Exploring Human Movements in Singapore: A Comparative Analysis Based on Mobile Phone and Taxicab Usages

Chaogui Kang*
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA, USA
cgkang@mit.edu

Peking University
5 Yiheyuan Road
Beijing, P.R. China
chaoguikang@pku.edu.cn

Stanislav Sobolevsky
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA, USA
stanly@mit.edu

Yu Liu
Peking University
5 Yiheyuan Road
Beijing, P.R. China
liuyu@urban.pku.edu.cn

Carlo Ratti
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA, USA
ratti@mit.edu

ABSTRACT

Existing studies extensively utilized taxicab trips and individuals' movements captured by mobile phone usages (referred as "mobile phone movements" hereafter) to understand human mobility patterns in an area. However, all these studies analyze taxicab trips and mobile phone movements separately.

In this paper, we: (1) integrate mobile phone and taxicab usages together to explore human movements in Singapore and reveal that mobile phone movements as a general proxy to all kinds of human mobility has substantially different characteristics compared to taxicab trips, which are one of the frequently used means of transportation; (2) investigate the ratio of taxicab trips and mobile phone movements between two arbitrary locations, which not only characterizes taxicab demands between these locations but also sheds light on underlying land use patterns.

In details, we quantify the distinct characteristics of mobile phone movements and taxicab trips, and particularly confirm that the number of taxicab trips decays with distance more slowly compared to mobile phone movements. From a spatial network perspective, taxicab trips largely reflect interactions between further-separating locations than mobile phone movements, resulting in emergence of larger spatial communities (delineated based on people mobility) in Singapore.

The contribution of this research is two-fold: (1) we clarified the divergences between observed human mobility patterns based on taxicab and mobile phone data; (2) we implemented an integrated approach of taxicab and mobile phone usages for gaining more informative insights in population dynamics, transportation and urban configuration.

*Please direct all correspondence to Chaogui Kang at cgkang@mit.edu.

UrbComp '13, August 11-14, 2013, Chicago, Illinois, USA.

General Terms

Trajectory, Human Movements

Keywords

Mobility Patterns, Taxicab, Call Detail Records

1. INTRODUCTION

Understanding human mobility patterns has gain extraordinary attentions due to the rapid development of location acquisition technologies, complex network sciences and human dynamics. The hidden patterns of human movements in space and time are of great importance for traffic forecasting [9], tourism management [1], disease spread [22] and evolution of social relationships [4]. Ongoing studies are also trying to unify a framework for smart city [17], urban computing [28], personalized recommendation [26] and transportation intelligence [12][10][6] by tracking individuals' spatiotemporal activities.

Revisiting the early age of human mobility studies, which is well known as Hägerstrand's (1970) time geography [8], researchers extensively utilized travel diaries to analyze individuals' travel activities in space and time due to limited data collection ability. Fortunately, to date Information and Communication Technology (ICT) almost enables ubiquitous sensing of individuals' spatiotemporal movements in real-time. Diverse tracking techniques have been adopted, including circulation of bank notes [2], handheld GPS devices [27], Call Detailed Records (CDRs) [7], smart cards [13] and check-ins [3].

On the other hand, these distinct techniques capture substantially different mobility patterns, raising the challenge of activity interpretation and modeling. For instance, by tracking the geographic circulation of bank notes in the United

States, [2] inferred that human movements in space in general follows the Lévy flight. However, those bank notes were carried by different people during the study period, making the inherent mechanism of human movements still unclear. In 2008, [7] adopted the mobile phone data for mobility analysis, and concluded that the (truncated) power-law of collective human mobility emerged from a convolution of population heterogeneity and individual Lévy trajectories. Meanwhile, mobility patterns derived from taxicab trips are better fitted by the exponential law than the Lévy flight model [11].

Different techniques also bring ambiguity of the definition of human movements due to their inconsistent spatiotemporal data qualities. There exist at least three distinct strands: trip-, shape- and sampling-based methods to define a single displacement. Taxicab trajectory is a typical illustration of the trip-based method in that the origin and destination (OD) of individual movement is given explicitly. In handheld GPS trajectory, OD is unknown and thus applicable for the shape-based method. In this case, three different preprocessing methods, namely rectangular-, angle- and pause-based models, have been widely applied to extract trips [19]. Mobile phone positioning captures individual's discrete locations randomly in time as well as with low spatial resolution. It thus follows the sampling-based framework.

