Journal of Transport Geography 70 (2018) 78–90

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Do different datasets tell the same story about urban mobility — A comparative study of public transit and taxi usage

Xiaohu Zhang[a], Yang Xu[b,*], Wei Tu[c,d], Carlo Ratti[d]

[a] Singapore-MIT Alliance for Research and Technology, 1 Create Way, Singapore
[b] Department of Land-Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[c] Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, School of Architecture and Urban Planning & Research Institute for Smart Cities, Shenzhen University, Shenzhen, China
[d] Senseable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

ARTICLE INFO

Keywords:
Human mobility
Urban spatial structure
Smart card transactions
Taxi data
Spatiotemporal analysis

ABSTRACT

Understanding human movements and their interactions with the built environment has long been a research interest in transport geography. In recent years, two important types of urban mobility datasets — smart card transactions and taxi GPS trajectories — have been used extensively but often separately to quantify travel patterns as well as urban spatial structures. Despite the fruitful research outcomes, the relationships between different types of transport flows in the same geographic area remain poorly understood. In this research, we propose an analytical framework to compare urban mobility patterns extracted from these two data sources. Using Singapore as a case study, this research introduces a three-fold comparative analysis to understand: (1) the spatial distributions of public transit and taxi usages and their relative balance; (2) the distance decay of travel distance, and (3) the spatial interaction communities extracted from the two transport modes. The research findings reveal that the spatial distributions of travel demand extracted from the two transport modes exhibit high correlations. However, more in-depth analysis (based on rank-size distribution and log odds ratio) reveals a higher degree of spatial heterogeneity in public transit usage. The travel distance of trips from public transit decays faster than that of taxi trips, highlighting the importance of taxis in facilitating long-distance travels. Both types of trips decay much faster when travel distance is beyond 20 km, which corresponds to the average distance from the urban periphery to the center. The spatial interaction communities derived from public transit are different on weekdays and weekends, while those of taxis show similar patterns. Both transport modes yield communities that reveal the city's polycentric structure, but their differences indicate that each of the transport modes plays a specific role in connecting certain places in the city. The study demonstrates the importance of comparative data analytics to urban and transportation research.

1. Introduction

The past two decades have witnessed an exponential growth of scientific research that characterizes human mobility and their interactions with the built environment. The rapid developments of information and pervasive sensing technologies have produced – especially in urban settings – a wide spectrum of human mobility datasets, empowering researchers to tackle critical questions in transport planning (Santi et al., 2014; Alexander et al., 2015; Tu et al., 2016), disease control (Bengtsson et al., 2011; Wesolowski et al., 2012), and social dynamics (Cho et al., 2011; Xu et al., 2017; Sun et al., 2013). The big data evolution has spurred "a new science" or many new sciences of cities, from which urban environments can be better understood as systems of networks and flows (Batty, 2013).

The networks and flows embedded in cities are defined by researchers through different types of datasets, resulting in a multi-faceted view of urban mobility patterns. For example, many studies have been conducted in recent years to quantify intra-urban mobility patterns based on taxicab usages (Wang et al., 2015; Liu et al., 2015; Kang and Qin, 2016), public transit data (Zhong et al., 2015; Liu et al., 2009), and mobile phone records (Gao et al., 2013; Ahas et al., 2010; Xu et al., 2016). Despite the fruitful research outcomes, most of the existing studies focus on a single type of human mobility dataset, which yields into insights that are somewhat isolated. The relationships between different types of networks and flows in the same geographic area – such as a city – remain poorly understood (Tu et al., 2018). It is, therefore, important to combine different data sources to obtain a

* Corresponding author.
E-mail address: yang.ls.xu@polyu.edu.hk (Y. Xu).

https://doi.org/10.1016/j.jtrangeo.2018.05.002
Received 31 October 2017; Received in revised form 25 March 2018; Accepted 3 May 2018
0966-6923/

comprehensive view of the spatial structures and organizations of cities. This will shed light on the bias when each data type is used alone to represent the dynamics of urban systems. More importantly, it would generate a deeper understanding of the interplay among different socio-economic processes.

In this research, we propose an analytic framework to compare urban travel patterns and the associated urban spatial structures extracted from smart card transactions and taxi GPS trajectories. The two types of datasets are widely used but often separately in revealing urban mobility patterns. Using Singapore as a case study, this work aims to fill the research gap by answering the following research question – do public transit and taxi usages in a city produce similar patterns of travel demand, travel distance, and urban spatial structures?

Based on origin-destination (OD) trips extracted from smart card transactions and taxi GPS trajectories — both cover a one-week period in Singapore — this study performs a three-fold comparative analysis. First, we analyze trip origins and destinations separately by focusing on the spatial distributions of outgoing and incoming trips. Two measures, namely rank-size distribution and log odds ratio, are introduced to quantify and compare spatial heterogeneity of travel demand extracted from the two datasets. We then examine spatial variations and statistical properties of travel distance (e.g., distance decay effect) to better understand the service radius of public transport and taxis in different parts of the city. Finally, we apply a community detection algorithm to the OD matrices to uncover the hidden spatial structures embedded in these two transport modes.

The remainder of this article is organized as follows. Section 3 provides an overview of related work of this research. Section 4 introduces the study area and the two mobility datasets. In Section 5, we introduce the approaches and measures for conducting the three-fold comparative analysis. We then present analysis results in Section 6. Finally, in Section 7, we conclude our findings and discuss future research directions.

2. Interplay between public transit and taxi services

Urban travel patterns are the outcome of the complex interactions between land use configuration and individual characteristics. The land use system governs spatial distribution of opportunities (in commercial, industrial areas) and the demand for these opportunities (in residential areas) (Geurs and van Wee, 2004). It determines intra-urban movement of people and goods from a macro level, which is modeled by trip generation and trip distribution in the “four-step” urban transportation planning process (Pas, 1995). The classical urban transport model assumes the trip amount between different zones is proportional to the number of households at the origin and the number of opportunities at the destination. However, when zooming into local neighborhood, the share disparity of different transport modes is more sensitive to individual characteristics such as income, education and vehicle ownership. This share disparity is captured in mode choice model which is used to forecast individuals' travel behaviors based on microeconomic theory (Ben-akiva and Bierlaire, 2003).

Public transit and taxi — two typical transport modes — are different in its service deployment and customer behaviors. The public transportation system is only considered by city government when population density reaches a certain level. Accessibility to transit service is higher around bus stops or metro stations than other places. It means intra-urban connectivity by public transit is enhanced at areas around bus stops or metro stations but not uniformly along bus/metro route. Unlike public transit, deployment of taxi service does not require high population density. It can be taken from anywhere, not constrained by a limited number of pick-up locations, although the service may be more accessible in some areas. In the areas where no public transit is provided, taxi complements its service; while it also competes with public transit as an alternative transport means with high flexibility in time and at a higher cost.

Human mobility datasets generated by public transportation and taxi can be used to reveal urban travel patterns. Mining the variances of the travel patterns will provide hints related to the local land use, socio-demographics, and the city structure. For example, a temporal profile of outgoing and incoming trips can be used to infer land use (Liu et al., 2012a). The mode share disparity may suggest differences of socio-demographics in different areas. Also, the flow of people and goods connects urban spaces which may have implications for spatial interactions in different regions. These spatial interactions reflect economic activities and reveal the underlying urban structures (Sun et al., 2014).

3. Related work

Understanding human mobility patterns has been a long standing research interest in areas such as urban planning, transportation, and geography. Before information and communication technologies (ICT) pervaded, travel surveys were used as the primary data source to support studies of human travel and daily activities. These studies cover important subjects such as trip chaining analysis (Hanson, 1980; Kitamura, 1984), characterization of human activity space (Newsome et al., 1998; Dijst, 1999; Schönfelder and Axhausen, 2003; Tu et al., 2017), and relationships between travel behavior and socio-economic characteristics (Hanson and Hanson, 1981; Kwan, 1999). The sampling schemes in these survey-based studies are often carefully designed, and the datasets usually contain detailed information of respondents. On the downside, however, the sample sizes are usually limited by the human and financial resources available.

With rapid developments of information and pervasive sensing technologies, researchers nowadays are able to access bigger and more diversified datasets, leading to a new paradigm of data-intensive science (Hey et al., 2009). The ways human and urban mobility can be measured are greatly enriched by datasets such as smart card transactions (Liu et al., 2009), taxi GPS trajectories (Li et al., 2011), and mobile phone records (Blondel et al., 2015). These datasets have both pros and cons when used in human mobility research, and they often reflect different or sometimes overlapping dimensions of human activities. For instance, taxi GPS trajectories record movements of taxicabs as well as their occupancy status over space and time. Such data have been used in previous studies to gain insights into cabdrivers' operation strategies (Li et al., 2011; Kang and Qin, 2016; Liu et al., 2010), urban traffic conditions (Castro et al., 2012), hot spots of taxi pick-up and drop-off points (Wang et al., 2009), and benefits of ridesharing in cities (Santi et al., 2014). Smart card transactions, on the other hand, often collect information about people's usage of public transit (e.g., card id as well as location and time for boarding/alighting). Due to the abilities to capture public transit usage for large populations, such data have been widely used to derive ridership statistics and performance indicators, and to guide transit planning and service improvements (Pelletier et al., 2011). In recent years, mobile phone data have gained increasing attention on human mobility research. Call detail records (CDRs) – a typical type of mobile phone data – have been widely used to characterize intrinsic properties of human moevements (Gonzalez et al., 2008; Song et al., 2010a, 2010b), people's use of urban space (Becker et al., 2013; Xu et al., 2015), and the interplay between human travel and social relations (Cho et al., 2011). Unlike taxi tracking and smart card transactions that are tied to specific means of transportation, mobile phone data capture snapshots of activities for large phone user pools, enabling a broader but mixed view of travelers' mobility patterns.

