city and specially observed venues. Yu, et al. [63] Applied DBSCAN algorithm to observe Weibo locations in Shanghai and compared with k-means.

Location based datasets have now been used in many studies for urbanization and its environmental effects [64], development and prediction [65–67], travel and activity patterns [68,69] and emergency response [70–72] and urban sustainability [73]. Hong [74] Highlighted the use of an LBSN data to observe the willingness of buyers to pay for various factors. Visit frequencies can represent opinions and the geographical preferences of the individuals for places and given different motivations. Liu, et al. [75] Identified the factors that might cause the outbreak of Ebola and investigated the reaction by China, using big data analysis and explored differences in check-in behavior by gender. For example, Blumenstock, et al. [76] analyzed call detail record (CDR) data from Rwanda to observe population

density and mobile phone use behavior by different genders. Wu, et al. [77] Highlighted the importance of big data as a tool to observe users' daily movement patterns and demographics specifically for housing prices. Preoţiuc-Pietro and Cohn [78] Studied the relationship between shared geo-locations and structured the nature of social connections. Kylasa, et al. [79] Introduced a new novel technique "activity correlation spectroscopy" for deriving connections by using the spectral and distributional structure of activity correlation within a set. Presently, there are some LBSNs available, including the focal ones in the present study. We infer that current research is helpful to understand gender differences and check-in behavior without considering gender equality.

3. Study Area and Data Source

In China, finding open and dependable data that describe geo location–based gender segregation is very hard. The LBSN dataset we are using in the current study comes from Chinese microblog Weibo during January–March 2016.

Shanghai, China (lying between 30°40′–31°53′ N and 120°52′–122°12′ E [80]) is located on the eastern edge of the Yangtze River Delta [81]. According to Gu, et al. [82] in 2015, Shanghai had a total area of 8359 km^2, with a gross domestic product of 366 billion USD. The disposable income per capita of Shanghai is 7333 USD, where the income per capita of urban residents is 7788 USD and the income per capita of rural residents is 3412 USD [83]. As of 2015, the agricultural land area in Shanghai was 317,926 ha. The construction land area was 301,709.27 ha, and the unused land area was 193,564.46 ha. Shanghai is considered to be the most populated and dense community in the world (by urban area inhabitants), and a significant international center for trade, trade, tourism and fashion with a population of around 24.15 million people. In 2016, Shanghai is 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) [84]. Seven of the districts (Changning, Hongkou, Huangpu, Jingan, Putuo, Xuhui, and Yangpu) are located in Puxi (literally Huangpu West). These seven districts are referred as downtown Shanghai or the city center [85,86], as shown in Figure 1.
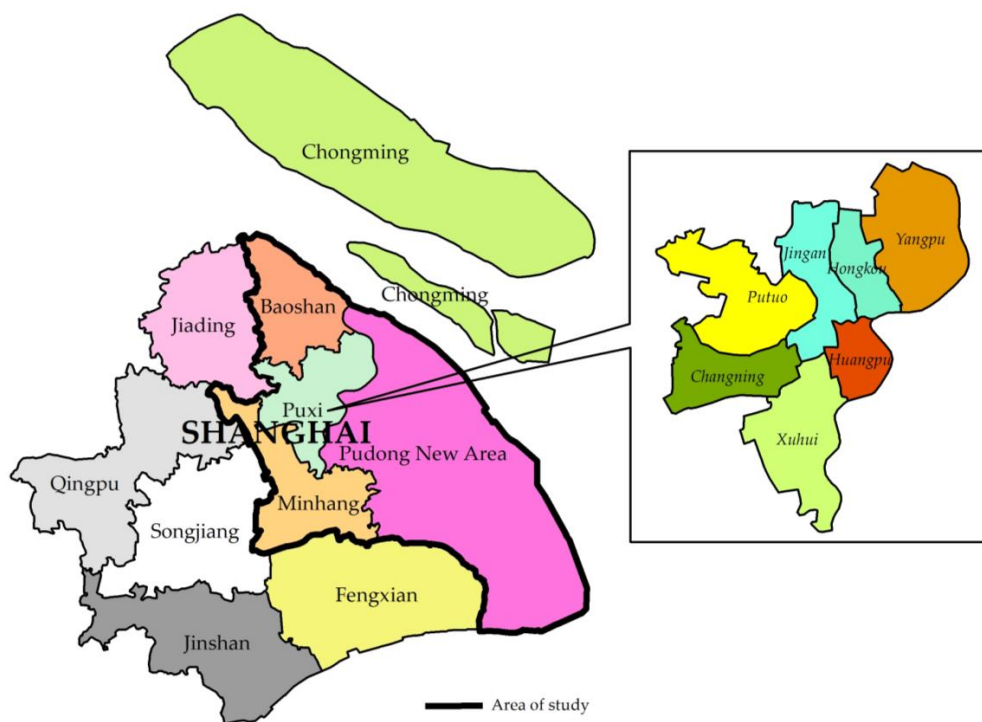

Figure 1. District map of Shanghai.

In addition to the information available in Weibo dataset like user id, date, and time, we also have additional metadata like gender, geo-location (longitude and latitude), venue name, and category, but no personal information like the name is available. Therefore, check-in data records the daily life patterns and user's behaviors towards the services, and it reflects the average person's day-to-day operations. Table 1 describes the necessary information about Shanghai dataset.

Table 1. Shanghai dataset used in current study.

Study Sample	
Total check-ins	**852,560**
Total users	**20,634**
Date range	January–March 2016
City of study	Shanghai, China

4. Methodology

In this paper, we analyzed geo-location data that includes the user(s) ID, time, geo-coordinates (longitude and latitude), and the venue name and category. Figure 2 presents the process flow of data collection and check-in behavior analysis.
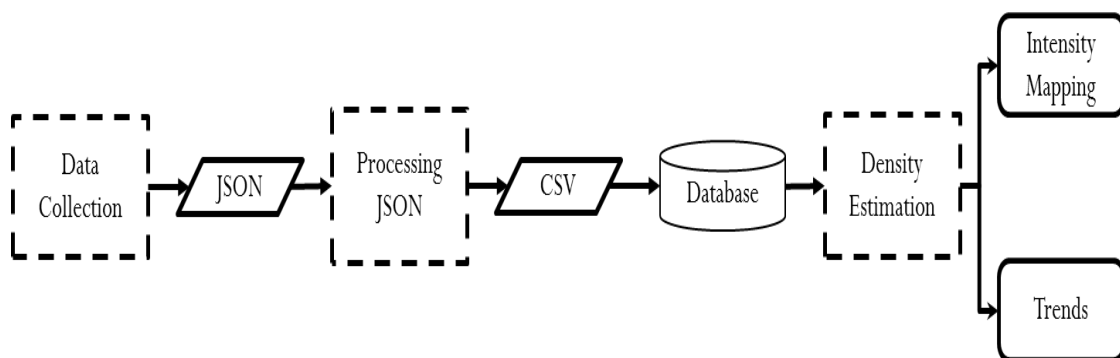

Figure 2. The process flow for data collection and analysis.