In general, divergences of existing observations of human mobility patterns are attributed to four reasons: different transport modes (walking, bus, train, flight and etc.), different population groups (age, racial, occupation and etc.), different geographical environments (road network, region, geology and etc.), and different spatial scales (intra-urban, inter-urban and etc.). As known, transport tools play substantially distinct roles in transportation systems and individuals' travel behaviors. Data associated with different transport modes generally depict human movements with different purposes and at different spatial scales. People with different social-economic backgrounds always utilize public transport modes and urban spaces distinctly. Geographical environments such as land uses and accessibility also have remarkable influences upon human mobility patterns. Understanding constraints of aforementioned factors upon observed human mobility patterns is meaningful for human mobility modeling.

In this article, we focus on exploring human movements in Singapore using both mobile phone and taxicab data, and target to clarify the distinct patterns derived as well as the underlying mechanism. Intuitively, mobile phone movement is a general proxy to all kinds of human mobility while taxicab trip is one of the frequently used means of transportation. Taxicab trip captures only a segment of individual's movements. Mobile phone trajectory, though deviating from real human trajectory, can capture individual's continuous and longitudinal movements. Thus, a comparative analysis of taxicab trip and mobile phone movements can

1. Deepen our understanding of peoples' taxicab usages and general activity patterns, which is informative for land use classification;

2. Uncover similarities and differences between human mobility patterns derived from taxicab and mobile phone data, respectively;

3. Unveil the relationship between dynamic population distribution (based on mobile phone usages) and taxicab usage in an area.

The contribution of this research is two-fold: (1) we clarified the divergences between observed human mobility patterns based on taxicab and mobile phone data; (2) we proposed an integrated approach of taxicab and mobile phone usages for gaining more informative insights in population dynamics, transportation and urban configuration.

The remainder of this article is organized as follows. Section 2 describes the data preprocessing procedure for discretizing the study area and trip extraction. In Section 3, we present the differences between taxicab trips and mobile phone movements in terms of spatial pattern, distance decay, and community structure. Section 4 provides a brief discussion about potential directions of this research. In final, Section 5 summarizes the contribution of this paper and concludes the divergences between the mobility patterns derived from taxicab and mobile phone usages.

2. DATA PREPROCESSING

2.1 Data Description

In this research, we extracted GPS logs of $15,000+$ taxicabs (**D1**) from February 27th, 2011 to March 5th, 2011 and CDRs of $2,000,000+$ mobile subscribers (**D2**) from March 27th, 2011 to April 2th, 2011 in Singapore over a whole week, respectively.

In **D1**, taking February 28th, 2011 as an example, the total number of GPS points for all taxicabs is 17,653,329. Each point is associated with longitude, latitude, time, velocity, and status (occupied/ unoccupied) of the given taxicab. The time intervals between consecutive GPS points vary along with different taxicabs as well as time of the day, while 1 second, 3 seconds and 5 seconds are the most common values.

In **D2**, the number of mobile subscribers accounts for approximate 40% of the total population in Singapore. For each subscriber, the CDRs contain voice, SMS and data logs. The antennas routing the mobile phone usages roughly construct the spatiotemporal trajectory of each subscriber. In total, there are 9,979 antennas across entire Singapore (Figure 1). In the downtown area (the south part of Singapore), the distance between two neighboring antennas is about 50 meters on average, which is capable to positioning mobile subscribers in very fine spatial resolution. In the surrounding area (excluding the open spaces in central and western Singapore), the corresponding value is about 200-500 meters.

2.2 Discretizing The Study Area

Technically, the precision of GPS log is usually higher than meters, while the precision of mobile positioning is much lower. To make taxicab and mobile phone trajectories comparable, we transform these two types of trips into an identical resolution. Though the Theissen partitioning is a most intuitive way to discretize the study area, it has several drawbacks, particularly in the case study of area with dense antennas like Singapore. First, the size of resulting Voronoi polygons varies with the density of antennas. In area with sparse antennas, the Voronoi polygon might largely exceed the service area of the antenna, making the estimation of trip origin and destination biased. Second, the service ar-

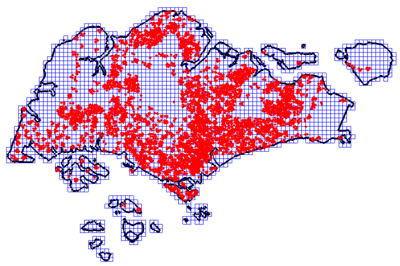


Figure 1: The 500 m by 500 m grids (blue-rectangle) and the antennas (red-dot) covering entire Singapore. The total number of grids is 2,640 and the antennas 9,979.

eas of antennas might overlap with each other in area with dense antennas due to the service capability of a single antenna. Thus, a subscriber might not be within the Voronoi polygon of the antenna routing the mobile phone activity. Considering that the average distance between two neighboring antennas is less than 500 meters across entire Singapore, we thus discretize the studies area into 500 m by 500 m grids (Figure 1). This uniform partitioning makes the origin and destination of each trip at the same scale and largely eliminates the biases discussed above.

