The Possibility of Big Data Spatio-Temporal Analytics for Understanding Human Behavior and Their Spatial Patterns in Urban Area

LI Miaoyi

September 2017

Dissertation

The Possibility of Big Data Spatio-Temporal Analytics for Understanding Human Behavior and Their Spatial Patterns in Urban Area

Graduate school of

Natural Science& Technology

Kanazawa University

Major subject: Environment Design

Course: Environment Planning

School registration No.: 1424052011

Name: 李 苗裔 (LI Miaoyi)

Chief advisor: Prof. Zhenjiang Shen

Date of Submission: September, 2017

Abstract

The aims of this dissertation are two-fold: first, to develop a comprehensive framework to integrate multiple dimensions and scales, and systematically compare human behavior patterns and current urban form; second, to further the understanding of the role spatial and temporal effects might play in human behavior patterns analyzing. The objective of this dissertation is the application of spatio-temporal analysis employing big data on human behavior patterns for identifying urban problems hidden in current urban form.

Firstly, we develop a framework to systematically measure and visualize human mobility reflected by trajectory data. We propose a framework that combines various spatio-temporal and network analysis units. Two case studies of Chinese cities are carried out to evaluate the usefulness of proposed conceptual framework. Our results suggest that the proposed framework can comprehensively quantify the variation of human mobility across various scales and dimensions. This part of work has been published by Environment Planning Part B: Urban Analytics and City Science. Meanwhile, under this framework thinking, we continue to use Taxi O-D data which belong to the type of urban trajectory dataset combining with POIs data to create the spatial and temporal entropy model to identify the degree of urban function mix. In this part, we considered individual trip behaviors happening in different functional zones when we had evaluated urban function mix. This research work has been published by China City Planning Review. Furthermore, we discuss the relationship between the spatio-temporal distribution of population and urban function mix, that help me to deep understand different urban functions and their mix of different attraction to people, which has been published by GeoJournal. Finally, we apply the spatio-temporal network analysis method for the comparative analysis of subway and taxi ridership pattern and their interactions in urban area. This chapter aims to examine the spatial variation of urban taxi ridership due to the impacts of a new subway line which has been published by Sustainability. We examine the spatio-temporal patterns and interactions of ridership in Wuxi by integrating taxi O-D trips from Taxi GPS data and subway smart card data from continuously collected fare transactions.

As proved by published research articles, the application of the framework of spatio-temporal analytics employing big data we introduced in this PhD research would focus on two essential aspects of the urban system: urban form and human behaviors. In this dissertation, human behavior refers to the mobility of human beings taking advantage of various transportation modes in urban area, which distribution generates spatial patterns during their daily life. Their spatial patterns can reflect how the mobility of human beings relative to the urban form. We tested human behaviors using their moving tracks recorded in mobile devices and traffic cards, these types of datasets belong to spatio-temporal big data. The analysis results can provide support for planner and designer to find urban problems hidden in current urban form.

Key Words: Human Mobility; Urban Form; Urban Function Mix; Spatio-Temporal Distribution of Population; Transit Ridership; Spatio-Temporal Entropy Model; Spatio-Temporal Network Analysis

Acknowledgement

This dissertation was impossible without the collaboration and insight of my committee members. I would like to take this opportunity to gratefully acknowledge my advisor, Prof. Zhenjiang Shen, who inspired me to start the research on spatio-temporal analysis and urban "big data" analysis, gave me his valuable comments, and has guided me through the process of writing this dissertation. I would also like to thank my committee members, Profs. Anrong Dang (Tsinghua University), CHIKATA Yasuo (Kanazawa University), KOBAYASHI Fumihiko (Kanazawa University) and NISHINO Tatsuya (Kanazawa University) from the bottom of my heart. I could not have asked for a better committee in my research endeavors. I truly admire all of my committee members for their scientific brilliance, and appreciate their time and patience.

My thanks also go to my parents in China, who have always been there for me throughout all these years with endless patience and encouragement. And to all my friends, colleagues, and classmates in Kanazawa and Beijing, thank you for your support and help throughout my doctoral study. Acknowledgements must be given to the Japan Government for sponsoring my degree at Kanazawa University. I would not have been able to finish this research without its financial support.

Finally, I must acknowledge the support of my wife, Yan, who has always been by my side throughout the dissertation process, providing aid, encouragement, and plenty of needed study breaks. We had a very lovely daughter during my Ph.D. course. Their endless love is essential to the success of my Ph.D. study.

List of Figures

List of Tables

List of Equations

Table of Contents

Chapter1

1. Introduction

1.1 Research Background

In contemporary city, urban form is not just the physical thing, we should think it as the dynamic system. locations, interactions, flows and networks can shape the current urban form, that means people's mobility and distribution in the urban space across different time are a very important for understanding dynamics of urban form. As Miller states, "all activities have both spatial and temporal dimensions that cannot be meaningfully separated" (Miller, 2004a, p648). However, it is challenging to explore and compare the potential interactions among space and time across scales in spatio-temporal data. Temporal Geographic Information Systems (temporal GIS) research has been an active and growing area of intellectual inquiry since the late 1980s and it aims to process, manage, and analyze spatio-temporal data (Goodchild, 1987; Egenhofer and Golledge, 1997; Yuan, 1999; Christakos et al., 2001). Due to the expanding universe of spatio-temporal data and the potential for gaining valuable insights into a number of phenomena based on spatio-temporal analysis, the demand for innovative visualization and analytical approaches has grown. Scholars have made much progress in developing new conceptual models, algorithms, and analytical methods to address a wide range of spatial and temporal dynamics across disciplines. Many exciting new research developments continue to push the conceptual, theoretical, and technological boundaries of our ability to identify and analyze spatio-temporality (Peuquet, 1984; Goodchild, 1987; Haining et al., 2001; Takatsuka and Gahegan, 2002; Anselin and Rey, 2002; Peuquet, 2002; Andrienko et al., 2003; Guo et al., 2006).

On the other hand, in planning practice, especially in China, after development, planners and designers have to pay more attention to transportation modes carrying people between their houses, working and shopping places because of over heavy traffic jam. Thus, researchers have to understand how people move, active and try to find out solutions for solving the planning issues. For this, how to analysis the increasingly available spatio-temporal big data which store detailed human mobility information become a very hot topic in China and it is important to understand the methodology of how to check the human mobility in urban area for their daily life

1.2 Research Purpose

In this dissertation, spatio-temporal big data refers to a large dataset which stores spatially referenced time-series with single or multiple thematic attributes, such as cell phone trip data, taxi O-D trip data, location-based service (LBS) data and so on. Thus, human behavior refers to the mobility of human beings taking advantage of various transportation modes in urban area, which distribution generates spatial patterns during their daily life. Their spatial patterns can reflect how the mobility of Human beings relative to the urban form. This PhD research aims to discuss the possibility of spatio-temporal analytics employing big data could contribute to analyzing human behavior patterns dynamically and comparatively for identifying urban problems hidden in current urban form that otherwise would be very difficult to detect.

Firstly, we develop a framework which can integrate spatio-temporal and network considering different scale, it sheds new insights on possible data fusion and comparative analysis. It can be applied to different data sources and methods to visualize human mobility reflected by trajectory data which exists in the current urban form. Secondly, according to this framework, we continue to mainly employ spatio-temporal big data which stores rich human mobility information to analyze spatial distribution pattern of urban function and its mixture, the correlation between the spatio-temporal distribution of population and urban function, spatio-temporal patterns and interactions between the subway and taxi ridership.

1.3 Literature Review

1.3.1 Existing Research for analyzing Human Behavior

There are many research reports on spatial analysis and modelling with Geographic Information System (GIS) in the world (Maguire et al, 2005). In this vein, many studies examined the capacity of GIS modelling in aid in understandings of household choices (Ettema, 2011; Gaube & Remesch, 2013; Q. Huang, Parker, Sun, & Filatova, 2013), residential demand (Fontaine & Rounsevell, 2009), and land-use changes (Jokar Arsanjani, Helbich, & de Noronha Vaz, 2013; Kocabas & Dragicevic, 2013). The applications of GIS modelling range from simulation to prediction to monitoring (Bernard & Kruger, 2000; Carrara, Guzzetti, Cardinali, & Reichenbach, 2000; Q. H. Huang & Cai, 2007) , from modelling human-related issues to modelling natural hazards (Al-Sabhan, Mulligan, & Blackburn, 2003; Perry, Sparrow, & Owens, 1999; Thornton, Pearce, & Kavanagh, 2011; Xu, Nyerges, & Nie, 2014), from modelling real-world problems to modelling virtual-world activities (Croitoru, Wayant, Crooks, Radzikowski, & Stefanidis, 2014; Turgeon, 2013), from using data collected by digitally instrumented devices (e.g. remote sensing images, GPS coordinates, sensor network data) to using data collected by humans (e.g. survey data, geo-social media data, mobile phone data) (Blaschke, Hay, Weng, & Resch, 2011; Liu, Sui, Kang, & Gao, 2014; Patino, Duque, Pardo-Pascual, & Ruiz, 2014; Widener & Li, 2014). In fact, the study of individual behaviors has a long history in GIS field. Survey data, such as the survey of people's daily activities, is initially exploited as the main data source(M. Kwan, 1998).

There are many national surveys on industrial activities and commercial activities and housing in urban spaces. But conventional surveys are impossible to show the tracks of all urban activities.

Big data including limited attributes of individual person and his/her behavior's, helps us to understand the diverse urban activities and complicated urban structures in urban spaces (Nabian, Offenhuber, Vanky, & Ratti, 2013). Some initiatives in this area include studying Beijing's function zones using human mobility and POI data(Yuan et al., 2012) and evaluating the effectiveness of a future road plan based on taxi trajectory (Zheng, Liu, Yuan, & Xie, 2011). Recently, many reports attempted to identify, characterize, and develop geospatial frameworks and algorithms for individual behavioral models for spatio-temporal knowledge discovery. Geospatial frameworks were developed to analyze people's activities patterns over time and geography (M. Kwan, 2004). Geo-visualization methods were introduced to aid in understandings of those patterns by delivering interactive visual outputs (M. Kwan, Lee, & Oxford, 2003). More recently, the increasing use of volunteered geographic information (VGI) and open data has further spurred the study of individual behaviors. In this thread, researchers built upon the previous work of spatio-temporal data mining, and developed methods to discover spatio-temporal patterns of human activities from geotagged photos and mobile phone data.(Andrienko, Andrienko, Mladenov, Mock, & Pölitz, 2010; Sagl, Delmelle, & Delmelle, 2014; YT Zheng, Li, Zha, & Chua, 2011).

1.3.2 Exploratory Spatial Data Analysis for Revealing Saptio-Temporal Pattern

It is the research question, not the methodology, that should drive the design of scientific studies. Most empirical studies are motivated by certain well defined research questions. The researcher chooses the appropriate analytical methods while also obtaining the data needed for application of the methods. In other words, it is the principal concern for most researchers that the data set and relevant software are available before stepping into the investigation. To raise a new research question might involve the "problem" of using some methods/tools which have not been developed. Only at the end of the process does the analyst interpret and evaluate the results. However, traditional research methods tend not to be very useful in revealing spatio-temporal patterns in new, large, and complicated spatio-temporal data sets. Hence, the analyst must become acquainted with the data before formulating valuable research questions.