Figure 3 presents a general framework for check-in frequency analytics. The frequency analytics methodology is divided into two stages: LBSN data collection and data analysis. The primary task of data collection phase is to download a large number of Weibo data in JavaScript Object Notation (JSON) format by using a python-based Weibo API as shown in Figure 2. However, in the data analysis stage, the critical task is to extract and analyze the feature of check-in data by considering location, time and gender. The analysis phase uses statistical and network analysis and data visualization to produce density maps and trends.

Weibo data is pre-processed to avoid noise and invalid records are filtered using the following criteria:

a.　Each check-in must have following information available: user id, date, time, gender, geo-location (longitude and latitude);

b.　The location of check-in is in Shanghai based on geo-coordinates as shown in Figure 1;

c.　The check-in lies within the date and time for the sampled data set;

d.　User(s) must have checked-in at least twice in a month, and the users with only one check-in record are considered invalid.

Before detecting hot-spots for check-in behavior, we analyzed check-ins by using a kernel density estimation (KDE) for estimating density function used in [79,87–89] to produce a smooth density surface of check-in hot-spots in geographic space [90].

In our study, we considered the data available in the form of geo-tagged check-in.

Let "C" be a set of historical check-in data i.e.,

$$C = \{c^1, \ldots \ldots, c^n\}$$

where $c^i = <x,y>$ is a geo-location of the check-in $1 < i < n$, of individual "i" and on time "t", where "C" is referred as the data set used.

$$f_{KD}(c|C, h) = \frac{1}{n} \sum_{i=1}^{n} K_h\left(c, \ c^i\right) \tag{1}$$

$$K_h(c, \ c^i) = \frac{1}{2\pi h} \exp\left(-\frac{1}{2}\ (c - c^i)^t \sum_h^{-1} (c - c^i)\right) \tag{2}$$

where "c" refers to the location of check-in in training dataset "C" with bandwidth "h". It is assumed that the value of "h" is dependent on the resulting density estimate "f_{KD}" which generates smooth density surface around "C" on data point "c^i."
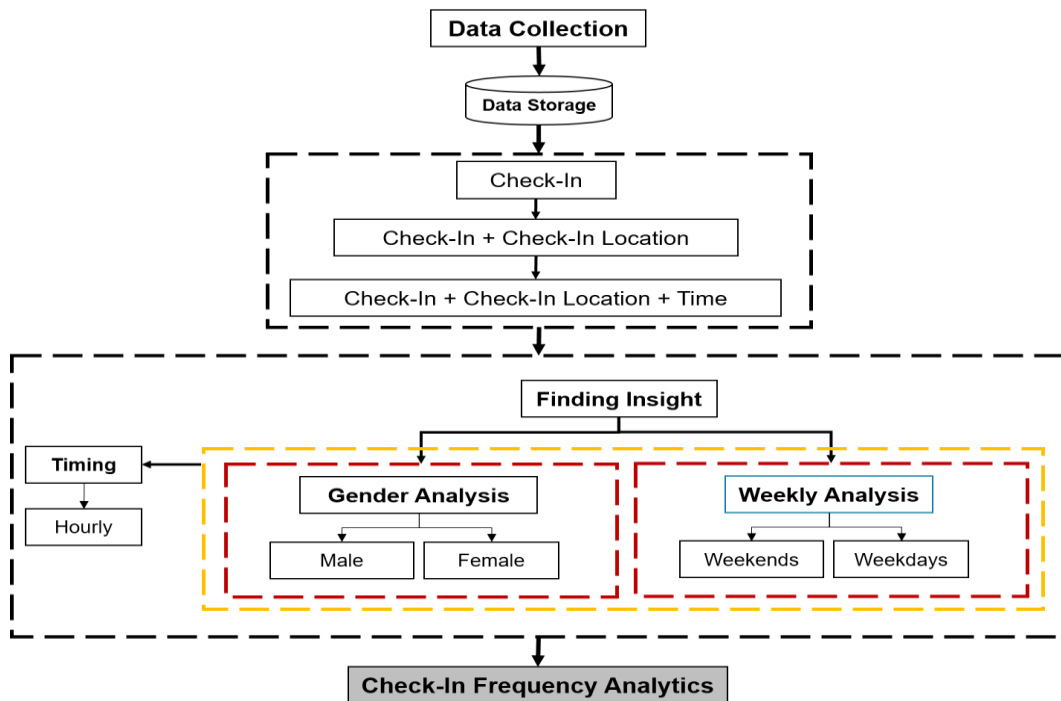

Figure 3. The general framework of check-in frequency analytics.

Compared with the grid maps, kernel density estimation provides smooth distributions by eliminating the local noise to a certain degree by providing a non-parametric probability distribution with optimal bandwidth used to minimize the error. From the kernel density results, we reveal the dynamic of the city in both space and time in different days of the week in various districts of Shanghai.

We hope our results are useful for a behavioral study of users in regions by analyzing their check-in frequency. Through density maps and trend graphs, we can show the check-in frequency of LBSN users in different districts of Shanghai and their behavior of check-in during different hours of the day, weekdays, and weekends.

5. Results and Discussion

For our experiments, we utilized the Weibo check-in data set and used KDE to analyze the density of check-in data. The overall density of check-ins during January–March 2016 can be observed in

