



Article Investigating "Locality" of Intra-Urban Spatial Interactions in New York City Using Foursquare Data

Yeran Sun

GIScience Research Group, Institute of Geography, Heidelberg University, Berliner Straße 48, Heidelberg 69120, Germany; yeran.sun@geog.uni-heidelberg.de; Tel.: +49-6221-54-5528; Fax: +49-6221-54-4529

Academic Editor: Wolfgang Kainz Received: 10 February 2016; Accepted: 11 March 2016; Published: 24 March 2016

Abstract: Thanks to the increasing popularity of location-based social networks, a large amount of user-generated geo-referenced check-in data is now available, and such check-in data is becoming a new data source in the study of mobility and travel. Conventionally, spatial interactions between places were measured based on the trips made between them. This paper empirically investigates the use of social media data (*i.e.*, Foursquare data) to study the "locality" of such intra-urban spatial interactions in New York City, and specifically: (i) the level of "locality" of spatial interactions; (ii) the impacts of personal characteristics on "locality" of spatial interaction and finally; (iii) the heterogeneity in spatial distribution of "local" interactions. The results of this study indicate that: (1) spatial interactions show a high degree of locality; (2) gender does not have a considerable impact on the locality of spatial interactions and finally; (3) "local" interactions likely cluster in some places within the research city.

Keywords: human mobility; spatial interaction; Foursquare; complex network; community detection

1. Introduction

Location-based social network (LBSN) products such as Foursquare, Gowalla, Google Latitude, and Facebook Places are becoming an important source of volunteered geographic information (VGI). As the most popular LBSN, Foursquare has nine venue categories, including food, travel, and transport, and these cover hundreds of sub-categories, including cafes, bus stations, and train stations [1]. Foursquare has over 50 million users worldwide, and over 6 billion check-ins had been made using the website by May 2014, with millions of new check-ins taking place every day [2]. Although check-in data has some limitations in terms of how it represents human mobility-for example, it shows age group and place category bias—such data has the ability to identify human mobility patterns in accordance with certain mechanisms [3,4]. When compared to some other data sources (e.g., survey data and mobile phone data), LBSN check-in data has some advantages as an indicator of human activity categories such as dining, working, and shopping, as it provides a fine-grained resolution, and is readily available. Therefore, user-generated geo-referenced check-in data has excellent potential when wishing to study human mobility, as some researchers have already demonstrated [3,5–11]. Since spatial interactions are measured through the use of human mobility patterns, so check-in data has significant potential within the study of spatial interaction. Therefore, in this study we will use LBSN check-in data to study intra-urban spatial interactions. Before introducing how the paper is structured, we will review the previous research on intra-urban mobility and spatial interactions.

1.1. Brief Overview of the Previous Research on Intra-Urban Mobility and Spatial Interactions

Over the last two decades, intra-urban mobility has become a popular research topic across the research community, including among geographers, urban planners, computer scientists, and physicists. Geographers are interested in spatial distribution of intra-urban mobility, and urban planners want to improve transport efficiency by investigating spatial and temporal variations of travel time and travel flows [12–15]. To quantitatively understand mobility, computer scientists and physicists are involved in modelling distribution of travel distance in a mathematical way [16–18]. A few studies have been able to identify human mobility patterns using mobile phone data, automobile GPS traces, and social media data [3,5,15–20]. One of the most common findings among such studies is that the distribution of trip distances tends to follow a power law [12,16] or an exponential law [4,21,22]. Another common finding is that humans follow simple, reproducible patterns of mobility [15,16]. Some other studies have attempted to interpret intra-urban mobility distributions using models [12,23] such as the gravity model [23] and its modified forms—mainly based on the distance-decay effect [12], and the population distribution model—which is the most popular simulation model used [12,22].

Some researchers [17,18] have used empirical studies to demonstrate high potential predictability levels with respect to user mobility [17], while others have attempted to uncover how urban form characteristics (e.g., land area, land use mix, road and population densities) and personal or household characteristics (e.g., age, gender, education levels, employment, income, and car ownership levels) impact upon intra-urban human mobility [11,15,20,24–27]. Furthermore, from a network perspective, some studies have focused on intra-urban travel networks [28,29], revealing that, like some typical connection networks such as the World Wide Web, the internet, friendship networks, and scientific collaboration networks, intra-urban travel networks are "small world" networks [30]. Finally, some studies have revealed underlying patterns of intra-urban mobility to understand urban structure and functions of sub regions [31,32], and to better measure popularity of places (e.g., retails) for site selections [33,34].