Since all the three types of data mentioned above include useful information about how people move from one place to another, they have all been used in previous studies to understand dynamics of population flows and intra-urban spatial structures. However, most of the

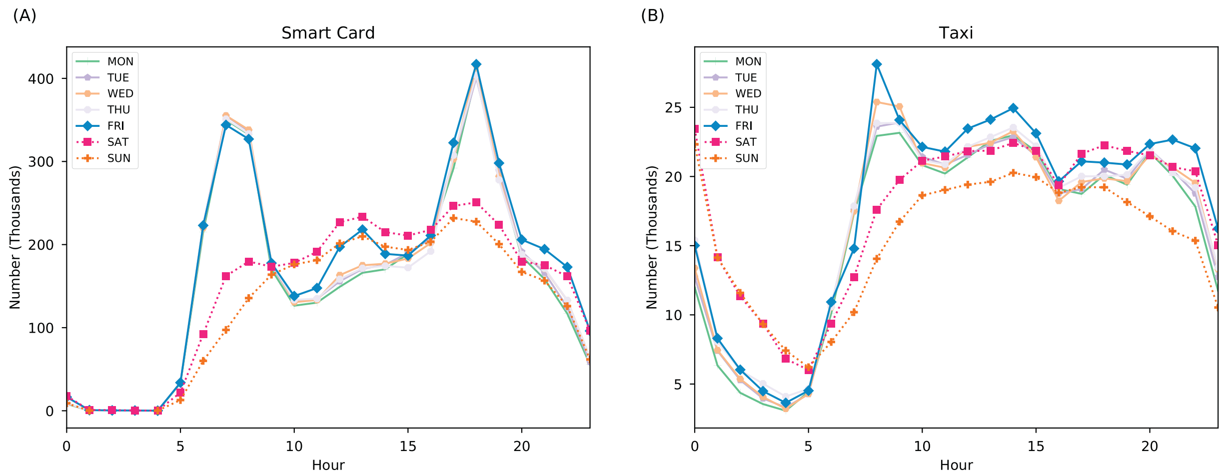

Fig. 1. Hourly variation of trips extracted from (A) smart card transactions and (B) taxi GPS trajectories.

studies involve only one single data type. Using a seven-day taxi trajectory data in Shanghai, the authors in Liu et al. (2012a) identify six types of traffic "source-sink" areas by analyzing the balance between taxi pick-ups and drop-offs at different locations in the city. The same dataset is used in another two studies to quantify the distance decay effect of taxi trips (Liu et al., 2012b) and sub-regional city structure (Liu et al., 2015). Both papers point out the issue of data representativeness and claim that "future studies should combine different data sources to show comprehensive patterns of urban spatial interaction" (Liu et al., 2015, p. 86). By analyzing smart card transactions collected in Singapore, a group of researchers apply community detection and network-based methods to understand the short-term dynamics (Zhong et al., 2015) and long-term evolution (Zhong et al., 2014) of urban spatial structures. Similarly, the authors mention that the analysis is limited by involving only the public transportation data (Zhong et al., 2015, p. 2197). By studying mobile phone data collected in 31 Spanish cities, the authors introduce several indices to characterize the morphological properties of cities (Louail et al., 2014). The comparative analysis is able to distinguish cities that are relatively monocentric from those that exhibit polycentric urban structures. However, since mobile phone data captures a mixed view of human mobility patterns, whether different types of population flows, for example, the ones derived from public and private transport usages, would produce similar spatial structures in a city cannot be investigated.

With that being said, we still lack a good understanding of the relationships among the flows and networks produced by different types of human mobility datasets. Note that some researchers have made their efforts toward this direction. For example, by analyzing taxicab usages and mobile phone data in Singapore, the authors in Kang et al. (2013) highlight the divergence in human mobility patterns – reflected by both the distance decay effect and spatial interaction communities – observed from the two types of datasets. Since movements observed from mobile phone data are claimed to be a general proxy for all kinds of human mobility, this research cannot fully disentangle the effects of different transportation modes on the observed divergence. By combining taxi GPS trajectories and smart card transactions in Wuxi, China, the authors in Li et al. (2017) examine the influence of a newly opened subway station on taxicab usages, and they find that the subway had more influence on places adjacent to subway stations and travels between urban center and suburban areas. Since the data covers four weeks before and after the opening of the subway (p. 4), whether the relationships between two types of travel demand would persist, or how they evolve, is worth a further investigation.

4. Study area and datasets

Singapore is a city-state that covers a total area of 719 km^2. It has a total population of 5.6 million as of 2016. The country has achieved rapid economic growth in the past century, and it is now a global finance and transport hub. The city is deployed with efficient mass transit services, allowing people to travel among destinations conveniently across the whole territory. According to the household interview travel survey (HITS) in 2012, mode share of public transit at peak period increased to 63%.

In this study, we investigate human mobility in Singapore using two different datasets – the smart card transactions and taxi GPS trajectories. The smart card data was provided by the Land Transport Authority of Singapore. It consists of 3,348,628 users observed in a one-week period from April 11th to 17th, 2011. The dataset includes all the tap-in and tap-out events of the smart card users for two public transit modes (i.e., bus and metro). During a tap-in or tap-out event, the boarding or alighting station as well as the time was recorded. Note that the smart card system also identified transit travels as one trip. For example, if one commuting trip includes two stages – the first one as a (first-mile) bus ride to a nearby metro station, and the second one as a metro ride to the user's work place – the dataset would store them as two separate records but label them with same trip id. In this research, we define trips as those individual travels that were labeled with the same trip id. This enables us to capture the real travel demand more precisely and prevent overestimating OD trips, especially to or from metro stations.

The taxi GPS trajectories were collected from one of the largest taxi companies in Singapore. The dataset tracks the GPS coordinates and vehicle status of over 15,000 taxis at a high temporal frequency from February 21st to 27th, 2011. The data collection period is very close to that of the smart card transactions, and neither of the two datasets was inclusive of public holidays. It is thus reasonable to assume that urban mobility patterns observed from the same data type (i.e., smart card data or taxi data) exhibit similar patterns during these two time periods, which ensures our comparison of the two datasets.

By arranging GPS records of each taxi in chronological order, we extract the OD trips using two important vehicle status – *FREE* and *POB* (passenger on board). Specifically, the status of a taxi would change between *FREE* and *POB* when the taxi picks up or drops off passengers. Thus, by identifying all the consecutive pick-up and drop-off events of each taxi, we successfully extracts 2,817,367 origin-taxi trips from the whole dataset. The average number of taxi trips per day is 402,481.

Fig. 1 illustrates the hourly variation of trips extracted from the two datasets during the observation days. For smart card transactions (Fig. 1A), the trips on weekdays exhibit a high level of regularity, with two peaks on each day during morning and afternoon rush hours. On Friday, there is a small peak around noon, which indicates a slight increase in travel demand during lunch time. On weekends, the temporal variations are less obvious and the travel demands during noon and in late afternoon are higher than those of other time periods. For the taxi data (Fig. 1B), the temporal variations of trips show similar patterns on weekdays, with travel demands being highest during morning rush hours. Such peaks are not observed during weekends. The temporal variations of trips on Saturday and Sunday show similar patterns, but the total demand for taxi trips is higher for the former (i.e., Saturday).

5. Methods

Smart card transactions and taxi GPS trajectories capture different aspects of human mobility patterns. Both of them are well studied but separately in previous research. In this study, we propose an integrated framework to analyze and compare mobility patterns extracted from the two types of datasets. The comparison mainly focuses on the following perspectives: (1) the spatial distribution of travel demand; (2) the statistical properties of collective travel behavior (e.g., distance decay effect); (3) the urban structures revealed from trip OD based spatial network.

Since the two datasets are collected at different spatial granularities, to make them comparable, we first map trip origins/destinations onto subzones (514 in total). The subzones are demarcated by the Urban Redevelopment Authority (URA) of Singapore. Specifically, the URA divides Singapore into 55 planning areas, with each of them further divided into subzones. Each planning area consists of about ten subzones on average. The size of subzones varies and depends on the land use configuration.

After mapping the trips onto subzones, we perform a three-fold comparative analysis. The first part analyzes trip origins and destinations separately and compares the spatial distribution of incoming and outgoing trips extracted from the two datasets. The second part focuses on the travel distance of trips and its spatial distribution in the city. The third part examines the spatial interactions of urban spaces observed from OD matrices. Since mobility patterns might be different on weekdays and weekends, we perform the analysis separately for the two periods.

5.1. Spatial distributions of travel demand

Public transit and taxis are two important transport modes in urban transportation systems. They provide passenger services with different levels of convenience, flexibility, and economic benefits. The travel demands generated from these two modes are both shaped by factors such as population size, land use patterns, and other socio-demographic characteristics. However, there exist different socio-economic processes that govern individual and collective choices over one mode to the other. It is thus essential to compare the spatial distribution of travel demand extracted from the two datasets to better understand how the two transport modes complement or compete with each other in facilitating human movements.

For each dataset (i.e., transport layer), we generate two directed graphs which summarize the trips that have been made between different origin and destination subzones on weekdays and weekends, respectively. Specifically, we use $G^s(V^s, E^s)$ to denote the directed graph extracted from the smart card transactions on weekdays, with each node $v^s \in V^s$ representing a subzone in Singapore. For an ordered pair of nodes (v_i, v_j), an edge $e_{i,j}{}^s \in E^s$ is constructed with its weight being the average daily trips, i.e., the total number of trips on weekdays from v_i to v_j divided by the total number of observation days. Given a node (subzone) i, the indegree $d^-(v_i^s)$ and outdegree $d^+(v_i^s)$ denote the average daily incoming and outoging trips at this subzone, respectively. Note that:

$$\sum_{v^s \in V^s} d^-(v^s) = \sum_{v^s \in V^s} d^+(v^s) = \left|E^s\right| \tag{1}$$

here $|E^s|$ denotes the average daily trips extracted from the smart card transactions on weekdays. The indegree and outdegree of the nodes are then used to describe the spatial distribution of travel demand from two different perspectives. Similarly, we generate another directed graph $G^{s'}(V^{s'}, E^{s'})$ to describe the travel patterns observed from the smart card transactions on weekends. We then repeat the above procedure and derive two directed graphs from the taxi GPS trajectories on weekdays and weekends, i.e., $G^t(V^t, E^t)$ and $G^{t'}(V^{t'}, E^{t'})$.

