

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: http://www.tandfonline.com/loi/raag21

Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data

Yang Xu, Shih-Lung Shaw, Ziliang Zhao, Ling Yin, Feng Lu, Jie Chen, Zhixiang Fang & Qingquan Li

To cite this article: Yang Xu, Shih-Lung Shaw, Ziliang Zhao, Ling Yin, Feng Lu, Jie Chen, Zhixiang Fang & Qingquan Li (2016): Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data, Annals of the American Association of Geographers

To link to this article: <u>http://dx.doi.org/10.1080/00045608.2015.1120147</u>



Published online: 09 Feb 2016.

(

Submit your article to this journal oxdot S



View related articles 🗹



則 🛛 View Crossmark data 🗹

Full Terms & Conditions of access and use can be found at http://www.tandfonline.com/action/journalInformation?journalCode=raag21

Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data

Yang Xu,* Shih-Lung Shaw,* Ziliang Zhao,* Ling Yin,[†] Feng Lu,[‡] Jie Chen,[‡] Zhixiang Fang,[§] and Qingquan Li[¶]

*Department of Geography, University of Tennessee, Knoxville

[†]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen

[‡]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing

[§]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University [¶]Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University

Activity space is an important concept in geography. Recent advancements of location-aware technologies have generated many useful spatiotemporal data sets for studying human activity space for large populations. In this article, we use two actively tracked cellphone location data sets that cover a weekday to characterize people's use of space in Shanghai and Shenzhen, China. We introduce three mobility indicators (daily activity range, number of activity anchor points, and frequency of movements) to represent the major determinants of individual activity space. By applying association rules in data mining, we analyze how these indicators of an individual's activity space can be combined with each other to gain insights of mobility patterns in these two cities. We further examine spatiotemporal variations of aggregate mobility patterns in these two cities. Our results reveal some distinctive characteristics of human activity space in these two cities: (1) A high percentage of people in Shenzhen have a relatively short daily activity range, whereas people in Shanghai exhibit a variety of daily activity ranges; (2) people with more than one activity anchor point tend to travel further but less frequently in Shanghai than in Shenzhen; (3) Shenzhen shows a significant north–south contrast of activity space that reflects its urban structure; and (4) travel distance in both cities is shorter around noon than in regular work hours, and a large percentage of movements around noon are associated with individual home locations. This study indicates the benefits of analyzing actively tracked cellphone location data for gaining insights of human activity space in different cities. Key Words: active cellphone location data, association rules, human activity space, spatiotemporal patterns.

活动空间是地理学中的重要概念。晚近能感知位置的科技革新,已为研究大型人口的人类活动空间生产 了诸多有用的时空数据集。我们于本文中,运用两组动态追踪的手机位置数据集,这两组数据集涵盖了 一整个工作日,用以描绘人们在中国上海与深圳的空间使用。我们引介三项能动性指标 (每日活动范围, 活动定着点数量,以及移动的频率) 来呈现个人活动空间的决定因素。我们透过将关联规则运用至数据 挖掘,分析这些个人活动空间的指标,如何能够相互结合,以获得这两座城市中的能动性模式之洞见。我 们进一步检视这两座城市中的累计能动性模式的时空变异。我们的研究结果,显示出这两座城市中的人 类活动的若干特徵: (1) 在深圳,有高比率的人口具有相对而言较短的每日活动范围,而在上海,人们则展 现出多样的每日活动范围; (2) 在上海,具有两个以上活动定着点的人们,较具有两个以上活动定着点的 深圳人更倾向拥有较为远距、却较不频繁的移动; (3) 深圳展现出显着的南—北活动空间对比,并反映出 其城市结构; (4) 两座城市中的旅行距离,在中午期间较正常上班时间为短,而午间移动有大幅的比例与 个人家户位置有关。本研究指出,分析动态追踪的手机位置数据,有助于获得不同城市中的人类活动空 间之洞见。关键词: 动态手机位置数据,关联规则,人类活动空间,时空模式。

El espacio de actividad es un concepto importante en geografía. Los recientes avances en tecnologías inteligentes de localización han generado muchos conjuntos de datos espaciotemporales útiles para estudiar el espacio de actividad humana para poblaciones grandes. En este artículo usamos dos conjuntos de datos de localización del teléfono celular activamente rastreado que cubren un día de la semana para caracterizar el uso del espacio por la gente en Shanghai y Shenzhen, China. Introdujimos tres indicadores de movilidad (ámbito cotidiano de actividad, número de puntos de anclaje de la actividad, y frecuencia de los movimientos) para representar los principales determinantes del espacio de actividad individual. Aplicando reglas de asociación en la minería de datos, analizamos la forma como estos indicadores del espacio de actividad de un individuo pueden combinarse entre sí para ganar entendimiento sobre los patrones de movilidad en estas dos ciudades. Adicionalmente examinamos variaciones espaciotemporales de patrones agregados de movilidad en las dos ciudades. Nuestros hallazgos revelan algunas características distintivas del espacio de la actividad humana en las dos urbes: (1) Un alto porcentaje de la gente de Shenzhen tiene un ámbito de actividad cotidiana relativamente corto, mientras la gente de Shanghai exhibe una variedad de ámbitos de actividad cotidiana; (2) la gente que tiene más de un punto de anclaje de la actividad tiende a viajar más lejos pero menos frecuentemente en Shanghai que en Shenzhen; (3) Shenzhen muestra un contraste significativo de espacio de actividad en sentido nortesur que refleja su estructura urbana, y (4) la distancia de viaje en ambas ciudades es más corta alrededor del mediodía que en las horas regulares de trabajo, y un alto porcentaje de los movimientos alrededor del mediodía están asociados con las localizaciones individuales de los hogares. Este estudio señala los beneficios de analizar los datos de la localización del teléfono celular activamente rastreado para ganar entendimiento del espacio de la actividad humana en diferentes ciudades. *Palabras clave: datos de localización del teléfono celular activado, reglas de asociación, espacio de actividad humana, patrones espaciotemporales.*