1.2. Motivation for This Study

In some studies [35,36], trips between geographic units such as cities and neighborhoods, are used to measure the interaction between two geographic units. Similarly, trips between locations such as restaurants, apartments, and bus stops, have been used to measure the interactions between locations in a city. The number and length of such trips can be used to represent the strength and length of interactions between locations. Spatial interactions have also been analyzed by researchers to better understand human mobility patterns and human-mediated dynamic phenomena, such as the spread of infectious disease, the results of which may be beneficial to urban planners and decision makers [28,37,38]. As one sub-field of urban planning, transportation planning will be improved by a better understanding of the spatial distribution of human mobility in cities. Besides, "Human behaviour plays an important role in the spread of infectious diseases, and understanding the influence of behaviour on the spread of diseases can be key to improving control efforts." [39]. Obviously, decision makers are concerned about how to improve control over the spread of infectious diseases. On the one hand, investigating spatial interaction patterns may be beneficial to businesses, e.g., by helping to identify a good location based on personalized user preferences, or selecting a good site for a new shop. The former has been explored by studies through the use of social media data [4,6], while the latter is a new field within mobility studies using social media data. Investigating spatial interaction distances may help when selecting the site of a new shop, restaurant, or other service facility. For instance, if a new restaurant is established to serve customers of fitness and sports centers, it should be located in close proximity to them. However, it may not always be the most profitable decision to open a restaurant extremely close to a gym, so when selecting a site, one needs to examine to what extent a restaurant close to a gym is able to attract a large number of gym users. In addition to business location decisions, management of the spread of infectious diseases can also benefit from spatial interaction or human mobility analysis [37,40,41]. For instance, if a healthy person visits a venue over the same time period as a sick person, such as over the course of a day, he or she is more likely to become infected than those who do not visit the venue. Therefore, mobility trajectories may be closely associated with the spread of a disease; with a strong interaction implying a higher possibility of disease spread

and with long-distance interactions implying a greater potential for a larger spatial range of spread. The strength and length of interactions between venues can be taken into account within disease spread simulations.

This study was explicitly motivated by the following. The distance decay effect [42] seemingly implies that human mobility or spatial interactions have a large degree of "locality" associated with them, so this study attempts to investigate the "locality" of spatial interactions in a quantitative way. Moreover, since the existing literature reveals that intra-urban mobility is influenced by personal characteristics, it will be interesting to examine if and how personal characteristics impact upon the "locality" of interaction. Furthermore, since spatial heterogeneity of population and physical activities likely result in spatial heterogeneity of human mobility, this study attempts to investigate the heterogeneity in spatial distribution of "local" interactions by identifying clustering of "local" interactions to understand travel behavior and activities of residents. Based on this, functions or land use patterns of sub regions covered by the clusters will be somewhat discussed.

There is some research that leverages LBSN data or other mobility data to analyze intra-urban mobility, spatial interactions, and activity transitions [4,6]. One the one hand, the majority of this research has not accounted for the impacts of personal characteristics on spatial interactions. In contrast, this study attempts to incorporate personal characteristics into the analysis of intra-urban spatial interactions. On the other hand, some other research utilizes LBSN data or other mobility data to divide a city or identify functions of regions by analyzing spatial patterns of human activities [43–45]. The majority of these researchers consider activities (points) separately, but ignore the relationships between different activities in terms of trips (activity transitions). Accordingly, complex network approaches are used to identify clusters of trips. Afterward, we will somewhat discuss functions or land use patterns of sub regions covered by the noticeable clusters. This study doesn't aim to discuss the functions or land use patterns of all sub regions in a city.

The remainder of this paper is organized as follows. Section 2 introduces how we analyze the "locality" of spatial interaction, while Section 3 presents the empirical analysis and the relevant results. Lastly, we will present the conclusion and make suggestions for future research.

2. Methodology

In this section, we will introduce the methodology used for this study. First, the study will investigate the level of "locality" of spatial interactions. Second, the study will explore if gender as a personal characteristic has a considerable impact on the "locality" of spatial interactions. Finally, this study will investigate the heterogeneity in spatial distribution of "local" interactions. Particularly, this study will identify clustering of "local" interactions (links), *i.e.*, area of strong intra-interactions. Additionally, functions or land use patterns of sub-regions covered by the clusters will be somewhat discussed. Land use data will be used to somewhat validate the discussions.

2.1. Depiction of Spatial Interactions

In this section we will introduce how to use check-in information to depict human mobility and spatial interaction patterns. Within location-based social networks such as Foursquare, each venue corresponds to a physical location (see Figure 1). Common types of venues include restaurants, offices, apartments, hotels, bus stops, shops, and gyms. Imagine a user checks in at venue *A* (house) and venue *B* (office) consecutively. In this situation, the "trip" is from venue *A* to venue *B*, irrespective of the specific route taken by the user travelling between these two venues. In the context of this study, a "trip" is a single journey represented by a line in geometrical terms. Sometimes, a user might travel twice between two venues in opposite directions, e.g., going from an office to a restaurant for lunch and then returning after lunch. If a user travels between two venues, these two venues are considered to be "spatially interacted", in other words, there is an "interaction" between these two venues.