2.3 Trip Extraction

In **D1**, statuses of individual taxicab, such as FREE, Passenger on Board (POB), and PAYMENT at an instant time are explicitly recorded. More specifically, when the status is FREE, the taxicab is available for taking a passenger. When the taxicab is occupied, its status turns from FREE to POB at the pick-up location. When the taxicab drops off the passenger at the destination, its status firstly turns from POB to PAYMENT, and then to FREE again to search passengers. The trajectory between the first POB status appearing at the pick-up point and the successive PAYMENT status at the drop-off point captures a taxicab trip in space and time. Table 1 lists the number of taxicab trips in different days over a whole week. On average, the total number of daily taxicab trips is about 400,000 in Singapore, while in weekends the number is slightly smaller than weekdays. Additionally, the number of trips increases day by day over weekdays, implying the raising demands of taxicabs when approaching the weekends. In this paper, we reassign the pick-up and drop-off points of taxicab trips into the 500 m by 500 m grids within which they fall. Thereafter, all taxicab trips are simplified as the displacements from the originating grid to the terminating grid.

In **D2**, two distinct approaches are widely applied to extract individual's movements as well as OD information. The first one is simply taking the change of locations between two consecutive mobile phone activities as a trip, while the second one requires an additional procedure of identifying the anchor points of people's call activities, and then defines movements between different anchor points as trips. Though the former method covers majority of individuals' movements, it is highly dependent on where, when and the frequency subscribers use their mobile phones. Moreover, the major drawback is that numerous short trips will be detected due to localization errors and users making consecutive network connections in the same area. In this paper, we reassign position of each mobile phone usage into the grid within which the antenna locates following the framework presented in Section 2.2. Then, if the user moves between two different grids when making two consecutive network connections, we define this movement as a trip. Note that in this paper we refer them as trips in parallel with taxicab trips for convenience. In general, this approach follows the latter method as discussed above. Considering that the minimum distance between two grids is 500 m, the approach: (1) considerably reduces short-distance trips within an identical grid as well as localization errors; (2) largely captures all kinds of individuals' movements. However, it is also noteworthy that these obtained trips can be still biased in that their ODs might deviate from individuals' real movements. The number of mobile phone movements also generally increases from Monday to Friday (Table 1).

3. DIVERGENCES BETWEEN TAXICAB TRIPS AND MOBILE PHONE MOVEMENTS

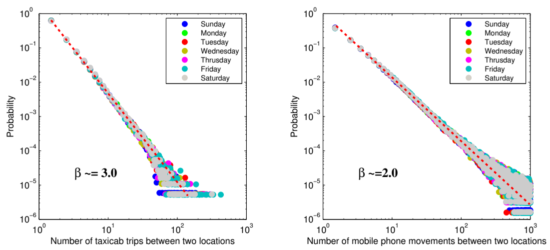
3.1 Spatial Distributions

Taxicab is a typical transport mode relying upon underlying urban configuration. On the one hand, at city scale taxicabs are almost crowded in urban areas, particularly the commercial districts. On the other hand, taxicabs are the dominantly used transport mode between certain specific places, such as airport and downtown, workplaces and residential sections. Figure 2a demonstrates the probability distribution of the number of taxicab trips between paired locations in Singapore per day over a whole week. For the majority of location pairs, the number of taxicab trips between them is less than 5 in each day. However, between certain locations the number of taxicab trips is 10 or 100 times higher (see the tail in Figure 2a). Under closer scrutiny, we find that the probability distribution well follows a power-law with an exponent close to 3.0, which reveals a highly spatial heterogeneity of taxicab trips distributed in the city. The number of mobile phone movements between two locations is generally much higher than the number of taxicab trips. As mentioned before, this is because taxicab trips capture only a very small part of people's daily movements in Singapore. The distribution of mobile phone movements between different locations also well follows the power-law (Figure 2b). However, the exponent of mobile phone movements, which is about 2.0, is smaller than the exponent of taxicab trips. It tells that mobile phone movements are distributed in space more evenly than taxicab trips. The low correlation between taxicab trips and mobile phone movements between paired locations, with a Pearson's Correlation Coefficient (PCC) about 0.13, further confirms the differences between taxicab trips and mobile phone movements. Figure 2 also implies that the collective human daily movements are very regular in different days over the whole week in terms of both taxicab trips and mobile phone movements.