Here we use two approaches to compare the spatial distribution of travel demand extracted from two the datasets. First, we use the rank-size distribution to reflect the spatial heterogeneity of travel demand and explore how the distributions vary between the two transport layers. Rank-size distribution has been widely used in previous studies to assess the hierarchy of objects based on their magnitudes, for example, to quantify the polycentric structure of cities Burger and Meijers (2012).

Since the magnitudes of the two datasets in this research are different, A normalization is needed to facilitate the comparison. For example, given the graph $G^s(V^s, E^s)$ extracted from the smart card transactions on weekdays, we derive two rank-size distributions using incoming and outgoing trips, respectively. When generating the rank-size distribution for incoming trips, the size of a subzone v_i^s is defined as the percentage of incoming trips:

$$size_{in}(v_i^s) = d^-(v_i^s)/|E^s| \tag{2}$$

For outgoing trips, the size of v_i^s is defined as:

$$size_{out}(v_i^s) = d^+(v_i^s)/|E^s| \tag{3}$$

Following this logic, for each combination of day (weekdays vs. weekends) and trip type (incoming vs. outgoing), we can derive two rank-size distributions that reflect the spatial heterogeneity of travel demand extracted from the two datasets. The slopes of the regression lines are then compared to assess their similarities. A line with a flat slope indicates that the travel demand tends to distribute uniformly across different areas. A steep slope, however, suggests that the travel demand is mainly concentrated in a few subzones.

The comparison of rank-size distributions provides a holistic view of how each type of travel demand distributes across space. However, it does not explicitly link the two transport layers in the same geographic areas (e.g., subzones). For example, given two rank-size distributions of incoming trips derived from the smart card and taxi data, even if the slopes of both regression lines are steep, it does not reveal whether or not the two types of travel demand are concentrated in the same geographic areas. Hence, we introduce another measure, log odds ratio, to describe the relative balance of the usage of two transport layers. Using incoming trips on weekdays as an example, given the two directed graphs $G^s(V^s, E^s)$ and $G^t(V^t, E^t)$, the log odds ratio of a given subzone i is computed as:

$$R_i^- = log_{10}\left(\frac{d^-(v_i^s)/|E^s|}{d^-(v_i^t)/|E^t|}\right) \tag{4}$$

The value of R_i equals to zero if the relative usage of public transit and taxis is the same in subzone i. A value of $R_i^- > 0$ suggests that the usage of public transit is higher than that of taxis, and vice versa. Similarly, the log odds ratio of subzone i based on outgoing trips on weekdays is computed as:

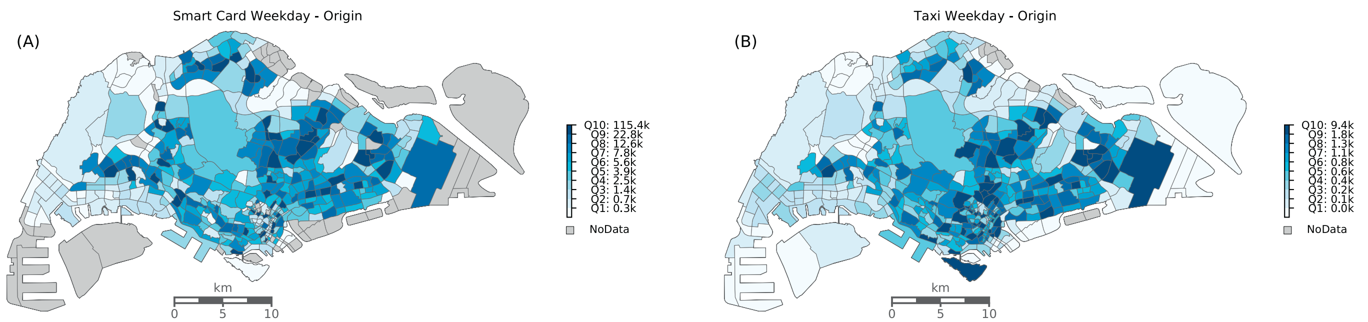

Fig. 2. Spatial distributions of average daily outgoing trips extracted from (A) smart card transactions and (B) taxi GPS trajectories on weekdays. The intensities of trips are rendered using gradient colors based on quantile classification.

$$R_i^+ = log_{10}\left(\frac{d^+(v_i^s)/|E^s|}{d^+(v_i^t)/|E^t|}\right) \quad (5)$$

We then repeat the same procedure and calculate the measures on weekends based on $G^{s'}(V^{s'}, E^{s'})$ and $G^{t'}(V^{t'}, E^{t'})$.

5.2. Spatial patterns and statistical properties of travel distance

Travel distance is another important metric that reflects the spatial structure of human movements. Considering the differences in people's usage over public transit and taxis, the statistical properties of travel distance extracted from these two modes may be different from each other. For example, people who live in the urban periphery might commute to the city center through public transit for day-to-day work while use taxis for some other trips. To better understand how public transit and taxis facilitate human movements, we first compute, for each transport mode, the median travel distance for both incoming and outgoing trips at the subzone level. Similar to the travel demand analysis, trips that occurred on weekdays and weekends are analyzed separately. We then map these values onto subzones and compare the geographic variations of travel distance for the two modes. For each type of day (weekdays vs. weekends), we further explore the distance decay effect of human movements by computing the probability of trips at different distance values.

5.3. Spatial interaction communities

Given trips extracted from smart card transactions or taxi GPS trajectories, the way they are distributed across space reflects the spatial interactions between different locations. In this research, we use community detection — a technique that has been widely used to uncover hidden structures in networks (Clauset et al., 2004; Blondel et al., 2008) — to assess whether the two transport layers produce similar or different spatial interaction patterns. Specifically, for each combination of data (smart card transactions vs. taxi GPS trajectories) and type of day (weekdays vs. weekends), we first derive an undirected graph summarizing the number of trips that occurred between each pair of subzones. Using trips extracted from smart card transactions on weekdays as an example, we use $A_{i,\ j}$ to denote the average daily trips observed between two subzones i and j. Each subzone i serves as a unique node in the graph with degree $k_i = \Sigma_j A_{ij}$. An algorithm based on modularity optimization is used to partition the trip network into communities of densely connected nodes while nodes belonging to different communities are sparsely connected (Blondel et al., 2008).

Given two subzones i and j, let $c_{i\_}$ and c_j denote the communities they belong to, function $\delta(c_i, c_j)$ is introduced to describe whether the two subzones belong to the same communities. It takes the value of 1 if $c_i = c_j$ and 0 other wise. Thus, the fraction of trips that occurred within communities are calculated as:

$$\frac{\sum_{ij} A_{ij}\delta(c_i, c_j)}{\sum_{ij} A_{ij}} = \frac{1}{2m}\sum_{ij} A_{ij}\delta(c_i, c_j) \quad (6)$$

Note that m denotes the total number of trips in the undirected graph $m = \frac{1}{2}\sum_{ij} A_{ij}$. If the trips are randomly placed in the graph by respecting the degrees of the nodes, the fraction of trips that would occur within communities would be $k_i k_j \delta(c_i, c_j)/4m^2$ (Clauset et al., 2004). The modularity Q is thus defined as the difference between observed fraction of within-community trips and the expected fraction in a randomized network:

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \quad (7)$$

Modularity optimization is a computationally intractable task and approximation algorithms are often used. Unfortunately, greedy searching may produce modularity value which is far below optimal. Thus we employ the iterative algorithm to detect communities which is proposed in Blondel et al. (2008). The algorithm initializes by setting each node of the network into a separate community, and determine whether to move one node from its current community into another one by evaluating the maximum gain that would take place. After that, it aggregates these meta-communities to build a new network of communities on which the same optimization strategy can be applied. The processes are repeated until no increase of modularity.

6. Results

6.1. Spatial distribution of travel demand

Fig. 2 illustrates the spatial distributions of average daily outgoing trips extracted from smart card transactions and taxi GPS trajectories on weekdays. From a visual perspective, it can be seen that the two types of travel demand match relatively well in geographic space. For example, subzones with a high level of public transit usage (Fig. 2A) tend to also produce a high number of taxi trips (Fig. 2B). This also happens in other combinations of trip category (incoming vs. outgoing) and day type (weekdays vs. weekends) (Fig. A.1). The Spearman's rank correlations of travel demand under the four combinations are 73.5% (outgoing trips on weekdays), 77.3% (outgoing trips on weekends), 71.6% (incoming trips on weekdays), 76.2% (incoming trips on weekends), respectively. That means public transit and taxi usages are highly comparable in many areas and this is possibly driven by factors such as population distribution and spatial configuration of land use. Note that we also observe some discrepancies of travel demand in particular regions. For example, in the central part around downtown core, pubic