Figure 1. Trips taken by Foursquare users between venues.

2.2. Length and Strength of Spatial Interactions

The *length* of an interaction is equal to the distance between the pairwise venues (*i.e.*, the length of trip), while the *strength* of an interaction is measured by the number of trips taken between interacted pairwise venues. The more trips there are between pairwise venues, the stronger the interaction between them.

2.3. "Locality" of Spatial Interactions

Since this study investigates the "locality" of spatial interactions, two "locality" indicators are introduced. Two indicators are used to characterize locality of interaction at the venue scale. To eliminate the effect of spatial heterogeneity among venues, relative distance is used instead of real distance. Therefore, supposing venue *j* is the *k*th nearest neighbor (*k*th NN) of venue *i*, then *k* is used to measure the relative distance between venue *i* and its neighboring venue *j*. The lower *k* is as a value, the shorter the relative distance will be. *K*'s nearest neighbor (*K*NNs) is the venue set composed of *i*th nearest neighbors. Two indicators are defined here to characterize the locality of interaction.

(1) The percentage of links a venue has with KNNs is used to measure the relative possibility of a venue being linked (interacted) with its neighbors. This value is used to measure the "locality" range of the interactions that take place. The higher the value is, the more the venue is likely to be interacted with neighboring venues than with distant venues. This value can be defined as:

percentage of links a venue has with KNNs
$$(P_i) = \frac{link \ count \ of \ a \ venue \ with \ KNNs \ (P_i)}{link \ count \ of \ venue \ (P_i)}$$
 (1)

where the *link count of venue with* $KNNs(P_i)$ is the number of links venue P_i has with its KNNs, and *link count of venue* (P_i) is the number of links venue P_i has.

(2) *The kth NN linked with a venue* value is used to represent that a venue is interacted with the *k*th nearest venue among its *K*NNs. This value is used to measure the "locality" strength of interactions between venues.

2.4. Interaction Network

In regards to investigating the heterogeneity in spatial distribution of "local" interactions, this study will use a complex network analysis method to identify clustering of "local" interactions (links). First, this sub-section introduces how to build a "local" interaction network on the basis of nodes and edges. After that, this subsection presents how to identify clustering of "local" interactions in a "local" interaction network.

2.4.1. Nodes and Edges

Trips within a city can be used to build an "interaction network". In this network, a node represents a venue and an edge represents an "interaction". The weight of an edge is measured by the number of trips taken between interacted pairwise nodes.

2.4.2. Network Structure and Community Analysis

In complex networks, links are not evenly distributed among nodes. The nodes of a network can be divided into groups of nodes with dense connections internally and sparser connections between groups [46]. Such a group is considered as a community, which is actually a sub-network. The community based approach is widely used to analyze the structure of complex networks [46,47].

A good partition of a network into communities must comprise many intra-community links and few inter-community links. Various community detection algorithms have been proposed and used to divide complex networks [48–51], however, there are not many empirical studies revealing what is the best algorithm. In this case, four widely used community detection algorithms—*i.e.*, FastGreedy [48], Spinglass [49], Walktrap [50], and Infomap [51]—are all used to divide the network in this study. Among the four partitions of the network, the partition with the smallest number of communities detected is then selected as the best partition for analysis of clustering of "local" interactions. A small number of communities detected means the network is divided into a small number of parts. Accordingly, the communities detected likely have a relatively large number of nodes and intra-community links (trips).

3. Empirical Analysis

The methods for spatial intersection described in Section 2 are applied for Foursquare Check-in data in New York City. In this section, the test data set will be briefly described at first, then the "locality" of spatial interactions will be empirically investigated.

3.1. Study Case and Empirical Data

This study uses New York City (NYC) in the United States as the research city. There are a large number of active social media users in NYC. New York City is composed of five boroughs: Brooklyn, Queens, Manhattan, the Bronx, and Staten Island.

In this paper, Foursquare check-in data is used to generate the empirical mobility data. Since Foursquare has a strict privacy policy, the check-ins were collected from Twitter, with which some Foursquare users share their check-ins. Within NYC's municipal boundary, 148,169 check-ins were acquired over the period 3 March 2014 to 27 April 2014 (12 continuous weeks). To make sure the trip generation process is reasonable, any noise—*i.e.*, where a check-in could not reasonably be used to constitute a trip—found in the four situations detailed below, is filtered-out, as follows:

Situation (1): Among the consecutive daily check-ins, more than one check-in is generated in the same position. In this situation, only the first generated check-in is retained; the others are discarded.

Situation (2): The time difference between consecutive check-ins for an identical user is greater than 8 h.

Situation (3): The speed of travel between two consecutive daily check-ins for an identical user is faster than extremely high (e.g., 250 km/h).