uman activities and movements generate the pulses of our cities. Studying human activity space could yield important insights into many socioeconomic phenomena and facilitate our understanding of human behavior and its relationships with the built environment. Activity space is an important concept in geography that describes the spatial extent, frequent locations, and movements of people's daily activities (Golledge and Stimson 1997; Schönfelder and Axhausen 2003). In the past several decades, studies of human activity space were mainly based on travel surveys and Global Positioning System data (Hanson 1980; Dijst 1999; Kwan 1999, 2000; Axhausen et al. 2002; Shoval and Isaacson 2007; Shaw, Yu, and Bombom 2008; Zheng et al. 2008; Chen et al. 2011; Shen, Kwan, and Chai 2013). Recent advancements of location-aware technologies have made it possible to collect large individual tracking data sets for studying the whereabouts of people over space and time. These newly emerging data sources, like social media and cellphone location data, provide us with opportunities to investigate human activity space for large populations. Although various methods have been suggested to measure people's use of space (Candia et al. 2008; Isaacman et al. 2010; Song et al. 2010; Cheng et al. 2011; Cho, Myers, and Leskovec 2011; Becker et al. 2013; Silm and Ahas 2014), several research challenges remain to be better addressed. For example, many previous studies examined important determinants of human activity space independently. It remains unclear how different determinants of an individual activity space are related to each other. Although some studies have used methods such as clustering to identify mobility patterns based on multiple characteristics of individual activity space, it can be difficult to interpret the major characteristics of each population group. In this study, we develop some intuitive individual mobility indicators (IMIs) to

represent individual activity space from three critical perspectives (i.e., spatial extent, frequent locations, and movements). We then introduce several approaches to uncover the interrelationships of these mobility indicators and compare activity space patterns among different cities or population groups.

We use two large, actively tracked cellphone location data sets collected in two major Chinese cities, Shanghai and Shenzhen, on a workday to investigate and compare human activity space patterns between these two cities as an example to illustrate the usefulness of our proposed IMIs. Different from call detail records (CDRs) that are passively collected when people engage in communication activities such as phone calls and text messages (Song et al. 2010; Becker et al. 2013; Xu et al. 2015), actively tracked cellphone location data provide locations of each cellphone at a regular time interval by detecting where a cellphone is located. Because many people make infrequent use of their cellphones, in a day and cellphone usage tends to have a natural biased spatiotemporal pattern (e.g., more cellphone communications after work than before work in a day), actively tracked cellphone location data generally offer better spatiotemporal coverage of individual activity space than CDR data. The main objective of this article is to develop a method that can measure the major characteristics of individual activity space based on actively tracked cellphone location data such that we can effectively compare aggregate activity space patterns among different cities. To achieve the objective, we develop three IMIs-daily activity range, number of activity anchor points, and frequency of movement-to answer critical questions of individual activity space (i.e., how far, how many, and how frequent). We then apply association rules in data mining (Han, Kamber, and Pei 2011) to examine how the three indicators are related to each other among the activity spaces of different individuals. We further investigate spatial and temporal variations of major characteristics of aggregate human activity patterns between Shanghai and Shenzhen.

Literature Review

Activity space and its related concepts (Lynch 1960; Brown and Moore 1970; Horton and Reynolds 1971; Lenntorp 1977; Golledge and Stimson 1997) have been widely used in geography to examine people's use of space. Various approaches including, but not limited to, standard deviational ellipse (Yuill 1971), confidence ellipse (Schönfelder and Axhausen 2003), and daily potential path area (Kwan 1998) have been proposed to measure individual space usage from perspectives of spatial extent, frequent locations, and movements. Over the past several decades, many studies have applied these approaches to study human activity space and its relationships with sociodemographic characteristics (Hanson and Hanson 1981; Newsome, Walcott, and Smith 1998; Dijst 1999; Kwan 1999; Axhausen et al. 2002; Buliung and Kanaroglou 2006). Most of these studies involved activitytravel surveys that can be expensive to collect and often limited in sample size. As we move into the big data era, many new data sources have emerged. For example, there have been several studies that used actively tracked mobile phone location data to solve problems related to mobility prediction (Gao, Tang, and Liu 2012), recognition of place categories (Zhu et al. 2012), and estimation of demographic attributes (Brdar, Culibrk, and Crnojevic 2012). Such data sets provide new opportunities for understanding people's use of space in their daily lives. Large data volumes present new challenges to the study of human activity space, however. In recent years, research has been conducted to study human activity space using cellphone location data. Measures such as radius of gyration (Gonzalez, Hidalgo, and Barabási 2008; Song, Blumm, and Barabási 2010), activity anchor points (Phithakkitnukoon et al. 2010; Cho, Myers, and Leskovec 2011), and daily activity range (Becker et al. 2013) have been used to reflect major characteristics of individual activity space. Most of the studies analyzed these characteristics separately, which could lead to a partial view of individual activity space.

Although clustering methods have been applied to address some research issues, such as identifying individuals with similar location sequence (Li et al. 2008), commuting flexibility (Shen, Kwan, and Chai 2013), and spatiotemporal activity patterns (Chen et al. 2011), it sometimes can be difficult to interpret the major characteristics of each population group derived from the clustering algorithms. Moreover, these clustering methods (e.g., hierarchical clustering) are computationally intensive and often perform inefficiently over very large data sets. This study attempts to develop some easy-to-compute and yet effective approaches to gain insights into activity space patterns. We build three mobility indicators to represent the most important determinants of individual activity space. By combining activity space theory and association rules in data mining, this study aims at providing a multidimensional view of individual activity spaces across different cities.