The process of “getting acquainted with data” is the basis of exploratory data analysis (EDA) (Andrienko and Andrienko, 2006). EDA, a philosophy of conducting data analysis, originates from Tukey’s seminal work (Tukey, 1977). According to Tukey, EDA is to analyze data for the purpose of interactively formulating hypotheses instead of testing hypotheses. Under the framework of EDA, exploratory spatial data analysis (ESDA) is defined as “detecting spatial patterns in data, formulating hypotheses based on the geography of the data, assessing spatial models” (Haining and Wise, 1997). Le Gallo and Ertur (2003) consider EDA as “a set of techniques aimed at describing and visualizing spatial distributions, at identifying a typical localizations or spatial outliers, at detecting patterns of spatial association, clusters or hot spots, and at suggesting spatial regimes or other forms of spatial heterogeneity”. ESDA can reveal complex spatial phenomenon not identified otherwise (Anselin, 1993), and it forms the basis for formulating spatially explicit research questions. The development of new methods of ESDA has stimulated a number of research efforts (Anselin and Getis, 1992; Longley et al., 2001; Getis et al., 2004; Rey and Anselin, 2006). Geometric methods have been suggested to summarize spatial patterns and analyze geographic process, such as Weber’s Triangle, the Gravity Model, and Central Place Theory, among others (Mu, 2004). Widdows (2004) reveals geometric patterns from applications in astronomy, music, and biology in ancient times to the design of today’s user interfaces and search engines. Recent progress in statistical shape analysis (Goodall and Mardia, 1999) reveals great potential for ESDA from studying shape variations at the micro scale such as those found in human brains (Mardia and Dryden, 1999) to the Voronoi polygons generated during the examination of central place theory (Dryden and Mardia, 1998).

As commented by Goodchild (2006) (P4-5), “. . .GIScience is applicable to varying degrees in any space, . . . such as the three-dimensional space of the human brain,. . .At the same time, advances made in the study of other spaces may be suitable sources of cross-fertilization in GIScience. Perhaps the next decade will see a much greater degree of interaction between GIScience and the sciences of other spaces, and much more productive collaboration.” While computational geometry and analytical cartography can generate in-depth visualization and create summaries of complex spatial patterns (Eppstein, 2005), they largely ignore dynamic effects. To consider spatial and temporal attributes jointly, Rey (2007) calls for extending exploratory temporal data analysis to space, while also incorporating time into exploratory spatial data analysis.

In many disciplines, researchers are asked to compare and contrast two things, such as two theories, two temporal trends, two spatial processes, and so on. When one or two space-time income data sets are presented, it is interesting to detect crucial differences or surprising commonalities between two regions or across two groups of regions, which can be refined to generate many important research questions. For example, the role of space in convergence dynamics can be compared among various national systems. Faced with a daunting list of differences and similarities, it is necessary to design research questions logically and comprehensively. However, existing exploratory approaches to spatio-temporal analysis, from data mining to visualization, are limited to building a framework combining spatio-temporal and network for comparative studies. Comparative spatio-temporal analysis suggested in this dissertation aims to create analysis unit to compare spatio-temporal patterns within one data set, as well as across two or more data sets.

Through reviewing the rich literatures, the main research question of this dissertation will be raised: how to build a comprehensive framework for comparative spatio-temporal data analysis for analysing spatial pattern of human behavior? To illustrate the function of this comprehensive framework, some spatio-temporal analysis methods will be developed to compare spatio-temporal pattern of human behaviors and current urban form. Through the above comparative analysis, we argue that it is possible to identify problems hidden in urban form.

1.4 Organization

Research topics for the dissertation are organized into four areas (see Fig 1.1). The first component is the introduction of a framework for comparative spatio-temporal analysis. Next, the spatio-temporal entropy model is designed based on combination POIs(Point of Interests) data and one type of spatio-temporal big data named taxi O-D data to evaluate the degree of urban function mix. The third component suggests a method reveal the distributional relationships of the population and urban functions. Lastly, will apply the spatio-temporal network analysis method to empirical studies for analyzing spatiotemporal patterns and network interactions between subway and taxi ridership based on taxi and subway O-D trip dataset.

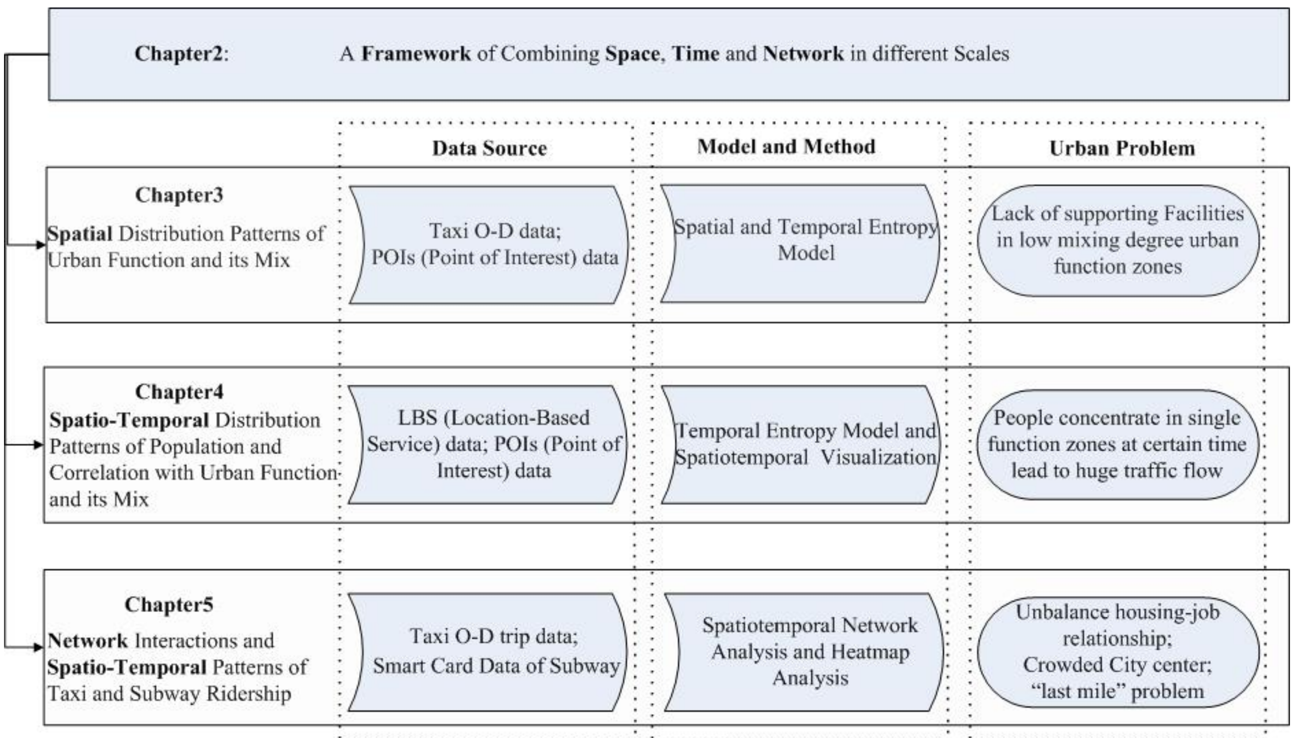


Fig 1.1 The framework of this Ph.D. research contents

1.4.1 A Framework for Comparative Spatio-Temporal Analysis

Chapter 2 will stress the need to study multiple dimensions and multiple scales underlying spatio-temporal income big datasets. We will propose a framework that combines various spatial, temporal and network analysis units. By customizing the combination of analysis units, the application of a comparative exploratory analysis framework will be suggested for use in analysis of spatio-temporal big data series to measure human mobility pattern reflected by trajectory data dynamically and comparatively. Two case studies of Chinese cities are carried out to evaluate the usefulness of proposed conceptual framework. Our results suggest that the proposed framework can comprehensively quantify the variation of human mobility across various scales and dimensions. I have published one paper on Environment Planning Part B: Urban Analytics and City Science.(SSCI) whose title is “A Framework of Comparative Urban Trajectory Analysis”, it will be the main content of Chapter2.

1.4.2 Evaluating Urban Function Mix based on POIs and Taxi O-D Data

Urban functions exist in urban form, it also occur interaction with human behaviors. That means identifying different functional zones and understanding their spatial pattern within the city is benefit for finding out some issues taking place in current urban form. In this Chapter, we will analyze the spatial distribution patterns of urban functions based on the data of points of interest (POIs). Then according to taxi O-D trip which is extracted from the taxi GPS data, we identified different functional zones. Finally, we establish a spatial entropy and a temporal entropy based on POIs data and taxi O-D data which belongs to one type of spatio-temporal big data, using this methods to calculate the degree of urban function mix and analyze the spatial pattern of the degree of urban function mix, we find out the typical areas where are lower mixing degree and what's reason lead to this situation. The research work of Chapter3 has been published by China City Planning Review (CSCD) whose title is "Application of Spatial and Temporal Entropy Based on Multivariate Data for Measuring the Degree of Urban Function Mix".

1.4.3 Uncovering the Relationship between Spatio-Temporal Distribution of Population and Urban Function with Location-Based Service Data

On basis of Chapter3, Chapter4 examine the spatio-temporal distribution of the population in an urban area and its relationship with urban functions using an unprecedented high-resolution and broad-coverage-crowd Location-Based Service (LBS) dataset. The analysis of the spatio-temporal population distribution based on temporal entropy indicates that population distributions of employment, commercial, and scenic areas have larger temporal fluctuations than those in residential and mixed-use areas. The relationship which is important in terms of knowing the potential demand for people in different urban functions and the current situation of services supply that urban facility provides. For Chapter4 we further published another paper on Geojournal (EI/SCImago) whose title is "Revealing the relationship between spatio-temporal distribution of population and urban function with social media data".

1.4.4 Examining Taxi Ridership Impacts from Newly Opening Subway Line with Taxi Trip Data

In addition to urban function discussed above Chapter3 and Chapter4, Transportation Network in big cities is also a crucial part of urban form. Nowadays, smart card and mobile device for transportation are widely used in many Chinese cities. We can try to use this type of spatio-temporal big datasets to visualize the human mobility of transportation network in urban area. Chapter5 will apply the spatio-temporal network analysis method for the comparative analysis of subway and taxi ridership and their interactions in urban area, that is to say, the interdependent transportation networks. Incorporating new data, This Chapter aims to examine the spatial variation of urban taxi ridership due to the impacts of a new subway line operation opened in 2014 in Wuxi, China. We examine the spatio-temporal patterns and interactions of ridership in Wuxi by integrating taxi O-D trips from GPS data and subway smart card data from continuously collected fare transactions. The contents of Chapter5 have been published by Sustainability (SSCI/SCI) whose title is "Examining the interaction of taxi ridership and subway for sustainable urbanization".

In the following chapters, spatial and temporal entropy model, spatio-temporal network analysis, spatio-temporal visualization and heatmap analysis methods will be set as the specific analysis methods under the analytics framework, and the urban trajectory dataset will mainly be used for case studies of different cities in China. They are Taxi O-D trip data which is extracted from Taxi GPS dataset, Cell Phone trip data and smart card data of subway. Meanwhile, we also combine POIs (Point of Interest) and LBS (Location-Based Service) data into our research work. It is important to understand the methodology of how to check the human mobility in urban area through the spatio-temporal big datasets.

Chapter2

2. A Framework for Comparative Spatio-Temporal Analysis

2.1 Introduction

Advancements in sensing and computing technologies have created urban trajectory datasets of human and vehicle movement at an unprecedented scale and speed. For example, the prevalence of GPS, Wi-Fi, Cellular, and RFID devices has enabled human dynamics to be recorded through the movement of taxis, fleets, public transits, and mobile phones at the individual level. Understanding and analyzing human dynamics can help assess transportation infrastructure and urban planning policies, and further inform better strategies to optimize urban and transportation planning, improve human life quality, and amend city operations.

Liu et al. (2012) studied the intra-urban daily human mobility using taxi trajectory data in Shanghai. Online check-in data recorded mostly by GPS-enabled mobile devices, can also be good indicators of human activities. To enable data-driven urban studies using these increasingly rich trajectory data and facilitate better decision-making for domain experts and policy-makers, the following two issues need to be addressed. First, trajectory data sets are often collected with diverse formats and at various scales and second, it is often challenging to incorporate social connection data with geographic analysis (Andris, 2016).