Situation (4): The distance between two consecutive daily check-ins for an identical user is short (e.g., 100 m or less). This might suggest that when visiting a venue a user will probably check in, not only at this venue, but at other venues nearby. Those check-ins made at other venues, but not actually visited by the user, are considered fake. Therefore, with lengths of less than 100 m, such fake short trips should be discarded. In this situation, the two consecutive check-ins do not constitute a trip.

Furthermore, active users who are likely local residents are selected. In this study, active local users are regarded as users who: (1) are likely to check in at locations as much as possible whenever

they actually visit these locations; who (2) are likely to have check-ins for a sufficient number of days (at least 28 days); and who (3) have a certain number of trips (at least 21 trips) between different venues.

As a result, 50,758 travel flows (trips) covering 18,333 venues and 40,111 links (interactions) are included. These trips are generated by 843 active sampled users (443 male users and 400 female users).

The land use data used in this empirical study were collected in 2010. The land use data were open data and downloaded from the Department of City Planning (DCP), NYC [52]. Within the land use dataset there are 11 urban land use categories: *One & Two Family Buildings, Multi-Family Walk-Up Buildings, Multi-Family Elevator Buildings, Mixed Residential & Commercial Buildings, Commercial & Office Buildings, Industrial & Manufacturing Buildings, Transportation & Utility, Public Facilities & Institutions, Open Space & Outdoor Recreation, Parking Facilities, and Vacant Land.*

3.2. Locality of Spatial Interactions

This section presents an empirical investigation of the locality of spatial interactions, and specifically, to what degree a venue is likely to be interacted with its neighbors. First of all, this sub-section presents the empirical distribution of two locality characteristics: *the percentage of links a venue has with KNNs,* and *the kth NN linked with a venue*. It should be noted that "*K*-nearest neighbors" (*K*NNs) means the *K*-nearest neighbors of the relevant venue. In this empirical investigation, a specific number has had to be set as the value of *K*. The value 200 is assigned to *K*, since 200*NNs* of a venue represents a considerable number of neighbors for a venue. 23,390 trips (46% of the total trips) connect the start venues to the end venues which are the 200NNs of the start venues. This indicates 200NN can be used to represent an appropriate neighborhood range.

In this sub-section, we empirically investigate how venues interact with their 200NNs. First, the distribution of *percentage of links a venue has with KNNs* is analyzed, with the results shown in Figure 2. Approximately 50% of the venues interact with their 200NNs. Figure 2 also shows that when the value is more than 0, the *percentage of links a venue has with 200NNs* values seems to exhibit a uniform distribution. This suggests that the relative possibility of a location being linked (interacted) with its neighbors seems to follow a uniform law.



Figure 2. Complementary cumulative distribution function (CCDF) of percentage of links a venue has with KNNs.

Second, the distribution of the *kth NN* linked to each venue is analyzed. In Figure 3, using a Kolmogorov-Smirnov (KS) *test* (for more detail see [53]), the distribution of the *kth NN* linked to each

venue follows an exponential law. This suggests that the distribution of relative interaction lengths follows an exponential law. In addition, among the venues' links with their 200NNs, approximately 80% are connected to their 100NNs. Therefore, the analytical results reveal that spatial interactions have a high degree of locality. Furthermore, although the relative possibility of a location being linked (interacted) with its neighbors seems to follow a uniform law, locations are more likely to be interacted with nearer neighbors than those further away.



Figure 3. CCDF of the *kth* NN linked with a venue among 200NNs ($P(x) \sim e^{-\lambda x}$, $\lambda = 0.022$).

3.3. Impact of Gender on "Locality"

This section presents the empirical results of research into the impact of gender on "Locality" characteristics (*i.e., percentage of links a venue has with 200NNs*, and *the kth NN linked with a venue*). Specifically, this section presents the impact of gender on the two locality characteristics, with Table 1 giving the average values found. As an alternative to the *T*-test when it is not guaranteed that samples are normally distributed, the Wilcoxon test is used to test if a sample set has a statistically significant higher average value than the other.

In the results from the *Wilcoxon* test, the *p*-values corresponding to the two characteristics are all much more than 0.05 (see Table 1). This means that the average *percentage of links a venue has with 200NNs* value for male users is not statistically different from that for female users, and the average *the kth NN linked with a venue* value for male users is not statistically different from that for female users at the 0.05 level. This indicates that gender does not have a considerable impact on the locality of spatial interactions.

Table 1. Average values of the two locality characteristics for male and female users.

Locality Characteristics	Male	Female	Wilcoxon Test <i>p</i> -Value
Percentage of links a venue has with 200NNs	0.36	0.36	>0.05
The <i>kth</i> NN linked with a venue	52	51	>0.05

3.4. Clustering of Local Interactions

This section presents the empirical results of research into identifying clustering of local interactions. In this study, local links connect the start venues to the end venues which are the

200NNs of the start venues. Accordingly, the local interaction network is composed of 13,112 nodes (venues) and 15,944 edges (local links). The weight of an edge equals the number of trips connecting the pairwise venues of the edge.