Study Area and Data Sets

Shanghai and Shenzhen are two major cities in China with their gross domestic products ranked the first and fourth, respectively, among all Chinese cities (National Bureau of Statistics of China 2012). Shanghai is a centuryold metropolis, with a population of 24 million as of 2013. It has eighteen administrative districts and covers an area of 6,340 km² (Figure 1A and 1B). Shenzhen, which is located in southern China adjacent to Hong Kong, has six administrative districts covering $1,952 \text{ km}^2$ and a population of 15 million as of 2012 (Figure 1C). Shenzhen was a small fishing village when it was chosen as China's first Special Economic Zone (SEZ) in 1979. Fast economic growth and urbanization have transformed Shenzhen into a major migrant city. As of 2011, the migrant population accounted for more than 70 percent of the total population in Shenzhen (Gazette of the People's Government of Shenzhen Municipality 2011). According to recent travel surveys (Lu and Gu 2011; Urban Planning Land & Resources Commission of Shenzhen Municipality 2013), nonmotorized trips accounted for a large percentage of total trips in Shanghai (walking: 26.2 percent; bicycle or moped: 28.7 percent) and in Shenzhen (walking: 50.0 percent; bicycle or moped: 6.2 percent). Comparing people's daily activity space in these two cities can help us better understand their urban dynamics that could be useful for urban design, transportation planning, business studies, and other applications.

This article uses two actively tracked cellphone data sets¹ collected on a weekday in Shenzhen (23 March 2012) and Shanghai (3 September 2012), respectively. The Shenzhen data set covers 5.8 million cellphones, with their locations reported approximately once every



Figure 1. Study areas: (A) administrative districts of Shanghai; (B) inset map of the central part of Shanghai; (C) administrative districts of Shenzhen.

hour as (x, y) coordinates of the cellphone tower to which a cellphone is assigned. This data set includes cellphone location records between 00:00 and 23:00 during the study day; each cellphone therefore has twenty-three observations. The Shanghai data set consists of 0.69 million cellphones. To be comparable, we removed records of the 23:00–24:00 time window in Shanghai's data set. Table 1 shows an example of the data format of the two cellphone data sets. The average nearest distance among the cellphone towers in Shanghai is 0.21 km, as compared to 0.19 km in Shenzhen.

Method

This section first introduces three IMIs, followed by estimation of each individual's home location that will be used as a reference point when we analyze individual activity space. We then describe how association rules are used to summarize and compare people's activity spaces in Shanghai and Shenzhen.

Individual Mobility Indicators

As shown in Table 1, an individual's cellphone trajectory T can be represented as

$$T = \{P_1(x_1, y_1, t_1), P_2(x_2, y_2, t_2), \dots, P_i(x_i, y_i, t_i)\},$$
(1)

where P_i denotes the *i*th (*i* = 1, 2, ..., 23) cellphone location record; x_i and y_i denote the longitude and latitude of a cellphone tower; and t_i represents a one-hour time window in which each location was recorded. We develop three IMIs, which are the number of activity anchor points, daily activity range, and frequency of movement, to capture the major characteristics of an individual activity space represented by *T*.

Measures such as standard deviational ellipse (Yuill 1971) and radius of gyration (Gonzalez, Hidalgo, and Barabási 2008) have been used in previous studies to represent the spatial dispersion of an individual's daily activities. In our study, we introduce *daily activity range*, which is defined as the maximum distance

 Table 1. Example of an individual's cellphone records in both data sets

User ID	Record ID	Time window in which location was reported (<i>t</i>)	Longitude of cellphone tower (x)	Latitude of cellphone tower (y)
932****	1	00:00-01:00	113.****	22.****
932****	2	01:00-02:00	113.****	22.****
932****	3	02:00-03:00	113.****	22.****
			113.****	22.****
932****	23	22:00-23:00	113.****	22.****

between all pairs of cellphone towers in T, to describe the spatial extent of an individual's activity space.²

Activity anchor points have been frequently used in the literature (e.g., Dijst 1999; Schönfelder and Axhausen 2003; Ahas et al. 2010) to denote a person's major activity locations such as home, workplace, favorite restaurants, and so on. The meaning of an activity anchor point could vary due to the context of each study, however. In this article, we define an activity anchor point as a set of cellphone towers that are geographically concentrated and where an individual spent a certain amount of time. One challenge of using cellphone location data to determine an individual's activity anchor points and movements among the anchor points is that an individual's cellphone location could switch among adjacent cellphone towers due to either cellphone load balancing (Csáji et al. 2013) or cellphone signal strength variation (Isaacman et al. 2012). Hence, to derive activity anchor points for T, we first extract all cellphone towers traversed by T, and calculate the frequency (i.e., number of time windows) each cellphone tower was visited. We then select the most visited cellphone tower, and group all the cellphone towers that are located within 0.5 km of the selected tower into a cluster. We then select the next most visited cellphone tower and perform the same grouping process. The process is repeated until all of the cellphone towers in T are processed. Finally, we calculate the number of cellphone location records (i.e., observations) assigned to each cluster. In this study, any cluster with two or more cellphone location records is identified as an activity anchor point. Figure 2 gives an example of an individual's trajectory in a three-dimensional space-time system proposed by Hägerstrand (1970). This individual's cellphone tower



Figure 2. An individual's cellphone trajectory T and key concepts in individual mobility indicators represented by a space–time system proposed by Hägerstrand (1970). (Color figure available online.)

locations are grouped into four clusters, with three of them (clusters A, B, and C) being identified as activity anchor points.

Note that we choose a constant distance threshold of 0.5 km to derive individual activity anchor points for the two cellphone data sets, and the reasons are as follows. First, although we are aware that cellphone tower densities could vary within a city, choosing a constant threshold enables us to consistently evaluate the space usage of individuals in a city. Second, as Shanghai and Shenzhen share a similar average nearest distance among cellphone towers (0.21 km and 0.19 km, respectively), choosing 0.5 km can not only address the problem of signal switches among nearby cellphone towers but also facilitates the comparison of human activity space between these two cities.

Movement is another important characteristic of human activity space. When deriving frequency of movement in T, we only consider the movements that occurred between clusters (green lines in Figure 2) because it is difficult to determine whether the movements within clusters (i.e., red lines in Figure 2) represent an individual's actual movements or are simply caused by load balancing or signal switches. By choosing a threshold of 0.5 km, we minimize the impact of load balancing and signal switches while maintaining all major movements in an individual's trajectory. Here, the frequency of movement is defined as the number of intercluster movements in T. This indicator measures how actively an individual travels among different activity locations in a day. For example, the individual in Figure 2 has a frequency of movement of 5 (i.e., the number of green segments in Figure 2).