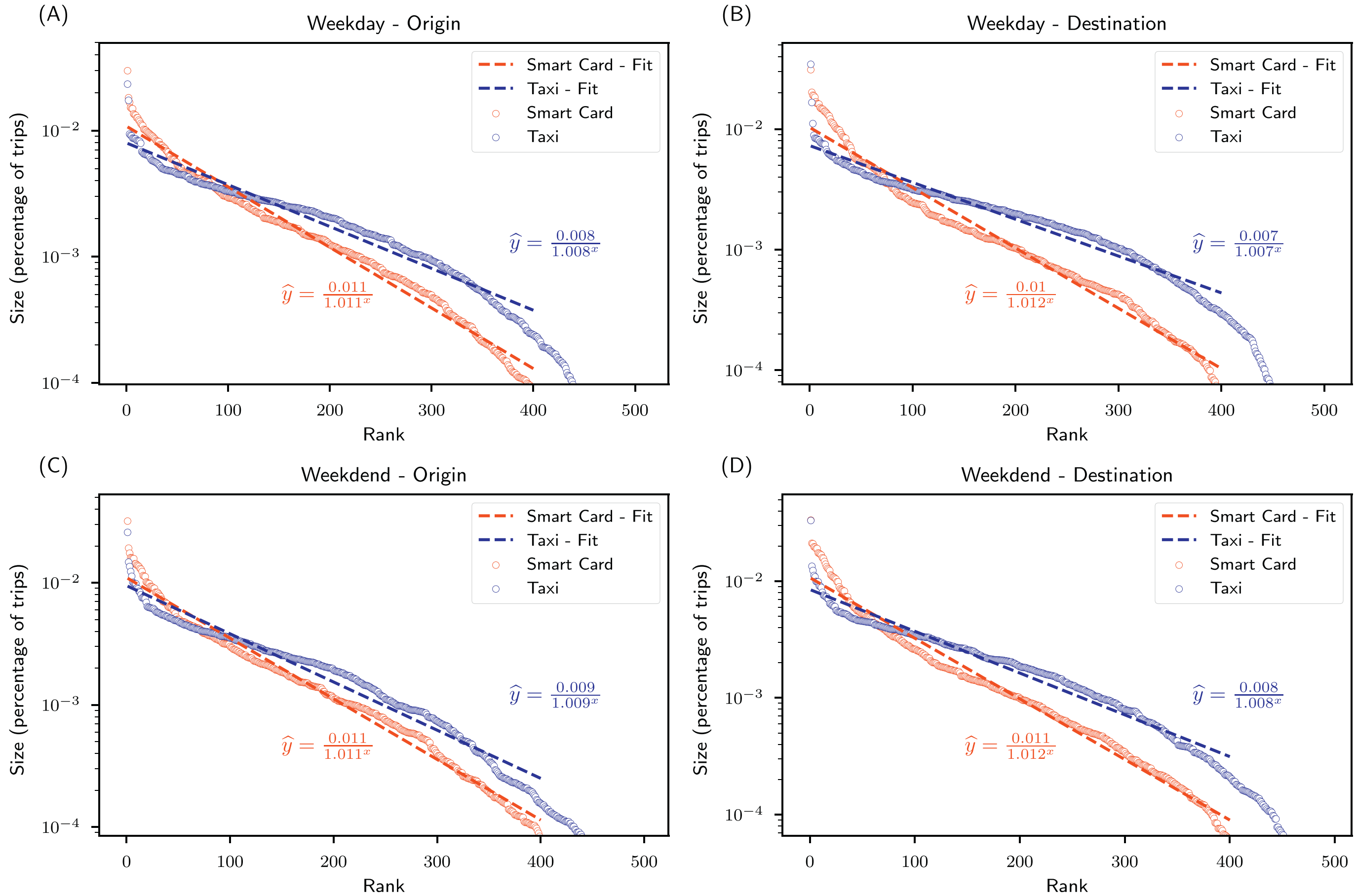

Fig. 3. Rank-size distributions based on each combination of trip category (outgoing vs. incoming) and type of day (weekdays vs. weekends).

transit usage concentrates in a few areas, while the demand for taxis is distributed more evenly. Another example is Sentosa island, a tourist attraction in the southern part of Singapore. The island is not well connected with public transit, which results in a high demand for taxis on both weekdays and weekends.

We further investigate the spatial patterns of travel demand using the two measures proposed in Section 5.1. Fig. 3 illustrates the rank-size distribution of trips observed on weekdays and weekends (with y-axis at log scale). As shown in Fig. 3A, on weekdays, the percentage of trips extracted from both datasets decays quickly for the first few subzones, indicating that these subzones produced a considerable amount of trips (i.e., the "king effect"). As the rank increases, we observe that the curve generated from taxi GPS trajectories tends to decay more slowly than the one from smart card data. To quantify the differences in the decay effects, we fit the two curves using the exponential model $\hat{y} = \frac{N_0}{R^x}$, of which the parameter R describes the rate of decay. The fitted lines for smart card and taxi data are $\hat{y} = \frac{0.011}{1.011^x}$ and $\hat{y} = \frac{0.008}{1.008^x}$, respectively. We then repeat this procedure by fitting all the curves based on the combination of trip category (outgoing/incoming) and type of day (weekdays/weekends). As illustrated in Fig. 3, given each combination of trip category and day type, the value of R for smart card data is always higher than that of taxi GPS trajectories, which suggests a higher degree of spatial heterogeneity of public transit usage at the subzone level.

As mentioned in Section 5.1, the rank-size distribution does not

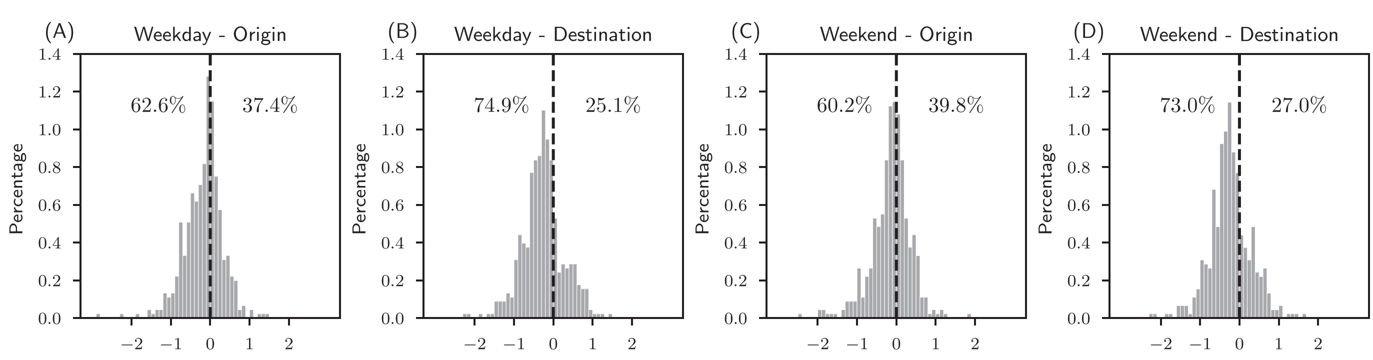

Fig. 4. Histograms of log odds ratio based on each combination of trip category (outgoing vs. incoming) and type of day (weekdays vs. weekends).

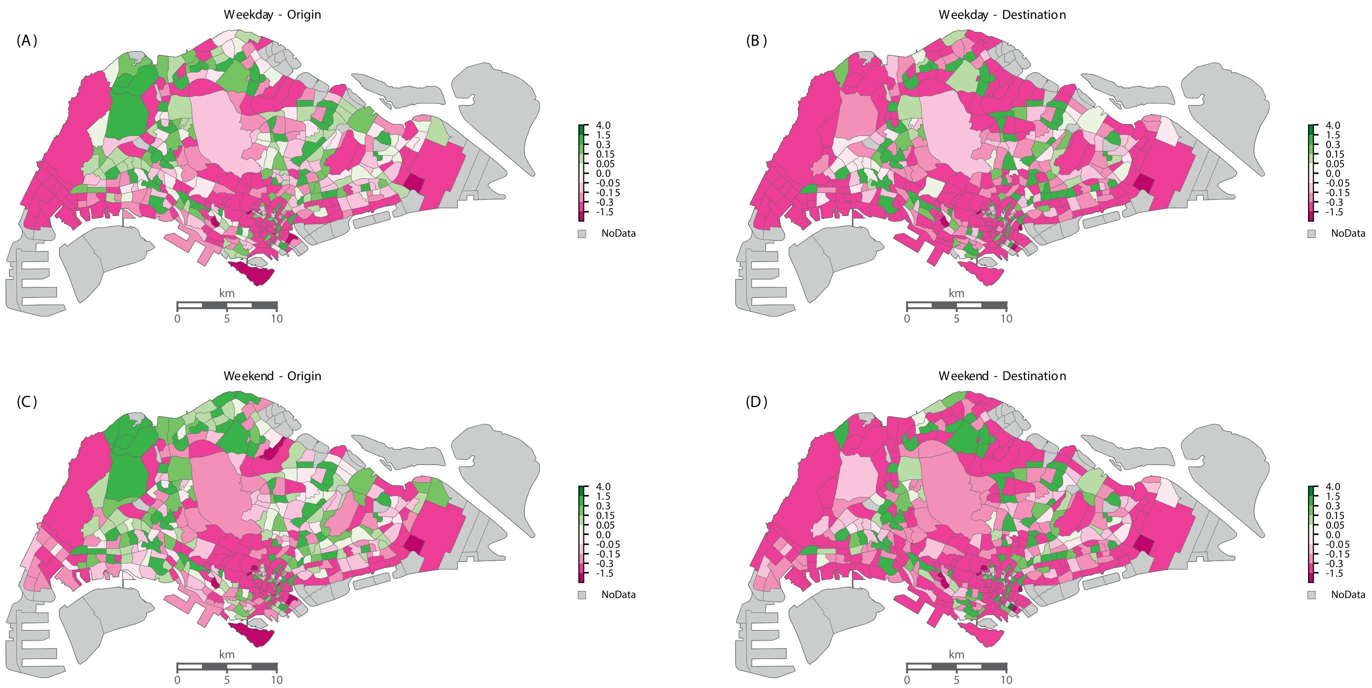

Fig. 5. Spatial patterns of log odds ratio based on each combination of trip category (outgoing vs. incoming) and type of day (weekdays vs. weekends).

explicitly link the two transport layers in same geographic regions. We thus derive log odds ratio to better understand the relative balance of their usages over space. Fig. 4 illustrates the histogram of log odds ratio based on outgoing and incoming trips on two types of days (weekdays vs. weekends). In general, outgoing trips on weekdays and weekends produces similar distributions (Fig. 4A and C), from which roughly 60% of the subzones are observed with a negative log odds ratio. That means the relative usages of taxis in the majority of subzones are higher than that of public transit. When it comes to incoming trips, the distributions become even more skewed, and > 70% of the subzones have a log odds ratio smaller than zero (Fig. 4B and D). The result, which is consistent with the findings from Fig. 3, indicates that public transit usage tends to be more concentrated in particular areas.