Existing methods and frameworks mostly focus on analyzing human dynamics at a specified scale (e.g. individual level, intra-urban level, or inter-urban level), instead of comparing human mobility patterns across various scales. Such comparative studies are particularly needed in the urban planning context, given that assessment and comparisons of different planning scenarios are important for decision-making. Taxi trajectory data can be used to investigate urban functional zones at different scales (Zhao et al. 2016). They may be studied at an individual level to identify human mobility patterns, or aggregated by officially defined boundaries or trajectory similarities to compare administrative spatial layout with human-demarcated community structures and explore gaps between public's transportation needs and the current state of transportation planning scenarios (Li et al. 2017). Experiments with different analysis units are often needed as well in that researchers or planning specialists may critically examine the effects of aggregation, both spatially and temporally, to identify the proper analysis unit (Ye and Rey 2013)

Trajectory data contains rich information about both spatial and social processes within urban space, and the analytical methods vary drastically from spatio-temporal analysis to network analysis at various scales. In this regard, we propose a framework that can accommodate various analysis needs by considering different combinations of spatio-temporal, and networks for analyzing trajectory data. Although there are no one-size-fits-all solutions for analyzing trajectory data, the proposed framework aims to provide researchers and planners a foundation to customize their analysis. The synthesis of various analysis scales and dimension allows users to uncover spatio-temporal and societal implications of human activity patterns dynamically and comparatively. In the rest of the chapter, we first provide a comprehensive literature review on theoretical foundations for developing the proposed framework, following by an elaboration of the framework. Two case studies are then presented to showcase the feasibility of the suggested framework. This chapter concludes with a discussion on potential applications of the framework and future research directions.

In this chapter, we propose a comprehensive framework that can systematically measure and visualize the human mobility pattern in urban area. In addition, this framework will be applied to different data sources (taxi and cell phone) and various urban settings (Beijing and Chongqing).

2.2 A Framework-Dimensions and Scales for Spatio-Temporal and Network

We establish an analytical framework that considers three dimensions and four scales. Table 2.1 conceptualizes the according to 12 basic analysis units. The social network here is represented by the relationship among urban objects connected through the trajectory. Human mobility patterns reflected by trajectory data are associated with places (Zhang et al. 2016). Therefore, trajectory among places evidences relationship among individuals, groups, and places (Andris, 2016). That is, the trips people travel to work, visit friends, and shopping all reflect their social connections with people and places. Analyzing the networks formed by trajectory data thus provides an additional lens to examine the social proximity among places (Huang et al. 2015).

The analysis unit at the individual scale signifies the geographical location of an attribute (A1, Table 2.1), the temporal label of an attribute (A5, Table 2.1), or a single trajectory that connects two places (A9, Table 2.1).

The analysis unit at the local scale explores a group of units formed by the focal observation and its neighboring observations. A focal area and its neighboring areas, for example, can be considered as the unit of analysis from the perspective of the spatial dimension (distribution) at the local scale (A2, Table 2.1) (Al-Dohuki et al., 2017). The specific time such as a focal hour, an hour before, and the hour after can be considered as the unit of analysis from the perspective of the temporal dimension at the local scale (A6, Table 2.1) (Huang et al., 2016). A local network can be considered as for how a focal community is connected to its related area (A10, Table 2.1). For example, how a shopping center is connected to its related communities through various transportation means (Wang et al., 2016a).

A meso-scale analysis studies a group of entities which shares similar features in spatio-temporal or network. The spatial distribution of areas with certain features can be treated as the meso-scale (A3, Table 2.1). The time period after an event or policy can also be considered as the meso-scale (A7, Table 2.1) (Wang et al., 2016b). Meso-level networks depict connections among a group of similar communities (A11, Table 2.1) (Al-Dohuki et al. 2017).

The analysis at the global scale examines distributions of all spatial entities, times, or trajectories. Spatial distribution of all spatial entities, for example, can be considered as the global scale from the perspective of the spatial dimension (distribution) (A4, Table 2.1) (Li et al. 2017). The entire study period can be considered as the global scale from a temporal point of view (A8, Table 2.1) (Wang et al. 2016b). The global-scale network takes into account all the trajectories and their associated places (A12, Table 2.1) (Huang et al. 2015). Limiting attention to only one of these dimensions or scales may result in a misguided or partial understanding of urban dynamics.

Table 2. 1 The framework for spatio-temporal network analysis

		Scales			
		Individual	**Local**	**Meso**	**Global**
Distributions	**Space**	A1	A2	A3	A4
(Dimensions)	**Time**	A5	A6	A7	A8

	Network	A9	A10	A11	A12

The framework developed based on Table 2.1 leads to a general task topology for analyzing the human behavior and their spatial pattern by integrating spatial, temporal and network distributions at the individual, local, meso, and global scales. It allows the behavior of a dynamic system to be reconstructed from a group of analysis units. The key aspect of this framework is to integrate the three dimensions of urban trajectory dataset in a four-scale environment. In total, 64 possible combinations of space and time, space, and network, time and network can be derived from the task design.

Spatio-temporal movement patterns at the individual level can adopt some methods from time geography, such as spatio-temporal prisms and choice set of individual′s activities (Miller 2005, Kuijpers *et al*., 2010, Chen and Kwan, 2012). At the local scale, we can examine how a focal agent interacts with others (Winter and Raubal, 2006; Neutens *et al*., 2010), and whether two moving objects have physically met (Kuijpers *et al*., 2011). The meso-scale analysis includes clustering and generalization of trajectories toward studying the community structures and social proximity among places (Andrienko and Andrienko, 2011; Guo *et al*., 2012a; Murray *et al*., 2012). The global scale considers the overall pattern, such as the social interaction potential in a city and predictions of future movements (Farber *et al*., 2012; Song *et al*., 2010; Horner *et al*., 2012).

2.3 Case Studies

2.3.1 Study Area

Trip flows reflect human mobility patterns and social flows (Andris 2016). These flows can be captured by various sources including GPS traces (e.g., Liu et al. 2010), pedestrian activities (e.g., Girardin et al. 2008), mobile phone records (e.g., Woodard et al. 2017), and subway cards (e.g., Lathia and Capra 2011). In the following section, we present case studies in Beijing and Chongqing, two large metropolitan areas in China, to showcase the proposed conceptual framework. Cell phone and taxi trajectory data are used for Chongqing and Beijing respectively to derive trip flows.

2.3.2 Methodology

Data source

Cell Phone Signaling Data

The mobile phone signaling data from GSM (Global System for Mobile Communication) network operated by China Unicom takes up approximately 20% of the market share in China. The data used in this chapter was collected from the A and E interface in GSM network. Each record is generated when a mobile device connects to the cellular network in the following instances:

-when a call is placed or received (both at the beginning and end of a call);

-when a short message is sent or received;

-when a location update occurs;

-when handover occurs;

Every record includes seven fields, including the IMSI (International Mobile Subscriber Identification Number), timestamp, location area code, base station ID, MSC (Mobile Switching Center) id and BSC (Base Station Controller) ID, event type (including random location update, periodic location update procedure, connect management service request, paging response, BSC handover and so on).

The data used in the case study of Chongqing has generated from five hundreds mobile phones over one month in 2013. It consists of 230 millions of records per day for 4.7 million devices and covers 38 counties of 82,400km^2 in Chongqing, with an average of 26 records per device.

Taxi GPS Data

The mobile devices equipped with a GPS receiver chip can collect information about the mobility of people, vehicles, and other mobile objects. The taxi GPS data used for the case study of Beijing was collected from GPS devices installed in taxis, a current wide-used floating car technology to track traffic conditions. According to the report from the Beijing Traffic Bureau, taxis accounted for 12 percent of Beijing's total ground traffic (Yuan et al., 2012), therefore the taxi datasets can well reflect people's activities and reveal the urban functional structure.

Trip flow matrix estimation based on Cell Phone Signaling Data

The methodologies for trip flow matrix estimation comprise five steps including data pre-processing, trajectory cluster and spurious points filtering, trip identification, commuting identification and trip flow aggregate.

Data pre-processing

Data pre-processing includes scanning records one by one to remove those fail to track IMSI number, grouping records by users, and sorting every user's records in chronological order to get each user's daily activities trajectory.

The base station positioning technology was adopted here. Phone users are geo-located by combining their current base station ID with the station's coordinates. In urban areas where base stations are densely distributed, the accuracy of geo-locating is about 200 to 800 meters. In rural areas where base stations are sparse, the accuracy of geo-locating differs from a few hundred to a few thousands of meters.

Users' daily trajectory is presented in an ordered sequence $M = \{m_1, m_2, ..., m_n\}$.Each location measurement $m_i \in M$ is characterized by a position expressed p_i in latitude and longitude and a timestamp t_i (Fig 2.1).

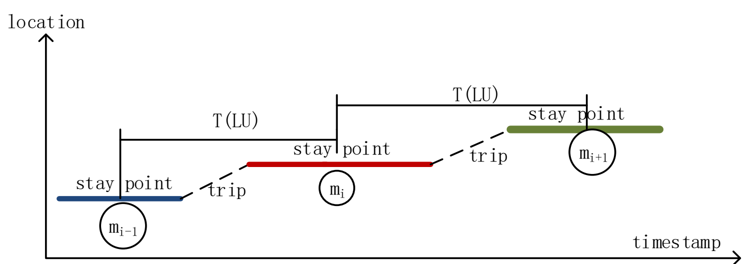


Fig 2. 1 Spatio-temporal trajectory for user's activities

The enter time t_i^{in} for base station m_i is $t_i^{in} \in (t_{i-1}, t_i)$

The departure time t_i^{out} for base station m_i is $t_i^{out} \in (t_i, t_{i+1})$

It is assumed that:

$$t_i^{in} = t_i \quad (2.1)$$

$$t_i^{out} = t_{i+1} \quad (2.2)$$

Mobile phone signaling records are sparse and irregular, in that users' displacements (consecutive non-identical locations) are often observed with long travel intervals. For example, the first location may be observed at 8:00 and next location may be observed after an hour or more. Therefore, we can only approximate when a user enters or leaves a base station, and estimate errors in calculating enter time and departure time. The max error of entering time or departure time is $(-T, 0)$ where T represents the max interval between two consecutive records and minus sign means the true enter or departure time is ahead of the calculated values.

The stay time T_i^{stay} for each point m_i equals $t_i^{out} - t_i^{in}$. The max error for stay time is $(-T, T)$.

Trajectory cluster and spurious points filtering

In practice, the base station that provides service for a user is not necessarily the one nearest to the user and may be constantly changing in spite of no actual displacement. This is because the operator often balances call traffic among adjacent towers by allocating a new call (or shifting an ongoing call) to the tower that is handling lower call volumes at that moment. To reduce the number of false displacements, we, therefore, take the following measures.

We first analyze two consecutive trajectory points; if the linear distance between two points is less than the threshold, trajectory point m_i and m_{i+1} will be fused together:

$$\mathrm{dis}(p_i, p_{i+1}) < \Delta S \quad (2.\ 3)$$

where $\mathrm{dis}(m_i, m_{i+1})$ represents the linear distance between point p_i and p_{i+1}, and ΔS represents the spatial threshold, for whom recommended value is 1 km to take into account the localization errors for the base station.

After this measure, duplicate trajectory points and trajectory points nearby will be merged into one point. The short distance trip (trip distance less than the threshold) will also be eliminated. Therefore, the total number of trips will reduce.

The fact that base stations a user connected to for communication may vary in a wide range make the illusion that a user moved a long distance in a short period of time. To diminish this effect, we then analyze three consecutive trajectory points successively. If the following three conditions are met at the same time, trajectory points m_i, m_{i+1} and m_{i+2} will be fused together.

$$\mathrm{dis}(p_i, p_{i+1}) \geq \Delta S \quad (2.\ 4)$$

$$t_{i+2} - t_i < \Delta T \quad (2.\ 5)$$

$$\mathrm{dis}(p_{i+2}, p_i) < \Delta S \quad (2.\ 6)$$

where ΔT is the temporal threshold, whose value is recommended to fifteen minutes.

This measure can effectively eliminate the false displacements of single drift phenomenon of the base station. However, we may lose some quick return trips in this process. That is, when a user makes a short stop at the destination and return to the origin quickly, the total time cost of the trip may be less than the time threshold.

When executing the fusion of trajectory points, the former point m_i will be reserved and the second point m_{i+1} will be removed. The departure time and stop time of m_i need to be updated synchronously as follows:

$$t_i^{out} = t_{i+1}^{out} \quad (2.\ 7)$$

$$T_i^{stay} = T_i^{stay} + T_{i+1}^{stay} \quad (2.\ 8)$$

After this process, all points left in the sequence are spatially dispersed.