The four community detection algorithms were all used to identify clustering of local interactions (area of high dense interactions). Among the four partitions of the network, the partition made by using the Walktrap algorithm was selected as the best partition for this study since the number of communities detected is the smallest. The spatial interaction network in NYC was divided into 1515 communities by using the Walktrap algorithm. The majority of the communities have a small number of intra-community trips and nodes. Among the 1515 communities detected, only four communities have more than 500 intra-community trips and 100 nodes. In this case, typical communities were further selected for analysis of clustering of local interactions. Typical communities have a large number of intra-community trips (more than 500 trips), and a ratio of more than 20 for the number of intra-community trips and number of inter-community trips. The four most typical communities, which in total have 790 venues and 3778 intra-community trips, were selected since they contained only 6% of the total venues but 16% of the total trips. This indicates the heterogeneity in spatial distribution of "local" interactions.

Furthermore, we focused on the four typical communities to analyze significant clustering of local interactions. Figure 4 displays the four most typical communities (significant clusters of local interactions). In Figure 4, convex hulls of communities were used to measure the spatial sizes of communities. Table 2 shows number of nodes and intra-community trips, and the most predominant link categories in the typical communities selected. Table 3 shows the most predominant land use categories within convex hulls of typical communities selected and percentages of the most predominant land use categories.

Community 1 is located within Brooklyn. The most predominant link categories in Community 1 are *Eating -> Eating, Shopping -> Eating,* and *Eating -> Shopping* (see Table 2). This indicates that there is clustering of links (trips) between *Shopping* venues and *Eating* venues within Community 1. This implies that some residents cluster in this area where they live, and are more involved in eating and shopping activities than working activities or entertainment activities. It seems that this area is likely a residential area. *One & Two Family Buildings* and *Multi-Family Elevator Buildings* are the 1st and 2nd most predominant land use categories within the convex hull of Community 1 (see Table 3), indicating this area is a residential area.

Community 2 is located within Queens, and around the Queens College and St. John's University. The most predominant link category in Community 2 is *University -> University* (see Table 2). This indicates that there is clustering of links (trips) from University venues to University venues within Community 2. This further implies that some residents cluster in this area, and are likely students in the Queens College. It seems that this area is likely an educational area. *Public Facilities & Institutions* is one of the most predominant land use categories within the convex hull of Community 2 (see Table 3), indicating this area is an educational area.

Community 3 is located within Manhattan and covers SoHo, Little Italy, Chinatown, and the New York University. There are no extremely predominant link categories in Community 3 (see Table 2). This further implies that some residents cluster in this area for various activities with a high mixed level. It seems that this area is a mixed area. *Mixed Residential & Commercial Buildings* and *Public Facilities & Institutions* are the 1st and 2nd most predominant land use categories within the convex hull of Community 3 (see Table 3), indicating this area is a mixed area.

Community 4 is located within Manhattan and around the Empire State Building. The most predominant link categories in Community 4 are *Office -> Office* and *Office -> Shopping* (see Table 2). This indicates that there is clustering of links (trips) between *Office* venues and *Office* venues or *Shopping* venues within Community 4. This further implies that some residents cluster in this area for participating in commercial work. It seems that this area is likely a commercial area. *Commercial & Office Buildings* is the 1st most predominant land use category within the convex hull of Community 4 (see Table 3), indicating this area is a commercial area.



Figure 4. Typical communities detected by using the Walktrap algorithm.

Table 2. Number of nodes and intra-community trips, and the most predominant link categories in the typical communities selected.

Community	Na la Caurt	Number of	Ratio *	Predominant Link Category		
Community	Node Count	Intra-com Trips		Start Category	End Category	Per (%) *
1	224	1900	633	Eating	Eating	14
				Shopping	Eating	9
				Eating	Shopping	9
				Shopping	Shopping	8
				Services	Services	7
2		729	243	University	University	16
				University	Office	8
	134			Office	University	7
				University	Eating	7
				Eating	University	5
3	309	581	22	Shopping	University	6
				Eating	Eating	6
				Nightlife	Eating	6
				Shopping	Eating	6
				Eating	Nightlife	6
4	123	568	568	Office	Office	19
				Office	Shopping	10
				Eating	Office	7
				Public Transport	Public Transport	5
				Shopping	Services	5
	4 1			1 1 ()	· · · · · · · · · · · · · · · · · · ·	

Note: Ratio * represents the ratio of number of intra-com trips and number of inter-com trips; Per (%) * represents the percentages of link categories.