By exploring the geographic patterns, as shown in Fig. 5, we find that the central part of Singapore, where the CBD (i.e., downtown core) locates, produced higher percentages of taxi trips than that of public transit, while many subzones with a negative log odds ratio (i.e., in green) are scattered around the outskirt of the city (Fig. 5A and C). By further comparing the spatial patterns of incoming and outgoing trips, it is found that certain subzones with a lower level of taxi usages (Fig. 5A and C) attracted higher percentages of taxi trips on both weekdays and weekends (Fig. 5B and D). This potentially suggests an asymmetry of human movements regarding usages of different transport modes. For example, some people might commute to work using public transit in the morning but go to shopping after work and then head back home using other transport modes (e.g., taxis or walking). The characteristics of individual mode choice and daily trip chains (Strathman et al., 1994) seem to influence the relative balance of public transit and taxi usages across Singapore. Validating this hypothesis, however, requires further examination of travelers' movement patterns.

We hypothesize that the spatial variations of log odds ratio is the outcome of complex interactions of socioeconomic factors, urban form, and transportation development. While a comprehensive understanding is beyond the scope of this research, for each combination of trip category (outgoing vs. incoming) and type of day (weekdays vs. weekends), we built a linear regression model with the log odds ratio (at subzone level) as the dependent variable, and population density, average household income, and distance (from the subzone's geometric center) to the nearest MRT as three independent variables:

$$\log OR = \beta_0 + \beta_1{}^*Density(pop) + \beta_2{}^*Income + \beta_3{}^*Distance(MRT) + \varepsilon \tag{8}$$

Note that the average household income of subzones are calculated from the Household Interview Travel Survey (HITS) collected by the Singapore Land Transport Authority (LTA) in 2012. The Singapore 2012 HITS collects 1-day travel diary of 35,715 individuals (about 1% sampling rate) along with other socio-demographic attributes – such as monthly income — that are self-reported by the respondents. The population of each subzone is acquired from the department of statistics in Singapore. We divided it by the subzone area to get the density metrics.

The regression results, as shown in Table 1, suggest that the average household income and distance to MRT are negatively associated with log odds ratio. More specifically, subzones with lower average household income and better accessibility to MRT tend to rely more on public transit. Population density is positively associated with log odds ratio. However, the coefficients in the four models are not significant. The R^2 of the four models range between 0.124 and 0.159, suggesting that the three variables only explain a limited proportion of variations in log odds ratio. There exist other relevant factors (e.g., land use type) that influence the relative balance of the public transit and taxi usages in Singapore.

6.2. Spatial variations of travel distance

Fig. 6 shows the spatial patterns of the median travel distance of outgoing trips extracted from the two datasets on weekdays. In general, the distance of taxi trips exhibit a smooth variation over space (Fig. 6B). The trips that started in the central part of Singapore are relatively short as compared to the ones that originated from the outer subzones. This suggests that the median distance of taxi trips in Singapore is accordant with the distance from the originating subzone to the city center. The outgoing trips extracted from smart card data, however, produces a

Table 1
Results of linear regression.

	Weekday - Origin beta (t-stat)	Weekday - Destination beta (t-stat)	Weekend - Origin beta (t-stat)	Weekend - Destination beta (t-stat)
Population Density (1000/km^2)	0.002 (1.429)	0.002 (1.202)	0.001 (0.346)	0.000 (0.061)
Average Income (1000 SGD)	−0.029 (−3.004)**	−0.035 (−3.207) **	−0.038 (−3.948)*	−0.043 (−4.112)*
Distance to MRT (km)	−0.190 (−3.769)*	−0.255 (−4.460)*	−0.198 (−3.907)*	−0.271 (−4.898)*
Adjusted R^2	0.124	0.145	0.129	0.159

** p-value < 0.001.
* p-value < 0.01.

different spatial pattern (Fig. 6A). The subzones with a high median distance are scattered across the whole Singapore and these areas play an important role in facilitating long-range human movements through public transport. Note that we also explore the results for other combinations of trip category (e.g., incoming trips) and day type (e.g., weekends). We find that by fixing the day type (weekdays vs. weekends), the incoming and outgoing trips extracted from taxi GPS trajectories yield very similar spatial patterns, and their Pearson correlations on weekdays and weekends are 0.84 and 0.78, respectively. However, the correlations between incoming and outgoing trips extracted from smart card transactions are lower (i.e., 0.54 for weekdays and 0.49 for weekends). It is likely that the usage of public transit in Singapore is more asymmetrical; trip chains that people take to workplaces in the morning is not highly identical with those returning home after work. Readers could refer to Appendices 9 and 10 for detailed information about the spatial patterns and correlations.

We further explore the distance decay effect of human movements observed from the two transport layers. On weekdays, as shown in Fig. 7A, for both public transit and taxi usages, the probability of trips observed at a given distance can be well characterized by two segments that follow a power-law distribution. When travel distance is within 20 km, the probability of trips extracted from both datasets decay relatively slowly. Fitting these two segments (within 20 km) yields a distribution of $p(d) \sim d^{-0.917}$ for public transit, and $p(d) \sim d^{-0.691}$ for taxi trips. Interestingly, when travel distance becomes 20 km or higher, the two types of trips start to decay quickly, with public transit characterized by $p(d) \sim d^{-14.078}$ and taxi trips by $p(d) \sim d^{-7.532}$. Weekends also show similar patterns (Fig. 7B). Here, two important findings can be summarized from the results. First, on both weekdays and weekends, trips observed from public transit always decay faster than taxi trips, which highlights the importance of taxis in facilitating long-distance travels. Second, both types of trips decay much faster when travel distance is beyond 20 km. By further examining the study area, we find that 20 km corresponds to the average distance from the urban periphery to the center (i.e., CBD), where a lot of employment opportunities are offered. The trips made by public transit and taxis within this 20 km-radius account for a large proportion of travel demand in the city, while long-distance trips, for example those across the island, are relatively scarce.

6.3. Spatial interaction communities

Fig. 8 shows the results of the community detection derived from smart card transactions and taxi GPS trajectories. In general, both datasets produce spatially cohesive communities that are mainly formed by adjacent districts. The communities, especially those derived from taxi GPS trajectories, are in line with Singapore's administrative regions. That means the administrative division of Singapore plays an important role in shaping the spatial structure of human mobility. The community patterns exhibit similarities in north and east parts and dissimilarities more evident in the center and west parts.

The communities derived from taxi trips are more coherent across space than that from smart card. Also, its clustering pattern is consistent on weekday and weekend (Fig. 8B and Fig. 8D), which may imply taxi usage is less variant in different days compared with public transit usage. The communities derived from smart-card trips are however more disaggregated, especially in city center where subzones being clustered into the same community are seen non-adjacent (Fig. 8A and Fig. 8C). This phenomenon arises because public transit strategically connects places where lots of commuting trips are demanded while unstops at places amid with less demand. The clustering pattern from smart-card trips is not consistent on weekday and weekend, and the main differences are evident in central and western regions. This community difference in weekday and weekend reflects a dynamic interaction of human travel behavior and land use. A majority of trips served by public transit are commuting on weekday and leisure on weekend. This shift of trip purposes refashions interactions among different places, leading significant impact on high-density employment areas, such as downtown in the center region and industrial parks in the west region.

In addition, some discrepancies revealed in community detection may reflect special urban phenomena. For example, Changi area/

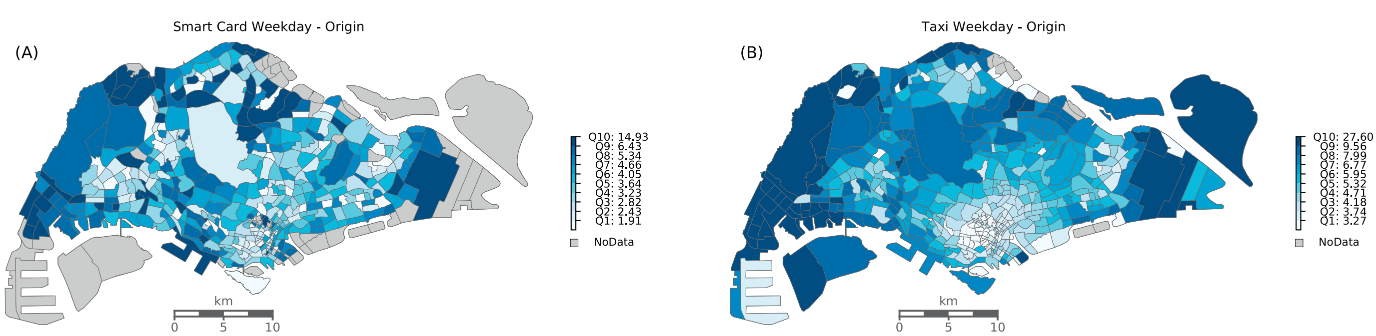

Fig. 6. Median travel distance of outgoing trips extracted from (A) smart card data and (B) taxi GPS trajectories on weekdays. The distance values are rendered using gradient colors based on quantile classification.

