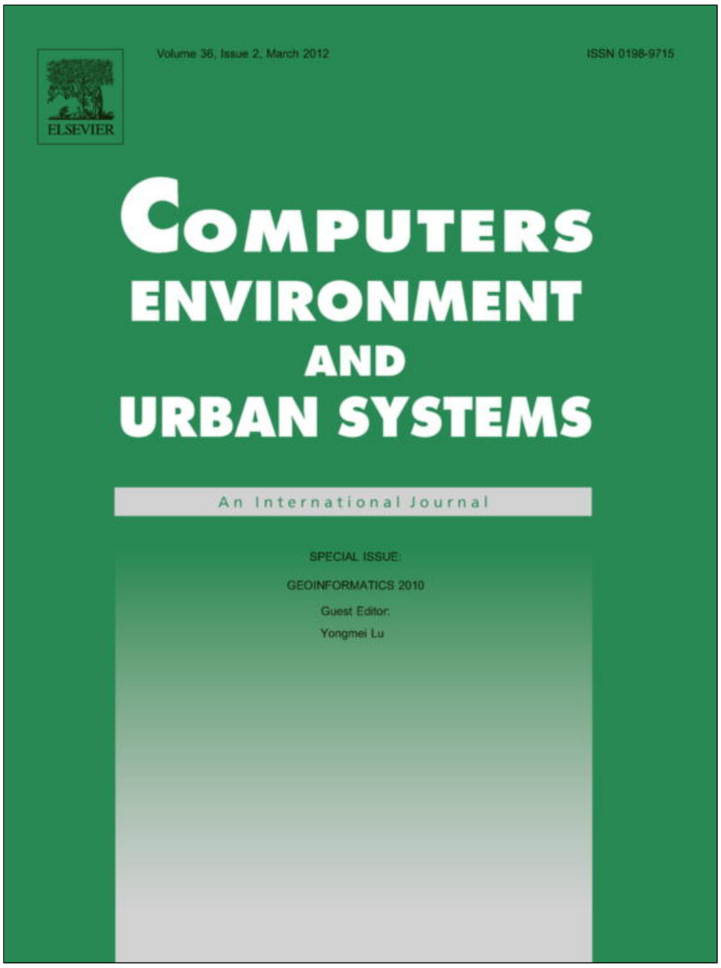
Volume 36, Issue 2, March 2012
ISSN 0198-9715
ELSEVIER
COMPUTERS
ENVIRONMENT
AND
URBAN SYSTEMS
An International Journal
SPECIAL ISSUE:
GEOINFORMATICS 2010
Guest Editor:
Yongmei Lu

Computers, Environment and Urban Systems 36 (2012) 118–130

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

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

Correlating mobile phone usage and travel behavior – A case study of Harbin, China

Yihong Yuan [a,*], Martin Raubal [a,b], Yu Liu [c]

[a] Department of Geography, University of California, Santa Barbara, CA 93106, USA
[b] Institute of Cartography and Geoinformation, ETH Zurich, 8093 Zurich, Switzerland
[c] Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Available online 6 August 2011

Keywords:
Information and communication technologies (ICTs)
Mobile phone
Human mobility
Geographic knowledge discovery

ABSTRACT

Information and communication technologies (ICTs), such as mobile phones and the Internet, are increasingly pervasive in modern society. These technologies provide new resources for spatio-temporal data mining and geographic knowledge discovery. Since the development of ICTs also impacts physical movement of individuals in societies, much of the existing research has focused on examining the correlation between ICT and human mobility. In this paper, we aim to provide a deeper understanding of how usage of mobile phones correlates with individual travel behavior by exploring the correlation between mobile phone call frequencies and three indicators of travel behavior: (1) radius, (2) eccentricity, and (3) entropy. The methodology is applied to a large dataset from Harbin city in China. The statistical analysis indicates a significant correlation between mobile phone usage and all of the three indicators. In addition, we examine and demonstrate how explanatory factors, such as age, gender, social temporal orders and characteristics of the built environment, impact the relationship between mobile phone usage and individual activity behavior.

1. Introduction

Information and communication technologies (ICTs), such as mobile phones and the Internet, have become increasingly pervasive in modern society. These technologies provide their users with more flexibility with respect to when, where, and how to travel. Understanding the influence of ICTs in our mobile information society (Raubal, 2011) will be essential for updating environmental policies, and maintaining sustainable mobility and transportation (De Souza e Silva, 2007). Moreover, ICTs have provided a wide range of spatio-temporal data sources, which can be used for geographic knowledge discovery and data mining in studies on geographic dynamics, such as human travel behavior and mobility patterns (Miller, 2009; Song, Qu, et al., 2010; Yuan, 2009). Since several spatio-temporal datasets (e.g., georeferenced mobile phone data) only provide incomplete data with relatively low resolution and few individual attributes, it is important to determine how much and to what extent we can extract knowledge from sparse data sources, as well as dealing with uncertainty in sparse datasets.

Due to the widespread use of mobile communications, there have been several studies focusing on extracting geographical knowledge from georeferenced mobile phone data. For example, Ahas' social positioning method (SPM) combines both location data and social attributes of mobile phone users to study the dynamics of urban systems (Ahas & Mark, 2005; Ahas et al., 2007). Gonzalez, Hidalgo, et al. (2008) studied the individual trajectories of 100,000 mobile phone users based on tracked location data in six months, providing new insights to understanding the basic law of human motion. Moreover, a lot of existing research has focused on the impact of mobile phone usage on human mobility patterns (Kwan, 2007; Nobis & Lenz, 2009; Thulin & Vilhelmson, 2007). However, due to the lack of sufficient data and the complicated nature of such interaction, there is still a continuing debate on how it works in everyday life. It has been proved by many researchers that we cannot simply conclude the ICT-travel connection as substitution or amplification (Kwan, Dijst, & Schwanen, 2007). Our research aims to provide a deeper understanding of how usage of mobile phones correlates with individual activity space, which is one of the most important characteristics of travel behavior. We use three indicators to represent three aspects (scale, shape, and randomness) of activity behavior: (1) radius, (2) eccentricity, and (3) entropy. The first two measure the basic descriptive characteristics of individual activity space, whereas the third one depicts the internal structure of activity space by measuring the regularity of individual trajectories. Although the three indicators have been applied in previous travel behavior studies, our research aims to explore the correlation between these indicators and the usage of mobile phones. Moreover, we focus on examining explanatory factors such as age, which impact the relationship between mobile phone usage and individual activity behavior. It is important to specify individual attributes, such as age, gender and income level,

* Corresponding author.
E-mail address: yuan@geog.ucsb.edu (Y. Yuan).

0198-9715/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compenvurbsys.2011.07.003

as well as supra-individual attributes, such as social-institutional and physical aspects, when investigating this problem (Yuan & Raubal, 2010). Since the results of the presented analysis told us little about causality, the term "relationship" in this research refers to correlation rather than causality.

The remainder of this paper is organized as follows. Section 2 provides a summary of previous work and emphasizes the novelty of our research. Section 3 introduces the mobile phone dataset from Harbin city that is used in this work, and explains the methodological background for the presented analysis. In Section 4, we explore the correlation between mobile phone usage and individual activity behavior based on the empirical results. Section 5 discusses the main results from the data analysis. In Section 6 we present conclusions and propose directions for future research.

2. Previous work

2.1. Activity space

As Jones, Koppelman, and Orfueil (1990) stated, activity analysis is a framework for analyzing travel as daily or multi-day patterns of behavior derived from differences in life styles and activity participation among the population. Among all the activity-based research, the measurement of activity space is an important topic when studying the spatial distribution of individual behavior. Activity space is defined as the local areas within which people travel during their daily activities (Mazey, 1981). There are several related concepts and terms, such as action space (e.g. Horton & Reynolds, 1971), awareness space (e.g. Brown & Moore, 1970), or space–time prisms (e.g. Hägerstrand, 1970). Previous research has focused on measuring the size, geometry, and inherent structure of human activity space (e.g., the randomness of activity patterns). For example, Schönfelder and Axhausen (2002) introduced the concept of intensity or density estimation to measure the probability of areas visited by a certain person. They used confidence ellipses to represent ellipse-shaped travel probability fields resulting from travel demands. In Song et al. (2010) and Gonzalez et al. (2008), activity space is calculated based on the rotation of user trajectories and the radius of gyration.

The approximation and measurement of activity space depicts its basic characteristics (e.g., size, shape, etc.). Moreover, there are other studies emphasizing on the reasons of how activity space forms. Generally, there are three determinants of activity space for a given individual (Golledge & Stimson, 1997, pp. 279):

- *Home location:* The position of the individual's home location.
- *Regular activities:* Regularly visited activity locations (Points of interest, POIs) such as the work location, grocery stores, gym, and cinemas.
- *Travel between and around the pegs:* Such as the duration of movements between the regularly visited places and the accessibility of public transport in the vicinity of home.

The combination of the above three internal determinants can be used to describe the development of activity space, as well as studying the cause and effect of human daily activities. As discussed in Section 1, in this research we use three indicators (radius, eccentricity, entropy) to measure the activity space of mobile phone users. Radius and eccentricity measure the descriptive characteristics of activity space, whereas entropy depicts the internal regularity of activity space. Moreover, in Section 4.2.4, we also extract the home and work locations of phone users. This provides us with further analysis on the internal determinants of activity space.