Figure 4, and it can be observed that the center of the city has a high density of check-ins, which is a normal behavior for a big city due to easy accessibility of transport (i.e., subway) and living facilities (i.e., food, entertainment). Moreover, the high density of check-ins can be observed near the district borders of Baoshan, Changning, Minhang, Putuo, and Pudong New Area as compared to the center of these district.
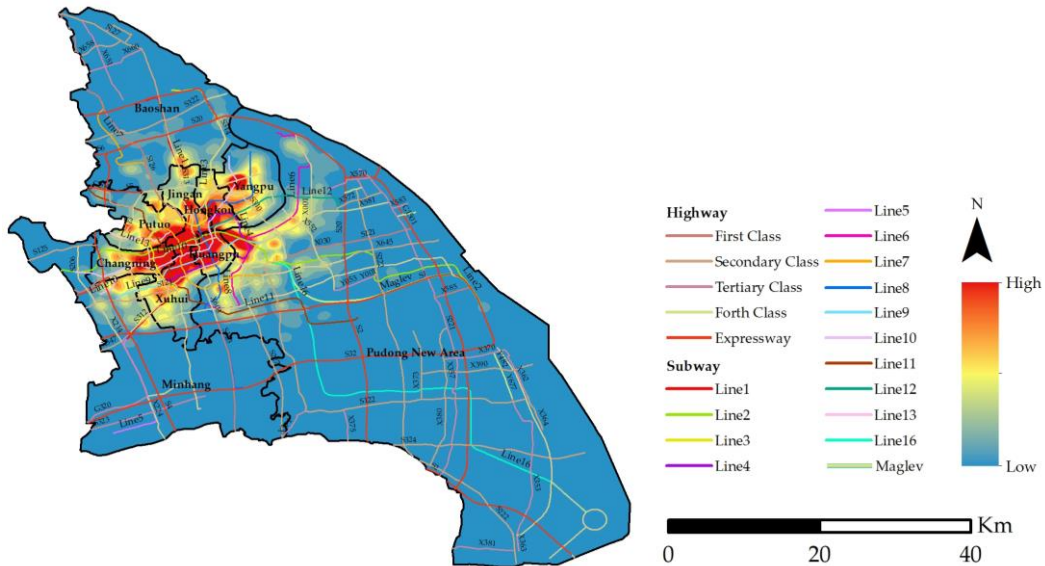

Figure 4. Overall check-in density in Shanghai.

To investigate the check-in frequency and behavior, we analyzed the data regarding gender (male and female) in 10 districts of Shanghai. Figure 5a,b shows the overall weekly check-in trend; which depicts that female users prefer to use Weibo more as compared to male users during the whole week as well as during weekdays and weekends in all districts of Shanghai. It is also observed that check-in frequency increases during Saturday and Sunday. Moreover, Figure 6a,b shows the check-in density in Shanghai. It is observed that female users prefer to use Weibo as compared to male users and hence justifies the results of Figure 5.
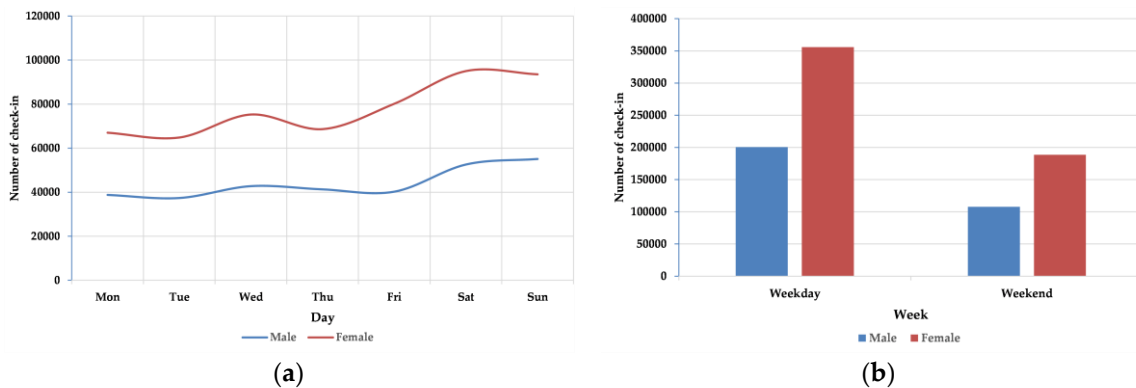

Figure 5. (**a**) Check-in trends of male and female users during a week; (**b**) check-in distributions of male and female users during weekday and weekend.
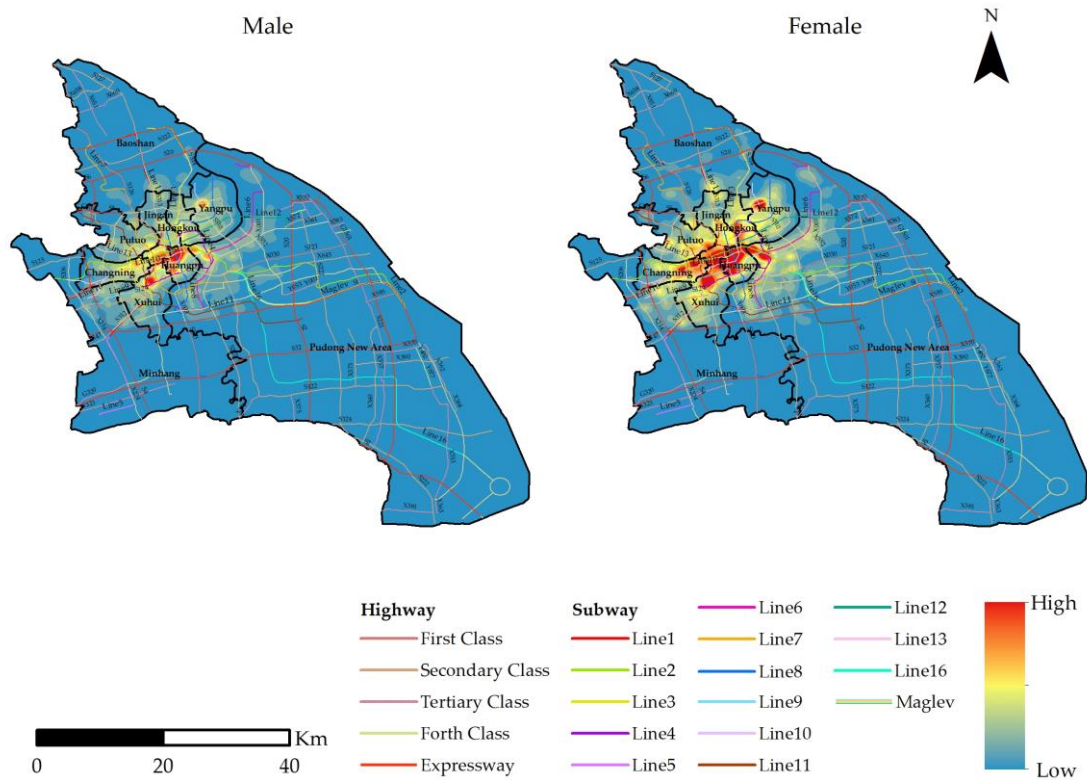
Figure 6. Overall check-in density in 10 districts of Shanghai for male and female.