Estimation of Individual Home Location

Considering people's daily routines in most big cities in China (Long, Zhang, and Cui 2012), we define home location of an individual as the activity anchor point with a minimum of four hours of stay at the location before 7:00 a.m. Based on this rule, we are able to estimate the home location for 97 percent of the sampled population in both Shanghai and Shenzhen. For each city, we compare our estimated home locations by administrative districts against the most recent census data (Gazette of the Sixth National Population Census for Shanghai Municipality 2010; Gazette of the People's Government of Shenzhen Municipality 2011). We find that our estimates are in agreement with the census data according to the population distribution by administrative districts (with Pearson coefficients of 0.95 and 0.99 for Shanghai and Shenzhen, respectively).

Building Association Rules

Association rules have been widely used in business research to uncover items that are frequently purchased together. They also have been used to describe associations between quantitative items or attributes (Han, Kamber, and Pei 2011). To uncover the major characteristics of individual activity space in the two cities, one key challenge is to analyze how the three IMIs are associated with each other to characterize each individual's activity space. Note that the IMIs can be represented as

$$IMIs \rightarrow (N, R, F),$$
 (2)

where N denotes the number of activity anchor points, R describes an individual's daily activity range, and F denotes the frequency of movement. The value of each indicator can be partitioned into several intervals:

$$N \to (N_1, N_2, \dots, N_a) \tag{3}$$

$$R \to (R_1, R_2, \dots, R_b) \tag{4}$$

$$F \to (F_1, F_2, \dots, F_c), \tag{5}$$

where *a*, *b*, and *c* represent the number of intervals or classes defined for each corresponding indicator. For each individual X, the IMIs can be represented by their specific characteristics based on the defined intervals:

$$X \to (N_i, R_j, F_k) \text{ (where } 0 \le i \le a, 0 \le j \le b, \ 0 \le k \le c).$$
(6)

We then introduce association rules to summarize the characteristics of human activity space for each city. The association rules are formulated as

$$(X, "N_i") \Rightarrow (X, "R_i," and "F_k").$$
 (7)

These rules describe how different intervals of the three IMIs are associated with each other in each individual's activity space. For each city, the support and the confidence of the association rules are calculated as follows:

support(X, "N_i")
$$\Rightarrow$$
 (X, "R_j," and "F_k")
= $\frac{\text{number of individuals with } (X, "Ni")}{\text{total population in the cellphone data set}}$
(8)

confidence(X, "N_i")
$$\Rightarrow$$
 (X, "R_j," and "F_k")
= $\frac{\text{number of individuals with } (X, "N_i" \text{ and "Rj" and "Fk")}{\text{number of individuals with } (X, "Ni")}$ (9)

The *support* of a rule denotes the amount of individuals meeting the left-hand-side (LHS) condition divided by the total population of the data set. The *confidence* of a rule denotes the amount of individuals meeting both sides of the rule divided by the number of individuals meeting the LHS condition. Both support and confidence indexes describe important characteristics of human activity spaces extracted from a particular data set. Note that we use N_i as the LHS of the association rules because the number of activity anchor points for an individual X is a discrete variable, which can be directly derived from individual cellphone trajectories.

Analysis Results

General Statistics

We first derive the general statistics of IMIs for the two cities. As shown in Figure 3A, the majority of people in Shenzhen had only one or two activity anchor points in the study day (38.8 percent and 38.5 percent of the population, respectively), whereas people in Shanghai were more diversified regarding the number of activity anchor points (N). For daily activity range (R), a large percentage of people in Shenzhen traveled within a very short distance during the day, as illustrated in Figure 3B. The cumulative distribution shows that nearly 50 percent of the people in Shenzhen traveled within 1.0 km and about 82 percent traveled within 5.0 km, as compared to 26 percent of people in Shanghai who traveled within 1.0 km and 60 percent who traveled within 5.0 km. The medians of R in Shenzhen and Shanghai are 1.1 km and 3.1 km, respectively.³ For frequency of movement (F), people on average made 3.76 movements in Shenzhen, as compared to 4.34 in Shanghai. The results indicate that (1) people in Shanghai had more major activity locations (i.e., N) in a day than people in Shenzhen; (2) the spatial extent



Figure 3. Distribution patterns of (A) number of activity anchor points and (B) daily activity range in Shanghai and in Shenzhen. (Color figure available online.)

of people's activities in Shanghai was generally larger than that of people in Shenzhen; and (3) people in Shanghai were more "active" in terms of the movements among their daily activity locations. It is still unclear, however, how the three determinants (N, R, and F) are related to each other in an individual's activity space. For example, do people with the same number of activity anchor points in Shenzhen and Shanghai have similar daily activity range or movement frequency? In the next section, we discuss the interrelationships of (N, R, F) based on the association rules to further understand the differences and similarities of individual activity space in the two cities.

Association Rules of IMIs

To generate the association rules, we first partition the three IMIs into intervals. As shown in Table 2, we partition N, R, and F into four, five, and five intervals, respectively. Each interval $(N_i, R_{j}, \text{ or } F_k)$ represents a particular value or range of values for the corresponding indicator. By mining the associations among the three indicators using the defined intervals, we are able to uncover the major characteristics of individual activity space for particular population groups within each city.

The support and confidence of the association rules are calculated to compare individual activity spaces in the two cities. As illustrated in Figure 4A and 4B, although there are more people with one activity anchor point in Shenzhen (support = 38.8 percent) than in Shanghai (support = 23.6 percent), people in these two subsets (N=1) had very similar activity space characteristics. The two subsets are dominated by individuals with very short daily activity range (R_1) and low movement frequency (F_1 and F_2). Only a very small percentage of people traveled very far (R_4 and R_5) and frequently (F_4 and F_5). The result indicates that quite a few people in both cities stayed around one particular location during the day. The "immobility" of these individuals reflects an interesting perspective of human activity spaces in the two cities and calls for further investigation of its driving force and related societal implications.