2.2. The role of ICTs in human mobility

The correlation between ICTs and human mobility has been a continuing theme in the field of transportation modeling. Janelle (1995) distinguished four types of communication modes based on different spatio-temporal constraints: Synchronous Presence (SP), Asynchronous Presence (AP), Synchronous Tele-presence (ST), and Asynchronous Tele-presence (AT). ICTs enhance the ability of tele-presence for human beings. As recognized by many researchers, the four communication modes interact with each other rather than being independent of each other. Therefore, the development of ICTs also impacts individual mobility in societies. Previous studies have focused on the interaction between ICT and human activity-travel behavior. Researchers recognized three main types of interaction: substitution, amplification, and synergy (Abler, 1975; Mokhtarian & Meenakshisundaram, 1999). In Pendyala, Goulias, and Kitamura (1991) and Saxena and Mokhtarian (1997), a strong substitution effect was found between telecommuting and physical activity in people's daily lives. However, other studies provided evidence for a significant amplification effect after reviewing a considerable number of cases in empirical studies. Mokhtarian and Meenakshisundaram (1999) argued that the usage of ICT actually generated, rather than reduced extra travel.

Undoubtedly, these efforts enhanced our understanding of the connection between ICT usage and travel behavior. However, due to the complex nature of the interaction between ICT use and human activity-travel behavior, many scholars have been skeptical of simple and universal conclusions of how ICTs affect daily activities (Kwan et al., 2007). Mokhtarian and Salomon (2002) recognized the differences between long-term studies and short-term studies. They stated that short-term empirical studies usually observed substitution effects of ICT. However, long-term studies often found significant amplification between ICT and activity behavior. Viswanathan and Goulias (2001) observed that there existed a positive correlation between phone call usage and physical mobility, whereas Internet use correlated negatively to time spent on travel. Couclelis (2004) described the spatio-temporal fragmentations and regrouping of daily activities in the age of instant access. In more recent studies, researchers also focused on the development of transportation problems in the age of instant access (Hjorthol, 2008; Mokhtarian, 2009; Salomon & Mokhtarian, 2008). Mokhtarian (2009) explained the paradox that transportation problems are becoming worse in spite of the substitution of telecommunication from five perspectives: (1) not all activities have an ICT counterpart; (2) ICT is not always a feasible alternative to physical travel (e.g., the limitation of cyber-infrastructures); (3) ICT is not always a desirable substitute (e.g., hanging out with friends in a bar); (4) Travel carries some positive utility; (5) Not all ICT activities can replace travel. The authors also discussed seven aspects, where ICT can be a substitute for physical travel. This research helps to understand the complex relationship between ICT usage and travel behavior. Moreover, as stated by Nobis, Lenz, and Vance (2005), substantial differences in the use of ICTs exist, resulting from many social factors, including age, gender, culture, and socioeconomic distribution. Therefore, it is also important to take the influence of population heterogeneity into account when investigating the relationship between ICTs and people's daily activities and mobility patterns. They consider communication and travel behavior as two facets of human interaction and examined the correlation under a larger context of individual behavior change and social conditions. The connection between ICT usage and mobility was observed through age groups. Young and educated people were found to be highly mobile and communicative. The effects of age, gender, and income level proved to be significant in determining both ICT usage and human mobility.

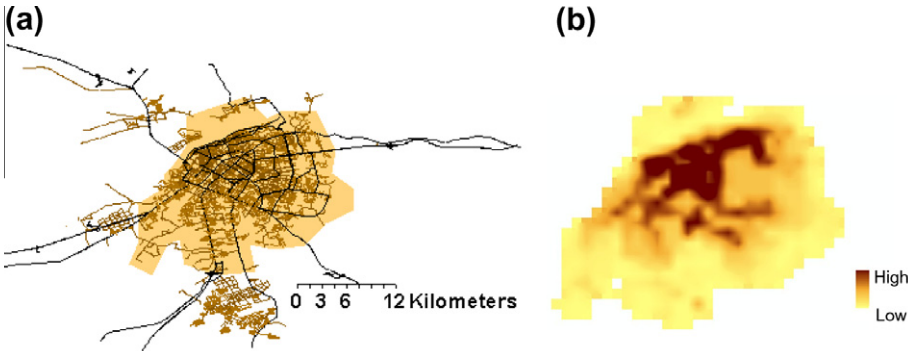

Fig. 1. (a) The basic road map of Harbin city and (b) population density data from the LandScan dataset (resampled raster).

Hence, as stated in Section 1, a general conclusion is insufficient to represent the complicated nature of this research question; therefore, the existing theories need to be extended. Moreover, various aspects of activity space have not been fully addressed when discussing the correlation between mobile phone usage and activity behavior. In this research, we will examine the correlation between three indicators of activity behavior and mobile phone usage, as well as investigate the impact of explanatory factors. This will provide new insights for answering the questions raised by the complex nature of the relationship between ICT and individual travel behavior.

3. Research design

3.1. Dataset

Harbin city is a major commercial, industrial, and transportation center situated in northeast China. It was ranked as one of the top ten populated cities in China. Fig. 1 shows a basic road map of Harbin city, as well as the population density distribution across the city based on the LandScan dataset.[1]

The dataset used for this research has been acquired from China Mobile Company for research purposes. The dataset covers 0.87 million mobile phone users from Harbin city and includes mobile phone call records for a time span of 9 days (4 weekend days and 5 weekdays). The data include the starting time, duration, and approximate location of mobile phone calls, as well as the age and gender attributes of the users. The ratio of male versus female users is approximately 3:5. Table 1 provides sample records of our dataset. The real phone number, longitudes and latitudes of the phone user are not shown in the table due to privacy issues.

For each user, the location of the nearest mobile phone tower is recorded both when the user makes and receives a phone call, resulting in a positional data accuracy of about 300–500 m. Fig. 2 visualizes the distribution of mobile phone towers in Harbin city. As can be seen, the center of the city has better coverage and therefore higher data accuracy.

Note that the location records in the dataset cannot represent the accurate moving trajectory of each user, since the locations are recorded only when a phone call connection has been established. However, as shown in Gonzalez et al. (2008), human mobility indicates a high level of regularity based on a time span of ten days, since there is a high probability that an individual returns to the position, where he/she was first observed within the following 240 h. Therefore, based on a summary of 9 days' records, the data in this research are still useful for depicting the general characteristics of individual mobility.

[1] http://www.ornl.gov/sci/landscan/.

Table 1
Sample records from the data set of Harbin city.

Phone number	Longitude	Latitude	Time	Duration
1350*******	126.*****	45.*****	14:36:24	12 min
1350*******	126.*****	45.*****	15:21:43	44 min
1350*******	126.*****	45.*****	15:47:04	22 min
Phone number		Gender		Age
1350*******		Male		45

3.2. Methodology

The described dataset offers a typical mobile phone data source, which provides us with three types of directly recorded information for each mobile phone user: (1) Mobile phone usage; (2) Individual attributes (age, gender); and (3) Spatio-temporal points for a given time interval. All the data mining and knowledge discovery tasks are based on the combination and interaction of the above three information categories. In this research, we use individual phone call frequencies (including both initiated and received) to represent mobile phone usage. Additionally, in order to depict individual travel behavior, three indicators are calculated as follows:

(1) **Movement radius**: For each individual, we approximate the physical movement area based on the rotation of user trajectories (Gonzalez et al. 2008; Fig. 3). The eigenvectors of trajectories determine the principal axes $\hat{e}_1$ and $\hat{e}_2$, and then the trajectories are approximated as ellipses, where $\hat{e}_1$ and $\hat{e}_2$ are the major and minor axes. For a given ellipse, the average value of the length of the semi-major and semi-minor axes is considered as a measurement of moving radius: $R = \frac{|\hat{e}_1|+|\hat{e}_2|}{4}$

(2) **Movement eccentricity**: Since user trajectories are approximated as ellipses, the movement eccentricity ($e = \sqrt{1-\left(\frac{|\hat{e}_2|}{|\hat{e}_1|}\right)^2}$, $e \in [0, 1]$) represents how much a particular trajectory deviates from being circular ($e = 0$).

For instance, if $e \approx 1$, it is highly possible that the particular person always moves between work and home, therefore, the trajectory is close to a straight line.

(3) **Movement entropy**: Based on Song et al. (2010), movement entropy is calculated as $E = -\sum_{i=1}^{N} p_i \log_2 p_i$, where p_i is the probability that location i is visited by the user. N stands for the total number of distinct locations visited in a given trajectory. Entropy characterizes the heterogeneity of visitation patterns. For example, if a given person A has only

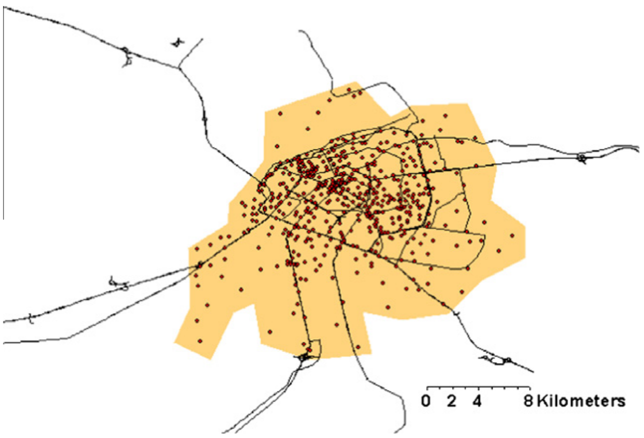

Fig. 2. The distribution of mobile phone towers in Harbin city.