Trip identification

The stop points can be identified based on stay time that is greater than a time threshold, which is determined by the area size of analysis and the time interval between consecutive records.

A user's trajectory can be cut into several consecutive trips by stop points. The stop point becomes the destination of the last trip as well as the origin of next trip like $s_1, \ldots s_2, \ldots s_i, \ldots s_n$, where s_i represents a stopping point.

Each trip for a user u can be characterized by an origin o, a destination d, departure time t_o and arrival time t_d as follows:

$\mathrm{trip}(u, o, d, t_o, t_d)$

$o = p(s_i)$
$d = p(s_{i+1})$
$t_o = t_{out}(s_i)$
$t_d = t_{in}(s_{i+1})$

In this chapter, the basic analysis units are counties, whose area ranging from tens to hundreds of square kilometers. We only care for the long distance trips that cross the county. So we can simplify the methodology to calculate the total stay time in each county instead of the base station. The location measurement in one county can be fused together and their stay time can be added together to get user's total stay time in that county. Stay time for county R can be calculated as follows:

$$T_R^{stay} = \sum_{p_i \in R} T_i^{stay} \quad (2.\ 9)$$

According to the length of stay time, we can determine whether the user stopped in the county for an activity or not. If the stay time surpasses the threshold, which was 2 hours in our case, the county can be regarded as a stopping point, which becomes a trip origin or destination. Otherwise, users are considered to be through traffic or make a temporary stop (e.g. stops for car fuel or driver rest).

Commuting identification

Most people have regular daily activity patterns that can be extracted through long-term observation (González, M. C. 2008). To extract users' mobility patterns, we first identify candidate home and workplace locations using a Location Stability Index (LSI) similar to what presented by Hao et al. (2010). Commuting trips between user's home and work places can then be identified accordingly.

We first divide each day into 48-time windows, with each covering 30 minutes. For each stop, the time window that first overlaps the stay period (t_i^{in}, t_i^{out}) over 50% was chosen as the enter time window and the window that latest overlap the stay period (t_i^{in}, t_i^{out}) over 50% was chosen as the departure time window. For each time window, we then overlay users' trajectory with stop points of multi-weekdays, and calculate the Location Stability Index (LSI) of each stay point using following formula:

$$LSI_i = \frac{\text{cover}(i,r)}{N_{day}} \quad (2.10)$$

Where N_{day} is the total days for analysis, $\text{cover}(i,r)$ is the number of stop points that be covered by the circle shape centered m_i with radius r.

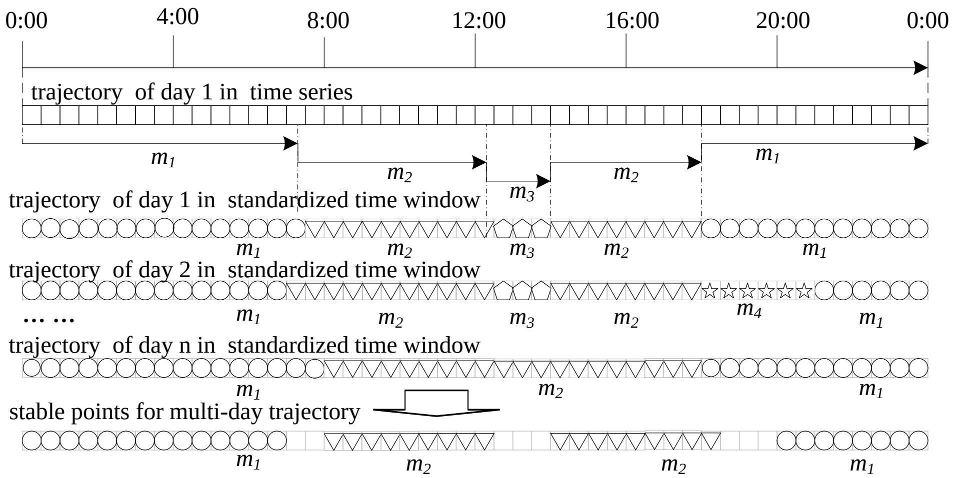


Fig 2. 2 Extracting spatio-temporal stable points from user's multi-day trajectory

The stay points with max *LSI* and is no less than the minimum threshold (value of 0.6 is suggested) was chosen as the multi-weekdays stay point in that time window(see Fig 2.2). A fusion action will be taken if the linear distance of two multi-weekday points in adjacent time windows is less than the spatial threshold, which can refer to (2) in this section.

We consider points with the longest stay time at night (22:00-7:00) as users' home locations and points with longest stay time in the day (7:00-22:00) as workplaces. The cumulative stay time of home or workplace needs to be more than the minimum threshold, (i.e. 2 hours) to be considered as valid.

Trip flow matrix of mobile phone users

All trips are then aggregated by their origin and destination to form a trip flow matrix. For each time window tw, the trip flow between area i and area j can be calculated as follows:

$$OD(i, j, tw) = \sum_{o \in i, d \in j, t_o \in tw} \mathrm{trip}(u, o, d, t_o, t_d) \quad (2.\ 11)$$

The spatial statistical unit for trip flow, estimated by trajectory mobile phone users should not be too small. The static results error will increase sharply with the unit size reduction because of base station positioning errors. It is recommended that the size of the statistical unit is no smaller than four square kilometers.

Taxi trajectory extracting and visualization based on Taxi GPS data

Taxi trajectory data is also pre-processed given its unstructured nature. The processing steps include:

①Taxis can only reflect people's purposeful activities and commute in a city when they carry passengers. Therefore, trajectories of taxis carrying passengers are obtained according to the traveling state.

②According to the trajectories of taxis that carry passengers, data on starting points, arrival points, and time of each trajectory is acquired.

③The starting area, arrival area, and time (hours) are attained by plotting each point on the map, getting the area which the coordinates are in, and dividing time by hour.

As a result, this section basically achieves the identification and spatialization of the taxi GPS data-based origin (O) points, destination (D) points, and the corresponding time (T), with a total of 2,253,626 O points and 2,253,551 D points.

2.4 Framework Implementation

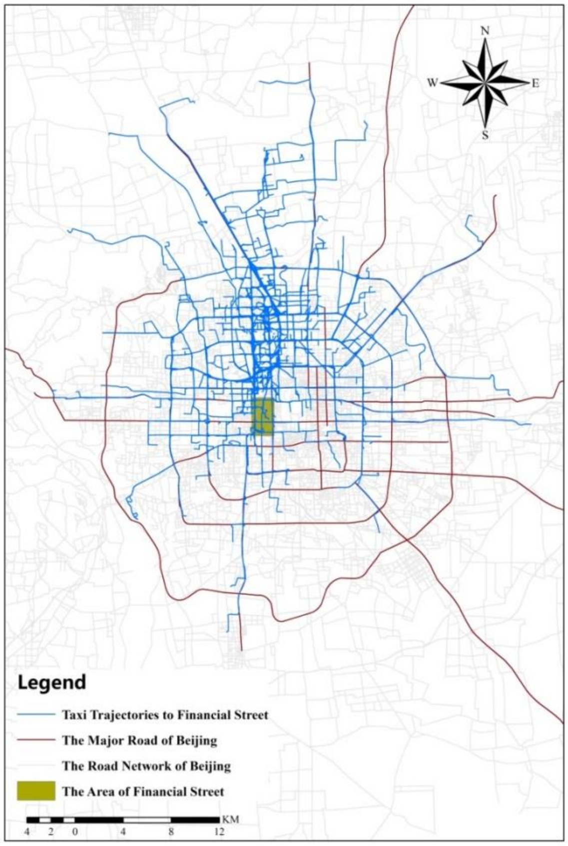
With pre-processed taxis trajectory and cellular phone data, we utilize the proposed framework to showcase how different combination of spatio-temporal and network analysis can help investigate human mobility patterns. Two different combinations of analysis units were used for Beijing taxi trajectory data (Example 1 and 2 in Table 2.2). In example 1, we combine individual-scale space, individual-scale time and local-scale network. That is, we used individual-level taxi trajectory data, and form them into origin-destination (OD) trips among places. As a result, a network is derived at a local scale that represents how a place is connected to other parts of the city. Figure 2.3 (a) and (b) demonstrate the resulting network that shows taxis trips going to and leaving from the Financial Street in Beijing within an hour period on Feb.27th, 2013. The two maps show similar patterns of in- and out- trips in that most people travel between Financial Street and the area to the north of the street. Although the trips were spread out through the city, the Financial Street is more connected to the northern city than the southern counterpart through taxi. In example 2, we aggregated all the taxi trajectory data of Beijing in 2011 and derived a network at a global level that shows interactions among different city places (see Fig 2.4). This network provides an overview of to what extent any of the two areas within the city are connected through taxi and which parts of the city are more socially connected. Vibrant areas that are more connected than others, such as Zhongguancun business district, the financial street of the Xicheng District, CBD, and Wangjing area, can also be identified from the network.

Example 3 is represented using cellphone data from Chongqing. Cell phone data are aggregated at the county level (Example 3 in Table 2.2) to examine the connectedness of the city. Although similar county-level OD data can be derived from railway or bus records, they are limited to certain types of transportation modes. Moreover, the high cost makes this data less accessible to researchers. Estimating trip volumes from cell phone data provides an additional channel to examine travel patterns within the city and may also supplement to less-frequently collected transportation census data. Figure 2.5 (a) shows that counties adjacent to the main city have the maximum amount of trip volume (more than 40 thousand per day) and the highest share (more than 50%). This indicates that these counties have very close relationship with the main city. The county of Qijiang and Fulin also have significant amounts of trips related to the main city, but with a lower share (less than 35%). This contrast may result from the fact that these two counties are economically independent of each other. Counties in the northeast or southeast have very few trip flows to or from the main city because of long distance and inconvenience of travel.

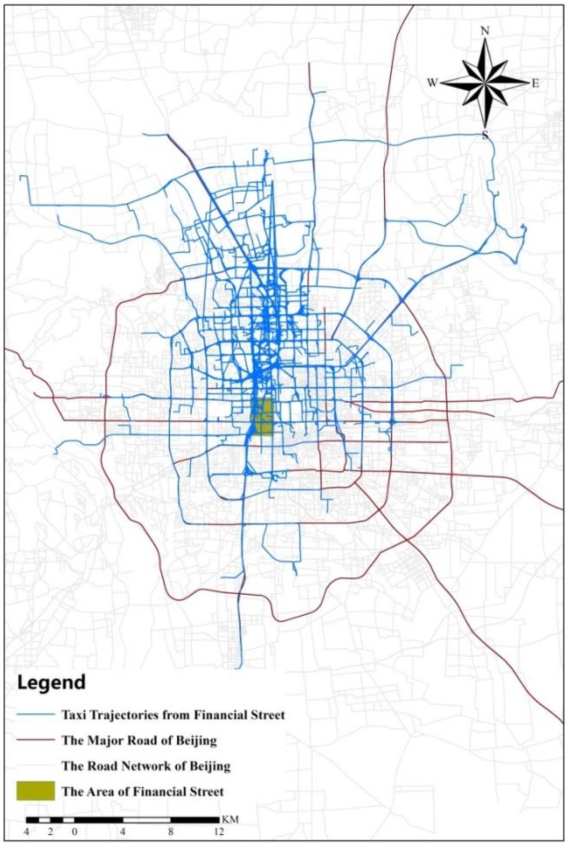
Figure 2.5 (b) (Example 3 in Table 2.2) takes each county's external trip flow volume (except these trips related to the main city) as an indicator to measure its regional dominance. High dominance means strong economic impacts, attraction to circumjacent counties, and more trips to or from external areas. The results indicate that in the northeast of Chongqing, Wanzhou showed a very significant regional dominance, and can be taken as the regional center. Yongchuan and Fulin, similarly, can be considered as the regional center on the west and east of the main city. However, in the southeast, there is no significant regional center, though Qianjiang has been defined as the regional center in the previous master plan. Jiangjin, Bishan, Hechuan have lost their regional dominance and no longer need to be defined as a regional center. Instead, they become integrated parts of the main city.