When N = 2 (Figure 4A and 4B), the percentages of people with different travel ranges distribute relatively evenly within each interval of *R* in Shanghai, whereas Shenzhen shows a decay with increasing travel range. The subset of Shenzhen is dominated by people with short daily activity ranges (e.g., 63.2 percent of people with $R \le 2$ km), whereas in the

Number of activity anchor points (N)		Daily activity range (R)		Frequency of movement (F)	
Intervals	Values	Intervals	Values	Intervals	Values
N ₁	N = 1	R_1	$R \le 1 \text{ km}$	F_1	F = 0
N_2	N = 2	R_2	$1 \text{ km} < R \leq 2 \text{ km}$	F_2	$1 \le F \le 3$
$\overline{N_3}$	N = 3	R_3	$2 \text{ km} < R \leq 5 \text{ km}$	$\overline{F_3}$	$4 \le F \le 7$
N ₄	$N \ge 4$	R_4 R_5	$5 \text{ km} < R \le 10 \text{ km}$ $R > 10 \text{ km}$	F_4 F_5	$8 \le F \le 11$ $F \ge 12$

Table 2. Intervals (classes) defined for the association rules



Figure 4. Association rules of individual activity space: (A_1-A_4) confidence of rules with left-hand-side (LHS) organized by N_i in Shanghai; (B_1-B_4) confidence of rules with LHS organized by N_i in Shenzhen. For each individual graphic, the percentages next to interval labels denote the sum of confidence for the corresponding rows or columns. (Color figure available online.)

subset of Shanghai, many people traveled very far in a day (e.g., 36 percent with R > 5 km). The observed difference could be potentially explained by the home–work relationships of people in the two subsets considering that home and workplace are two primary activity locations for most people. How frequently people moved serves as an important indicator of urban dynamics. According to our observation, although the majority of people in both subsets fall within F_2 and F_3 , people in Shenzhen traveled more frequently (39.7 percent and 48.7 percent within F_2 and F_3) than people in Shanghai (53.8 percent and 37.9 percent within F_2 and F_3). Note that we analyze the temporal variations of people's movement patterns in the two cities later in this section to further examine when (and where) people were more active in their daily activity spaces.

There is a notable change of the distribution of association rules as *N* increases from 2 to 3 for people in Shanghai. The majority of people with N = 3 (Figure 4A) in Shanghai had a very large daily activity range (54.5 percent with R > 5 km), which is quite different from the relatively even distribution of *R* when N = 2 (Figure 4A). The result indicates that the third activity anchor point might have a significant effect on people's travel range in Shanghai. Shenzhen, however, still has a large proportion of people with a short daily activity range (64.6 percent with $R \le 5$ km, as shown in Figure 4B), which is similar to what we observe when N = 2 (Figure 4B).

According to the comparisons among the three subgroups in the two cities, we can see that in Shenzhen, a small activity space was usually enough to fulfill various purposes of people's daily activities such as work, dining, recreation, and so forth. In Shanghai, activity locations were more widely distributed in an individual's activity space. People were more likely to travel far from their primary activity locations (i.e., home and workplace) for certain travel and activity purposes. For $N \ge 4$ (Figure 4A and 4B), we see an increase in both travel range and movement frequency in both cities as compared to the previous three subgroups. People in Shanghai still traveled further but less frequently, as compared to the same population group in Shenzhen. Note that we have tested other partition schemes to generate different intervals for IMIs, and the corresponding association rules reveal similar patterns of people's activity spaces in the two cities.

Spatial Variations of Human Activity Space

Analyzing the geographic patterns of people's activity space within the context of the built environment could produce an improved understanding of their daily activity patterns. For example, it would be meaningful to explore the geographic distributions of people with a small daily activity range (R), which is an important feature of individual activity space in both cities, especially in Shenzhen. Figure 5 illustrates the geographic distributions of people with a daily activity range $R \le 2$ km. Specifically, we divide the study areas into 2-km grids and aggregate individuals based on their estimated home locations. Each grid cell represents the number of individuals with $R \le 2$ km, normalized by the total number of individuals in that grid cell.

As shown in Figure 5A, many grid cells in the core areas of Shanghai have a higher percentage of people with R < 2 km (i.e., green cells in Huangpu, Luwan, and Jingan districts; readers can refer to the inset map in Figure 1B) as compared to the grid cells in suburbs (i.e., orange and red cells) such as Jinshan, Songjiang, Qingpu, Jiading, and Pudongxingu districts. Note that we also observe grid cells with higher percentages (i.e., green cells) in certain suburbs such as Minxing and Fengxian districts. It is interesting to find that the observed patterns are in general agreement with the analysis results by Sun, Pan, and Ning (2008), who studied job-housing balance in Shanghai. They indicated that core areas such as Huangpu, Luwan, and Jingan have more job opportunities as compared to the number of residents, so more people would have a



Figure 5. Geographic distributions of individuals with daily activity range ≤ 2 km in the two cities: (A) Geographic patterns in Shanghai; (B) geographic patterns in Shenzhen. (Color figure available online.)

relatively shorter commuting distance. Some suburbs around the core areas are more housing-oriented, so more people would have a longer commuting distance. We avoid making any further statement because the daily activity range examined in this study does not reflect people's actual commuting distances.

Figure 5B shows that there is a general north–south divide in Shenzhen and the proportion of people with R < 2 km in most grid cells of the two northern districts (Baoan and Longgang) is larger than 60 percent or even 80 percent, which indicates that most people who live around these cells have a small daily activity range during the study day. In the southern part of Shenzhen, the percentages are generally lower. To explore potential causes of the identified patterns, we also display the locations of major factories in Shenzhen. It appears that grid cells with a high percentage are generally colocated with major factories in Shenzhen. Many factories in Shenzhen provide workers with dormitories adjacent to their workplace. In addition, many immigrants tend to rent apartments near their workplace to save commuting time and cost. The findings suggest that the geographic patterns of people's activity space in Shenzhen and in Shanghai are quite different and that the identified patterns are likely to be related to the underlying socioeconomic characteristics.