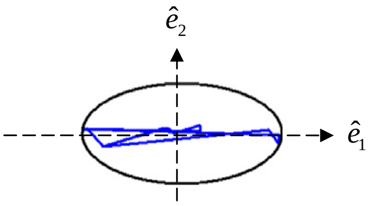

Fig. 3. Transformation of trajectories (Gonzalez et al., 2008).

visited the two locations 1 and 2 in the dataset, and each location has been visited five times, then the entropy value is calculated as

$$E_A = -(0.5 * \log_2 0.5 + 0.5 * \log_2 0.5) = 1.0$$

whereas if location 1 is visited nine times and location 2 is visited only once (a more homogeneous pattern), the entropy would be lower:

$$E_A = -(0.9 * \log_2 0.9 + 0.1 * \log_2 0.1) \approx 0.47$$

Furthermore, as stated by Beckman (2000) and Nobis et al. (2005), activity patterns are restricted by individual level factors (e.g., age, gender) and "supra-individual regime" (temporal order, social conditions, etc.). They also distinguished two types of temporal orders: *natural temporal order* (e.g., morning, afternoon, and evening) and *social temporal order* (e.g., weekends and weekdays). Due to the multitude of explanatory factors, it is not feasible to discuss all of them in this research. Therefore, we consider two individual level factors (age and gender) and two "supra-individual regimes" (social temporal order and the road network density in the built environment) when discussing the relationship between mobile phone usage and individual travel behavior. Our goal is to answer the following questions:

(1) What is the correlation between travel behavior (represented by the three indicators radius, eccentricity, and entropy) and mobile phone usage?
(2) How do individual-level attributes (age and gender) affect this correlation?
(3) How do supra-individual-level attributes (social temporal orders [weekdays, weekends] and built environment) affect this correlation?

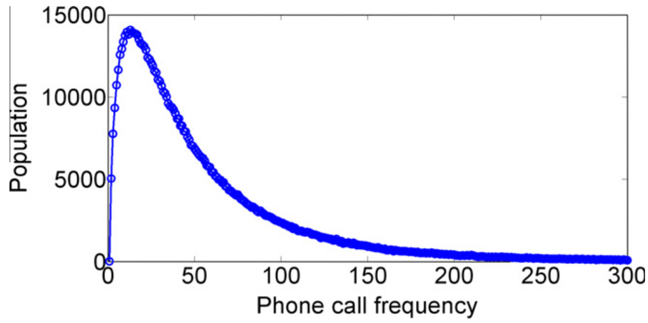

Fig. 4. The sample distribution based on phone call frequencies.

4. Data analysis

4.1. Generic analysis

This section discusses the generic questions within the context of the whole dataset. Here the whole sample is considered homogeneous. Therefore, the effects of individual social attributes and supra-individual-level attributes are not included in the generic analysis. We rather look at the overall picture of this research question. First, the whole dataset is divided into groups according to the frequency of mobile phone usage, i.e., people within the same group have the same phone call frequency. Fig. 4 shows the distribution of phone call frequencies. Note that the sample of 0.87 million people only includes mobile phone subscribers making calls during the 9-day study period and who provided their certified personal information (e.g., copy of citizen ID) when subscribing. This represents 17.6% of the total population of Harbin main city of 4.75 million (of whom 86.2% are over 18 years old).[2] Although there is no official statistics for the total number of mobile phone subscribers in Harbin city, we estimate the number to be around 2 million based on the fact that 40% of the total population in China were mobile phone users by the end of 2007.[3] Therefore, the sample data size (0.87 million) represents approximately 43.5% of the total and is therefore large enough to generalize the results to the whole population. Since there are only a few people who fall into groups, where phone call times >300, we remove these people from the data analysis in order to eliminate outliers. Moreover, note that only those who have $\geqslant$10 phone call connections are considered in the analysis, otherwise the user trajectories would be inaccurate due to insufficient spatial information. After the data cleaning process (eliminating people whose phone call times >300 or <10), the number of users covered by the analysis is 786,621. From these we calculate the average radius, average eccentricity, and average entropy for each group.

The correlation between mobile phone usage and each of the three indicators is investigated using linear regression. As can be seen in Figs. 5a, 6a and 7a the correlation appears to be logistic rather than linear; therefore, we calculate the logarithm values (base 10) for both the x and y axes and use the new logarithm values for linear regression. The results of the linear regression are displayed in Figs. 5b, 6b and 7b.

The results of the generic data analysis show a positive correlation in Figs. 5a and b and 7a and b, and a negative correlation in Fig. 6a and b, indicating the following statistical results for the whole sample in the dataset:

[2] Data Source: Chinese Academy of Social Sciences (http://www.cass.net.cn/file/20090311220659.html).

[3] Data Source: Xinhua Net (http://news.xinhuanet.com/english/2008-02/08/content_7582783.htm).

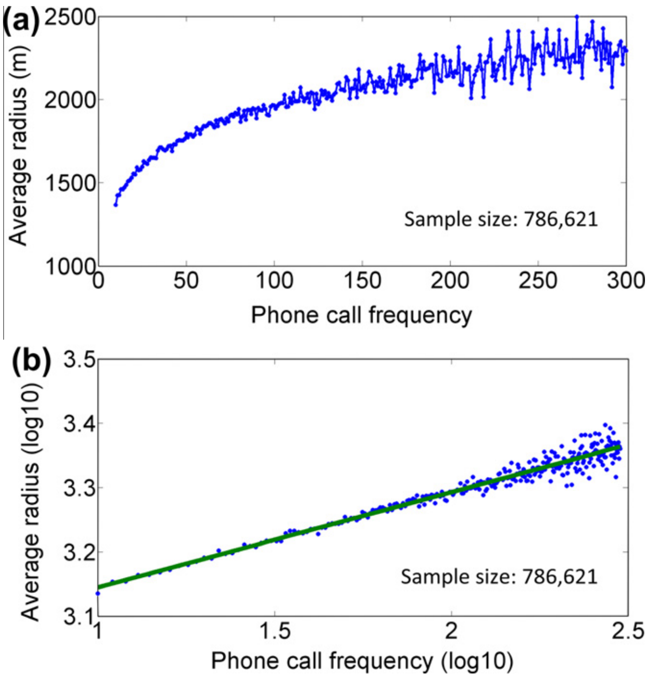

Fig. 5. The correlation between mobile phone usage and average movement radius, (a) original values and (b) logarithm values (base 10).

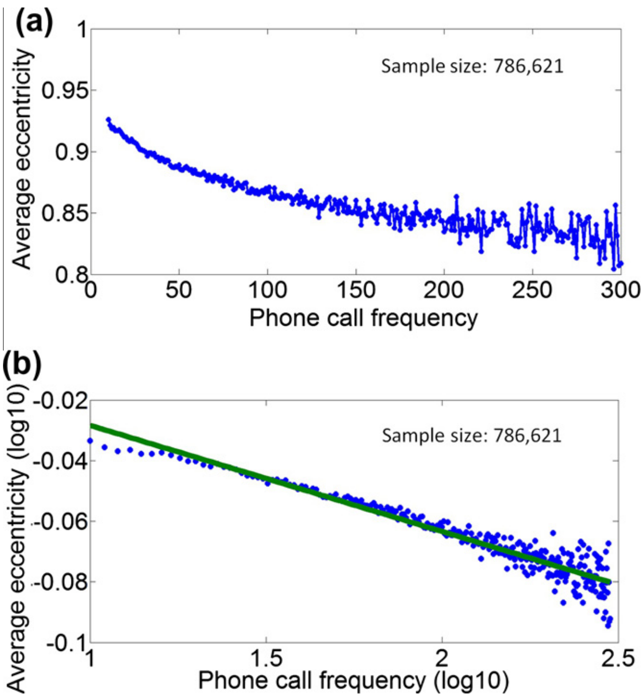

Fig. 6. The correlation between mobile phone usage and average eccentricity, (a) original values and (b) logarithm values (base 10).

(1) The average movement radius increases as mobile phone usage increases.
(2) The average movement eccentricity decreases as mobile phone usage increases.
(3) The average movement entropy increases as mobile phone usage increases.