In Shanghai, to observe the check-in trends of both male and female users, it is essential to measure the check-in frequency during weekends and weekdays over a period. In Figure 7a,b increasing trend can be observed during weekday from 07:00 a.m.–10:00 a.m. and 16:00 p.m.–22:00 p.m. Moreover, during the weekend, an increasing trend is observed from 08:00 a.m.–22:00 p.m. However, it also observed that the check-in frequency of male users is almost consistent with a slight increase during the weekend as compared to female. Furthermore, it is observed that during whole week check-in frequency increases a lot at night (20:00 p.m.–23:59 p.m.) as compared to morning (06:30 a.m.–09:30 a.m.).
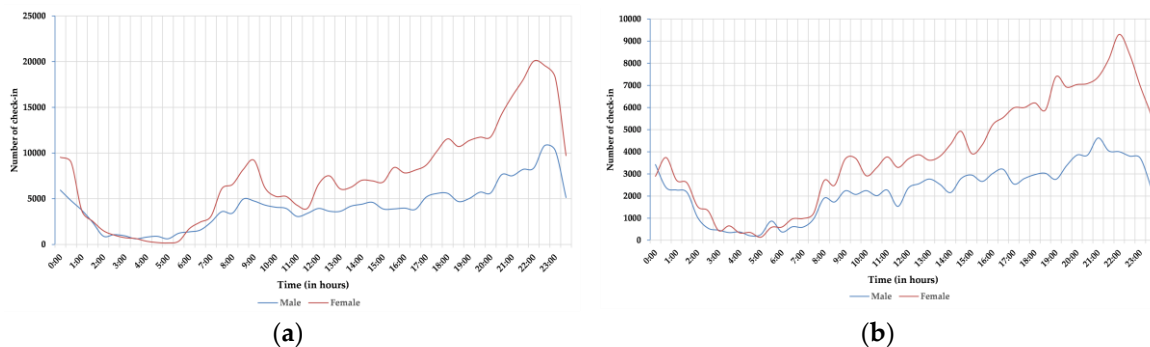

Figure 7. Hourly check-in trend of male and female users during (**a**) weekdays and (**b**) weekends.

Figure 8a presents the distribution of all the check-ins made in different districts of Shanghai. It is no surprise that Pudong New Area district (which is the most prominent district regarding size and is the business center of Shanghai) has the highest number of check-ins. However, from Figure 8b, we can observe the difference of check-in behavior during Saturday and Sunday in Huangpu, Xuhui,

Jingan, and Minhang districts as compared to other areas, where we have more check-ins made during Saturday as compared to Sunday.
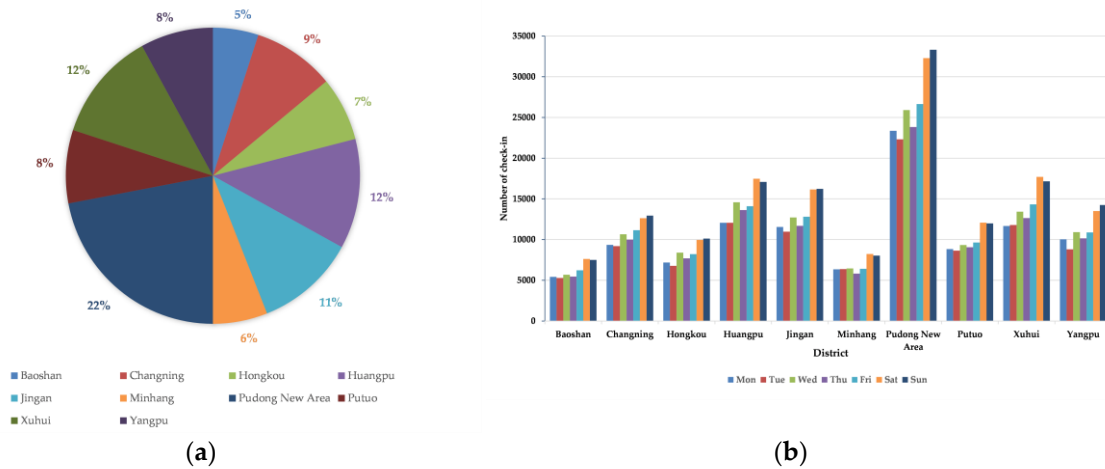

Figure 8. (**a**) Percentage distribution of check-in in different districts of Shanghai (**b**) overall weekly check-in distribution in 10 districts of Shanghai.

Data is analyzed to observe weekly check-in distribution by gender (male and female) and is presented in Figure 9. To our surprise, the difference of check-in behavior during Saturday and Sunday observed in Figure 8b is mainly due to change in check-in behavior by female users. Same check-in behavior can be observed from Figure 9 by the male users during Saturday and Sunday in Changning and Xuhui. However, from Figure 9 noticeable change in check-in behavior can be observed during Saturday as compared to Sunday in most of the districts, i.e., Baoshan, Hongkou, Huangpu, Jingan, Minhang, Putuo, and Xuhui.
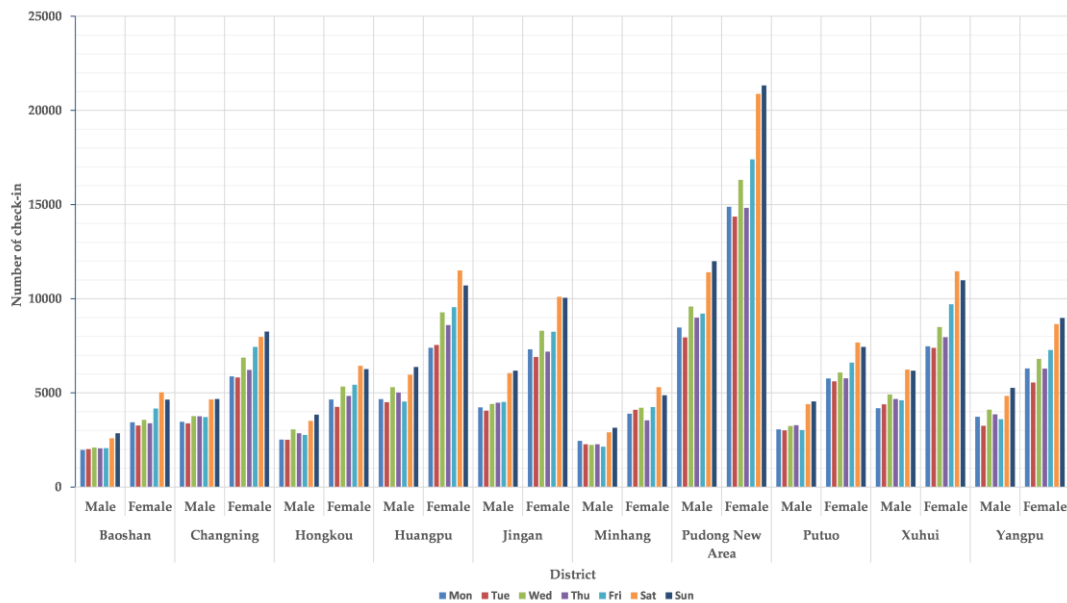

Figure 9. Check-in distribution in 10 different districts of Shanghai by male and female users.