Temporal Variations of Aggregate Movement Patterns

We further analyze the temporal variations of aggregate movement patterns to understand when people were more active in their daily activity spaces. Figure 6 shows the percentages of people who moved through a day in the two cities, organized by the number of individual activity anchor points (N). People with N = 1 in the two cities did not move much in a day. The percentages are relatively stable over time (less than 10 percent). As expected, movement patterns of the other three subgroups in these two cities exhibit two peaks during the morning and afternoon rush hours. There is a local peak around time intervals 12 and 13 for people in Shenzhen, however, which indicates that people in Shenzhen moved more frequently around noon than other work hours. More importantly, the difference of aggregate movement patterns around noon between the two cities explains our previous finding that people in Shenzhen generally move more frequently than people in Shanghai when controlling the number of activity anchor points (N).

We further explore the temporal variations of average movement distances in the two cities. As shown in Figure 7, people's movement distances in both cities are generally lower around noon than the work hours. By further analyzing movements around noon (i.e., time intervals 12 and 13), we find that 43 percent of individuals travel from or to their home locations around noon in Shanghai, as compared to 66 percent in Shenzhen. The shorter movement distance around noon reveals an interesting aspect of people's lifestyle in both cities. Although people in Shanghai have a longer travel distance in general, short-range movements still dominate in both cities. As described previously, the share of nonmotorized trips accounts for more than 50 percent of all trips in both Shanghai and Shenzhen. Our analysis results suggest that travel mode, such as walking and bicycling, should receive more attention in urban and transportation planning that has been mentioned in



Figure 6. Temporal variations of aggregate movement patterns in the two cities, organized by the number of individual activity anchor points: (A) Temporal patterns in Shanghai; (B) temporal patterns in Shenzhen. (Each time interval is associated with two consecutive time windows as shown in Table 1. For example, time interval 1 in this figure shows movement patterns from 00:00–01:00 to 01:00–02:00). (Color figure available online.)



Figure 7. Temporal variation of average movement distance in Shanghai and Shenzhen. (Color figure available online.)

some government reports in recent years (e.g., Urban Planning Land & Resources Commission of Shenzhen Municipality 2013). Appropriate transportation services should be deployed to accommodate medium- and short-range trips in cities such as Shenzhen and Shanghai.

Discussion and Conclusion

The emergence of big individual tracking data sets brings new opportunities and challenges to the understanding of human activity space in urban environments. In this study, we develop several intuitive IMIs using cellphone location data to describe the major determinants of individual activity space. Different from previous studies (Kang et al. 2010; Isaacman et al. 2012; Becker et al. 2013) that investigate determinants of human activity space independently, we analyze how these mobility indicators (e.g., daily activity range, number of activity anchor points, and frequency of movement) can be combined with each other to analyze an individual's activity space. The association rules of IMIs are able to uncover the complexities of individual activity space for a given population or a geographic region. The support and confidence of the derived rules serve as a signature of people's daily activity patterns and enable us to compare human activity spaces systematically across different geographic regions. By using active tracking cellphone location data sets collected in Shanghai and Shenzhen, we summarize and compare the major characteristics of activity space patterns in these two cities. The association rules and spatiotemporal analysis of aggregate human activity patterns allow us to better understand the socioeconomic

characteristics of these cities and yield some insights into transportation planning and urban design.

Our analysis results reveal several interesting aspects of human activity space in the two cities, and the implications are worth discussing. First, quite a few people in both cities stay around one particular location for the whole day. Such unique activity patterns might reflect some societal issues such as "urban villages" (Wei and Yan 2005) in cities that consist of lowincome communities of migrant population. Additional efforts are needed to further examine the "immobility" of these people and the potential driving forces related to land use planning (Pan, Shen, and Zhang 2009) and social segregation (Schönfelder and Axhausen 2003; Silm and Ahas 2014). Second, for the majority of people in Shenzhen, a small activity space was usually enough to fulfill the needs of people's daily activities, which is consistent with the government's goal of building a compact city with sustainable urban form. In Shanghai, however, activity locations are more widely distributed in an individual's activity space, and people are more likely to travel far from their home and workplaces for certain activity purposes. Shenzhen and Shanghai, one being a city with a large migrant population and the other being a century-old metropolis with many local residents, have very different sociodemographic characteristics and urban forms, which play an important role in shaping people's daily activity patterns. Third, the geographic disparity of people's travel range in Shenzhen is significant. The difference between the north and south could be partially explained by the socioeconomic divide in Shenzhen. In Shanghai, the geographic disparity is less obvious, and our analysis suggests that the identified patterns could be potentially explained by people's commuting patterns and the job-housing relationships in the city.

Currently, the research findings only reflect people's activity space in the two cities for a day. In the future, we plan to further investigate the temporal variations of individual activity space (e.g., seasonality, and difference between workdays and weekends) by using actively tracked cellphone data sets that cover longer time periods. It would also be meaningful to compare the analysis results derived from active and passive cellphone location data (e.g., CDRs), for example, to examine whether they reveal similar or different patterns of people's activity space. This will help us better understand the strengths and weaknesses of each data type and the intrinsic characteristics of human activity space. Nevertheless, the research findings in this article enhance our understanding of the geographies of human mobility in a space-time context. We believe that the proposed methods are useful to other types of large individual tracking data sets for dataintensive analyses of human activity space.

Notes

- 1. The mobile phone location data sets used in this study were acquired from research collaborators in China. The research was approved by the Institutional Review Board (IRB) at the University of Tennessee, Knoxville.
- 2. Radius of gyration (Gonzalez, Hidalgo, and Barabási 2008) is another frequently used measure that describes the spatial dispersion of an individual's activity space. In this study, we calculated both daily activity range and radius of gyration, and we found that they are highly correlated with each other (Pearson coefficient = 0.96). We thus use daily activity range in this study to represent the spatial extent of an individual's activity space due to its intuitive meaning.
- 3. The median of *R* in Shenzhen and Shanghai (1.1 km and 3.1 km, respectively) are much lower than that of the New York and Los Angeles regions (6.08 km and 8.0 km, respectively) computed by Isaacman et al. (2010) using cellphone location data. It is not surprising to see that the two U.S. cities have a larger daily activity range than Shanghai and Shenzhen because U.S. cities are more automobile oriented.