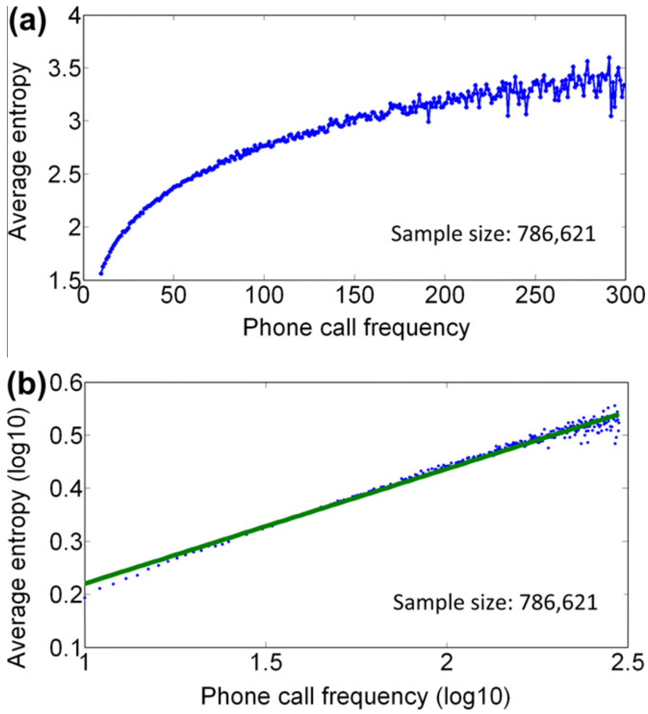

Fig. 7. The correlation between mobile phone usage and average entropy, (a) original values and (b) logarithm values (base 10).

Table 2
Correlation coefficients from the generic analysis.

	Fig. 5a	Fig. 6a	Fig. 7a
Spearman correlation coefficients	0.947	−0.958	0.973
	Fig. 5b	Fig. 6b	Fig. 7b
Pearson correlation coefficients (*R* value)	0.975	−0.972	0.992

Furthermore, the correlation significance tests are conducted for both the original data and the calculated logarithm data. Spearman coefficients are used for the original data, whereas Pearson coefficients are used for the logarithm data since Pearson coefficients can only be used for linearly correlated data. As indicated in Table 2, the correlations shown in Figs. 5–7 are all quite high (>0.94 in magnitude), and are significant at the 0.01 level (2-tailed tests). Therefore, we generally conclude that, for people with higher mobile phone usage: (a) their movement radii are larger; (b) their movement trajectories are closer to a circle; (c) their movement entropy is higher, therefore the movement is more random.

4.2. The role of explanatory factors

As discussed in Section 3.2, human activities are restricted by a multitude of factors, including individual level factors and "supra-individual regime". In this section we focus on analyzing the role of explanatory factors when studying the correlation between mobile phone usage and travel behavior. Since it is not feasible to discuss the impact of all explanatory factors, we consider two individual level factors (age and gender) and two "supra-individual regimes" (social temporal order and the road network density in the built environment) in this research.

4.2.1. Why explanatory factors must be considered

The statistical results in Section 4.1 reflect a significant correlation between mobile phone usage and all three movement

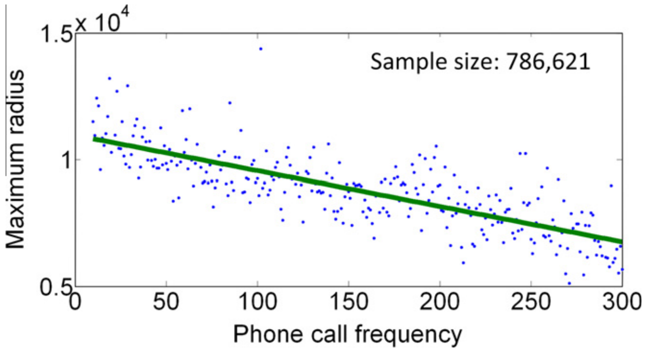

Fig. 8. The correlation between mobile phone usage and maximum radius (m).

indicators. Nevertheless, these conclusions are not comprehensive for explaining the entire problem. As an example, Fig. 8 depicts the correlation between maximum movement radius and mobile phone usage. Note that we do not use logarithm values in this figure, since the original data indicate a strong linear correlation. The data analysis in Section 4.1 shows a significant positive correlation between mobile phone usage and average movement radius (Fig. 5), whereas a negative correlation between mobile phone usage and maximum movement radius is demonstrated in Fig. 8.

To further explain the nature of the correlation, we extract four groups with different frequencies of mobile phone usage and analyze the distribution of individual movement radius in each group. Fig. 9a–d represent the histograms of movement radius for the four groups (the frequency of mobile phone calls = 10, 50, 100, 200 respectively). As shown in Fig. 9, when the frequency of mobile phone calls increases, we have the following statistical results:

(1) The average movement radius increases.
(2) The standard deviation of movement radius decreases.
(3) The maximum movement radius decreases.

Moreover, further analysis indicates that the distribution of age represents a similar pattern as the histograms of movement radius in Fig. 9. Therefore, we can also conclude that the standard deviation of age decreases when mobile phone usage increases.

The results above can be interpreted from two perspectives. First, since the average movement radius increases as mobile phone usage increases in both Figs. 5 and 9, we can generally conclude that mobile phone usage is positively correlated to movement radius. Second, for groups with fewer phone calls, the population heterogeneity is more significant due to the higher standard deviations of age and movement radius. Therefore, the groups with fewer phone calls correspond to a larger variety of phone users. People with more phone calls also have a more homogeneous social background. This explains the negative correlation between maximum movement radius and mobile phone usage. The groups with a larger population heterogeneity are more likely to have a large range of movement radius and more extreme values of movement radius, so the maximum values decline as the population heterogeneity declines. Therefore, the correlation between mobile phone usage and travel behavior will be different due to various individual backgrounds and life styles. For instance, the movement radius of a mailman is naturally large with little correlation to his mobile phone usage. Another example would be the distinctive activity patterns in different age groups. Fig. 10 shows the mobility hotspots for two age groups during 2–3 pm on a weekday: teenagers (age 12–17) and seniors (age > 60). As indicated in Fig. 10, the clustering of teenagers appears both in the center and in the Northwest of the city, whereas the density pattern of seniors is more widely distributed. Due to the different demands for living resources in various population groups, such analysis can provide helpful references for updating urban infrastructures for different population groups (e.g., high schools, hospitals, etc.). For example, there appears to be a cluster of seniors in the northern part of the city, the center of which is very close to a large park in Harbin city.

Hence, it is necessary to control for explanatory variables when discussing the correlation between mobile phone usage and activity behavior. In the remainder of Section 4 we will discuss two aspects of attributes: individual (age and gender) and supra-individual (temporal and spatial attributes).

4.2.2. The impact of individual level factors

4.2.2.1. The impact of age. Age is considered as an important control variable when exploring the effects of social attributes. To further investigate the correlation between mobile phone usage and travel behavior in different age groups, we divide the whole sample into the following five groups:

(1) Age < 18
(2) 18 ⩽ Age ⩽ 23
(3) 23 < Age ⩽ 39
(4) 39 < Age ⩽ 59
(5) Age > 59

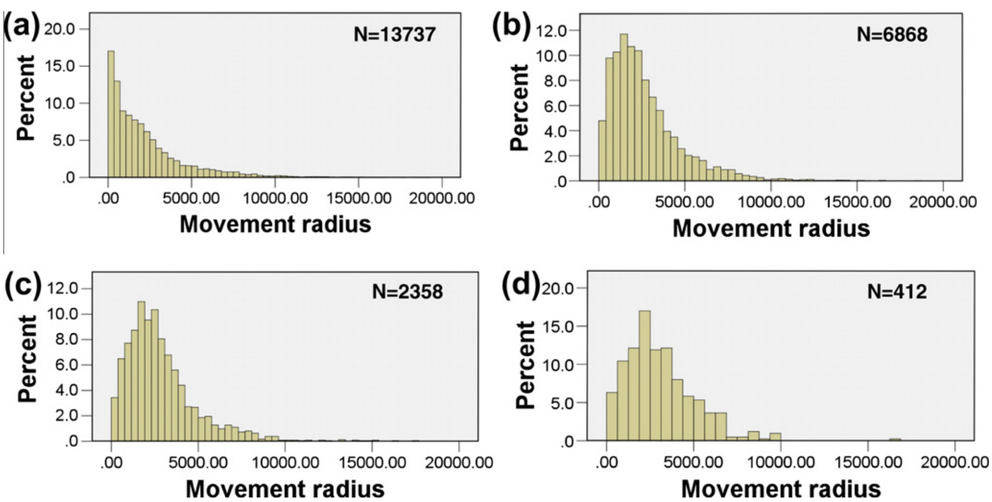

Fig. 9. Histograms of movement radius (m) for groups with different phone usage. (a) phone calls = 10; (b) phone calls = 50; (c) phone calls = 100 and (d) phone calls = 200.