Community	Predominant Land Use Category	Percentage (%)	
	One & Two Family Buildings	34	
1	Multi-Family Walk-Up Buildings	14	
	Multi-Family Elevator Buildings	8	
	Public Facilities & Institutions	8	
	Open Space & Outdoor Recreation	7	
2	One & Two Family Buildings	49	
	Public Facilities & Institutions	15	
	Multi-Family Walk-Up Buildings	12	
	Multi-Family Elevator Buildings	9	
	Open Space & Outdoor Recreation	6	
3	Mixed Residential & Commercial Buildings	28	
	Public Facilities & Institutions	20	
	Multi-Family Elevator Buildings	15	
	Commercial & Office Buildings	14	
	Multi-Family Walk-Up Buildings	11	
	Commercial & Office Buildings	74	
	Transportation & Utility	10	
4	Mixed Residential & Commercial Buildings	6	
	Multi-Family Elevator Buildings	4	
	Public Facilities & Institutions	2	

Table 3. Names and percentages of the most predominant land use categories within convex hulls of typical communities selected.

4. Conclusions and Future Work

In this study, the "locality" of intra-urban spatial interactions was empirically investigated using LBSN data. The empirical results generated indicate that: (1) spatial interactions have a high degree of locality; (2) as a kind of personal characteristic, gender does not have a considerable impact on the locality of spatial interactions; and (3) "local" interactions likely cluster in some places within NYC.

Compared to other data sources, check-in data have both advantages and limitations. On the positive side, when compared to census travel data, LBSN check-in data are low cost (they can be downloaded for free) and have a large spatial scale. Geo-referenced check-in can also be at the street level, whereas census travel data are usually publicly available at the census tract level only. However, check-in data also have some limitations when used to study mobility. First, compared to some other mobility data (e.g., mobile phone and taxi trace data), geo-referenced check-ins are relatively sparse in spatial terms because of their relatively low record frequency. For instance, normally one record per minute is taken in a taxi trace record set for an individual user, while there are less than 10 check-ins per day recorded in one historic record set for an individual user. The taxi trace is; therefore, able to represent a trajectory between distinct locations in much more detail than check-in records. Second, geo-referenced check-ins are heterogeneously distributed; for instance, user-generated check-in data is abundant in urban areas, but sparse in rural regions. The result of this is that most of the existing studies of mobility using check-in data only take place in large cities. Also, the more effective application of information and communications technology (ICT) in urban areas is more likely to result in a higher number of user-generated check-ins. Added to this, users tend to carry out more check-ins at certain venues, such as airports, restaurants, shops, and railway stations [54] than at home, partly because there are not so many home-based venues offered by Foursquare. Compared to check-ins made at restaurants, shops, or in work, check-ins made at home are relatively rare, though the majority of people travel to and from home several times a day. Despite having a fine-grained resolution, check-in data is only used in a limited way to characterize home-to-work travel behavior, and this is one of the most important research fields within transportation studies, since check-ins that can indicate whether users are at home or not with much certainty are relatively rare. The third issue

is data representativeness. The mobility of a young person is better represented by check-in data than that of an elderly person, since most elderly people do not check-in frequently, or do not use social media at all [54]. This is a potential shortcoming when wishing to use check-in data to represent the human mobility of all users in a city. Finally, traveler profiles are not available or incomplete when using check-in data; therefore, some mobility patterns at the individual level (e.g., age, profession, and so forth) cannot easily be extracted from check-in data sets. The availability of user profiles issue is also an obstacle to any improvement of data representativeness.

In the future, some further aspects should be taken into account when carrying out an analysis of intra-urban spatial interactions. First, within the social media, social relationships are a vital aspect, and so need to be considered. However, the use of such data raises a question: To what extent can social relationships affect spatial interactions? Second, it will be interesting to examine in the future if and how urban form characteristics, or socio-economic characteristics, impact upon the "locality" of interactions. Finally, a combination of check-in data and other data (e.g., mobile phone data) seemingly has potential in the study of spatial interactions. When undertaking an empirical study, however, some obstacles such as the inconsistency of positional accuracy and recording frequency will need to be removed or reduced.

Acknowledgments: The author is thankful to the China Scholarship Council (CSC) for financial support with a PhD scholarship. The author acknowledges the financial support of the Deutsche Forschungsgemeinschaft (DFG) and Ruprecht-Karls-Universität Heidelberg (Heidelberg University) within the funding programme Open Access Publishing.

Conflicts of Interest: The author declares no conflict of interest.