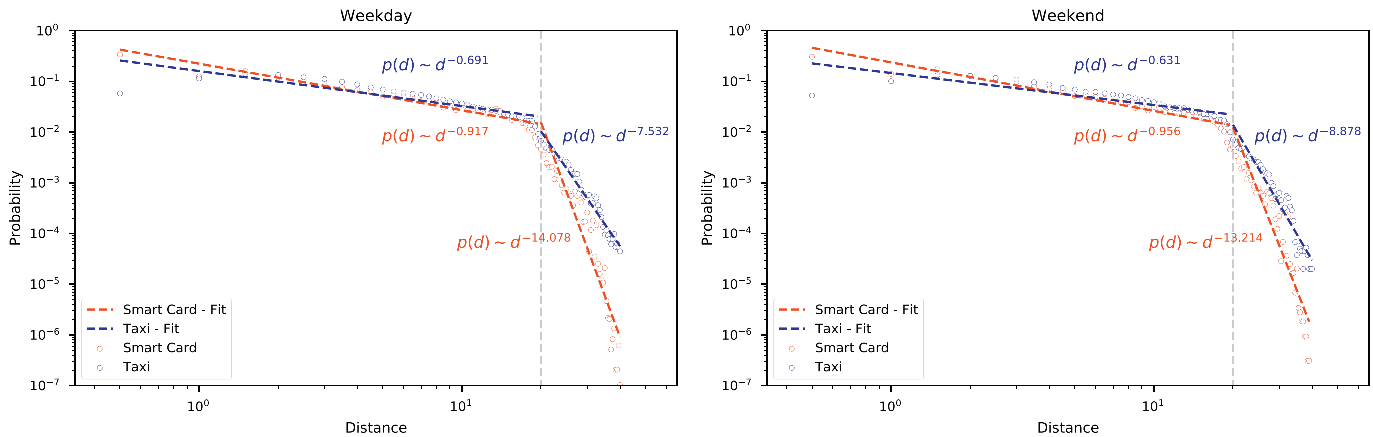

Fig. 7. Distance decay of trips extracted from smart card and taxi data on two types of days (weekdays vs. weekends).

airport behaves differently in communities derived from taxi trips that most areas belonging to the same community are adjacent to each other. The result, which is consistent with the finding from Kang et al. (2013), suggests intensive taxi trips between the downtown area and the airport. Nonetheless, in communities derived from the smart card, it is classified to the same community with nearby areas. It implies long travel time might hinder people from using public transit for a long-distance trip despite its low cost. Sentosa island is seen to connect with north-east part in communities derived from smart card (Fig. 8A and D), which is because both Sentosa and the north-east part are connected with several spots in downtown area. Sentosa is a top leisure destination for tourists coming to Singapore. We suspect those spots which it connects with downtown are also top destinations for the tourists.

7. Discussion and conclusion

Vast human mobility datasets have provided unprecedented opportunities for urban and transportation research. Such datasets have enabled the investigation of urban mobility patterns from a flow or network perspective. Existing studies attempt to reveal travel patterns and urban spatial structures from mobility data, but very few aim at

Fig. 8. Spatial interaction communities derived from smart card and taxi data on two types of days (weekdays vs. weekends).

comparing urban phenomena and processes from datasets of different types. In this research, we propose an analytical framework – by coupling smart card transactions and taxi GPS trajectories – to compare fundamental dimensions of urban mobility patterns. We first construct graphs of human movements through origin-destination (OD) matrices extracted from the two datasets, which provide the foundations for later analyses. We employ two measures, the rank-size distribution and log odds ratio, to assess the spatial distributions of two different types of travel demand as well as their relative balance. We then examine the distance decay effect of two types of trips to better understand how public transit and taxis facilitate human movements with different ranges. Finally, a community detection algorithm is used to uncover the hidden urban spatial structures. Using Singapore as the case study, we sysmetically compare and summarize the observed patterns from the above three perspectives. The commonalities and discrepancies revealed from the two transportation modes comprehend our understanding of people's use of urban space, which also shed light on the potential biases if one dataset is used alone to conclude urban mobility patterns.

This work reveals some noteworthy findings in Singapore, and some of them may also apply in other cities. The high correlations of travel demand for the two transport modes observed in this study may reflect good mixture of different income groups at the subzone level. Since travel demand is driven by population density and spatial configuration of land uses while the mode share is more sensitive to individual socioeconomic status, the high income neighborhoods are expected to rely more on taxis. Thus, the correlation of travel demands between public transit and taxis may become lower in cities with higher social-spatial segregations. The rank-size distributions of travel demand indicate that smart card usage at the subzone level tends to decay more rapidly than that of taxis, which may be induced by a higher spatial heterogeneity of public transit usage and the spontaneity and greater flexibility of taxi service. With that being said, public transit can only pick up or drop off passengers at bus stops or metro stations, which makes its service more concentrated within certain subzones. But the usage of taxis, as observed in this research, tends to distribute more evenly across the city. This is reflected by the skewness of log odds ratio distribution that more places are observed with negative values, suggesting a higher relative usage of taxis in those areas. Also, the mismatch of spatial patterns observed from incoming and outgoing trips implies an asymmetry of human movements with regard to the transport mode choices. Although public transit demand tends to decay faster at the subzone level, the decaying speed is not significantly different from that of taxis, partly because Singapore has an efficient public transport network covering most parts of the city. For other cities with less mature public transportation systems, a larger disparity in decaying speed shall be reflected in the slopes of the fitting curves. The regression analysis, although explaining a limited proportion of the variations (in log odds ratio), suggests that the relative balance of two types of travel demand is related to the underlying built environment and socio-demographic characteristics.

The mapping of median travel distance, especially in taxi usage, exhibits concentric spatial patterns in the city. In a monocentric city, job opportunities concentrate in the city center, which attracts a large amount of people in the day time. The urban commuting distance is excepted to increase from the city center to the periphery, exhibiting a concentric pattern of average travel distance. However, the travel distance pattern will become more complicated as cities become more polycentric. Some studies find that Singapore has been developing rapidly toward a polycentric urban form (Han, 2005). Such transition is partially supported from the travel distance patterns observed in this study (Fig. 6B). Suburban centers, although less dominant compared to the city center, have emerged in the east, west and north regions. It is also found that the distance distribution follows a truncated power law with the breaking point at around 20 km, which is the average distance from the urban periphery to the city center. We can anticipate that most trips shall be less than this distance, and this phenomenon could also exist in other cities.

The spatial interaction communities formed by taxi trips exhibit similar patterns on weekdays and weekends, indicating a higher regularity of taxi usage on different types of days. But when it comes to public transit, the communities exhibit notable changes from weekdays to weekends. These changes are likely to be related to the dynamic interactions of travel behavior and land use. The community detection also reveals the city's polycentric structure from a bottom-up view, and the results observed from both datasets match relatively well Singapore's administrative regions. Different from Louail et al. (2014) which investigates urban polycentricity using hot spot analysis, the community detection algorithm employed in this research reveals the polycentric urban structure from a network perspective. It captures more about the spatial interactions among places through individual travel patterns. The difference in the derived communities from public transit and taxis also suggest that each of the transport modes plays a specific role in connecting certain places in the city.

In this study, we investigate some important dimensions of urban mobility using smart card transactions and taxi GPS trajectories. The research does not attempt to exhaust all possibilities of comparative data analytics. The comparisons are refined in several representative aspects that we believe are of wide interest to urban and transportation scholars. Our findings suggest that different mobility datasets can be used together to depict a multi-faceted view of urban mobility and associated spatial structures. It also reminds us that different conclusions regarding the space-time structures of a city can be reached by looking at different types of mobility datasets. The current study only examines one particular city. In the future, we plan to apply our analytic framework across different cities around the globe. Such a hybrid approach with both intra- and inter-city comparisons would deliver a more comprehensive view on the universal mechanisms, or the social, economic and cultural differences that shape human movements in urban areas.

8. Acknowledgement

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions on earlier drafts of the manuscript. The research is supported in part by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under CREATE programme, Singapore-MIT Alliance for Research and Technology (SMART) Centre, Future Urban Mobility (FM) Interdisciplinary Research Group, and the Hong Kong Polytechnic University Start-Up Grant under project 1-BE0J.

Appendix A. Spatial distribution of travel demand extracted from smart card transactions and taxi GPS trajectories