To observe the daily check-in trend in 10 districts of Shanghai, we analyzed the trend in a 24 h period. Figure 10a presents the daily check-in trend in 10 districts of Shanghai; high usage trend is

observed during the morning (06:30 a.m.–09:30 a.m.), in Shanghai. It is also observed that the trend continues to rise till midnight after 23:00 pm for both male and female users as shown in Figure 10b,c.
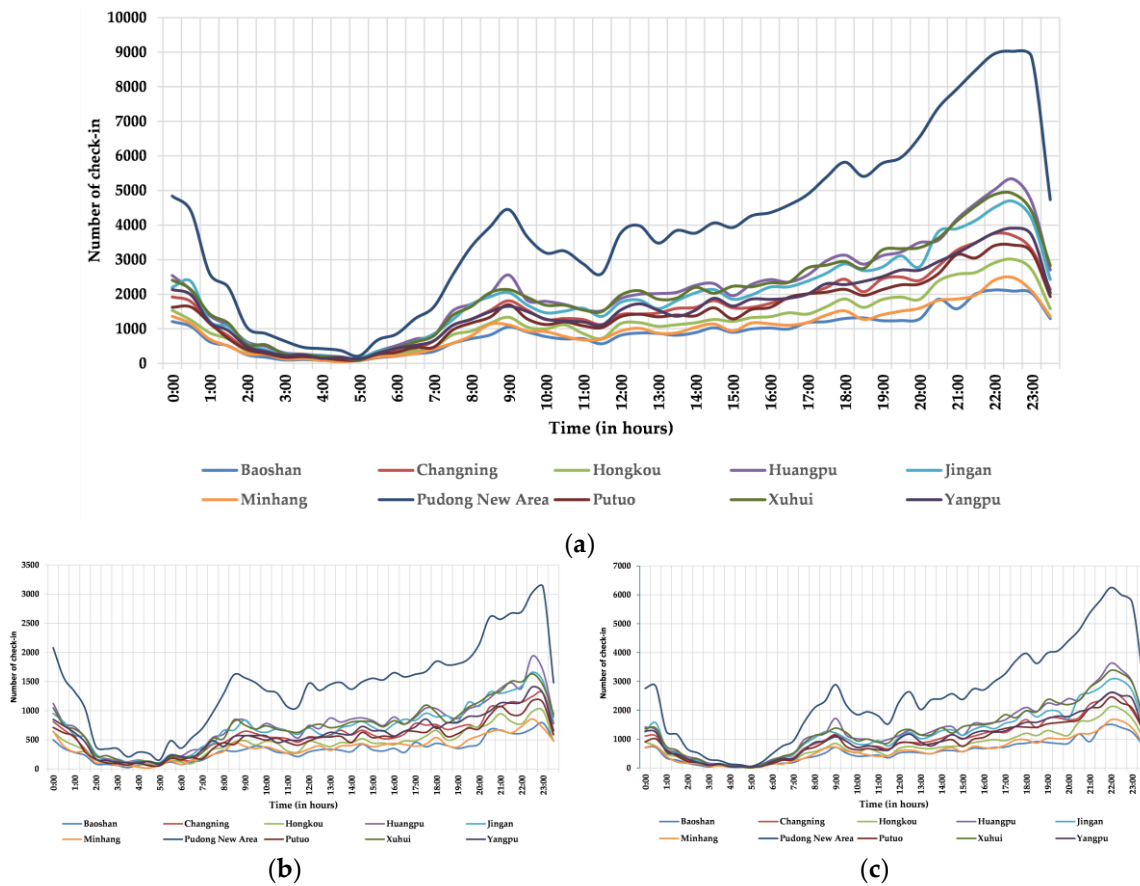

Figure 10. (**a**) Average daily check-in trend in 10 districts of Shanghai (**b**) average male users daily check-in trend in 10 districts of Shanghai (**c**) average female users daily check-in trend in 10 districts of Shanghai.

To further observe the change in check-in behavior, we used kernel density estimation and visualized the density maps for 10 districts of Shanghai. Figure 11 reveal the dynamic of the districts in both space and time in 10 districts of Shanghai. It can be clearly observed that the city center has more check-in density as well as more density is observed near the district borders.

The gender difference in 10 district of Shanghai is examined by the comparison of male and female users check-ins in 10 districts of Shanghai during January–March 2016. We use a relative difference [68,91] (d_r) to calculate the gender differences in 10 districts of Shanghai, it is often used as a quantitative indicator of quality assurance and quality control in the proportion of all check-ins and is expressed as follows:

$$d_r = \frac{\left| P_m - P_f \right|}{\left(\frac{|P_m| + |P_f|}{2} \right)} \tag{3}$$