Table 2. 2 Examples of unit of analysis framework (spatio-temporal dynamic of urban trajectory)

		Examples		
		Example 1	**Example 2**	**Example3**
Distributions	**Spatio-temporal network**	A1+A5+A10	A4+A8+A12	A4+A8+A11



(a) Counts of trips going to Financial Street on Feb.27th, 2013



(b) Counts of trips leaving from Financial Street on Feb.27th, 2013

Fig 2. 3 Taxi trajectories in Financial Street area of Beijing from 7 am to 8 am Feb. 27th, 2013

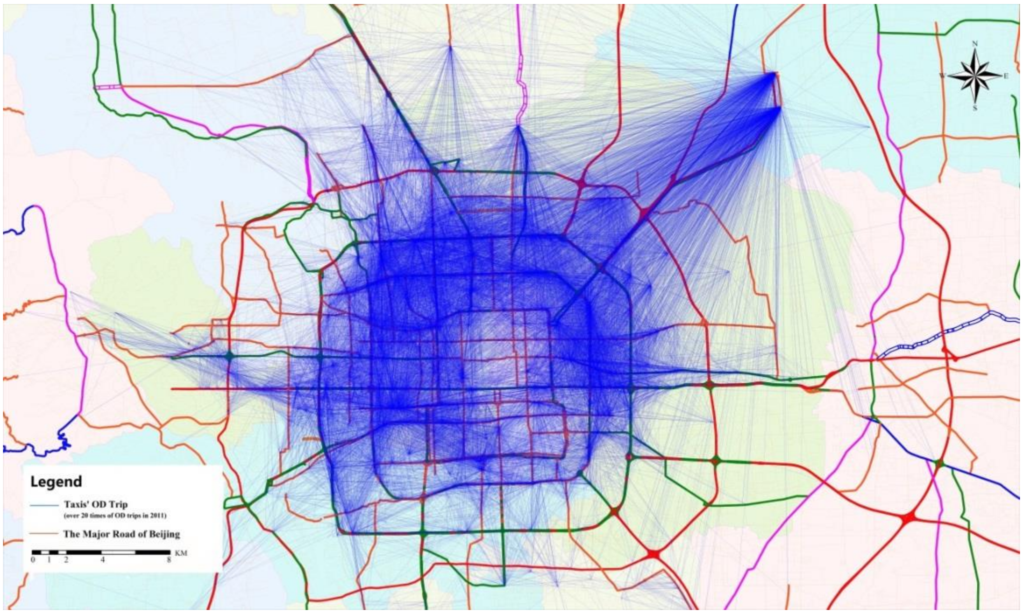
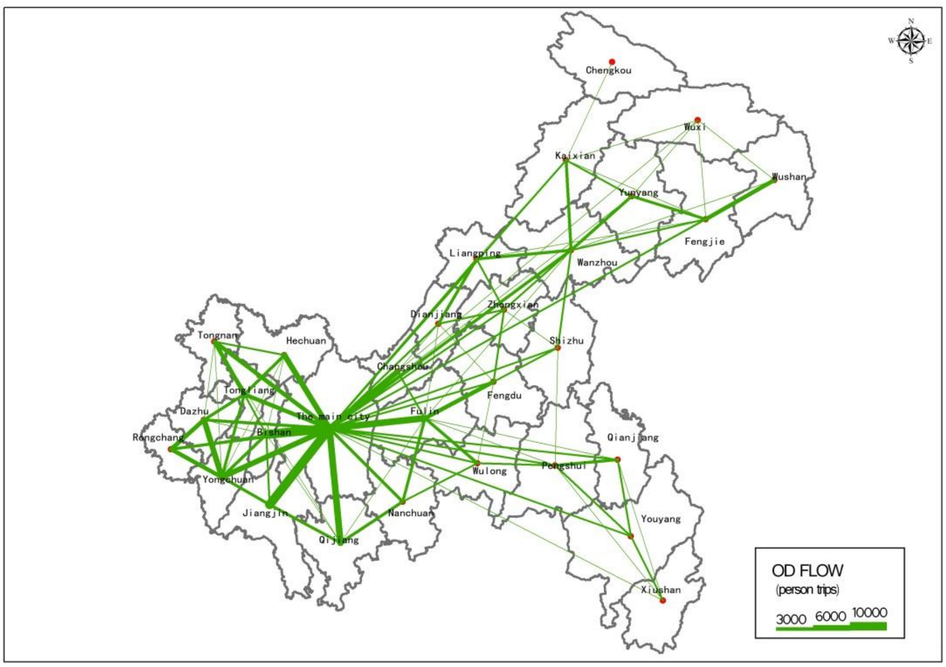
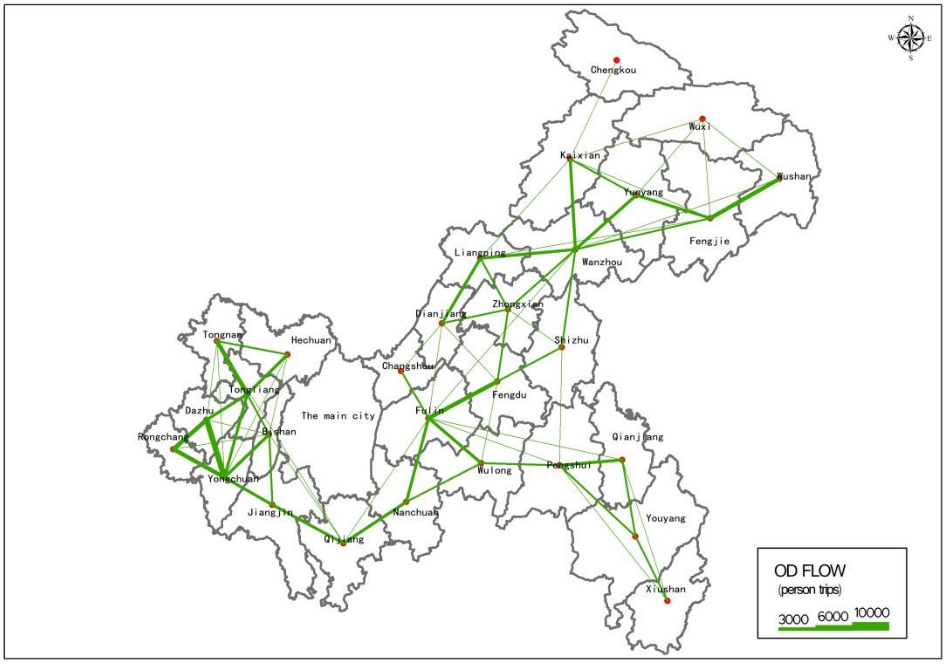


Fig 2. 4 Counts of OD trips within Beijing in 2011



(a) Trip flow distribution between the main city and suburban counties



(b) Trip flow distribution between suburban counties

Fig 2. 5 Trip flow distribution in Chongqing City within one day

As discussed above, cellular phone and taxi trajectory data can reflect human activity and thus form networks of human mobility patterns. In our case studies, we extracted human mobility patterns and incorporate them with the analysis of urban structure through an integrated spatio-temporal network analysis framework. The results suggest that social network of activities can help us understand urban functions (e.g. traffic demanding areas and domain regions) at different scales. Such an understanding can be used to help with better planning strategies.

2.5 Summary of This Chapter

With the fast growth of urban trajectory data, new opportunities are emerging for researchers to study urban form through real-time human activities. Nevertheless, how the dataset with unprecedented breadth and depth may facilitate the studies of human behavior and their spatial pattern and further support effective urban planning requires a new spatio-temporally explicit framework and methods that have not been fully addressed in current scholarships. This chapter aims to bridge this gap by suggesting a framework that considers different spatio-temporal and network scales through the notion of analysis units. By combining different analysis units, researchers can answer questions about urban space through the exploration and comparison of the spatio-temporal patterns of human activities and interactions across various scales and dimensions. The case studies indicate how such combination can help answering questions about spatio-temporal social network dimension in mixed scales of units. Specifically, the presented examples demonstrate the advantage of analysis unit in that it provides both flexibilities for analyzing data with different resolutions and comprehensiveness of understanding urban dynamics at various scales. Researchers can adopt similar approaches and develop their own combination of units according to the purpose of analysis.

Secondly, the proposed framework has its root in spatio-temporal analysis and network analysis. Researchers can combine methods from these two fast-growing domains and develop integrated approaches to address their research needs. Thirdly, the framework of multi-scale spatio-temporal network analysis enables access to a much wider thinking which addresses the role of dimensions and scales at different stages of urban dynamics for more in-depth study. In other words, the current work is mainly from an exploratory perspective, which can motivate urban scientists to design a series of tasks and formulate new hypotheses from theoretical and policy perspectives. This spatio-temporal work provides an important contribution to the current spatial science and urban studies literature, which lack frameworks of addressing integrated spatio-temporal network analysis. Although the proposed framework arose in the analysis of human activities and interactions, it can also be applied to a wide set of socioeconomic processes with geo-referenced data measured over time.

This chapter notes that the multi-scale and multi-dimension methods can expose some hidden patterns and trends that otherwise would be very difficult to detect. This research presents a general framework for pattern discovery and hypothesis exploration in urban trajectory datasets. On this basis, this framework and specific domain could benefit from each other in the following procedures: First, the analyst has the specific reason for investigating issues related to urban structure, which can be expressed as a general question or a set of general questions. Second, this nature of the investigation is checked against the task topology of the dataset. Third, the analyst carries out the matched tasks and detects something both interesting and relevant to this investigation. Fourth, new, more specific questions might appear, motivating the analyst to look for more details. These questions affect what details will be viewed and in what ways. Lastly, the general questions in step 1 are revised and the investigator goes through the procedures again. As such, explanations of various urban dynamics can be provided based on rigorous analysis, and policy interventions are then proposed in light of the understanding of the spatio-temporal-network dataset, which will open up a rich empirical context for social sciences (Yeet al. 2017).

Chapter3

3. Evaluating Urban Function Mix based on POIs and Taxi O-D Data

3.1 Introduction

The urban space has a significant functional structure, whereby the quantitative study and evaluation on this functional structure has always been one of the key issues of urban planning. However, as urban areas are relatively large and complex, it is costly to even get access to the limited available data, nevertheless, quantifying the degree of urban function mix has proven difficult.

Recently, several domestic studies in China have emerged to conduct a data-based quantitative research approach on urban functions. After obtaining the Sina Weibo points of interest (POIs) data and the check-in data through internet crawling technology, Long and Liu (2013) firstly managed to divide an entire research area into grid cells in consideration of the various land uses reflected by these data and the actual and planned land use data of Beijing. Then, borrowing the mixed land use index proposed by Frank et al. (2004), they were able to identify the land use mix of each grid cell. Yuan et al. (2012) evaluated urban functions of the traffic analysis zone (TAZ) by using the taxi trajectory and POIs data in Beijing, and they suggest combining the bus card data with the taxi trajectory data to achieve a more comprehensive evaluation on the urban functional structure. Long et al. (2013) identified the urban functional zones of Beijing by means of the bus card data and the POIs. Although all these studies attempt to reveal urban functions based on open source data and big data, or even a combination of both, the existing problems and shortcomings coming from them require being further addressed.

Previous studies mainly focus on the land use mix at a district level (a group of plots) in an urban area rather than at a plot level analyzing the mixing of retail stores, housing, catering and so on (Long and Liu, 2013). And the objective of previous studies based on big data and open source data is to identify or evaluate urban functional zones, yet lacking an in-depth quantitative exploration on each plot. In addition, the urban data are characterized by diversification and heterogeneity. Diversification of urban data here refers to the fact that data sources are different and researchers are organized different databases for research works. Heterogeneity refers to the fact that urban data consist of a lot of aspects, and it is difficult to thoroughly examine urban functions based on only one kind of data. Accordingly, the multivariate data can contribute to analyzing and working out complex urban issues, for example, analyzing on the urban functional structure are not only associated with POIs which can represent facility types, but also related to taxi O-D that are one kind of spatio-temporal big data reflecting transport activities.