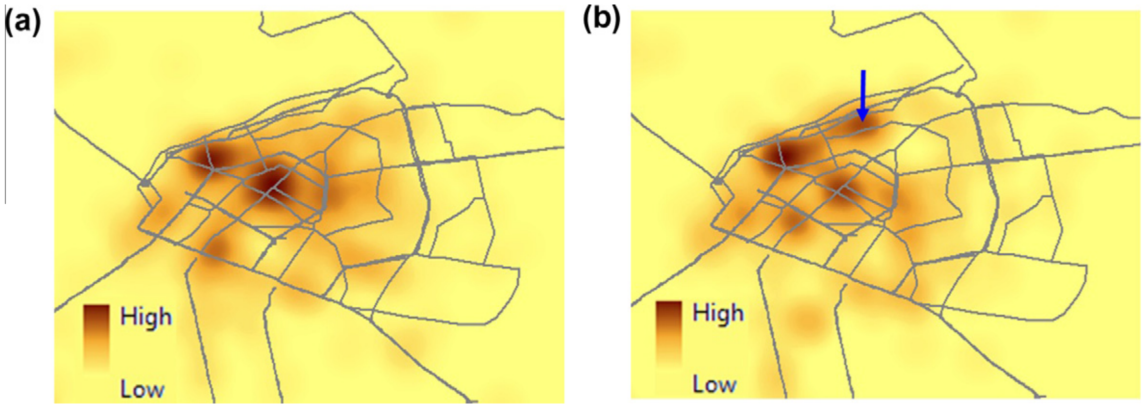

Fig. 10. The mobility clustering of (a) teenagers and (b) seniors based on kernel density estimation.

Table 3
The correlation values between movement indicators and mobile phone usage for five different age groups.

		<18	18–23	23 + –39	39 + –59	59+
	Sample size	12,286	240,145	379,013	147,976	7131
Radius	Spearman correlation (*R* value)	0.167	0.754	0.896	0.758	0.279
	Mean (*m*)	1502.502	1596.103	1851.839	1796.169	1641.112
Eccentricity	Spearman correlation (*R* value)	−0.240	−0.858	−0.942	−0.871	−0.252
	Mean	0.899	0.895	0.885	0.888	0.900
Entropy	Spearman correlation (*R* value)	0.857	0.988	0.997	0.994	0.822
	Mean	1.921	2.076	2.325	2.239	1.864

Group 1 stands for juveniles, while approximately 26.8% of the people in Group 2 are college students.[4] In Groups 3 and 4, the majority are employed people. Note that there is no information about employment status for each user in our dataset. However, the percentage of employed people in Groups 3 and 4 is higher than 90% based on the population census in Harbin city[5]. Group 5 stands for retired people, since the retirement age in China is 60.

Table 3 depicts the correlation values between mobile phone usage and the three movement indicators for the five different age groups. All the correlations shown in Table 3 are significant at the 0.01 level (2-tailed tests).

As demonstrated in Table 3, even though the correlation level in general differs, the significant correlation between mobility and use of mobile phones exists for all age groups. The correlation values in the five age groups appear in the following order:

Radius: Group 3 > Group 4 > Group 2 > Group 5 > Group 1
Eccentricity: Group 3 > Group 4 > Group 2 > Group 5 > Group 1
Entropy: Group 3 > Group 4 > Group 2 > Group 1 > Group 5

For all three movement indicators, Group 1 (“younger than 18”) and Group 5 (“older than 59”) appear to have the lowest correlation values. Hence, the travel behavior of children, teenagers, and seniors is least correlated with mobile phone usage. Moreover, Group 3 (young employed people) indicates the strongest correlation between mobility and mobile phone usage. The correlation values in Group 2 (college students) and Group 4 (middle-aged employed people) are similar and slightly lower than that of Group 3, demonstrating a high correlation in both groups.

Moreover, as shown in Fig. 11, Group 1 and Group 5 have the lowest average movement radii and average entropy, as well as

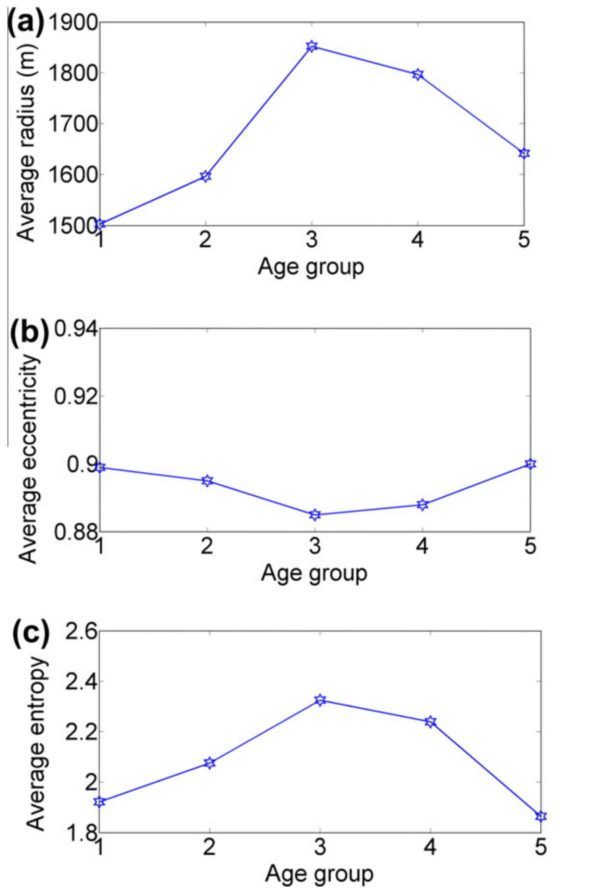

Fig. 11. The mean values of three movement indicators in different age groups, (a) average movement radii, (b) average eccentricity, and (c) average entropy.

[4] Data source: The Education Department of Heilongjiang Province: (http://www.hlje.net).
[5] Data source: “Harbin Statistics yearbook 2007” published by China Statistics Press.

Table 4
The correlation values for male and female.

		Male	Female
	Sample size	472,390	314,231
Radius	Spearman correlation (*R* value)	0.986	0.992
	Mean (*m*)	2519.284	2658.143
Eccentricity	Spearman correlation (*R* value)	−0.945	−0.942
	Mean	0.889	0.857
Entropy	Spearman correlation (*R* value)	0.963	0.954
	Mean	2.198	2.257

Table 5
Correlation values for weekdays and weekends.

		Weekdays	Weekends
	Sample size[a]	771,593	753,312
Radius	Spearman correlation (*R* value)	0.973	0.968
Eccentricity	Spearman correlation (*R* value)	−0.970	−0.961
Entropy	Spearman correlation (*R* value)	0.994	0.992

[a] Note that we further eliminated those who did not make phone calls on weekends (or weekdays). We also eliminated outliers whose daily phone call frequency > 30 on weekends or > 25 on weekdays.

the highest average eccentricity values. Group 3 is the counterpart with the largest average radius and the most unpredictable movement patterns. Therefore, Group 3 (young employed people) is considered the most mobile group of people.

4.2.2.2. The impact of gender. Similarly, we conduct the analysis for the gender attribute of phone users. Table 4 describes the correlation values for male and female groups. Both groups indicate a significant correlation at the 0.01 level (2-tailed tests) between phone call frequency and the three indicators of activity space. The average movement radius of females is slightly larger than that of males. However, there are no substantial differences between males and females regarding the mean values of the three indicators and the correlation coefficients values.

4.2.3. The impact of supra-individual-level factors

4.2.3.1. The impact of social temporal orders. As stated by Beckman (2000), social temporal orders (weekdays, business hours, etc.) are considered to have an important impact on individual travel behavior. Due to the inherent inconsistency of behavior patterns on weekdays and weekends, it is necessary to explore the correlation between mobility and mobile phone usage for different temporal orders. Table 5 partitions the temporal orders into weekdays and weekends. As can be seen, a significant correlation exists for both weekdays and weekends, but the connection between phone usage and mobility for weekends is slightly weaker than (almost the same as) that for weekdays. Hence, there are no substantial differences between weekdays and weekends regarding the correlation between mobile phone usage and human mobility.

As shown in Fig. 12a, the movement radii for weekends are larger than those for weekdays, reflecting the fact that people's movements on weekends cover a slightly larger spatial extent. Fig. 12b visualizes the correlation between average eccentricity and phone call frequency for both weekdays and weekends. As can be seen, eccentricity values for weekends are higher than those for weekdays, indicating that movements occurring on weekends are closer to straight lines. Similarly, as represented in Fig. 12c, entropy values for weekends are lower than those for weekdays, indicating that people's movements on weekends are more regular, which

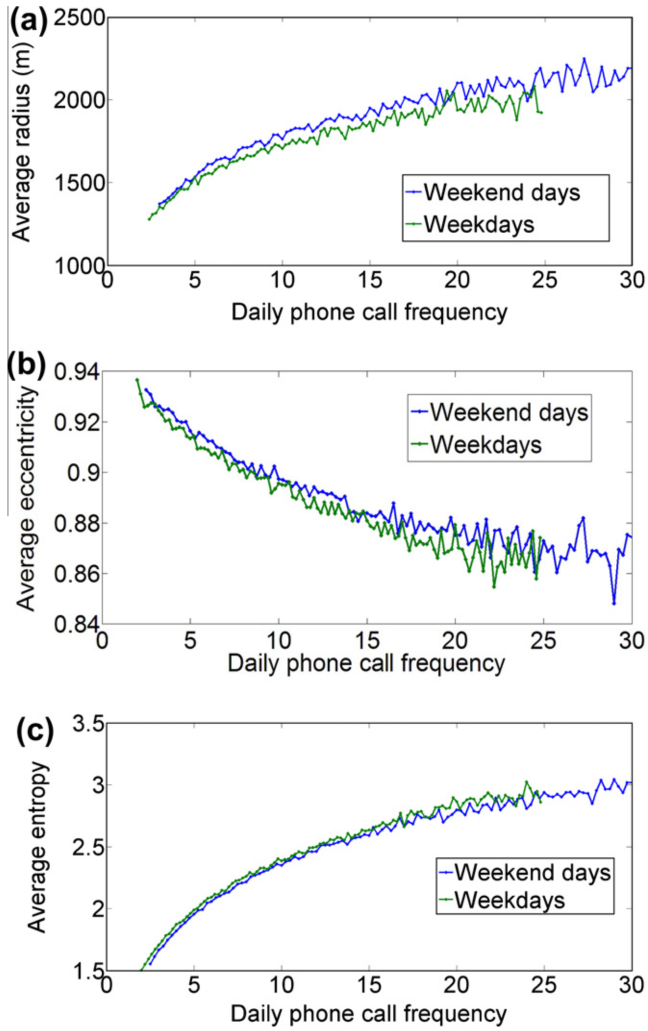

Fig. 12. Correlation between the three movement indicators and mobile phone usage for weekdays and weekends; (a) average movement radius, (b) average eccentricity, and (c) average entropy.

is consistent with the results from Fig. 12b. We will further discuss these results in Section 5.