To further explore the spatial distribution of taxicab trips and mobile phone movements, we aggregate the outgoing and the incoming trips for each location. Figure 3 shows the case for Sunday only considering that the daily patterns of

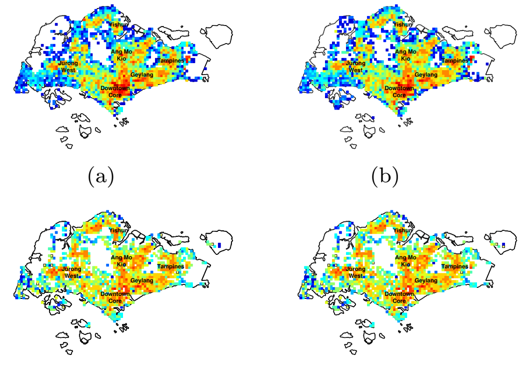
Table 1: Number of taxicab trips (D1) and mobile phone movements (D2) per day over a whole week

	SUN	MON	TUE	WED	THU	FRI	SAT
D1	382,720	407,405	411,874	414,755	420,948	439,937	432,125
D2	16,234,488	19,891,333	19,753,961	20,608,775	20,673,944	22,223,400	18,767,840



(a) (b)

Figure 2: Number of taxicab trips (a) and mobile phone movements (b) between two locations per day over a whole week. Both the distributions of taxicab trips and mobile phone movements overlap with each other and generally follow a power-law function with the scaling exponent 3.0 and 2.0, suggesting the high heterogeneity and regularity of collective human daily movements.



(a) (b)

(c) (d)

Figure 3: Distribution of the outgoing/incoming trips (top: taxicab; bottom: mobile phone) at each location per day over a whole week. In details, the pick-up locations of taxicab trips are concentrated in space (left), whereas the drop-off locations distribute more uniformly across Singapore (right). The mobile phone movements are normalized to the same scale as the taxicab trips.

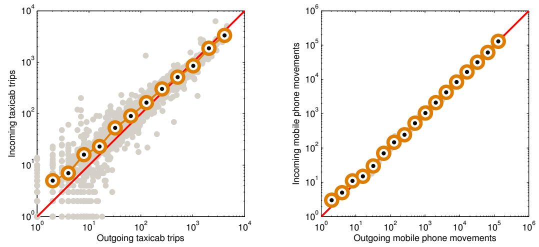
taxicab and mobile phone trips are highly homogenous over the whole week as discussed above. Obviously, the numbers of outgoing and incoming trips at different locations are quite different in the city. Both outgoing and incoming taxicab trips are highly clustered at the downtown Singapore and the airport (Figure 3a and 3b). It reveals that taxicabs are extensively taken transport tools serving the airport and the downtown area. In more details, the pick-up and drop-off points of taxicab trips decrease from the urban center/subcenter (Downtown Core, Jurong West, Yishun, Ang Mo Kio, Geylang and Tampines) to the surrounding areas, implying the high correlation between taxicab usages and underlying urban configuration. For mobile phone movements, the outgoing and incoming flows are more evenly spatially distributed in Singapore (Figure 3c and 3d).

Another interesting finding is the symmetric properties of taxicab trips and mobile phone movements in Singapore. Figure 4 illustrates the correlation between the outgoing and the incoming trips at each location. Generally, the outgoing and the incoming trips are highly symmetric. The PCC of the outgoing and the incoming taxicab trips is 0.95 and the value of mobile phone movements 0.99. On the one hand, this observation tells that taxicab trips and mobile phone movements are highly regular and repeatable in space. On the other hand, it demonstrates that taxicab trips are more stochastic than mobile phone movements. As shown in Figure 4a, at most locations the total number of taxicab trips is less than 100, and the number of incoming taxicab trips is larger than the number of outgoing taxicab trips. However, at places with large number of outgoing/incoming taxicab trips, the situation is reversed: the number of incoming taxicab trips is generally smaller than the number of outgoing taxicab trips. In other words, the drop-off points of taxicab trips distribute more uniformly in space compared to the pick-up points. We interpret that the pick-up locations of taxicab trips are largely concentrated in the downtown area, whereas the drop-off locations distribute more randomly in space. Meanwhile, the high symmetry of mobile phone movements reflects constraints of individuals' daily routines such as commutes between home and workplace.

3.2 Displacements

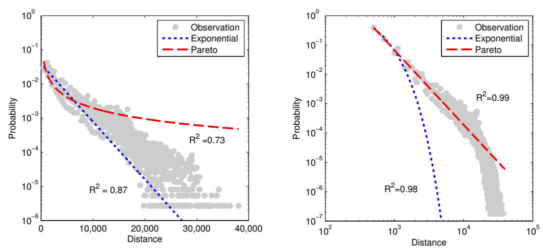
The distance decay effect of individual human movements is a fundamental and meaningful phenomenon in geography that gains interests of researchers in various fields. The distance decay of human travels relies on both geographical environment and transport modes. In the context of this study, we have two distinct types of trips, enabling us to decouple the impact of transport modes upon human movements in the same area.