This chapter, taking urban area of Beijing as an example, first of all conducts an analysis on the spatial distribution of major functions in 300m * 300m grid cells on the basis of the POIs. Then, relying on the idea of integrating multivariate data, the chapter comes up with a smaller grid space and initially realizes a more accurate identification and evaluation on urban function mix using both the taxi O-D which belongs to one type of spatio-temporal big data and the POIs data, or in other words, using the spatial entropy model set up based on the POIs data and the temporal entropy model developed based on the taxi O-D, which can be taken as a supporting tool to analyze the comprehensiveness of urban functions..

3.2 Data Sources

Taking Beijing as an example, this chapter is based on a multivariate data comprising of 649,359 POIs in total, obtained mainly via data mining techniques and the taxi O-D of 19,000 taxi in one week.

The data source - POIs

While maps are a basic way to describe urban form, which mainly consists of streets and buildings, the POIs data can describe the basic information of each building's function. Thus, the POIs data may, to a certain extent, show the distribution pattern of urban functional zones.

The data source – taxi GPS data

The mobile devices that are equipped with a GPS receiver chip can collect information about people, vehicles, and other mobile objects. The taxi GPS data used in this chapter is collected based on the currently widely-used floating car technology, by installing GPS devices in taxis as a sensor to track traffic conditions. According to the report from the Beijing Traffic Bureau, taxis account for 12 percent of Beijing's total ground traffic (Yuan, Zheng, et al., 2012), therefore the taxi datasets can well reflect people's activities and further reveal the urban functional structure

Taxi O-D identification and space matching based on the taxi GPS data

As the above data is unstructured, they are preprocessed for the research to deal with taxis' trajectory:

①Taxis carrying passengers can reflect people's purposeful activities and commute in a city, while empty taxis travel for the sake of passengers, which cannot represent urban traffic characteristics. As a consequence, the trajectories of taxis carrying passengers are obtained according to the traveling state.

②According to the trajectories of taxies carrying passengers, the data on starting points, arrival points, and time are acquired.

③The starting area, arrival area, and time (hours) are attained by plotting each point on the map, getting the area which the coordinates are in, and dividing time by hour.

As a result, this chapter basically achieves the identification of the taxi O-D based on taxi GPS database, and the corresponding time (T), with a total of 2,253,626 original (O) points and 2,253,551 destinations (D) points.

3.3 Urban Spatial Entropy and Temporal Tntropy

3.3.1 Spatial Entropy based on the POIs Data

Entropy is one of the most widely used metrics to measure the functional mix, which stems from Shannon's research on entropy and its widespread application in studies on biodiversity; Frank and Pivo (1994) are the first to apply entropy to urban planning and land use mix, though they are just preliminary studies on the mixture of retail stores, offices, recreation, public institutions, and other principal functions, as well as the interaction between single- and multi-family homes. This section will use the research method and extend it to the POIs data model.

Firstly, according to the basic principles, an entropy model is established based on different types of POIs data:

Assuming there are A POIs data in total, including N ($N = 14$) types, such as commercial, public services, and administrative institutions. If each type consists of A_1, A_2 ... A_n, then $A = A_1 + A_2 + ... + A_n = \sum_i Ai$ ($i = 1, 2 ... N$), and the probability can be defined as:

$$P_i = A_i/A = A_i/\sum_i^N Ai \quad (3.\ 1)$$

Obviously $\sum_i Pi = 1$, then the entropy of urban facility functions is:

$$H = -\sum_{i=1}^{N} Pi * logpi \quad (3.2)$$

Where, H ($H \geq 0$) is the entropy. The entropy can reflect the degree of urban function mix that is mixed urban functions of three commercial use (shops, convenience stores, and markets) and two public services (community service facilities and barbershops) in this chapter, the higher the entropy, the more various the facility functions are, and there will be fewer differences among different types of urban functions in terms of number. When $A_1 = A_2 = ... = A_n$, $P_1 = P_2 = \ldots = P_n = 1/N$, H reaches the maximum H_m, then,

$$H_m = logN \quad (3.3)$$

As can be seen, if $P_e = 1/N$, urban functions reach a state of equilibrium.

Since the entropy of urban functions has been obtained, this chapter will further explore the entropy of spatial form. It divides the urban space into grid cells: defining a rectangle which can just cover urbanized areas, with the scale r divided into $M * N$ grid cells, and then observe the spatial distribution pattern of different functional types. Providing that the total number of a certain type of POIs is A_k, and its number in the grid i row and j column is A_{ij}, then

$$P_{ij} = A_{ij}/A_K \quad (3.4)$$

Apparently $\sum_i \sum_j Pij = 1$, thereby the spatial entropy of urban functions is:

$$H_s = -\sum_i^M \sum_j^N Pij * logpij \quad (3.5)$$

Where, H_s is the spatial entropy. By dividing the urban space into grid cells (using such software as Arcgis), it is not difficult to get P_{ij} value based on different types of POIs data, so it is easy to obtain the value of H_s. Consequently, we can get the degree of urban function mix at different urban areas.

3.3.2 Temporal Entropy based on POIs and Taxi OD Data

The taxi O-D also belongs to one type of spatio-temporal big data, based on the taxi

O-D data in section 3.2 and the model for calculating spatial entropy in section 3.3.1, a calculation model of temporal entropy is established: supposing that there are *A* taxi O or D points in total, time is accurate to an hour in order to exclude particular time periods (e.g., morning peak and evening peak) during which vehicles are too concentrated to impact the results, thus N ($N = 7 * 24$) time periods are obtained depending on the attribute of time; then based on the formulas in section 3.3.1, this chapter acquires the destination (D) and origin (O) point data of each grid cell in different periods within one week.

Accordingly, it gets the temporal entropy model based on the taxi O-D T data for the purpose of calculating the frequency of occurrence of a taxi in a certain area. A higher temporal entropy indicates that taxis are more frequently appear in a certain area at different times; and a lower temporal entropy designates that taxis are more frequently appear only in the certain period of time a certain area; and a higher O-D number and a higher temporal entropy indicate an area with 24-hour vitality.

According to the characteristics of the taxi GPS data described in section 3.2, the spatial and temporal distribution pattern of the taxi O-D can, to some extent, reveal people’s behavior, based on which it can be concluded that the larger number of O-D points and a more evenly distribution in a certain area at different times signify the higher degree of urban function mix in this area, because only when there are many urban functions in a certain area, it can attract people who come to here at different times. This is a good supplement for the spatial entropy based on the POIs data in section 3.3.1.

3.4 Geospatial Analysis on Urban Area based on the Spatial and Temporal Entropy

3.4.1 Urban Functional Distribution Patterns in Beijing

3.4.1.1 Centralization, Decentralization, and Hierarchies

Based on the POIs data, the different functional distribution patterns were examined respectively (see Fig 3.1) which is a starting point to study on the degree of urban function mix and a guarantee for the rationality to apply the spatial entropy and temporal entropy. First of all, based on their names, the POIs can be divided into different types, such as commerce, public services, administrative agencies, parking lots, catering, recreation, primary and middle schools, cultural facilities, and business offices. Secondly, the 300m * 300m grid cells are used for the purpose of calculating the number of different types of POIs in each grid cell of Beijing, based on which the density of a certain type of POIs is obtained. Finally, each grid cell is given different color according to its density, among which red and blue represent high and low density respectively, thus approximately showing the spatial distribution pattern of a certain function (see Fig 3.1).

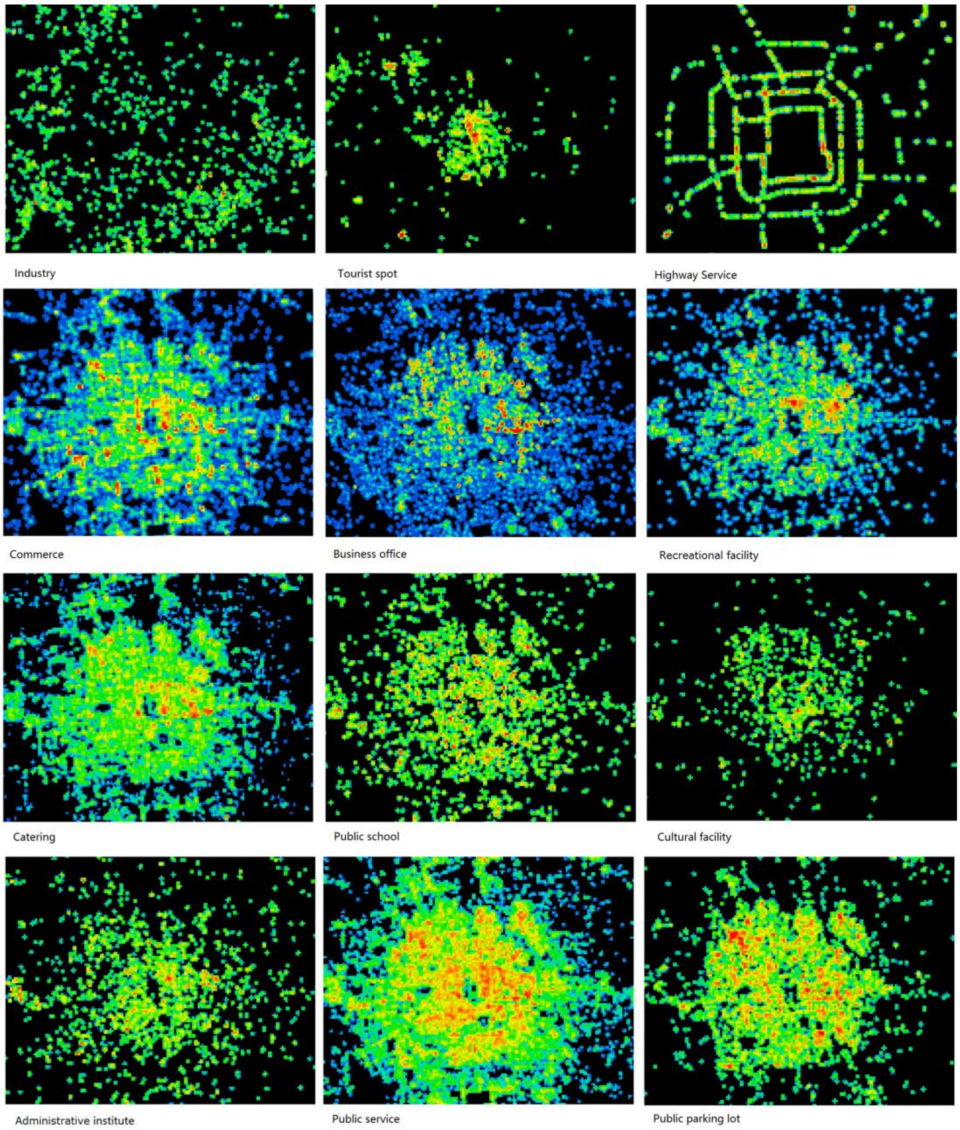


Fig 3. 1 Urban functional distribution patterns in Beijing in Beijing

Most functional facilities, except a small number of outliers like industry, tourist spot and highway service, primarily spread out from the center to the edge, which in general demonstrates a functional agglomeration effect. However, the degree to which the functional agglomeration varies with functional type. For example, commerce, business office, entertainment, and catering are more heterogeneously distributed, as they highly concentrate on a small number of places (denoted by red cells) in the central districts; and meanwhile, public facility, public school, public institution, cultural facility and public parking are scattered in a relatively homogeneous way, as the highly concentrated places (represented by red cells) are relatively smaller or more evenly distributed. It can be suggested that the profit-oriented facilities, such as the commercial, are benefited from the spatial agglomeration effect, but the public-service-oriented facilities give more emphasis on the spatial equity status of distribution.

More subtle characteristics can be found with respect to the spatial distribution of the profit-oriented and public-service-oriented facilities discussed in the last paragraph. For the profit-oriented facilities, the commercial, entertainment and catering more concentrate on the inner city of Beijing and forming larger active centers at the city level; and in particular, the various extent to which each type of functional agglomeration can be identified around the center of each larger active center, and for example, the density of each type of facility gradually and smoothly decreases from the center to the edge. By contrast, business office does not highly aggregate within the inner city, but focusing on the CBD (to the east of the inner city), the embassy district (to the northeast), Wangjing (to the northeast as well), and Zhongguancun (a high-tech district to the northwest); and for each business office area, the density of business office sharply drops down within its surroundings and this suggests a relatively clear boundary for those business office areas.