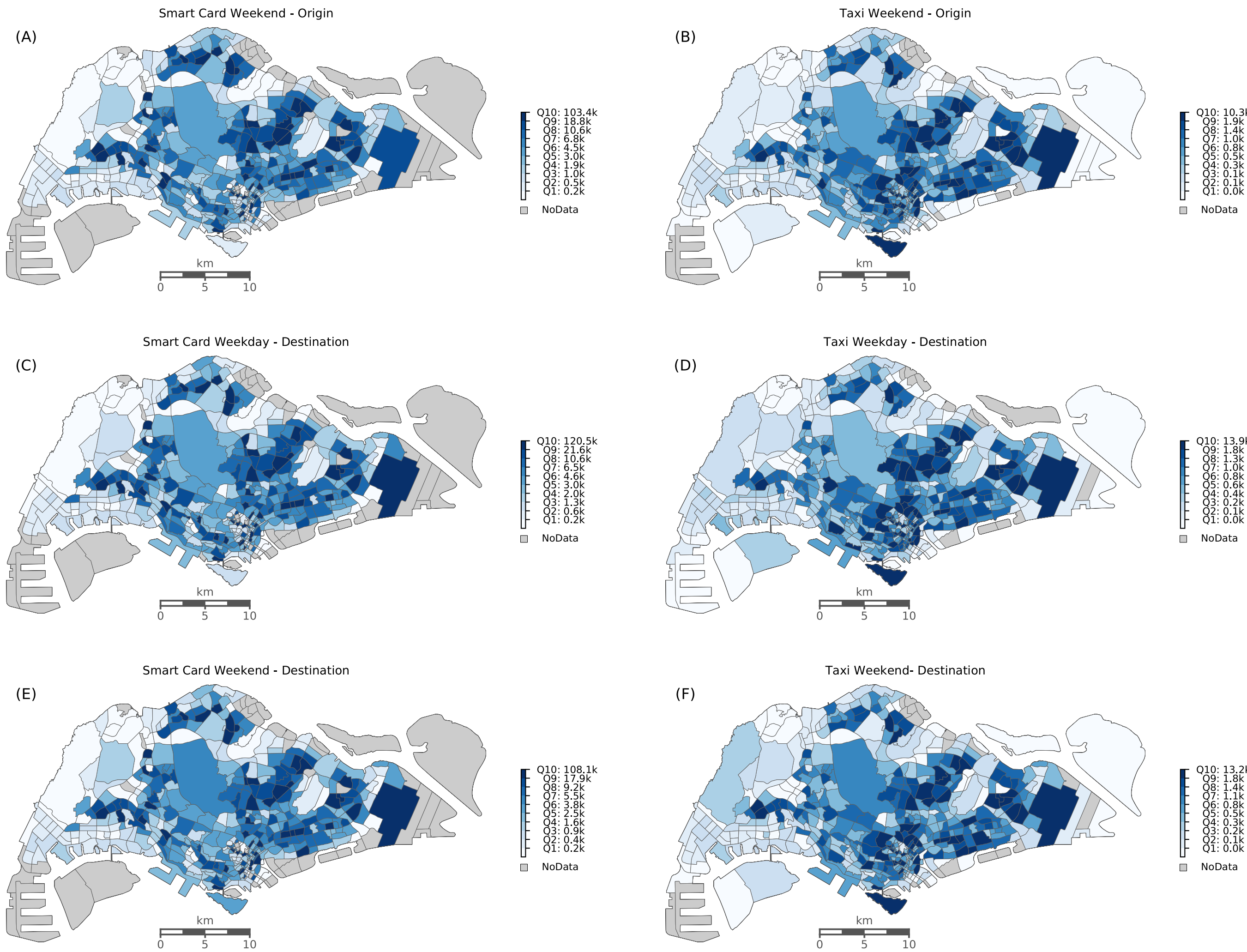

Fig. A.1. Spatial distributions of average daily trips extracted from smart card transactions and taxi GPS trajectories: (A-B) outgoing trips on weekends; (C-D) incoming trips on weekdays; (E-F) incoming trips on weekends. The intensities of trips are rendered using gradient colors based on quantile classification.

Appendix B. Median travel distance of trips extracted from smart card transactions and taxi GPS trajectories

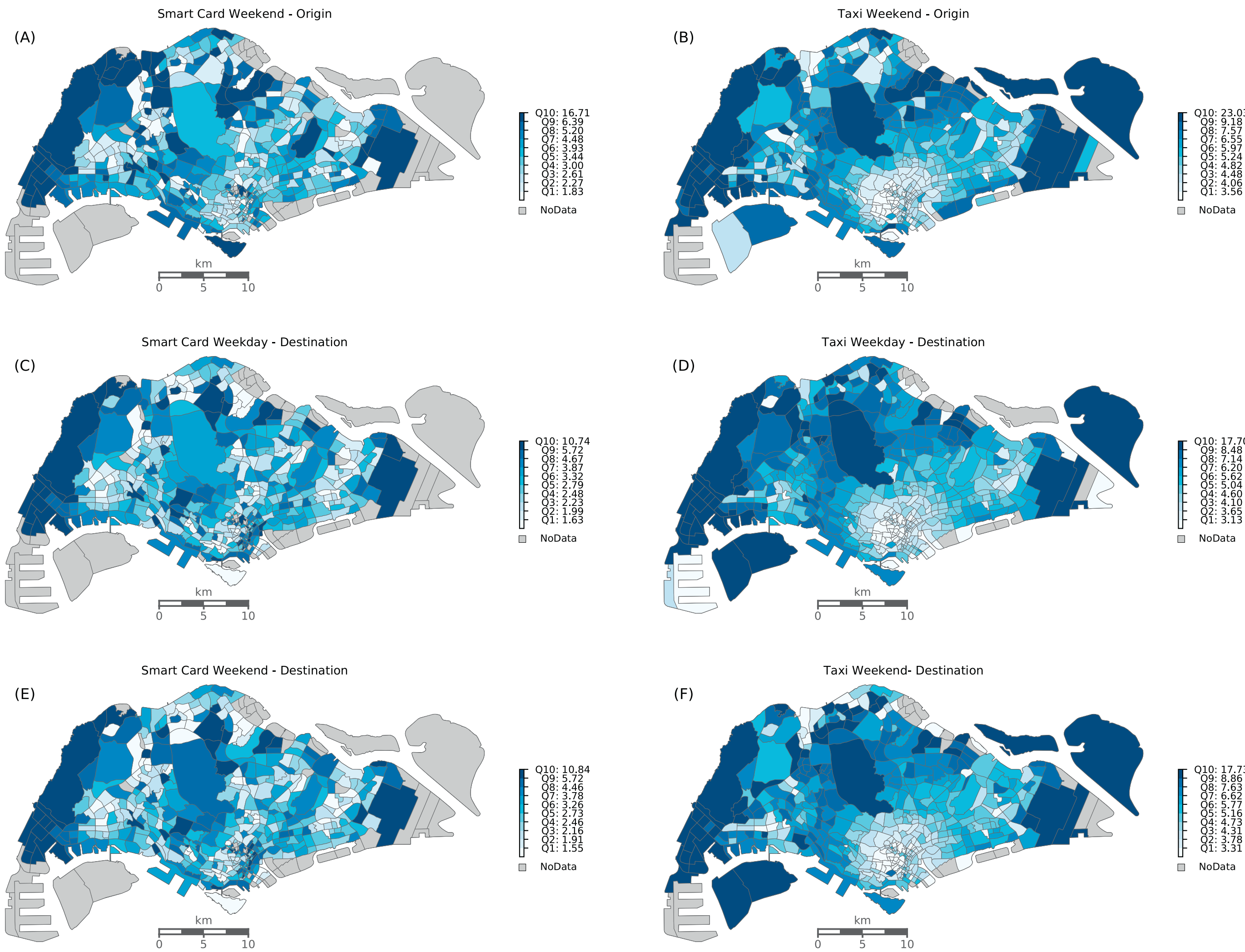

Fig. B.1. Median travel distance of: (A-B) outgoing trips extracted from smart card and taxi data on weekends; (C-D) incoming trips extracted from smart card and taxi data on weekdays; (E-F) incoming trips extracted from smart card and taxi data on weekends. The distance values are rendered using gradient colors based on quantile classification.

Appendix C. Correlations between median travel distance of incoming and outgoing trips extracted from the two datasets

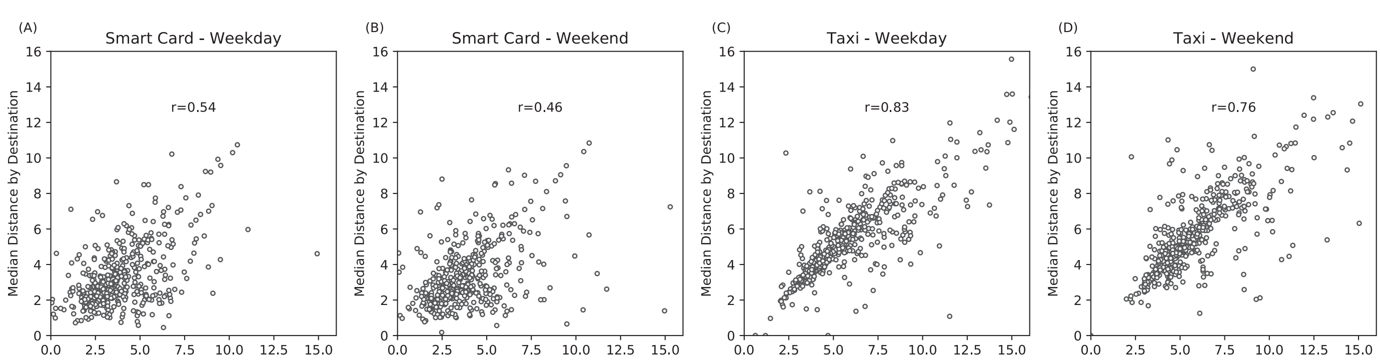

Fig. C.1. Correlations between median travel distance of incoming and outgoing trips extracted from: (A) smart card transactions on weekdays; (B) smart card transactions on weekends; (C) taxi GPS trajectories on weekdays; (D) taxi GPS trajectories on weekends.

References

Ahas, Rein, Aasa, Anto, Silm, Siiri, Tiru, Margus, 2010. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: case study with mobile positioning data. Transp. Res. C 18 (1), 45–54.

Alexander, Lauren, Jiang, Shan, Murga, Mikel, González, Marta C., 2015. Origin–destination trips by purpose and time of day inferred from mobile phone data. Transp. Res. C Vol. 58, 240–250.

Batty, Michael, 2013. The New Science of Cities. Mit Press.

Becker, Richard, Cáceres, Ramón, Hanson, Karrie, Isaacman, Sibren, Ji, Meng Loh, Martonosi, Margaret, Rowland, James, Urbanek, Simon, Varshavsky, Alexander, Volinsky, Chris, 2013. Human mobility characterization from cellular network data. Commun. ACM 56 (1), 74–82.

Ben-akiva, Moshe, Bierlaire, Michel, 2003. Discrete choice methods & their applications to short term travel decisions. Handbook of Transportation Science, (1985). pp. 29.

Bengtsson, Linus, Lu, Xin, Thorson, Anna, Garfield, Richard, 2011. Johan Von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. PLoS Med. 8 (8), e1001083.

Blondel, Vincent D., Guillaume, Jean-Loup, Lambiotte, Renaud, Lefebvre, Etienne, 2008. Fast unfolding of communities in large networks. J. Stat. Mech. 2008 (10) P10008.