Figure 5 demonstrates the distance distributions of taxicab trips and mobile phone movements in Singapore. As expected, the proportion of taxicab trips within 1,000 meters is much lower than its counterpart of mobile phone movements. However, due to the particular role of taxicabs in public transportation systems, the proportion of taxicab trips larger than 3,000 meters is much higher than its counterpart of mobile phone movements. Besides, considering that mobile phone movements are mixtures of various transport modes including walking, bus, subway as well as taxicab, its distance decay effect is inherently stronger than that of taxicab trips. Therefore, the distribution of displacements of taxicabs is more flat than the distribution of displacements of mobile phone movements. This observation explicitly confirms that taxicab trips, which are constraint by travel distance and time, decay more slowly than mobile phone movements. Quantitatively, the distribution of displacements of mobile phone movements within 10 km well



(a) (b)

Figure 4: Symmetric properties of the outgoing/incoming taxicab trips (a) and mobile phone movements (b) at each location. The pie-line represents the average of each scatter. At most places, the total number of taxicab trips is less than 100 and the number of incoming taxicab trips is generally larger than the number of outgoing taxicab trips. However, at those places with large number of outgoing/incoming taxicab trips, the situation is reversed. In contrary, mobile phone movements are highly symmetric across the study area.



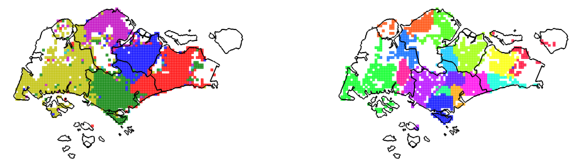
(a) (b)

Figure 5: The distance decays of taxicab trips (a) and mobile phone movements (b). In general, mobile phone movements follow the power-law distribution with an exponent about 2.5, while taxicab trips follow the exponential distribution with an exponent about 2.9 (when normalizing distance by its mean and standard deviation).

follows the power-law with an exponent about 2.5, while in the tail the exponent is much larger due to constraint of the boundary of Singapore (Figure 5b). The distribution of displacements of taxicab trips, however, well follows the exponential law with an exponent about 2.9 (Figure 5a).

The difference between the distributions of displacements of taxicab trips and mobile phone movements might be attributed to two factors. First and foremost, taxicab is a kind of transport tool mainly used in specific circumstances by people. It captures only a small part of individuals' daily movements, particularly the medium and long distance travels as well as those movements originating/terminating at specific areas. Besides, taxicab trips highly relies the underlying road network, making it less flexible and accessible in certain areas. Another potential reason is that mobile phone movements depict inherent biases regarding individual human movements. The temporal-sparse and spatial-coarse nature of mobile phone data as well as individuals' habits of mobile phone usages make it hard to derive human movements accurately.

3.3 Regionalization



(a) (b)

Figure 6: The communities derived from taxicab trips (a) and mobile phone movements (b) in Singapore. The 5 communities in the network of taxicab trips well match the 5 administrative regions in Singapore. Particularly, the airport is clustered together with the downtown (green). The network of mobile phone movements is partitioned into 16 small communities, which are generally sub-divisions of the 5 regions in Singapore.

From the network-based perceptive, human movements reflect spatial interactions between different locations and the volume of human movements represents the strength of their interaction. As discussed above, taxicab trips largely reflect interactions between further-separating locations than mobile phone movements. Following this framework, we uncover how taxicab trips and mobile phone movements can be applied to identifying different regions in Singapore.

In this analysis, each grid is abstracted as a node, v_i ($i = 1, 2, \ldots, 2640$), and the number of trips from node i to node j as weight of the link E_{ij} between them. Based on taxicab trips and mobile phone movements, we thus obtain two direct spatial networks $S1$ and $S2$ consisting of the $N = 2,640$ grids in Singapore. Note that E^1_{ij} for $S1$ and E^2_{ij} for $S2$ are the number of taxicab trips and the number of mobile phone movements from v_i to v_j respectively. But, for simplicity, hereafter we use E_{ij} to denote the weight from v_i to v_j as the general form. By applying the community detection algorithm to both $S1$ and $S2$, results as shown in Figure 6 confirm that both taxicab trips and mobile phone movements are capable for uncovering spatial cohesive communities inside the city. In $S1$, we find 5 large communities in Singapore as Central Region (green), East Region (red), North Region (purple), North-East Region (blue) and West Region (orange) (see Figure 6a). In $S2$, we find 16 more fine-grained spatial cohesive communities (Figure 6b) that are generally sub-regions of the 5 regions derived from taxicab trips.