The statistical results above are based on the aggregation of the whole data sample, but for a particular person, the movement patterns on weekdays and weekends can be completely different or unrelated. Therefore, it is difficult to determine the correlation between movements on weekends and weekdays for a particular person. For example, in Fig. 13, movement radii for weekends and weekdays are represented by parallel coordinates, which is a common way for exploring and visualizing multivariable data (Inselberg, 2009). Axis X_1 shows movement radii for weekends, whereas X_2 stands for movement radii for weekdays. As can be seen in Fig. 13, we select the first 100 people with the largest movement radii on weekends; however, there appears to be no regular pattern for their movement radii on weekdays (Fig. 13a). This is also the case when selecting the first 100 people for weekdays (Fig. 13b).

4.2.3.2. The impact of the built environment. Besides individual attributes and temporal orders discussed in the earlier part of this section, the characteristics of the built environment also have an inevitable impact on human mobility behavior. There are several control variables in the spatial environment that can be investigated when studying the correlation between mobile phone usage

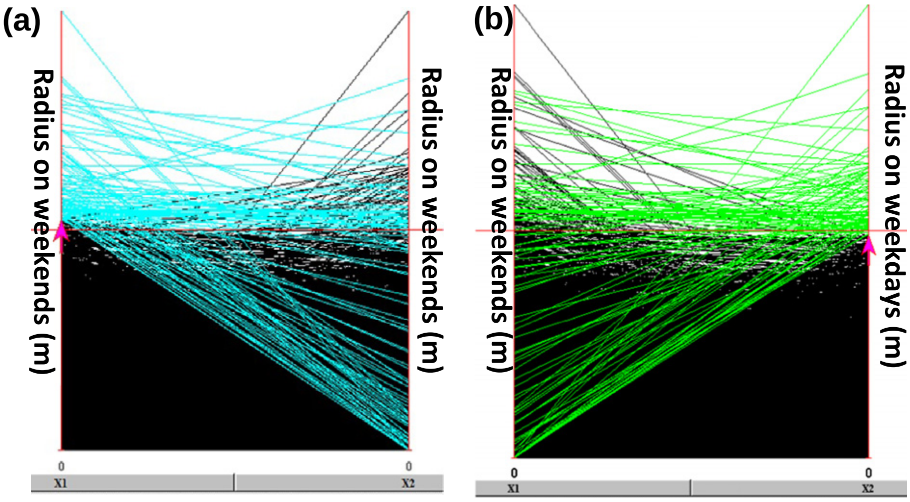

Fig. 13. Selected movement radii for weekends and weekdays; (a) the largest 100 movement radii on weekends are selected and (b) the largest 100 movement radii on weekdays are selected.

and travel behavior, such as the distributions of urban infrastructures and social-functionality divisions (e.g., Central Business District, CBD). In this section we focus on the influence of transportation network density and individual home/work locations when discussing the correlation between mobile phone usage and the travel behavior of phone users.

In a first step, the urban area of Harbin city is divided into two sub-areas based on the weighted road density. Table 6 describes the weights for different types of roads around Harbin city.

Fig. 14 visualizes the results of the road density analysis. As can be seen, the road network is densely distributed in the center of the city. There are several methods to identify the central area of a city. In this research we use road density as a measurement. Based on the results in Fig. 14a, we divide the urban area into two sub-areas (Fig. 14b): dark gray color stands for the central district with the top 5% road density distribution, whereas light gray stands for the other non-central area.

Table 6
Categories of roads.

Type	National road	Highway	Provincial road	Urban highway	Cantonal and streets
Weight	5	5	3	3	1

Due to privacy issues, it is difficult to acquire information such as the home/work locations of phone users. Therefore, as a next step, we extract the work and home locations of mobile phone

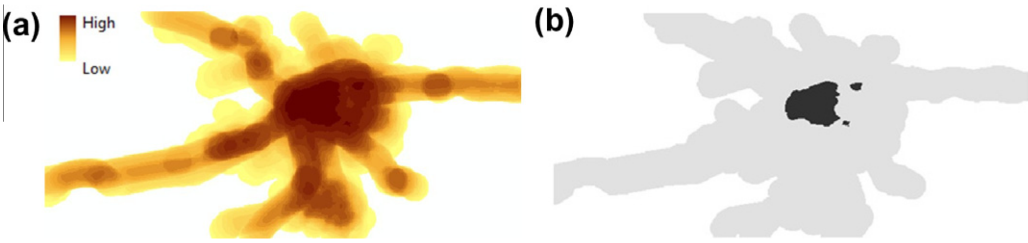

Fig. 14. (a) Road density map in Harbin city and (b) central and non-central urban area.

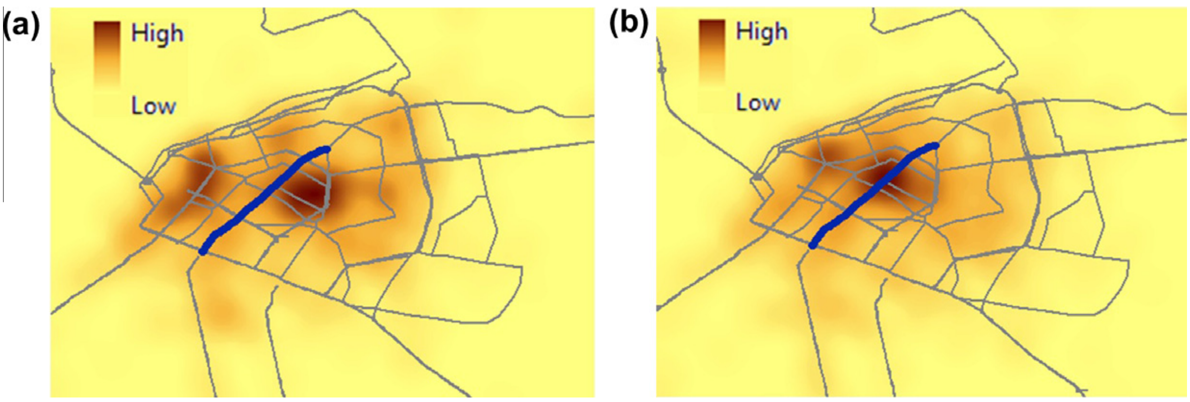

Fig. 15. Clustering of (a) home locations and (b) work locations based on kernel density estimation.

Table 7
Correlations with phone call frequencies for sample groups based on home/work locations.

	Sample Size	Both outside 56,980	One inside, one outside 15,750	Both inside 24,529
Radius	Spearman correlation (R value)	0.648	0.237	0.578
	Mean (m)	3041.847	3134.433	1916.662
Eccentricity	Spearman correlation (R value)	−0.605	−0.529	−0.543
	Mean	0.563	0.564	0.504
Entropy	Spearman correlation (R value)	0.870	0.761	0.824
	Mean	2.816	3.441	2.948

users based on individual trajectories. Such analysis offers valuable input for enriching the personal profiles of users and studying urban areas according to their functions. In order to identify the *stops* in trajectories the methodologies described in Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki, and Ratti (2010) are applied: The trajectory of a certain individual is identified as a sequence of chronological locations:

$$R = \{(p_1, t_1) \rightarrow (p_2, t_2) \rightarrow \ldots \rightarrow (p_n, t_n)\}$$

where the p_i refer to spatial locations and the t_i refer to time points. Then the trajectories are regrouped into sub-trajectories based on the restriction that any two consecutive points within a sub-trajectory are located within the cell of the same mobile phone tower. If the time duration of a sub-trajectory is longer than the temporal threshold ΔT, the sub-trajectory is identified as a *stop* for the particular user. Once the *stops* have been extracted, the home location of each user is estimated as the most frequent *stop* during the night hours (7 pm–7am) and the work location is the most frequent *stop* during day hours on weekdays (Monday–Friday 8am–5 pm). Note that we cannot extract home and work locations from all the users in the data sample, since the extraction of stops are based on the occurrence of phone calls. As a result, we successfully extracted work locations for 446,692 users, home locations for 116,083 users, as well as both work and home locations for 97,259 users. Fig. 15 demonstrates the distribution of home and work locations of users in Harbin city.