For the public-service-oriented facilities, public facility and public parking in general have higher density, and more concentrate within the district bounded by the 4th Ring Road (although the north part has higher density than the south part); public institution and cultural facility obviously have relatively lower density and more focusing on the smaller district surrounded by the 3rd Ring Road; and the spatial distribution of public schools are the most homogenous. Since public facilities and public parking are partly involved in some commercial development in Beijing but public schools are fully funded by the government, it implies that even for the public-service-oriented facilities, those partly relating to the profit seeking need to take account of the spatial aggregation effect.

3.4.1.2 Spatial mixture of urban functions

To get a more detailed analysis of the mixed urban functions, this chapter further studies three elements of commerce (shops, convenience stores, and markets) and two elements of public services (community service facilities and barbershops). Although Figure 3.1 clearly shows that the density of main urban functions is slowly decreasing from the city center to the periphery, those correlations are not strong, except the correlation between catering and haircut with the R-square above 0.5. (see Table 3.1). To a large extent, this demonstrates that the aggregation patterns or the dispersion patterns of those functions are not the same and this might affect the way of mixing up those functions.

However, catering and haircut seem to relate to some types of functions, if the R-square above 0.3 can be considered of as the kind of threshold beyond which the spatial distribution patterns of any two types can be treated to have some kind of relationship. For example, catering has a relationship with those functions like entertainment, public parking, public facility, commerce, convenience store, small & medium shop and community service; and some of them are the public-service-oriented facilities. Haircut relates to some functions like a convenience store, entertainment, public parking and community service; and some of them are the profit-oriented facilities. Perhaps this suggests that catering and haircut function as the kind of glue item that promotes mixed-use and functional diversity in Beijing.

	Commerce	Public services	Administrative agencies	Parking lots	Catering	Recreation	Primary and middle schools	Cultural facilities	Business office	Convenience stores	Shops	Trading markets	Community services	Barber shops
Commerce	1.000													
Public services	0.392	1.000												
Administrative agencies	0.088	0.114	1.000											
Parking lots	0.216	0.316	0.085	1.000										
Catering	0.393	0.404	0.145	0.449	1.000									
Recreation	0.194	0.227	0.130	0.226	0.457	1.000								
Primary and middle school	0.116	0.176	0.181	0.154	0.219	0.162	1.000							
Cultural facilities	0.092	0.090	0.062	0.134	0.162	0.093	0.072	1.000						

Business office	0.114	0.174	0.038	0.266	0.289	0.100	0.077	0.071	1.000					
Convenience stores	*	0.262	0.157	0.181	0.339	0.226	0.212	0.029	0.072	1.000				
Shops	*	0.177	0.050	0.196	0.383	0.158	0.069	0.101	0.134	0.135	1.000			
Trading markets	*	0.062	0.026	0.049	0.077	0.045	0.036	0.012	0.021	0.124	0.069	1.000		
Community services	0.280	*	0.228	0.229	0.352	0.249	0.237	0.094	0.138	0.358	0.146	0.125	1.000	
Barbershops	0.163	*	0.148	0.303	0.540	0.305	0.220	0.102	0.220	0.339	0.256	0.075	0.388	1.000

Table 3. 1 The R-square of the correlation between pair of functions
Note: light grey denotes higher values, and dark grey represents moderate values, and * means no correlation test was processed.

In addition, there is also the kind of relationship between commerce and public facility, convenience store and community service, public parking and public facility, respectively. To some extent, this implies a weak interaction between some profit-oriented and public-service-oriented facilities, although the correlation is not strong. However, the co-presence of those two types of facilities in one place will facilitate urban vitality.

3.4.2 The Degree of Urban Function Mix in Beijing

3.4.2.1 Spatial and Temporal Correlation of Urban Functions

According to the algorithm in section 3.3.1, this chapter gets the origin area (O point), the destination area (D point), and the corresponding time (T) based on the taxi GPS data, in accordance with which an OD (origin-destination) traffic network can be created. The network denotes not only the traffic flow between any two areas of the city, but also the connection between different functions.

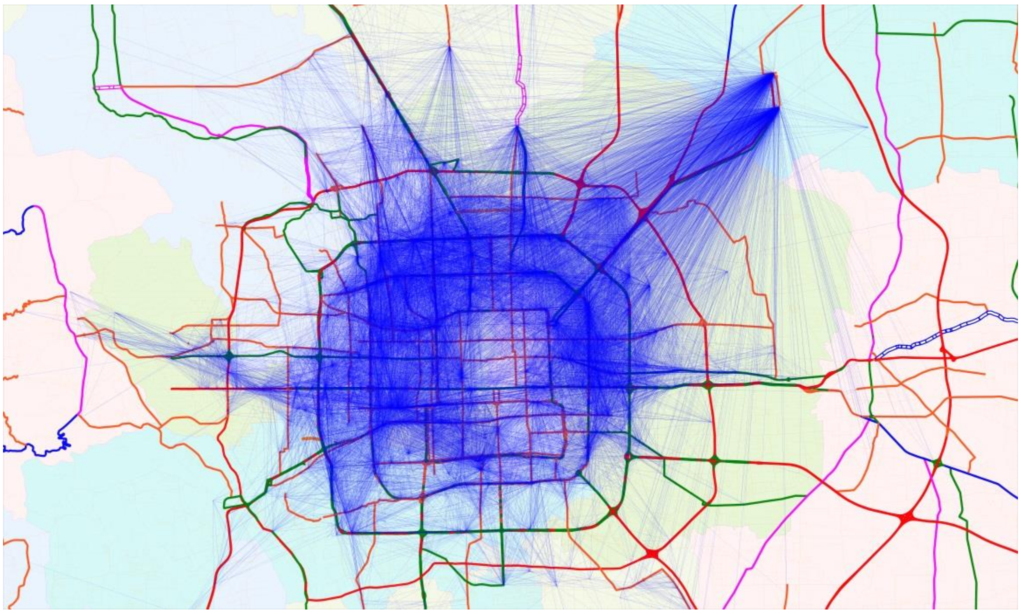


Fig 3. 2 Functional link network based on the taxi O-D in Beijing

Figure 3.2 is a functional link network based on the spatialization of taxi O-D within a week, from which we can distinguish the Zhongguancun business district, the financial street of the Xicheng District, CBD, Wangjing area, and other areas. These are vibrant areas closely linked with other areas. Then, are these areas of mixed functions?

The next section will focus on mixed urban functions based on the entropy model and the POIs and taxi O-D. First of all, from the viewpoint of the entire research area, it identifies the spatial distribution characteristics of mixed urban functions; then it further analyzes the spatial relationship between the spatial entropy based on POIs and the temporal entropy based on the taxi O-D data; and finally it studies Beijing’s traditional core areas including the Old Dongcheng District, the Old Xicheng District, the area within the Second Ring Road, CBD, and Zhongguancun area as well as the emerging representative functional areas, namely Wangjing, Huilongguan, the Airport area, etc., proving the complementarity of the two entropies in identifying the degree of urban function mix and finding some differentiation law.

3.4.2.2 Spatial Distribution Characteristics of the Degree of Urban Function Mix

According to the algorithm in section 3.3.1, the POIs-spatial entropy can be obtained to reveal the degree of urban function mix within a certain area. As shown in Figure 3.3, the general characteristics of Beijing’s POIs-spatial entropy is marked as high as 0.86 – 1.09 in the area within the3rd Ring Road; between 0.71 – 0.85 in the area between the 3rd Ring Road to the 4th Ring Road; lower to 0.62 outside the 4th Ring Road; and 0.86 – 0.93 in some satellite town areas, such as Tongzhou, Daxing, Fangshan, and Shunyi.

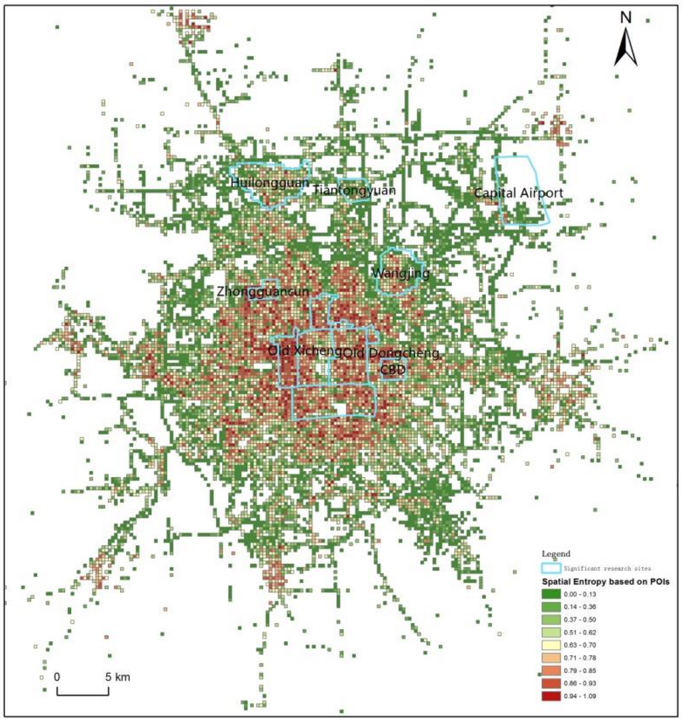


Fig 3. 3 Distribution of the POIs-spatial entropy of Beijing

The key sites are featured by the Old Xicheng District with the highest POIs-spatial entropy of more than 0.94; 0.71 – 0.93 in most of the CBD, Zhongguancun, and Wangjing areas; and below 0.78 in most of Huilongguan and Tiantongyuan. In addition, the spatial entropy of T3 Terminal in Beijing Capital International Airport can reach 0.71, for most other places around the airport it is less than 0.51.

The taxi O-D temporal entropy can be obtained in accordance with the algorithm in section 3.2. As shown in Figure 3.4, the temporal entropy of the taxi O points mainly gather within the 4th Ring Road and decrease progressively from the center to the surrounding areas. In detail, the temporal entropy of the starting points within the 3rd Ring Road is as high as 2.0; 1.5 – 2.0 in the area between the 3rd and 4th Ring Road; lower than 1.3 in most areas outside the 4th Ring Road; and even less than 1.1 outside the 5th Ring Road. It shows that the areas that are closer to the city center are in similar demands for taxis in different time periods.

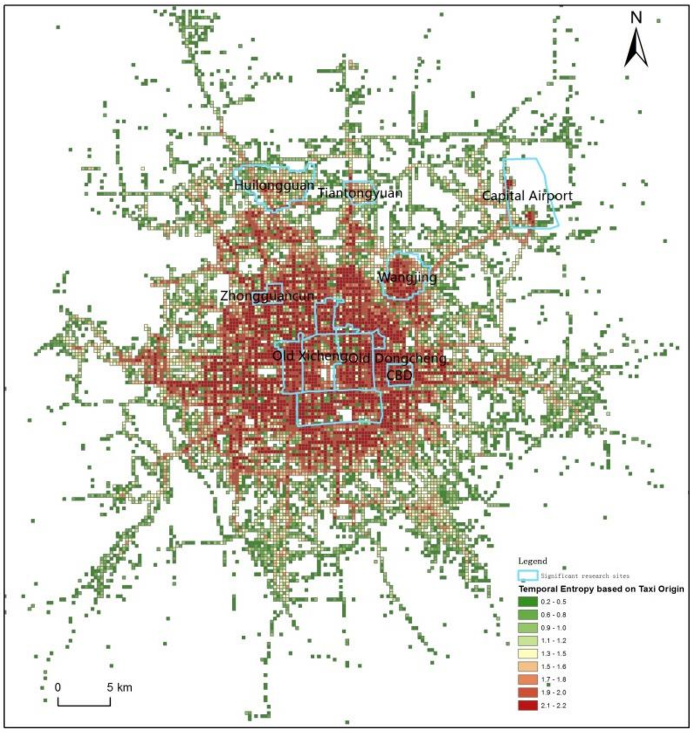


Fig 3. 4 Distribution of the taxi-Os temporal entropy in Beijing

In the key areas, the taxi-Os temporal entropy of the CBD, which is located between the 2ndto 3rd Ring Road is the highest (2.1 – 2.2). The temporal entropy of Zhongguancun that is between the 3rd and 4th Ring Roads is relatively higher (1.5 – 2.0). Although the Airport and Wangjing areas are outside the 4th Ring Road, their entropy can reach as high as 1.9 owing to the gathering of the taxi-Os. The entropy of Huilongguan and Tiantongyuan communities which are located outside the 5th Ring Road is around 0.9 – 1.6, significantly higher than the surrounding area, yet less than that of the area within the 3rd Ring Road.