Interestingly, the 5 regions derived from taxicab trips well match the 5 urban planning subdivisions demarcated by the Urban Redevelopment Authority of Singapore. Particularly, the airport at the east of Singapore is clustered together with the Central Region, revealing the heavy taxicab traffic between the downtown area and the airport in Singapore. Furthermore, we adopt the clustering comparison approach to quantify the similarity between the resulting spatial communities and the administrative divisions in Singapore [16][5]. The upper-triangular and the lower-triangular of Table 2 tabulate the Rand (RI) and the Fowlkes-Mallows (FM) similarity indices between the regions defined by taxicab trips, mobile phone movements and administrative divisions. The

Table 2: The similarity between derived communities and administrative regions in Singapore (upper-triangular: Rand Index; lower-triangular: Fowkles-Mallows Index)

	Taxicab	Mobile	5 Regions
Taxicab	1	0.8309	0.8326
Mobile	0.4633	1	0.8300
5 Regions	0.6222	0.5129	1

RI and FM both confirm the high coincidence between the 5 taxicab-based regions, the 16 mobile-based districts and the 5 administrative regions in Singapore. Note that RI only takes into account the matching part of two partitions, while FM relies on both the matching part and the mismatching part of two partitions.

The implication of this finding is two-fold: (1) human movements are significantly constraint by the underlying urban structure. In this research, trips within a given administrative region are much more than trips across different administrative regions in Singapore; (2) different types of trips can be applied to detect regions at different scales, implying that the constraint of urban structure is different for different transport modes. As discussed in Section 3.2, the distance decay of taxicab trips is slower than mobile phone movements. Taxicab trips mainly capture the travel-activities between further-separating locations than mobile phone movements. In Singapore, the analysis depicts that taxicab trips highlight the connections between districts within the identical higher-level region, resulting in the combination of certain districts into a large community in space. More generally, the emergence of communities in spatial-embedded networks depends on the strength of distance decay effect of the underlying spatial interaction systems.

4. AN INTEGRATED APPROACH

In a city, the ratio of taxicab trips and overall human movements between two given locations is of great importance for taxicab demand estimation and dispatch. We thus calculate the ratios of taxicab trips and mobile phone movements between two arbitrary locations in Singapore. For most paired places, the proportion of taxicab traffic is about 10%, demonstrating the important role that taxicabs play in the transportation system. The result also reveals that the proportion of taxicab trips is highly heterogeneous for different paired places. At certain places, the ratio is very low (0.1% - 1%), implying that taxicab is a rarely used transport mode in these areas. One possible reason is that other public transport modes such as subway and buses largely substitute the role of taxicabs at these places. Also, the average ratio of taxicab trips and mobile phone movements longer than 5,000 m is highly stable between 30% and 40%. It further confirms that taxicab plays a very dominant role in medium and long distance travels in Singapore. Besides, between some paired places the ratios of taxicab trips and mobile phone movements exceed 1.0, suggesting that mobile phone data have definitely exclude a part of people movements that are captured by the taxicab trips. This also provides strong evidence that mobile phone dataset has limits for inferring general people movements patterns.

It is also meaningful to testify whether the ratio of taxicab trips and mobile phone movements can reveal underlying urban environment at each location. As shown in Figure 7, the ratios at different locations in Singapore show distinctly different spatial patterns compared to the spatial distributions of taxicab trips and mobile phone movements (refer to Figure 3). Almost all the locations with a high ratio of taxicab trips are within the downtown core. Figure 8 demonstrates the strong positive correlation between the ratio of taxicab usage and underlying population (derived from aggregate mobile phone traffic) at each location in Singapore. In this sense, at macro scale the spatial distribution of ratios well follows the general urban structure ("core - urban - suburban"). Moreover, a preliminary and meaningful finding in this integrated analysis is that the ratio of taxicab and mobile phone trips is a good additional indicator for land use inference based on mobile phone and taxicab usages. We find that a hotspot emerges at the southwestern part of Singapore, which cannot be identified by taxicab trips or mobile phone movements separately. By checking with the official land use map, we confirm that this region (the Pioneer district in Singapore) is the largest developing area in Singapore, which depicts quite different activity patterns with other locations associating with small taxicab trips and mobile phone movements. The implication is that an integrated framework of multi-sourced mobility datasets is promising for better land use inference and deeper understanding of urban configuration.