References

- 1. Foursquare: Foursquare Category Hierarchy. Available online: https://developer.foursquare.com/ categorytree (accessed on 1 January 2016).
- 2. Foursquare: About Us. Available online: https://foursquare.com/about (accessed on 1 January 2015).
- Cheng, Z.; Caverlee, J.; Lee, K.; Sui, D. Exploring millions of footprints in location sharing services. In Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Spain, 17–21 July 2011.
- 4. Wu, L.; Zhi, Y.; Sui, Z.; Liu, Y. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE* **2014**, *9*, e97010. [CrossRef] [PubMed]
- Cho, E.; Myers, S.A.; Leskovec, J. Friendship and mobility: User movement in location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 21–24 August 2011; pp. 1082–1090.
- Noulas, A.; Scellato, S.; Mascolo, C.; Pontil, M. An empirical study of geographic user activity patterns in foursquare. In Proceedings of Fifth International AAAI Conference on Weblogs and Social Media, Palo Alto, CA, USA, 17–21 July 2011; pp. 570–573.
- Scellato, S.; Noulas, A.; Mascolo, C. Exploiting place features in link prediction on location-based social networks. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 21–24 August 2011; pp. 1046–1054.
- 8. Noulas, A.; Scellato, S.; Lambiotte, R.; Pontil, M.; Mascolo, C. A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE* **2012**, *7*, e37027. [CrossRef]
- Wei, L.Y.; Zheng, Y.; Peng, W.C. Constructing popular routes from uncertain trajectories. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 195–203.
- 10. Bao, J.; Zheng, Y.; Mokbel, M.F. Location-based and preference-aware recommendation using sparse geo-social networking data. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–9 November 2012; pp. 199–208.
- 11. 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]

- 12. Liu, Y.; Kang, C.; Gao, S.; Xiao, Y.; Tian, Y. Understanding intra-urban trip patterns from taxi trajectory data. *J. Geogr. Syst.* **2012**, *14*, 463–483. [CrossRef]
- 13. Stathopoulos, A.; Karlaftis, M. Temporal and spatial variations of real-time traffic data in urban areas. *Transp. Res. Rec. J. Transp. Res. Board* **2001**, *1768*, 135–140. [CrossRef]
- 14. Stathopoulos, A.; Karlaftis, M. A multivariate state space approach for urban traffic flow modeling and prediction. *Transp. Res. Part C Emerg. Technol.* **2003**, *11*, 121–135. [CrossRef]
- 15. Jiang, B.; Yin, Y.; Zhao, S. Characterizing the human mobility pattern in a large street network. *Phys. Rev. E* **2009**, *80*. [CrossRef] [PubMed]
- González, M.C.; Hidalgo, C.; Barabási, A.L. Understanding individual human mobility patterns. *Nature* 2008, 453, 779–782. [CrossRef] [PubMed]
- 17. Song, C.; Qu, Z.; Blumm, N.; Barabási, A.L. Limits of predictability in human mobility. *Science* **2010**, 327, 1018–1021. [CrossRef] [PubMed]
- Song, C.; Koren, T.; Wang, P.; Barabási, A.L. Modelling the scaling properties of human mobility. *Nat. Phys.* 2010, *6*, 818–823. [CrossRef]
- 19. Bazzani, A.; Giorgini, B.; Rambaldi, S.; Gallotti, R.; Giovannini, L. Statistical laws in urban mobility from microscopic GPS data in the area of Florence. *J. Stat. Mech. Theory Exp.* **2010**, *5*. [CrossRef]
- 20. Kang, C.; Ma, X.; Tong, D.; Liu, Y. Intra-urban human mobility patterns: An urban morphology perspective. *Phys. A: Stat. Mech. Its Appl.* **2012**, *391*, 1702–1717. [CrossRef]
- 21. Liang, X.; Zheng, X.; Lu, W.; Zhu, T.; Xu, K. The scaling of human mobility by taxis is exponential. *Phys. A: Stat. Mech. Its Appl.* **2012**, *391*, 2135–2144. [CrossRef]
- 22. Liang, X.; Zhao, J.; Dong, L.; Xu, K. Unraveling the origin of exponential law in intra-urban human mobility. *Sci. Rep.* **2013**, *3*. [CrossRef] [PubMed]
- 23. Zipf, G.K. The P1P2/D hypothesis: On the intercity movement of persons. *Am. Sociol. Rev.* **1946**, *11*, 677–686. [CrossRef]
- 24. Frank, L.; Pivo, G. Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking. *Transp. Res. Rec.* **1994**, *1466*, 44–52.
- 25. Van Acker, V.; Witlox, F. Commuting trips within tours: How is commuting related to land use? *Transportation* **2011**, *38*, 465–486. [CrossRef]
- 26. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012.
- 27. Gainza, X.; Livert, F. Urban form and the environmental impact of commuting in a segregated city, Santiago de Chile. *Environ. Plan. B Plan. Des.* **2013**, *40*, 507–522. [CrossRef]
- 28. Chowell, G.; Hyman, J.M.; Eubank, S.; Castillo-Chavez, C. Scaling laws for the movement of people between locations in a large city. *Phys. Rev. E* 2003, *68*, 661021–661027. [CrossRef] [PubMed]
- 29. Kang, C.; Liu, Y.; Wu, L. Delineating intra-urban spatial connectivity patterns by travel-activities: A case study of Beijing, China. In Proceedings of 23rd International Conference on Geoinformatics, Wuhan, China, 19–21 June 2015.
- 30. Batty, M. Cities as small worlds. Environ. Plan. B Plan. Des. 2001, 28, 637-638. [CrossRef]
- Noulas, A.; Blake, S.; Renaud, L.; Mascolo, C. Topological properties and temporal dynamics of place networks in urban environments. In Proceedings of the 24th International Conference on World Wide Web Companion, Florence, Italy, 18–22 May 2015; pp. 431–441.
- Fu, Y.; Liu, G.; Papadimitriou, S.; Xiong, H.; Ge, Y.; Zhu, H.; Zhu, C. Real estate ranking via mixed land-use latent models. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, 10–13 August 2015; pp. 299–308.
- Karamshuk, D.; Noulas, A.; Scellato, S.; Nicosia, V.; Mascolo, C. Geo-spotting: mining online location-based services for optimal retail store placement. In Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL, USA, 11–14 August 2013.
- 34. Georgiev, P.; Noulas, A.; Mascolo, C. Where businesses thrive: Predicting the impact of the Olympic Games on local retailers through location-based services data. In Proceedings of 8th AAAI International Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 2–4 June 2014.