Funding

This research was jointly supported by the National Natural Science Foundation of China (41231171, 41271408, 41371377), 41301440, 41571431, Research Shenzhen Fundamental Funding of (JCYJ20130401170306842), Shenzhen Scientific Research and Development Funding Program (ZDSY20121019111146499), and Shenzhen Dedicated Funding of Strategic Emerging Industry Development Program (JCYJ20121019111128765).

References

- Ahas, R., S. Silm, O. Järv, E. Saluveer, and M. Tiru. 2010. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology* 17 (1): 3–27.
- Axhausen, K. W., A. Zimmermann, S. Schönfelder, G. Rindsfüser, and T. Haupt. 2002. Observing the rhythms of daily life: A six-week travel diary. *Transportation* 29 (2): 95–124.
- Becker, R., R. Cáceres, K. Hanson, S. Isaacman, J. M. Loh, M. Martonosi, J. Rowland, S. Urbanek, A. Varshavsky, and C. Volinsky. 2013. Human mobility

characterization from cellular network data. *Communications of the* ACM 56 (1): 74–82.

- Brdar, S., D. Culibrk, and V. Crnojevic. 2012. Demographic attributes prediction on the real-world mobile data. Paper presented at the Nokia Mobile Data Challenge 2012 Workshop, Newcastle, UK.
- Brown, L. A., and E. G. Moore. 1970. The intra-urban migration process: A perspective. *Geografiska Annaler* Series B: Human Geography 52 (1): 1–13.
- Buliung, R. N., and P. S. Kanaroglou. 2006. A GIS toolkit for exploring geographies of household activity/ travel behavior. *Journal of Transport Geography* 14 (1): 35–51.
- Candia, J., M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási. 2008. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical* 41 (22): 224015.
- Chen, J., S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. 2011. Exploratory data analysis of activity diary data: A space–time GIS approach. *Journal of Transport Geogra*phy 19 (3): 394–404.
- Cheng, Z., J. Caverlee, K. Lee, and D. Z. Sui. 2011. Exploring millions of footprints in location sharing services. Paper presented at the 5th International Conference on Weblogs and Social Media, Barcelona, Spain.
- Cho, E., S. A. Myers, and J. Leskovec. 2011. 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, ed. C. Apte, 1082–90. New York: ACM. http:// dl.acm.org/citation.cfm?id=2020408&picked=prox (last accessed 13 January 2016).
- Csáji, B. C., A. Browet, V. A. Traag, J.-C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel. 2013. Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications* 392 (6): 1459–73.
- Dijst, M. 1999. Two-earner families and their action spaces: A case study of two Dutch communities. *GeoJournal* 48 (3): 195–206.
- Gao, H., J. Tang, and H. Liu. 2012. Mobile location prediction in spatio-temporal context. Paper presented at the Nokia Mobile Data Challenge 2012 Workshop, Newcastle, UK.
- Gazette of the People's Government of Shenzhen Municipality. 2011. Issue No. 17, Serial No. 741. http://www.sz.gov. cn/zfgb/2012_1/gb785/201204/t20120423_1844697.htm (last accessed 28 October 2014).
- Gazette of the Sixth National Population Census for Shanghai Municipality. 2010. http://www.stats-sh.gov.cn/sjfb/ 201105/218819.html (last accessed 28 October 2014).
- Golledge, R. G., and R. J. Stimson. 1997. Spatial behavior: A geographic perspective. New York: Guilford.
- Gonzalez, M. C., C. A. Hidalgo, and A.-L. Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453 (7196): 779–82.
- Hägerstrand, T. 1970. What about people in regional science? Papers in Regional Science 24 (1): 7–24.
- Han, J., M. Kamber, and J. Pei. 2011. Data mining: Concepts and techniques. 3rd ed. Amsterdam: Morgan Kaufmann.

- Hanson, S. 1980. The importance of the multi-purpose journey to work in urban travel behavior. *Transportation* 9 (3): 229–48.
- Hanson, S., and P. Hanson. 1981. The travel-activity patterns of urban residents: Dimensions and relationships to sociodemographic characteristics. *Economic Geography* 57 (4): 332–47.
- Horton, F. E., and D. R. Reynolds. 1971. Effects of urban spatial structure on individual behavior. *Economic Geography* 47 (1): 36–48.
- Isaacman, S., R. Becker, R. Cáceres, S. Kobourov, J. Rowland, and A. Varshavsky. 2010. A tale of two cities. In Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications, ed. A. Dalton, 50–51. New York: http://dl.acm.org/citation.cfm?id=1734583&picked= prox (last accessed 13 January 2016).
- Isaacman, S., R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. 2012. Human mobility modeling at metropolitan scales. Paper presented at the 10th International Conference on mobile Systems, Applications, and Services, Low Wood Bay, Lake District, UK. http://www.sigmobile.org/mobisys/ 2012/index.php (last accessed 13 January 2016).
- Kang, C., S. Gao, X. Lin, Y. Xiao, Y. Yuan, Y. Liu, and X. Ma. 2010. Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *Proceedings of the 18th International Conference on Geoinformatics*, ed. Y. Liu and A. Chen, 1–7. Beijing: IEEE. http:// ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?reload=true &punumber=5559273 (last accessed 13 January 2016).
- Kwan, M.-P. 1998. Space–time and integral measures of individual accessibility: A comparative analysis using a pointbased framework. *Geographical Analysis* 30 (3): 191–216.
- ——. 1999. Gender, the home–work link, and space– time patterns of non-employment activities. *Economic* Geography 75 (4): 370–94.
- Lenntorp, B. 1977. Paths in space-time environments: A time-geographic study of movement possibilities of individuals. *Environment and Planning A* 9 (8): 961–72.
- Li, Q., Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma. 2008. Mining user similarity based on location history. Paper presented at the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Irvine, CA, USA.
- Long, Y., Y. Zhang, and C. Cui. 2012. 利用公交刷卡数据 分析北京职住关系和通勤出行 [Identifying commuting pattern of Beijing using bus smart card data]. *Journal* of Geographical Sciences 67 (10): 1339–52.
- Lu, X., and X. Gu. 2011. The fifth travel survey of residents in Shanghai and characteristics analysis. Urban Transport of China 9 (5): 1–7.
- Lynch, K. 1960. The image of the city. Cambridge, MA: MIT Press.
- National Bureau of Statistics of China. 2012. 主要城市年 度GDP [Annual GDP for major cities in China]. http://data.stats.gov.cn/workspace/index?m=csnd (last accessed 28 October 2014).