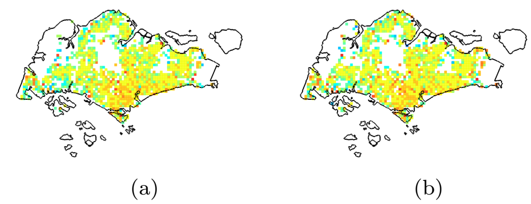


Figure 7: Distribution of the ratios of taxicab trips and mobile phone movements at each location across Singapore (a: outgoing, b: incoming). A hotspot emerges at the southwestern corner of Singapore, which cannot be identified by taxicab trips or mobile phone movements solely.

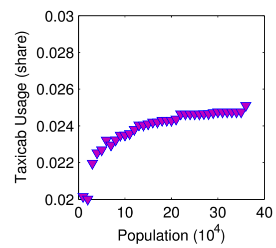


Figure 8: The relationship between ratio of taxicab usage and underlying population.

5. DISCUSSIONS

This article has uncovered substantial differences between taxicab trips and mobile phone movements in terms of spatial distribution, distance decay and community structure. However, all the analyses are based on the trip patterns aggregated over a whole day. Considering that the temporal pattern of trips in a day is more informative than the aggregated daily pattern, we will focus on the temporal variations of trips at each location as well as between two arbitrary locations in Singapore in future works. For instance, the temporal pattern of differences between the incoming and outgoing taxicab trips at a given location permits us to infer the underlying land use with very high accuracy [14]. Together with the POIs, these trips are applicable for differentiating functional regions in a city [23]. Similarly, the signatures of mobile phone usages along time also allow land use inference in a city [18]. In the context of this research, we can adopt this geographical classification framework and further leverage the incoming and outgoing trips from taxicab and mobile phone usages to infer the land uses in Singapore, respectively. Additionally, we will also verify whether the temporal pattern of ratios of taxicab trips and mobile phone movements can be applied for better land use inference.

In a second strand, it will be interesting to discuss how distance decay effect determines the formation of communities in spatial-embedded networks. As shown in Figure 6, networks of taxicab trips and mobile phone movements have substantial different community structures. The size of communities in a network with low distance decay is larger than the size of communities in a network with fast distance decay. All the communities are highly cohesive in space, while there is not any external spatial constraint in the community detection algorithm. This phenomenon is an explicit illumination of Tobler's first law of geography (TFL) that "Everything is related to everything else, but near things are more related than distant things" [21]. Tracing back to 2003, the TFL forum in 2003 AAG annual meeting has already treated the network sciences as an empirical validation of TFL, and implicitly tried to unify a framework of (spatial) network sciences and geography [20]. Particularly, the emergence of cohesive communities explicitly illustrates how TFL works in spatial-embedded networks. More importantly, in this paper we have not only revealed the correlation between distance decay and community structure, but also, at least partially, clarified the mechanism how distance decay determines the community structure of spatial-embedded networks. Considering that different travel activities such as flight travel, taxicab usage and walking depict different extents of distance decay, a comparative analysis of the distance decay effects of different transport modes as well as the resulting community-structures will be of great interest in the future. From a network-based perspective, it will provide an intuitive way to uncover the hierarchical structure of a city across various scales.

Finally, mobile phone activity as a proxy of underlying population dynamics has largely stimulated urban analysis in real-time. It provides a much-preferred tool than traditional methods based on static population distribution and land use maps. Taking transportation analysis as an example, mobile phone usages can be applied for estimating road traffic volume, transport demands and OD matrices. In past, cabdrivers almost adopt their passenger finding strategy according to personal experience (and knowledge) or recommenders based on, for instance, "hot" parking places [25] and ad-hoc route finding algorithm [24]. Passenger recommenders and carpool services [15] usually take taxicab or customer distribution into account exclusively. In the context of this paper, we further propose to testify the relationship between mobile phone activity and taxicab traffic. That is, if at a given place the mobile phone activities are highly correlated to the number of taxicab trips, we can use mobile phone data to predict the volume of potential taxicab passengers in real time. It will greatly benefit cabdriver's intelligence and real-time transportation system management.

6. CONCLUSION

In this article, we conduct a comparative analysis of human mobility patterns in Singapore based on two distinct datasets: the mobile phone usages and the taxicab trajectories. Firstly, this analysis provides an explicit clarification to the divergences in observed human mobility patterns based on taxicab and mobile phone data. Secondly, it also proposes an integrated approach of taxicab and mobile phone usages for gaining more informative insights in population dynamics, transportation and urban configuration.

Overall, mobile phone movements as a general proxy to all kinds of human mobility has substantial different characteristics to taxicab trip, which is just one of the frequently used means of transportation. Spatially, the distribution of ODs of taxicab trips and mobile phone movements are quite different. Taxicab trips are more concentrated (in the downtown area) and asymmetric than mobile phone movements. In terms of distance decay, taxicab trips are constraint by travel distance and time, and thus decay more slowly than mobile phone movements. In terms of community structure, taxicab trips largely reflect interactions between further-separating locations compared to mobile phone movements, resulting in emergence of larger spatially cohesive communities in Singapore. As a proposed integrated approach, the ratio of taxicab trips and mobile phone movements between two arbitrary locations in the city can shed light on underlying land uses and help to predict taxicab demands between the two locations.