- 35. Cranshaw, J.; Schwartz, R.; Hong, J.; Sadeh, N. The Livehoods project: Utilizing social media to understand the dynamics of a city. In Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (ICWSM-12), Dublin, Ireland, 4–7 June 2012; pp. 58–65.
- 36. Liu, Y.; Sui, Z.; Kang, C.; Gao, Y. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE* **2014**, *9*, e86026. [CrossRef] [PubMed]
- Balcan, D.; Colizza, V.; Goncalves, B.; Hu, H.; Ramasco, J.J.; Vespignani, A. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* 2009, *106*, 21484–21489. [CrossRef] [PubMed]
- 38. Thiemann, C.; Theis, F.; Grady, D.; Brune, R.; Brockmann, D. The structure of borders in a small world. *PLoS ONE* **2010**, *5*, e15422. [CrossRef] [PubMed]
- 39. Funk, S.; Salathé, M.; Jansen, V.A.A. Modelling the influence of human behaviour on the spread of infectious diseases: A review. *R. Soc.* **2010**, *7*, 1247–1256. [CrossRef] [PubMed]
- Belik, V.V.; Geisel, T.; Brockmann, D. The impact of human mobility on spatial disease dynamics. In Proceedings of 2009 IEEE International Conference on Computational Science and Engineering, Vancouver, BC, Canada, 29–31 August 2009; pp. 932–935.
- 41. Wesolowski, A.; Eagle, N.; Tatem, A.J.; Smith, D.L.; Noor, A.M.; Snow, R.W.; Buckee, C.O. Quantifying the impact of human mobility on Malaria. *Science* **2012**, *338*, 267–270. [CrossRef] [PubMed]
- 42. Miller, H.J. Tobler's first law and spatial analysis. Ann. Assoc. Am. Geogr. 2004, 94, 284–289. [CrossRef]
- 43. Toole, J.; Ulm, M.; González, M.; Bauer, D. Inferring land use from mobile phone activity. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; pp. 1–8.
- 44. Pan, G.; Qi, G.; Wu, Z.; Zhang, D.; Li, S. Land-use classification using taxi GPS traces. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 113–123. [CrossRef]
- 45. Yuan, Y.; Raubal, M.; Liu, Y. Correlating mobile phone usage and travel behavior—A case study of Harbin, China. *Comput. Environ. Urban Syst.* **2012**, *36*, 118–130. [CrossRef]
- 46. Guimerà, R.; Amaral, L.A.N. Cartography of complex networks: Modules and universal roles. *J. Stat. Mech. Theory Exp.* **2005**. [CrossRef]
- Guimerà, R.; Mossa, S.; Turtschi, A.; Amaral, L.A.N. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc. Natl. Acad. Sci. USA* 2004, 102, 7794–7799.
 [CrossRef] [PubMed]
- 48. Newman, M. Fast algorithm for detecting community structure in networks. *Phys. Rev. E* 2004, 69. [CrossRef] [PubMed]
- 49. Reichardt, J.; Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* 2006, 74. [CrossRef] [PubMed]
- Pons, P.; Latapy, M. Computing communities in large networks using random walks. J. Graph Algorithms Appl. 2006, 10, 191–218. [CrossRef]
- 51. Rosvall, M.; Bergstrom, C.T. Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. USA* **2007**, *105*, 1118–1123. [CrossRef] [PubMed]
- 52. Department of City Planning (DCP). City of New York: MapPLUTO. Available online: http://www.nyc.gov/html/dcp/ (accessed on 1 January 2016).
- 53. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
- 54. Factbrowser: The Research Discovery Engine. Available online: http://www.factbrowser.com/tags/ foursquare/ (accessed on 7 July 2010).



© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).