- Newsome, T. H., W. A. Walcott, and P. D. Smith. 1998. Urban activity spaces: Illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation* 25 (4): 357–77.
- Pan, H., Q. Shen, and M. Zhang. 2009. Influence of urban form on travel behaviour in four neighbourhoods of Shanghai. Urban Studies 46 (2): 275–94.
- Phithakkitnukoon, S., T. Horanont, G. Di Lorenzo, R. Shibasaki, and C. Ratti. 2010. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *Human behavior understanding*, ed. A. A. Salah, T. Gevers, N. Sebe, and A. Vinciarelli, 14–25. New York: Springer.
- Schönfelder, S., and K. W. Axhausen. 2003. Activity spaces: Measures of social exclusion? *Transport Policy* 10 (4): 273–86.
- Shaw, S. L., H. Yu, and L. S. Bombom. 2008. A space-time GIS approach to exploring large individual-based spatiotemporal datasets. *Transactions in* GIS 12 (4): 425-41.
- Shen, Y., M.-P. Kwan, and Y. Chai. 2013. Investigating commuting flexibility with GPS data and 3D geovisualization: A case study of Beijing, China. *Journal of Transport Geography* 32:1–11.
- Shoval, N., and M. Isaacson. 2007. Sequence alignment as a method for human activity analysis in space and time. Annals of the Association of American Geographers 97 (2): 282–97.
- Silm, S., and R. Ahas. 2014. Ethnic differences in activity spaces: A study of out-of-home nonemployment activities with mobile phone data. Annals of the Association of American Geographers 104 (3): 542–59.
- Song, C., Z. Qu, N. Blumm, and A.-L Barabási. 2010. Limits of predictability in human mobility. *Science* 327 (5968): 1018–21.
- Sun, B., X. Pan, and Y. Ning. 2008. 上海市就业与居住间 衡对交通出行的影响分析 [Analysis on influence of job-housing balance on commute travel in Shanghai]. Urban Planning Forum 1:77–82.
- Urban Planning Land & Resources Commission of Shenzhen Municipality. 2013. 深市步行和自行车交 通规及设计导 [Guidelines of transport planning and design for pedestrian and bicycle systems in Shenzhen]. http://www.szpl.gov.cn/xxgk/ztzl/zxcgh/jtghcgg. pdf (last accessed 31 August 2015).
- Wei, L., and X. Yan. 2005. Transformation of "urban village" and feasible mode. City Planning Review 7:9–13.
- Xu, Y., S.-L. Shaw, Z. Zhao, L. Yin, Z. Fang, and Q. Li. 2015. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation* 42 (4): 625–46.
- Yuill, R. S. 1971. The standard deviational ellipse; an updated tool for spatial description. Geografiska Annaler Series B: Human Geography 53 (1): 28–39.
- Zheng, Y., L. Liu, L. Wang, and X. Xie. 2008. Learning transportation mode from raw GPS data for geographic applications on the web. Paper presented at the 17th International Conference on World Wide Web, Beijing.
- Zhu, Y., E. Zhong, Z. Lu, and Q. Yang. 2012. Feature engineering for place category classification. Paper presented at the Nokia Mobile Data Challenge 2012 Workshop, Newcastle, UK.

YANG XU is a PhD candidate in the Department of Geography at the University of Tennessee, Knoxville, TN 37996. E-mail: yxu30@utk.edu. His research interests include geographic information science, space-time GIS, spatiotemporal analysis of human dynamics, and GIS for transportation.

SHIH-LUNG SHAW is Alvin and Sally Beaman Professor and Arts and Sciences Excellence Professor in the Department of Geography at the University of Tennessee, Knoxville, TN 37996. E-mail: sshaw@utk.edu. His research interests include transportation geography, geographic information science, space-time analytics of human dynamics, space-time GIS, and GIS for transportation.

ZILIANG ZHAO is a PhD student in the Department of Geography at the University of Tennessee, Knoxville, TN 37996. E-mail: zzhao7@utk.edu. His research interests include space-time GIS data model design, geovisualization, and big data analytics.

LING YIN is an Associate Professor at Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, 518055, PR China. E-mail: yinling@siat.ac.cn. Her research interests include spatiotemporal data analysis and GIS for transportation.

FENG LU is a researcher in the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, PR China. E-mail: luf@lreis.ac.cn. His research interests include spatial data modeling, GIS for transportation, navigation and location-based services, trajectory data mining, and complex network analysis.

JIE CHEN is an assistant researcher in the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, PR China. E-mail: chenj@lreis.ac.cn. Her research interests include urban GIS, time geography, space-time GIS, and human mobility analysis.

ZHIXIANG FANG is a Professor at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, 430079, PR China. E-mail: zxfang@whu.edu.cn. His research focuses on spatiotemporal modeling in transport geography.

QINGQUAN LI is a Professor at Shenzhen University, Guangdong, China, and Wuhan University, Wuhan, Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen, 518060, PR China. E-mail: liqq@szu.edu.cn. His research interests include 3D and dynamic modeling in GIS, location-based services, surveying engineering, integration of GIS, Global Positioning System and remote sensing, and intelligent transportation systems.