Besides, in this paper we also present a brief discussion about the potential works based on taxicab and mobile phone usages. It can broaden our understanding of land use inference and transportation analysis using taxicab and mobile phone data. Among, a promising direction is to utilize datasets with regard to human movements associating with various transport modes to uncover the hierarchical community structure of the city across various scales.

7. ACKNOWLEDGMENTS

The authors thank the National Science Foundation, the AT&T Foundation, the MIT SMART program, the MIT CCES program, Audi Volkswagen, BBVA, the Coca Cola Company, Ericsson, Expo 2015, Ferrovial, GE, all the members of the MIT SENSEable City Lab Consortium and the China Scholarship Council for supporting the research.

8. REFERENCES

[1] R. Ahas, A. Aasa, Ülar Mark, T. Pae, and A. Kull. Seasonal tourism spaces in estonia: Case study with mobile positioning data. *Tourism Management*, 28(3):898–910, 2007.

[2] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[3] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui. Exploring millions of footprints in location sharing services. In *Proceedings of the 5th International AAAI*

Conference on Weblogs and Social Media (ICWSM), pages 81–88, 2011.

[4] E. Cho, S. A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1082–1090, New York, NY, USA, 2011. ACM.

[5] E. B. Fowlkes and C. L. Mallows. A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association*, 78(383):553–569, 1983.

[6] Y. Gao, P. Xu, L. Lu, H. Liu, S. Liu, and H. Qu. Visualization of taxi drivers' income and mobility intelligence. In *Advances in Visual Computing*, pages 275–284. Springer, 2012.

[7] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[8] T. Hägerstrand. What about people in regional science? *Papers in Regional Science*, 24(1):7–24, 1970.

[9] B. Jiang. Characterizing the human mobility pattern in a large street network. *Physical Review E*, 80(2):021136, 2009.

[10] B. Li, D. Zhang, L. Sun, C. Chen, S. Li, G. Qi, and Q. Yang. Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 63–68, 2011.

[11] X. Liang, X. Zheng, W. Lv, T. Zhu, and K. Xu. The scaling of human mobility by taxis is exponential. *Physica A: Statistical Mechanics and its Applications*, 391(5):2135–2144, 2012.

[12] Z. Liao, Y. Yu, and B. Chen. Anomaly detection in gps data based on visual analytics. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, pages 51–58, 2010.

[13] L. Liu, A. Hou, A. Biderman, C. Ratti, and J. Chen. Understanding individual and collective mobility patterns from smart card records: A case study in shenzhen. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, pages 1–6, 2009.

[14] Y. Liu, F. Wang, Y. Xiao, and S. Gao. Urban land uses and traffic "source-sink areas": Evidence from gps-enabled taxi data in shanghai. *Landscape and Urban Planning*, 106(1):73–87, 2012.

[15] S. Ma, Y. Zheng, and O. Wolfson. T-share: A large-scale dynamic taxi ridesharing service. In *Proceedings of 29th IEEE International Conference on Data Engineering, ICDE 2013*, 2013.

[16] W. M. Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

[17] C. Ratti. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 33(5):727–748, 2006.

[18] J. Reades, F. Calabrese, and C. Ratti. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 36:824–836, 2009.

[19] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong. On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, 19(3):630–643, 2011.

[20] D. Z. Sui. Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers*, 94(2):269–277, 2004.

[21] W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46:234–240, 1970.

[22] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee. Quantifying the impact of human mobility on malaria. *Science*, 338(6104):267–270, 2012.

[23] J. Yuan, Y. Zheng, and X. Xie. 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*, KDD '12, pages 186–194, New York, NY, USA, 2012. ACM.

[24] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. T-drive: driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 99–108, New York, NY, USA, 2010. ACM.

[25] J. Yuan, Y. Zheng, L. Zhang, X. Xie, and G. Sun. Where to find my next passenger? In *Proceedings of the 13th International Conference on Ubiquitous Computing*, pages 109–118. ACM, 2011.

[26] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 1029–1038, New York, NY, USA, 2010. ACM.

[27] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma. Understanding mobility based on gps data. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 312–321, New York, NY, USA, 2008. ACM.

[28] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *Proceedings of the 13th International Conference on Ubiquitous Computing*, UbiComp '11, pages 89–98, New York, NY, USA, 2011. ACM.