Both home and work locations are clustered in the city center; however, there are slight differences between the locations of hotspots in Fig. 15a and b. The highlighted street in Fig. 15 is one of the main streets and it runs across the whole city. The home locations are mostly concentrated on the southeast side of the main street, whereas the work locations are evenly distributed on both sides of the street. Additionally, the home locations show two clustering centers in the study area (one in the middle, the other on the western side), indicating that Harbin city has multiple active sub-areas that function as residential districts.

After the home/work locations are extracted for each phone user, we examine whether these locations are in the central area of Harbin city (marked as dark gray in Fig. 14b). The whole population is divided into three groups:

Group 1: Both home and work locations are located outside the central area.
Group 2: Either home or work location is in the central area, whereas the other one is located outside the central area.
Group 3: Both home and work locations are in the central area.

Table 7 shows the correlation between phone call frequency and the three indicators (radius, eccentricity, entropy) in the three groups. All the correlations shown in Table 7 are significant at the 0.01 level (2-tailed tests).

As indicated in Table 7, Group 3 (those people who live and work in the city center) has the smallest average movement radius. Group 1 and 2 have similar average radii and eccentricity values. This demonstrates that people who have at least one of their home/work locations outside the urban central area have a larger activity space and longer commuting distance. However, Group 2 has a larger entropy value than the other two groups, indicating that those who live in the urban central area and work in the non-central area (or vice versa) have a more random mobility pattern.

Based on Spearman correlation, all three groups show a significant correlation between phone call frequency and the three indicators (radius, eccentricity, entropy). In Group 2 the correlation coefficients are smaller than in Group 1 and 3. This indicates a weaker correlation between mobile phone usage and daily activity behavior in Group 2.

5. Discussion

First, the generic analysis in Section 4.1 indicates a significant correlation between phone call frequency and all three indicators of travel behavior. Movement radius and entropy are both positively correlated with the usage of mobile phones, whereas a negative correlation is observed between eccentricity and phone call frequency. Therefore, we can generally conclude that, for people with higher mobile phone usage:

(a) Their movement radii are larger, therefore they cover a larger activity space in their daily lives.
(b) Their movement trajectories are closer to a circle. Hence, it is feasible to infer that their regularly visited locations are widely distributed rather than linearly ordered.
(c) Their movement is more random, since the entropy of their movement patterns is higher. Therefore, it is much easier to predict the movement patterns of people with low mobile phone usage.

Since Figs. 5b, 6b, and 7b all represent strong linear correlations, we can conclude that the distributions in Figs. 5a, 6a, and 7a can be approximated as logarithmic functions. In this research, we have not looked into the specific parameters of the logarithmic functions for each figure. However, the approximated limits of the functions can be seen directly in Figs. 5a, 6a, and 7a. For instance, the limit of average eccentricity gets close to 0.8 as phone usage increases, indicating that as phone call frequency increases, the shape of the phone users' activity space still stays close to a straight line instead of a circle.

Although it is difficult to infer the causality between mobile phone usage frequency and movement radii, it is highly possible that the high usage of mobile phones expanded the spatial scale of people's daily movements, as well as making their movements widely distributed and hard to predict. On the other hand, it is also possible that extensive movement generated lots of phone calls (e.g., the coordination of travel). However, since the correlation in this research reveals little about causality, further work on the causal relations between mobile phone usage and travel behavior is needed.

Second, we have also considered the impact of individual factors regarding the correlation between mobile phone usage and travel patterns. As indicated in Section 4.2.1, there exists a significant negative correlation between the maximum movement radii and mobile phone usage, whereas the average movement radii positively correlate with mobile phone usage. A possible explanation for this contradiction is that for groups with fewer phone calls, the population heterogeneity is more significant due to the higher standard deviations of age and movement radii. Therefore, the range of movement radii is larger for groups with lower mobile phone usage. This contradiction (Fig. 5a versus Fig. 8) also leads to the question of whether the correlation between mobile phone usage and travel behavior will be different due to various individual social backgrounds and life styles. In order to address this problem, we differentiated the correlation values among age and gender groups. Based on the statistical analysis, children, teenagers, and seniors have the lowest correlation values, whereas the young employed people show the highest correlation values. Therefore, people between 24 and 39 years of age have the highest connection between mobile phone usage and their daily traveling. It is interesting to note that, although slight differences exist, there is no obvious distinction among all five age groups when exploring the correlation between entropy and mobile phone usage. Therefore, the strong connection between mobile phone usage and the randomness (or predictability) of human mobility exists for all age groups. The results also indicate that people between 24 and 39 years of age are the most mobile people. This is consistent with the empirical results in Nobis et al. (2005), in which young and educated people are found to be the most mobile and communicative group. In addition, the gender analysis does not show any significant differences for males and females in terms of the average values of radius, eccentricity, entropy, as well as the correlation coefficients.

Third, besides individual level factors, we have also explored the role of social temporal orders in determining correlation values. As can be seen from the results, the correlation values for weekdays are almost the same as those for weekend days, indicating that, compared to the activities (mostly professional activities) on weekdays, the activities usually conducted on weekends (e.g., going to the movies, grocery shopping, etc.) have a similar connection strength to the use of mobile phones. Moreover, we also analyzed the effect of the temporal orders on mobility itself. The results appear to be interesting and contradictory to what one would expect. The movements on weekends tend to be slightly more regular and predictable, showing that most people live a more regular life style on weekends. Hence, it is reasonable to conclude that most people still tend to visit regular locations (e.g., preferred grocery stores) in their leisure time, although movements on weekends have less spatial constraints (e.g., work locations). This is different from what we common-sensically expect, i.e., that mobility on weekends is more random. In addition, the results indicate that it is difficult to predict the characteristics of individual mobility on weekends based on people's traveling on weekdays, and vice versa.

Fourth, we explored the impact of the built environment by estimating weighted road density and the home/work locations of phone users. Note that there are several other methods to identify the central area of the city. For example, we can also use the Landscan population density data (Fig. 1b) instead of the road network density data. The results demonstrate that those who live and work in the city center with high road density have a smaller activity space. On the other hand, those who live and work in non-central areas have a larger movement radius. One possible reason is the uneven distribution of urban infrastructures (e.g., grocery stores, parks, hospitals, entertainment sites, etc.) between the central and suburban area, so people who live and work outside the city center need to travel more for their daily living resources. Another interesting point is that, as shown in the correlation coefficients, people who live in the central area and work outside the city center (or vice versa) show the weakest connection between their mobile phone usage and physical travel. This implies that the life style (e.g., occupation type) of phone users may have an impact on the correlation; however, more control variables and more comprehensive data are needed to statistically test this hypothesis.

Finally, it is important to highlight that there are different aspects of uncertainty issues related to this study. These issues arise in our data mining process in different ways. As Xia (2005) argued, uncertainty exists not only in datasets to be mined but also in knowledge that is mined, as well as in the process of applying uncertain knowledge to new datasets. In this research we have the following three types of uncertainty sources:

(1) Natural variability of human mobility: Although human mobility seems to be highly predictable (Gonzalez et al., 2008; Song et al., 2010), randomness is an inevitable part of human motion.
(2) Inaccuracy and imprecision due to insufficient knowledge: the trajectories recorded in mobile phone datasets are neither accurate nor precise. First, the accuracy of positioning data often depends on the density of mobile phone towers in the study area. In this case, the estimation accuracy of people who move in the central area is higher than of those moving in the suburban area. Second, the location records in the dataset cannot represent the accurate moving trajectory of each user, since the locations are recorded only when a phone-call connection has been established. Third, the precision of spatial information varies for different datasets, e.g., a record such as "126.51551E, 45.15153 N" is more precise than "126.52E, 45.15 N".
(3) Imperfection of regression models: As Box and Draper (1987, p.424) stated: "Essentially, all models are wrong, but some are useful." Every regression model can be improved. In this research we used the Spearman correlation and linear regression model. The application of different models will inevitably impact the uncertainty of results.

6. Conclusions and future work

This research focused on the exploration of the correlation between mobile phone usage and different aspects of travel behavior. We used three indicators (movement radius, eccentricity, and entropy) to represent the scale, shape, and randomness of individual mobility. Moreover, we examined the impact of individual level factors (age, gender) and supra-individual level factors (social temporal orders and built environment). Based on the empirical results from a case study in Harbin city, China, several conclusions on the correlation between mobile phone usage and activity behavior are derived.