where "P_m" and "P_f" denote the check-in probability of male and female users in 10 districts of Shanghai during January–March 2016.
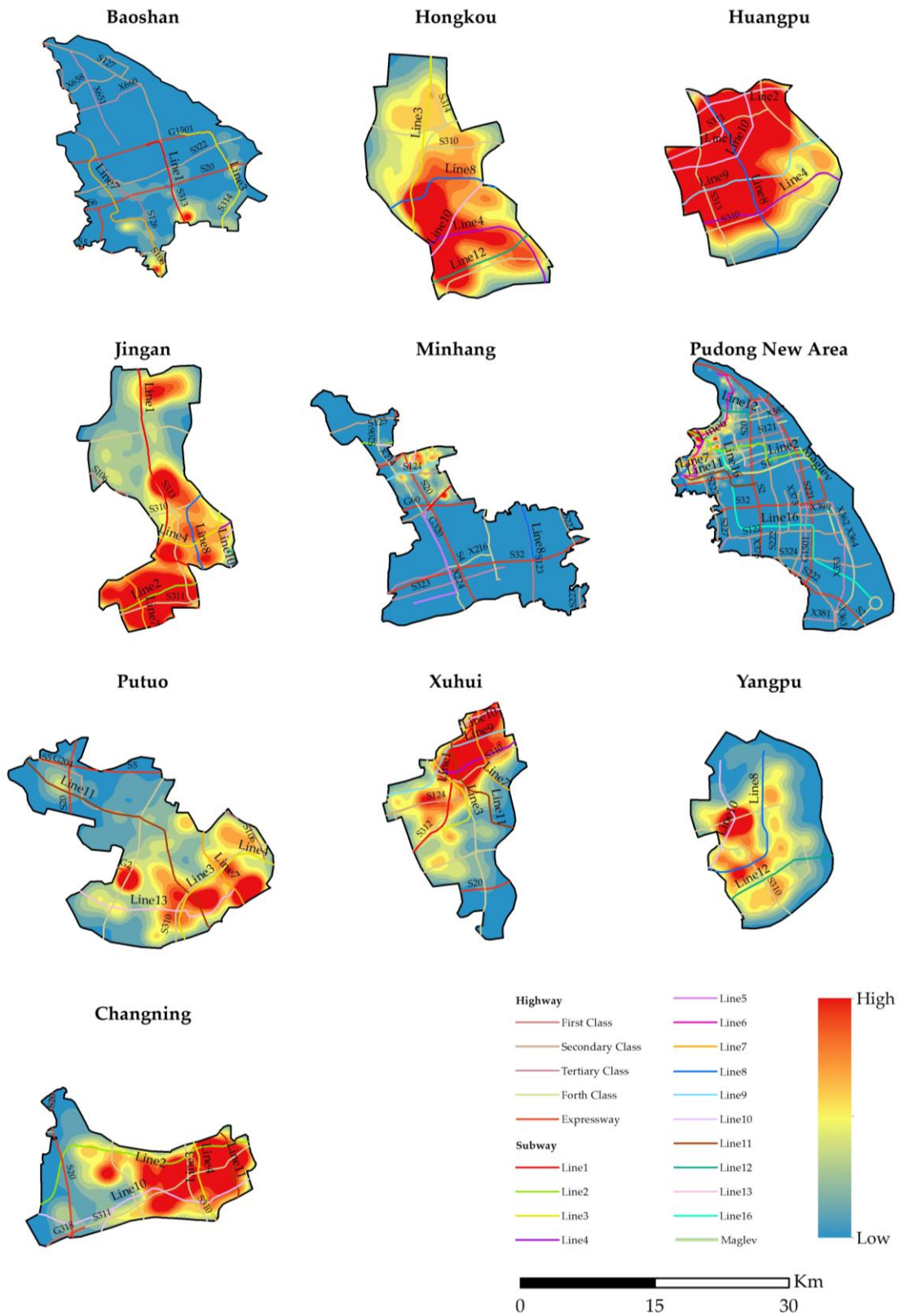
Figure 11. Check-in densities in the 10 districts of Shanghai.

Gender differences in 10 districts of Shanghai are pragmatically explored at the cumulative level. First, we calculated the gender differences of in check-ins in 10 districts of Shanghai as a percentile of total accumulated check-ins made during January–March 2016. Table 2 displays the results of the

relative difference calculated by using the Equation (3) during weekday and weekend. In Table 3, the relative difference values for the Saturday and Sunday are significantly larger than other days. Also, the relative difference values associated with Friday and Saturday are more than 0.55, while the values for the other days lies between 0.5. Results in Table 4 indicate that at the cumulative level, there are relatively significant gender differences in the number of check-ins in some districts (i.e., Huangpu, Pudong New Area, and Xuhui) by Weibo users in Shanghai. Results reveal that female users are more likely to use Weibo during the whole week, days and even in all 10 studied districts of Shanghai, whereas male users are apt to use Weibo during the weekday as compared to the weekend, as shown in Table 5.

Moreover, as observed from Figure 11, high values of check-ins are located in at the district boundaries, and the reason for this might be the significant proportion of financial and commercial activities. Finally, all the results imply that female users are more likely to use Weibo in 10 districts of Shanghai as compared to male users.

Table 2. Gender differences during weekday and weekend.

Week	Male	Female	d_r
Weekday	23.50%	41.75%	0.559
Weekend	12.63%	22.12%	0.546

Table 3. Gender differences during the whole week.

Day	Male	Female	d_r
Mon	4.55%	7.86%	0.534
Tue	4.38%	7.60%	0.538
Wed	5.01%	8.83%	0.551
Thu	4.84%	8.05%	0.498
Fri	4.72%	9.40%	0.663
Sat	6.17%	11.15%	0.575
Sun	6.46%	10.97%	0.517

Table 4. Gender differences in 10 districts of Shanghai.

District	(Check-In) Percentage		d_r
	Male	Female	
Baoshan	1.837%	3.23%	0.549
Changning	3.216%	5.69%	0.555
Hongkou	2.474%	4.37%	0.553
Huangpu	4.268%	7.58%	0.559
Jingan	3.982%	6.82%	0.526
Minhang	2.047%	3.54%	0.535
Pudong New Area	7.933%	14.08%	0.558
Putuo	2.884%	5.27%	0.586
Xuhui	4.129%	7.45%	0.573
Yangpu	3.363%	5.85%	0.540

Table 5. Gender differences during weekday and weekend in 10 districts of Shanghai.

District	Weekday (Check-in) Percentage			Weekend (Check-In) Percentage		
	Male	Female	d_r	Male	Female	d_r
Baoshan	1.198%	2.092%	0.544	0.639%	1.13%	0.558
Changning	2.122%	3.782%	0.562	1.094%	1.90%	0.540
Hongkou	1.610%	2.877%	0.564	0.864%	1.49%	0.532
Huangpu	2.820%	4.970%	0.552	1.448%	2.61%	0.571
Jingan	2.547%	4.458%	0.546	1.435%	2.36%	0.489
Minhang	1.336%	2.347%	0.549	0.711%	1.19%	0.507
Pudong New Area	5.188%	9.129%	0.551	2.745%	4.95%	0.573
Putuo	1.834%	3.501%	0.625	1.050%	1.77%	0.513
Xuhui	2.672%	4.814%	0.572	1.456%	2.63%	0.575
Yangpu	2.177%	3.780%	0.538	1.186%	2.07%	0.543

6. Conclusions

In the current study, we presented an in-depth empirical investigation of check-in behavior using intensity maps and trends using LBSN data. We investigated the check-in behavior from several different angles: the difference in gender, during weekdays and weekends, and daily and hourly patterns. In our results, we observe high rates of social media usage from female users and differences in check-in behavior during weekdays and weekends in all studied districts of Shanghai.

Apart from the inherent limitations of LBSN data, we discuss here to what extent LBSN data can be exploited to observer check-in behavior. More specifically, compared to other data sources (such as survey, census, GPS traces and call detail records), LBSN check-in data have some advantages, such as 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 of traditional data sources.

Based on the results of the empirical 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 weekends) and can help policymakers to define policies regarding the supply of services in urban areas within a city. It can also help to observe variations in population density over the period and act as a tool to estimate the supply of services in the city.

In the future, we plan to use LBSN data as a means to investigate the factors that influence the change in human check-in behavior within the city.

Author Contributions: Muhammad Rizwan, Wan Wanggen and Ofelia Cervantes conceived the research; Muhammad Rizwan designed the research, performed the simulations and wrote the article; Ofelia Cervantes and Luc Gwiazdzinski proof read the article for language editing. All authors read and approved the final manuscript.