Blondel, Vincent D., Decuyper, Adeline, Krings, Gautier, 2015. A survey of results on mobile phone datasets analysis. EPJ Data Science 4 (1), 10.

Burger, Martijn, Meijers, Evert, 2012. Form follows function? Linking morphological and functional polycentricity. Urban Stud. 49 (5), 1127–1149.

Castro, Pablo Samuel, Zhang, Daqing, Li, Shijian, 2012. Urban traffic modelling and prediction using large scale taxi gps traces. In: International Conference on Pervasive Computing. Springer, pp. 57–72.

Eunjoon Cho, Seth A Myers, and Jure 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, Pages 1082–1090. ACM, 2011.

Clauset, Aaron, Newman, Mark E.J., Moore, Cristopher, 2004. Finding community structure in very large networks. Phys. Rev. E 70 (6), 066111.

Dijst, Martin, 1999. Two-earner families and their action spaces: a case study of two dutch communities. GeoJournal 48 (3), 195–206.

Gao, Song, Liu, Yu, Wang, Yaoli, Ma, Xiujun, 2013. Discovering spatial interaction communities from mobile phone data. Trans. GIS 17 (3), 463–481.

Geurs, Karst T., van Wee, Bert, 2004. Accessibility evaluation of land-use and transport strategies: Review and research directions. J. Transp. Geogr. 12 (2), 127–140 ISSN 09666923. doi: 10.1016/j.jtrangeo. 2003.10.005.

Gonzalez, Marta C., Hidalgo, Cesar A., Barabasi, Albert-Laszlo, 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782.

Han, Sun Sheng, 2005. Polycentric urban development and spatial clustering of condominium property values: Singapore in the 1990s. Environ. Plan. A 37 (3), 463–481. ISSN 0308518X. https://doi.org/10.1068/a3746.

Hanson, Susan, 1980. The importance of the multi-purpose journey to work in urban travel behavior. Transportation 9 (3), 229–248.

Hanson, Susan, Hanson, Perry, 1981. The travel-activity patterns of urban residents: dimensions and relationships to sociodemographic characteristics. Econ. Geogr. 57 (4), 332–347.

Hey, Tony, Tansley, Stewart, Tolle, Kristin M., et al., 2009. The Fourth Paradigm: Data-intensive Scientific Discovery. 1 Microsoft Research Redmond, WA.

Kang, Chaogui, Qin, Kun, 2016. Understanding operation behaviors of taxicabs in cities by matrix factorization. Comput. Environ. Urban. Syst. Vol. 60, 79–88.

Kang, Chaogui, Sobolevsky, Stanislav, Liu, Yu, Ratti, Carlo, 2013. Exploring human movements in singapore: a comparative analysis based on mobile phone and taxicab usages. In: Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing. ACM, pp. 1.

Kitamura, Ryuichi, 1984. Incorporating trip chaining into analysis of destination choice. Transp. Res. B Methodol. 18 (1), 67–81.

Kwan, Mei-Po, 1999. Gender, the home-work link, and space-time patterns of none-mployment activities. Econ. Geogr. 75 (4), 370–394.

Li, Bin, Zhang, Daqing, Sun, Lin, Chen, Chao, Li, Shijian, Qi, Guande, Yang, Qiang, 2011. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In: In Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on. IEEE, pp. 63–68.

Li, Miaoyi, Dong, Lei, Shen, Zhenjiang, Lang, Wei, Ye, Xinyue, 2017. Examining the interaction of taxi and subway ridership for sustainable urbanization. Sustain. For. 9 (2), 242.

Liu, Liang, Hou, Anyang, Biderman, Assaf, Ratti, Carlo, Chen, Jun, 2009. 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. IEEE, pp. 1–6.

Liu, Liang, Andris, Clio, Ratti, Carlo, 2010. Uncovering cabdrivers' behavior patterns from their digital traces. Comput. Environ. Urban. Syst. 34 (6), 541–548.

Liu, Yu, Wang, Fahui, Xiao, Yu, Gao, Song, 2012a. Urban land uses and traffic 'source-sink areas': evidence from gps-enabled taxi data in shanghai. Landsc. Urban Plan. 106 (1), 73–87.

Liu, Yu, Kang, Chaogui, Gao, Song, Xiao, Yu, Tian, Yuan, 2012b. Understanding intra-urban trip patterns from taxi trajectory data. J. Geogr. Syst. 14 (4), 463–483.

Liu, Xi, Gong, Li, Gong, Yongxi, Liu, Yu, 2015. Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr. 43, 78–90.

Louail, Thomas, Lenormand, Maxime, Cantú, Oliva García, Picornell, Miguel, Herranz, Ricardo, Frias-Martinez, Enrique, Ramasco, José J., Barthelemy, Marc, 2014. From mobile phone data to the spatial structure of cities. Sci. Rep. 4, 1–12 (5276).

Newsome, Tracy H., Walcott, Wayne A., Smith, Paul D., 1998. Urban activity spaces: illustrations and application of a conceptual model for integrating the time and space dimensions. Transportation 25 (4), 357–377.

Pas, Eric, 1995. The urban transportation planning process. In: Hanson, S. (Ed.), The Geography of Urban Transportation. Guilford Press, New York, pp. 53–77.

Pelletier, Marie-Pier, Trépanier, Martin, Morency, Catherine, 2011. Smart card data use in public transit: A literature review. Transp. Res. C 19 (4), 557–568.

Santi, Paolo, Resta, Giovanni, Szell, Michael, Sobolevsky, Stanislav, Steven, H., 2014. Strogatz, and Carlo Ratti. Quantifying the benefits of vehicle pooling with shareability networks. Proc. Natl. Acad. Sci. 111 (37), 13290–13294.

Schönfelder, Stefan, Axhausen, Kay W., 2003. Activity spaces: measures of social exclusion? Transp. Policy 10 (4), 273–286.

Song, Chaoming, Zehui, Qu, Blumm, Nicholas, Barabási, Albert-László, 2010a. Limits of predictability in human mobility. Science 327 (5968), 1018–1021.

Song, Chaoming, Koren, Tal, Wang, Pu, Barabási, Albert-László, 2010b. Modelling the scaling properties of human mobility. Nat. Phys. 6 (10), 818–823.

Strathman, James G., Dueker, Kenneth J., Davis, Judy S., 1994. Effects of household structure and selected travel characteristics on trip chaining. Transportation 21 (1), 23–45.

Sun, Lijun, Axhausen, Kay W., Lee, Der-Horng, Huang, Xianfeng, 2013. Understanding metropolitan patterns of daily encounters. Proc. Natl. Acad. Sci. 110 (34), 13774–13779.

Sun, Lijun, Jin, Jian Gang, Axhausen, Kay W., Lee, Der-Horng, Cebrian, Manuel, 2014. Quantifying long-term evolution of intra-urban spatial interactions. J. R. Soc. Interface 12 (102). http://dx.doi.org/10.1098/rsif.2014.1089. (jan, ISSN 1742-5689).

Tu, Wei, Li, Qingquan, Fang, Zhixiang, Shaw, Shih-lung, Zhou, Baoding, Chang, Xiaomeng, 2016. Optimizing the locations of electric taxi charging stations: a spatial–temporal demand coverage approach. Transport. Res. C 65, 172–189.

Tu, Wei, Cao, Jinzhou, Yue, Yang, Shaw, Shih-Lung, Zhou, Meng, Wang, Zhensheng, Chang, Xiaomeng, Xu, Yang, Li, Qingquan, 2017. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. Int. J. Geogr. Inf. Sci. 31 (12), 2331–2358.

Tu, Wei, Cao, Rui, Yue, Yang, Zhou, Baoding, Li, Qiuping, Li, Qingquan, 2018. Spatial variations in urban public ridership derived from gps trajectories and smart card data. J. Transp. Geogr. 69, 45–57.

Wang, Handong, Zou, Haixiang, Yue, Yang, Li, Qingquan, 2009. Visualizing hot spot analysis result based on mashup. In: Proceedings of the 2009 International Workshop on Location Based Social Networks. ACM, pp. 45–48.

Wang, Wenjun, Pan, Lin, Yuan, Ning, Zhang, Sen, Liu, Dong, 2015. A comparative analysis of intra-city human mobility by taxi. Phys. A Stat. Mech. Appl. 420, 134–147.

Wesolowski, Amy, Eagle, Nathan, Tatem, Andrew J., Smith, David L., Noor, Abdisalan M., Snow, Robert W., Buckee, Caroline O., 2012. Quantifying the impact of human mobility on malaria. Science 338 (6104), 267–270.

Xu, Yang, Shaw, Shih-Lung, Zhao, Ziliang, Yin, Ling, Fang, Zhixiang, Li, Qingquan, 2015. Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. Transportation 42 (4), 625–646.

Xu, Yang, Shaw, Shih-Lung, Zhao, Ziliang, Yin, Ling, Lu, Feng, Chen, Jie, Fang, Zhixiang, Li, Qingquan, 2016. Another tale of two cities: understanding human activity space using actively tracked cellphone location data. Ann. Am. Assoc. Geograph. 106 (2), 489–502.

Xu, Yang, Belyi, Alexander, Bojic, Iva, Ratti, Carlo, 2017. How friends share urban space: an exploratory spatiotemporal analysis using mobile phone data. Trans. GIS 21 (3), 468–487.

Zhong, Chen, Arisona, Stefan Müller, Huang, Xianfeng, Batty, Michael, Schmitt, Gerhard, 2014. Detecting the dynamics of urban structure through spatial network analysis. Int. J. Geogr. Inf. Sci. 28 (11), 2178–2199.

Zhong, Chen, Ed, Manley, Arisona, Stefan Müller, Batty, Michael, Schmitt, Gerhard, 2015. Measuring variability of mobility patterns from multiday smart-card data. J. Comput. Sci. 9, 125–130.