(1) All three indicators (movement radius, eccentricity, and entropy) are significantly correlated with mobile phone usage. As shown in the statistical analysis, people with higher mobile phone usage have a larger spatial extent regarding their movement patterns, higher movement entropy, and their movement trajectories are less linear.
(2) Furthermore, we also examined the impact of explanatory factors on the correlation between travel behavior and mobile phone usage. Gonzalez et al. (2008) stated that the distribution of radius of gyration for mobile phone users captures a population-based heterogeneity. We extended their results by demonstrating that the distributions of all three indicators

(including the radius) are influenced by individual level and supra-individual social factors. As discussed in Section 5, both age and transportation network density indicate a strong impact on determining the correlation between mobile phone usage and travel behavior. However, there are no substantial differences between gender groups and social temporal order (weekends versus weekdays). We discovered that middle aged people ($23 < \text{Age} \leqslant 39$) have the highest correlation between cell phone usage and activity behavior among all age groups. Moreover, people whose work and home locations are both in the urban central areas (or both in the non-central areas) have a higher correlation between their mobile phone usage and daily activity behavior. As a result, the explanatory variables (both population heterogeneity and the characteristics of the built environment) need to be considered when discussing the correlation between human mobility and mobile phone usage.

This research provides us with new insights regarding the study of correlation between ICTs and travel behavior. Analyzing the mobility of mobile phone users offers valuable input to many mobile phone applications, such as individualized searching and advertising (currently leading edge business applications in the Location-Based Services market), real-time city modeling and simulation, early warning and emergency response systems, sustainable transportation planning, or air pollution and hot spot identification. Moreover, as indicated by the demonstrated analysis and results, ICT data such as mobile phone datasets offer new resources for geographic knowledge discovery and achieving inferential spatio-temporal information in the age of instant access, both important research areas in Geographic Information Science. Our future research will focus on analyzing the effect of additional explanatory factors, for example, income level, and occupation type once the data become available. A framework of how this research question is affected by different natural and social factors must be established. This would provide more specific results on how mobile phone usage impacts individual mobility, as well as offering new insights to policy makers. In addition, more work on the impact of "supra-individual regime" is needed. An important research direction would be exploring the correlation between mobile phone usage and the particular city structure. We will also look into the generalization of the results to other cities and countries, as well as the comparison among cities with various social, political, and cultural backgrounds. Another promising future work direction relates to differentiating the functionality of initiated and received phone calls. Moreover, in most of the correlation figures (e.g., Figs. 5–7 and 12), we can observe heteroscedasticity, with the variance of the error term increasing with frequency. We will further address this in our future analysis. Finally, it is very likely that the two indicators eccentricity and entropy are not independent of each other; therefore, we will continue exploring and clarifying the relationship between these two variables.

Acknowledgements

The data set used in this research was provided by the Geosoft Lab, Peking University, China. We thank Helen Couclelis and Kostas Goulias for providing insightful comments at various stages of this research. The three reviewers provided excellent feedback, which helped us to improve the content and clarity of this paper.

References

Abler, R. (1975). Effects of space adjusting technologies on the human geography of the future. In R. Abler, D. Janelle, A. Philbrick, & J. Sommer (Eds.), *Human geography in a shrinking world. North scituate* (pp. 35–56). MA: Duxburg Press.
Ahas, R., Aasa, A., Silm, S., Aunap, R., Kalle, H., & Mark, Ü. (2007). Mobile positioning in space – Time behaviour studies: Social positioning method experiments in Estonia. *Cartography and Geographic Information Science, 34*(4), 259–273.
Ahas, R., & Mark, Ü. (2005). Location based services – New challenges for planning and public administration. *Futures, 37*(6), 547–561.
Beckman, K. (2000). Umweltgerechtes Verkehrsverhalten beginnt in den Köpfen. In: TÜV Energie und Umwelt GmbH (Hrsg.): Mobilitätsforschung für das 21. Jahrhundert: Verkehrsprobleme und Lösungsansätze (pp. 213–238). Köln.
Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: Wiley.
Brown, L. A., & Moore, E. G. (1970). The intra-urban migration process: A perspective. *Geografiska Annaler, 52*(B), 1–13.
Couclelis, H. (2004). Pizza over the internet: E-commerce, the fragmentation of activity and the tyranny of the region. *Entrepreneurship and Regional Development, 16*(1), 41–54.
De Souza e Silva, A. (2007). Mobile phones and places: The use of mobile technologies in Brazil. In H. J. Miller (Ed.), *Societies and cities in the age of instant access* (pp. 295–310). Dortdrecht, The Netherlands: Springer.
Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: A geographic perspective*. New York: Guilford Press.
Gonzalez, M. C., Hidalgo, C. A., et al. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.
Hägerstrand, T. (1970). *What about people in regional science? Papers of the Regional Science Association, 24*, 7–21.
Hjorthol, R. J. (2008). The mobile phone as a tool in family life: Impact on planning of everyday activities and car use. *Transport Reviews, 28*(3), 303–320.
Horton, F., & Reynolds, D. R. (1971). Effects of urban spatial structure on individual behaviour. *Economic Geography, 47*, 36–48.
Inselberg, A. (2009). *Parallel coordinates: Visual multidimensional geometry and its applications*. New York: Springer.
Janelle, D. (1995). Metropolitan expansion, telecommuting, and transportation. In S. Hanson (Ed.), *The geography of urban transportation* (pp. 407–434). New York: The Guilford Press.
Jones, P., Koppelman, F., & Orfueil, J. P. (1990). Activity analysis: State-of-the-art and future. In P. Jones (Ed.), *Developments in dynamic and activity-based approaches to travel analysis* (pp. 34–55). Aldershot, UK: Avebury.
Kwan, M. P. (2007). Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *Professional Geographer, 59*(4), 434–446.
Kwan, M. P., Dijst, M., & Schwanen, T. (2007). The interaction between ICT and human activity-travel behavior. *Transportation Research Part A – Policy and Practice, 41*(2), 121–124.
Mazey, M. E. (1981). The effect of a physio-political barrier upon urban activity space. *Ohio Journal of Science, 81*(5-6), 212–217.
Miller, H. (2009). Geographic data mining and knowledge discovery: An overview. In H. J. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* (2nd ed., pp. 3–32). London: CRC Press.
Mokhtarian, P. L. (2009). If telecommunication is such a good substitute for travel, why does congestion continue to get worse? *Transportation Letters, 1*, 1–17.
Mokhtarian, P. L., & Meenakshisundaram, R. (1999). Beyond tele-substitution: Disaggregate longitudinal structural equations modeling of communication impacts. *Transportation Research Part C, 7*(1), 33–52.
Mokhtarian, P. L., & Salomon, I. (2002). Emerging travel patterns: Do telecommunications make a difference? In H. S. Mahmassani (Ed.), *Perpetual motion: Travel behavior research opportunities and application challenges* (pp. 143–182). Oxford, UK: Pergamon Press/Elsevier.
Nobis, C., & Lenz, B. (2009). Communication and mobility behaviour – A trend and panel analysis of the correlation between mobile phone use and mobility. *Journal of Transport Geography, 17*(2), 93–103.
Nobis, C., Lenz, B., & Vance, C. (2005). Communication and travel behaviour: Two facets of human activity patterns. In H. Timmermans (Ed.), *Progress in activity-based analysis* (pp. 471–488). Oxford, UK: Elsevier.
Pendyala, R. M., Goulias, K., & Kitamura, R. (1991). Impact of telecommuting on spatial and temporal patterns of household travel. *Transportation, 18*, 411–432.
Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. In A. A. Salah, T. Gevers, N. Sebe, & A. Vinciarelli (Eds.), *Proceedings of the first international workshop on human behaviour understanding (HBU 2010)* (pp. 14–25). Heidelberg: Springer.
Raubal, M. (2011). Cogito ergo mobilis sum: The impact of location-based services on our mobile lives. In T. Nyerges, H. Couclelis, & R. McMaster (Eds.), *The SAGE handbook of GIS and society* (pp. 159–173). Los Angeles, London: Sage Publications.
Salomon, I., & Mokhtarian, P. L. (2008). Can telecommunications help solve transportation problems? A decade later: Are the prospects any better? In D. A. Hensher & K. J. Button (Eds.), *The handbook of transport modeling* (2nd ed., pp. 519–540). Amsterdam: Pergamon.
Saxena, S., & Mokhtarian, P. L. (1997). The impact of telecommuting on the activity spaces of participants. *Geographical Analysis, 29*, 124–144.
Schönfelder, S., & Axhausen, K. W. (2002). *Measuring the size and structure of human activity spaces the longitudinal perspective, Arbeitsbericht Verkehrs- und Raumplanung*, 135, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich.
Song, C. M., Qu, Z. H., et al. (2010). Limits of predictability in human mobility. *Science, 327*(5968), 1018–1021.

Thulin, E., & Vilhelmson, B. (2007). Mobiles everywhere: Youth, the mobile phone, and changes in everyday practice. *Young, 15*(3), 235–253.

Viswanathan, K., & Goulias, K. G. (2001). Travel behavior implications of information and communications technology (ICT) in Puget Sound region. *Transportation Research Record, 1752*, 157–165.

Xia, Y. (2005). *Integrating uncertainty in data mining*. Unpublished Doctoral Dissertation, Los Angeles: University of California.

Yuan, M. (2009). Toward knowledge discovery about geographic dynamics in spatiotemporal databases. In H. J. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* (2nd ed., pp. 347–365). London: CRC Press.

Yuan, Y., & Raubal, M. (2010). On correlation between mobile phone usage and travel behavior – A case study of Harbin, China (extended abstract). In *Geographic information science – 6th International conference, GIScience 2010*, Zurich, Switzerland.