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

References

1. Kheiri, A.; Karimipour, F.; Forghani, M. Intra-urban movement flow estimation using location based social networking data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 781. [CrossRef]
2. Weibo. Available online: http://www.weibo.com (accessed on 21 March 2018).
3. Facebook. Available online: https://www.facebook.com/ (accessed on 21 March 2018).
4. Twitter. Available online: https://twitter.com/ (accessed on 21 March 2018).
5. 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.

6.   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]

7.   De Mauro, A.; Greco, M.; Grimaldi, M. A formal definition of big data based on its essential features. *Lib. Rev.* **2016**, *65*, 122–135. [CrossRef]

8.   Miller, H.J.; Goodchild, M.F. Data-driven geography. *GeoJournal* **2015**, *80*, 449–461. [CrossRef]

9.   Charalabidis, Y.; Loukis, E. Participative public policy making through multiple social media platforms utilization. *Int. J. Electron. Gov. Res.* **2012**, *8*, 78–97. [CrossRef]

10.   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. Spat. Inf. Sci.* **2017**, *42*, 59. [CrossRef]

11.   Criado, J.I.; Sandoval-Almazan, R.; Gil-Garcia, J.R. *Government Innovation through Social Media*; Elsevier: Amsterdam, The Netherlands, 2013.

12.   Zheng, L.; Zheng, T. Innovation through social media in the public sector: Information and interactions. *Gov. Inf. Q.* **2014**, *31*, S106–S117. [CrossRef]

13.   Sobaci, M.Z.; Karkin, N. The use of twitter by mayors in turkey: Tweets for better public services? *Gov. Inf. Q.* **2013**, *30*, 417–425. [CrossRef]

14.   Agostino, D. Using social media to engage citizens: A study of Italian municipalities. *Public Relat. Rev.* **2013**, *39*, 232–234. [CrossRef]

15.   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]

16.   Tursunbayeva, A.; Franco, M.; Pagliari, C. Use of social media for e-government in the public health sector: A systematic review of published studies. *Gov. Inf. Q.* **2017**, *34*, 270–282. [CrossRef]

17.   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; p. 48.

18.   Ellison, N.B.; Steinfield, C.; Lampe, C. The benefits of facebook "friends:" social capital and college students' use of online social network sites. *J. Comput. Med. Commun.* **2007**, *12*, 1143–1168. [CrossRef]

19.   Erl, T.; Khattak, W.; Buhler, P. *Big Data Fundamentals*; Prentice Hall: Upper Saddle River, NJ, USA, 2016.

20.   Vastardis, N.; Yang, K. Mobile social networks: Architectures, social properties, and key research challenges. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1355–1371. [CrossRef]

21.   Ahmed, A.M.; Qiu, T.; Xia, F.; Jedari, B.; Abolfazli, S. Event-based mobile social networks: Services, technologies, and applications. *IEEE Access* **2014**, *2*, 500–513. [CrossRef]

22.   Andreassen, C.S. Online social network site addiction: A comprehensive review. *Curr. Addict. Rep.* **2015**, *2*, 175–184. [CrossRef]

23.   Bao, J.; Zheng, Y.; Wilkie, D.; Mokbel, M. Recommendations in location-based social networks: A survey. *GeoInformatica* **2015**, *19*, 525–565. [CrossRef]

24.   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.

25.   Kumar, S.; Saravanakumar, K.; Deepa, K. On privacy and security in social media—A comprehensive study. *Procedia Comput. Sci.* **2016**, *78*, 114–119.

26.   Lowry, P.B.; Cao, J.; Everard, A. Privacy concerns versus desire for interpersonal awareness in driving the use of self-disclosure technologies: The case of instant messaging in two cultures. *J. Manag. Inf. Syst.* **2011**, *27*, 163–200. [CrossRef]

27.   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]

28.   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.

29.   Stefanone, M.A.; Huang, Y.C.; Lackaff, D. Negotiating Social Belonging: Online, Offline, and in-between. In Proceedings of the 2011 44th Hawaii International Conference on System Sciences (HICSS), Kauai, HI, USA, 4–7 January 2011; pp. 1–10.

30.   Boyd, D.M.; Ellison, N.B. Social network sites: Definition, history, and scholarship. *J. Comput. Med. Commun.* **2007**, *13*, 210–230. [CrossRef]

31. Huang, H.-Y. Examining the beneficial effects of individual's self-disclosure on the social network site. *Comput. Hum. Behav.* **2016**, *57*, 122–132. [CrossRef]

32. Wong, C. Smartphone Location-Based Services in the Social, Mobile, and Surveillance Practices of Everyday Life. Master's Thesis, University of London, London, UK, 2014.

33. 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.

34. Comito, C.; Falcone, D.; Talia, D. Mining human mobility patterns from social geo-tagged data. *Pervasive Mob. Comput.* **2016**, *33*, 91–107. [CrossRef]

35. Altomare, A.; Cesario, E.; Comito, C.; Marozzo, F.; Talia, D. Trajectory pattern mining for urban computing in the cloud. *IEEE Trans. Parallel Distrib. Syst.* **2017**, *28*, 586–599. [CrossRef]

36. Brimicombe, A.; Li, C. *Location-Based Services and Geo-Information Engineering*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 21.

37. Cheng, Z.; Caverlee, J.; Lee, K.; Sui, D.Z. Exploring millions of footprints in location sharing services. *ICWSM* **2011**, *2011*, 81–88.

38. Humphreys, L. Mobile social networks and urban public space. *New Media Soc.* **2010**, *12*, 763–778. [CrossRef]

39. Roche, S. Geographic information science i: Why does a smart city need to be spatially enabled? *Prog. Hum. Geogr.* **2014**, *38*, 703–711. [CrossRef]

40. Anthopoulos, L.G.; Vakali, A. Urban planning and smart cities: Interrelations and reciprocities. In *The Future Internet Assembly*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 178–189.

41. Scellato, S.; Noulas, A.; Lambiotte, R.; Mascolo, C. Socio-spatial properties of online location-based social networks. *ICWSM* **2011**, *11*, 329–336.

42. Li, N.; Chen, G. Sharing location in online social networks. *IEEE Netw.* **2010**, *24*, 20–25. [CrossRef]

43. Luo, F.; Cao, G.; Mulligan, K.; Li, X. Explore spatiotemporal and demographic characteristics of human mobility via twitter: A case study of Chicago. *Appl. Geogr.* **2016**, *70*, 11–25. [CrossRef]

44. 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; pp. 1–8.

45. Alharbi, B.; Qahtan, A.A.; Zhang, X. Minimizing user involvement for learning human mobility patterns from location traces. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 865–871.

46. Jin, L.; Long, X.; Zhang, K.; Lin, Y.-R.; Joshi, J. Characterizing users' check-in activities using their scores in a location-based social network. *Multimedia Syst.* **2016**, *22*, 87–98. [CrossRef]

47. Bao, J.; Lian, D.; Zhang, F.; Yuan, N.J. Geo-social media data analytic for user modeling and location-based services. *SIGSPATIAL Spec.* **2016**, *7*, 11–18. [CrossRef]

48. 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]

49. 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]

50. Ye, Y.; Zheng, Y.; Chen, Y.; Feng, J.; Xie, X. Mining individual life pattern based on location history. In Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, Taipei, Taiwan, 18–20 May 2009; pp. 1–10.

51. CHEN, B.-Y.; Kun, Y.; WANG, J.-S.; SUN, M.-Z. Research on evaluation of popularity of lijiang scenic area based on microblog data. *DEStech Trans. Comput. Sci. Eng.* **2017**. [CrossRef]

52. Zhen, F.; Cao, Y.; Qin, X.; Wang, B. Delineation of an urban agglomeration boundary based on sina weibo microblog 'check-in'data: A case study of the Yangtze River delta. *Cities* **2017**, *60*, 180–191. [CrossRef]

53. Shen, Y.; Karimi, K.; Law, S. Encounter and its configurational logic: Understanding spatiotemporal co-presence with road network and social media check-in data. In Proceedings of the 11th International Space Syntax Symposium, Instituto Superior Técnico, Portugal, 3–7 July 2017; Volume 11, pp. 111.111–111.122.

54. Wu, C.; Ye, X.; Ren, F.; Du, Q. Check-in behaviour and spatio-temporal vibrancy: An exploratory analysis in shenzhen, china. *Cities* **2018**, *77*, 104–116. [CrossRef]

55. Soliman, A.; Soltani, K.; Yin, J.; Padmanabhan, A.; Wang, S. Social sensing of urban land use based on analysis of twitter users' mobility patterns. *PLoS ONE* **2017**, *12*, e0181657. [CrossRef] [PubMed]

56. Chen, C.; Ma, J.; Susilo, Y.; Liu, Y.; Wang, M. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transp. Res. Part C Emerg. Technol.* **2016**, *68*, 285–299. [CrossRef] [PubMed]

57. Hesse, B.W.; Moser, R.P.; Riley, W.T. From big data to knowledge in the social sciences. *Ann. Am. Acad. Polit. Soc. Sci.* **2015**, *659*, 16–32. [CrossRef] [PubMed]

58. Gao, H.; Liu, H. Mining human mobility in location-based social networks. *Synth. Lect. Data Min. Knowl. Discov.* **2015**, *7*, 1–115. [CrossRef]

59. Ye, M.; Janowicz, K.; Mülligann, C.; Lee, W.-C. What you are is when you are: The temporal dimension of feature types in location-based social networks. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 102–111.

60. Ye, M.; Shou, D.; Lee, W.-C.; Yin, P.; Janowicz, K. On the semantic annotation of places in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 520–528.

61. Lian, D.; Xie, X. Learning location naming from user check-in histories. In Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Chicago, IL, USA, 1–4 November 2011; pp. 112–121.

62. Noulas, A.; Scellato, S.; Mascolo, C.; Pontil, M. An empirical study of geographic user activity patterns in foursquare. *ICwSM* **2011**, *11*, 2.

63. Yu, X.; Ding, Y.; Wan, W.; Thuillier, E. Explore hot spots of city based on dbscan algorithm. In Proceedings of the 2014 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 7–9 July 2014; pp. 588–591.

64. Cui, L.; Shi, J. Urbanization and its environmental effects in shanghai, china. *Urban Clim.* **2012**, *2*, 1–15. [CrossRef]

65. Han, B.; Cook, P.; Baldwin, T. Geolocation prediction in social media data by finding location indicative words. *Proc. COLING* **2012**, 1045–1062.

66. 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]

67. 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, NC, USA, 26–30 April 2010; pp. 61–70.

68. 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]

69. 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]

70. 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]

71. Yates, D.; Paquette, S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. *Int. J. Inf. Manag.* **2011**, *31*, 6–13. [CrossRef]

72. 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*; Springer: Cham, Switzerland, 2017; pp. 443–457.

73. 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]

74. Hong, I. Spatial analysis of location-based social networks in Seoul, Korea. *J. Geogr. Inf. Syst.* **2015**, *7*, 259. [CrossRef]

75. Liu, K.; Li, L.; Jiang, T.; Chen, B.; Jiang, Z.; Wang, Z.; Chen, Y.; Jiang, J.; Gu, H. Chinese public attention to the outbreak of ebola in west africa: Evidence from the online big data platform. *Int. J. Environ. Res. Public Health* **2016**, *13*, 780. [CrossRef] [PubMed]

76. Blumenstock, J.E.; Gillick, D.; Eagle, N. Who's calling? Demographics of mobile phone use in Rwanda. *Transportation* **2010**, *32*, 2–5.

77. 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]

78. Preoţiuc-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.

79. Kylasa, S.B.; Kollias, G.; Grama, A. Social ties and checkin sites: Connections and latent structures in location-based social networks. *Soc. Netw. Anal. Min.* **2016**, *6*, 95. [CrossRef]

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. Ross, C. *Regional China: A Business and Economic Handbook by Rongxing Guo*; Palgrave Macmillan: New York, NY, USA, 2013; p. 386.

82. Gu, X.; Tao, S.; Dai, B. Spatial accessibility of country parks in shanghai, china. *Urban For. Urban Green.* **2017**, *27*, 373–382. [CrossRef]

83. Jiang, Y.; Shi, X.; Zhang, S.; Ji, J. The threshold effect of high-level human capital investment on china's urban-rural income gap. *China Agric. Econom. Rev.* **2011**, *3*, 297–320. [CrossRef]

84. 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]

85. 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.

86. Shen, J.; Kee, G. *Development and Planning in Seven Major Coastal Cities in Southern and Eastern China*; Springer: Cham, Switzerland, 2016; Volume 120.

87. Zhang, X.; Butts, C.T. Activity correlation spectroscopy: A novel method for inferring social relationships from activity data. *Soc. Netw. Anal. Min.* **2017**, *7*, 1. [CrossRef]

88. 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.

89. Xie, Z.; Yan, J. Kernel density estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst.* **2008**, *32*, 396–406. [CrossRef]

90. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986; Volume 26.

91. Van Iddekinge, C.H.; Lanivich, S.E.; Roth, P.L.; Junco, E. Social media for selection? Validity and adverse impact potential of a facebook-based assessment. *J. Manag.* **2016**, *42*, 1811–1835. [CrossRef]