This is also the case with the taxi-Ds temporal entropy which mainly concentrates within the 5th Ring Road and decreases gradually from the urban center to the outskirts, though it is slightly lower than the taxi-Os temporal entropy (see Fig 3.5). In detail, the entropy can be as high as 1.9 – 2.0 within the 3rd Ring Road, about 1.5 – 1.8 in the area between the 3rd and the 4th Ring Roads, 1.6 outside the 4th Ring Road, and lower than 1.2 outside the 5th Ring Road. It is also confirmed that there is a relatively small fluctuation on the taxi frequency of occurrence in the urban center over time.

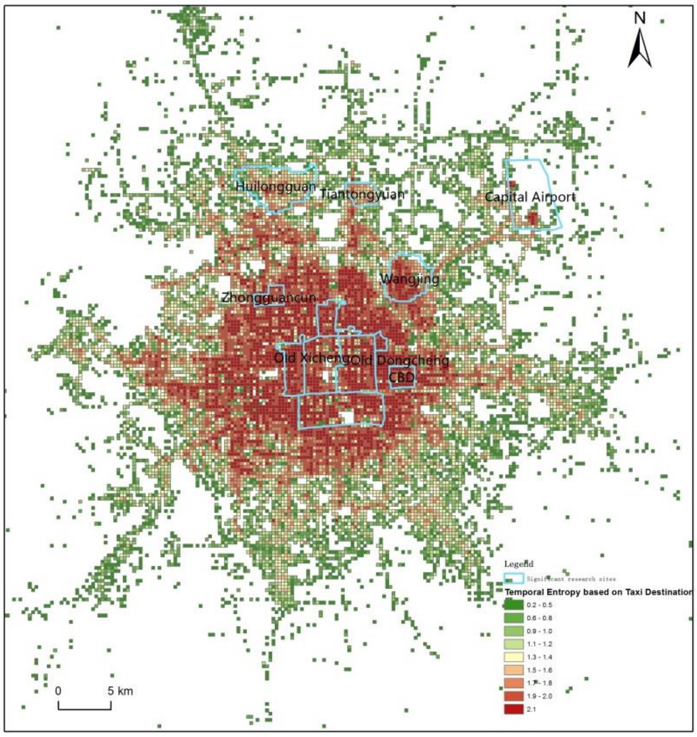


Fig 3. 5 Distribution of the taxi-Ds temporal entropy in Beijing

The taxi-Ds temporal entropy of the key sites is marked over 1.9 – 2.0 in the Old Dongcheng and Xicheng Districts, and in the CBD; 1.7 – 2.0 in Wangjing and Zhongguancun; 1.1 – 1.5 in Huilongguan and Tiantongyuan; generally lower than 1.1 in the Capital Airport area except for the Terminal whose entropy can reach 1.7 – 1.9.

3.4.2.3 Correlation between the POIs-spatial Entropy and the Taxi O-D Temporal Entropy

As to whether there is a correlation between POIs-spatial entropy and the taxi O-D temporal entropy, this chapter firstly gets POIs-spatial entropy and the temporal entropy of the taxi-Os, taxi-Ds, and then makes a table of three types of entropy by sampling. Accordingly, it figures out their correlation between any two entropies as shown in Table 3.2: the correlation coefficient between the taxi-Os and the taxi-Ds is 0.627, a higher degree of correlation; in addition, the correlation coefficients between POIs-spatial entropy and taxi-Os and taxi-Ds temporal entropy are 0.384 and 0.446 respectively, an ordinary degree of correlation.

Furthermore, the POIs-spatial entropy is taken as the dependent variable to do regression analysis on the taxi-Os and taxi-Ds temporal entropy. As shown in Table 3.3, the test values of both the POIs-spatial entropy and the taxi-Os, the taxi-Ds temporal entropy are less than 0.005, proving their significant correlation. This not only explains the reason why their entropy has similar spatial distribution characteristics, but also to some extent proves that it is reasonable to use the two kinds of entropies to assess the degree of urban function mix.

3.4.2.4 Correlation between the POIs-spatial Entropy and the Taxi O-D Temporal Entropy in the Key Sites

The following is the results of the analysis on the correlation between the POIs-spatial entropy and the average entropy of the taxi O-D (see Table 3.2-3.4) and the scattered plot analysis of the key sites (see Figs 3.6 – 3.8):

Table 3. 2 Correlation between the POIs-spatial entropy and the taxi O-Dtemporal entropy

Index	POIs[a]	TAXI-O[b]	TAXI-D[c]
POIs	1	0.384**	0.446**
TAXI-O	0.384**	1	0.627**
TAXI-D	0.446**	0.627**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 3. 3 Regression analysis on the POIs-spatial entropy and the taxi O-D temporal entropy

Index	Standardized Coefficients Beta	t	Sig.
TAXI-O[b]	0.384	45.651	0.000
TAXI-D[c]	0.446	59.22	0.000

Note：a. Dependent Variable: POIs[a]

Table 3. 4 Correlation between the POIs-spatial entropy and taxi-Os temporal entropy and taxi-Ds temporal entropy in key sites

Index	POIs[a]	TAXI-O[b]	TAXI-D[c]
POIs	1	0.915**	0.925**
TAXI-O	0.915**	1	0.996**
TAXI-D	0.925**	0.996**	1

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

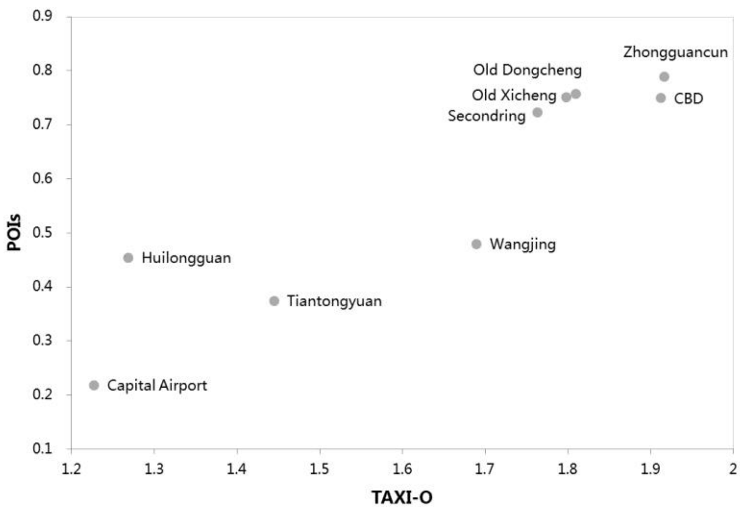


Fig 3. 6 Correlation between the POIs-spatial entropy and the taxi-Os temporal entropy in key sites

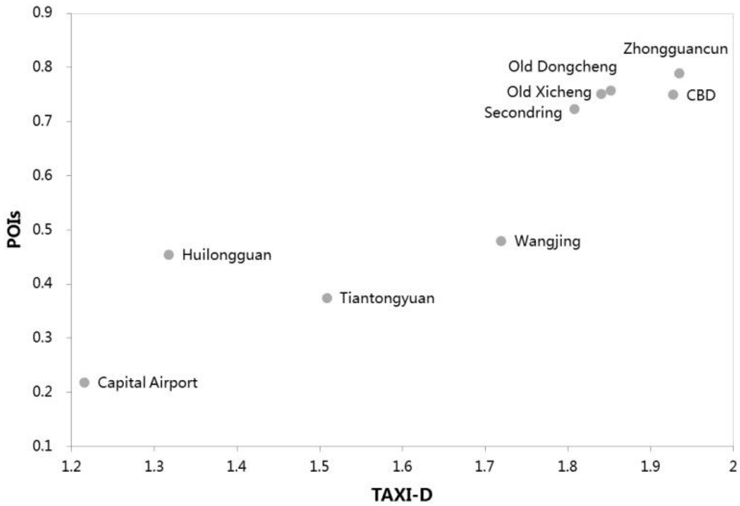


Fig 3. 7 Correlation between the POIs-spatial entropy and the taxi-Ds temporal entropy in key sites

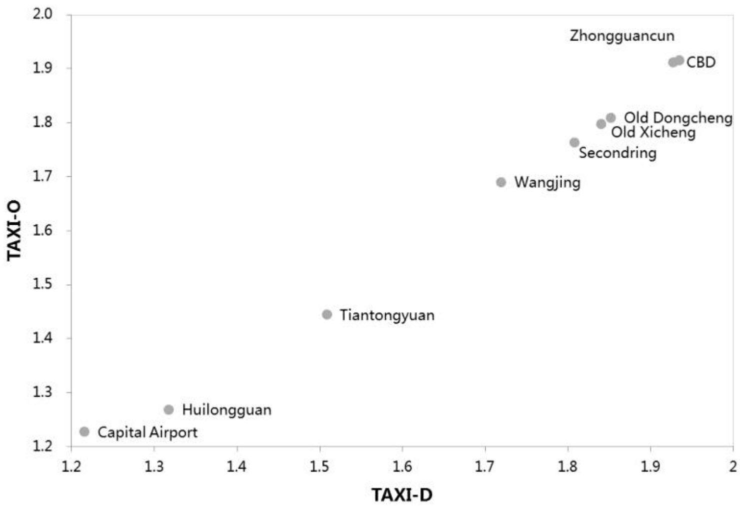


Fig 3. 8 Correlation between the taxi-Os temporal entropy and the taxi-Ds temporal entropy in key sites

①There is a strong correlation between the POIs-spatial entropy and the taxi O-D temporal entropy where the Pearson correlation coefficients are 0.915 and 0.925 respectively for the correlation between the POIs-spatial entropy and the taxi-Os temporal entropy and that between the POIs-spatial entropy and the taxi-Ds temporal entropy. In general, a higher POIs-spatial entropy indicates a higher taxi-Os temporal entropy and a higher taxi-Ds temporal entropy.

②Both the POIs-spatial entropy and the taxi O-D temporal entropy are high in Zhongguancun, CBD, the Old Dongcheng District, the Old Xicheng District, and the area within the Second Ring Road, which indicates these areas with a high degree of urban function mix, relatively mature urban amenities and functional development, concentrated residences, business offices, and recreation, high consumption, and high frequency of taking taxis. The middle-level POIs-spatial entropy, but a high taxi O-D temporal entropy in Wangjing demonstrate that it is a relatively new development urban area, focusing on residential and business office functions, and most of which are foreign companies, of which staff members have relatively high consumer purchasing power. In addition, Wangjing is a large area with only one subway station, lacking connectivity, far away from the city center and being of a relatively high taxi frequency of occurrence. Although Huilongguan and Tiantongyuan are mainly constituted by the residential function with a large population and relatively abundant community service facilities and recreational facilities, its average POIs-spatial entropy is at a middle level and its taxi O-D temporal entropy is low. The key reason is that most residents in Huilongguan and Tiantongyuan communities are ordinary workers who often take bus and subway, of which residents in Huilongguan usually take subway line 13 to the Shangdi business area and to the Xizhimen transfer station; it is convenient for residents living in Huilongguan to take subway line 5 to southern Beijing. The low POIs-spatial entropy and low taxi O-D temporal entropy in the Capital Airport is as a result of its single function, comparatively mature and frequent airport bus and airport express system, and people who take a taxi only concentrated in the Terminal area.

③With the Pearson correlation coefficient being up to 0.996, the Taxi-Os temporal entropy and the taxi-Ds temporal entropy are of the strongest correlation. There is a high taxi O-D temporal entropy in Zhongguancun, CBD, Old Dongcheng District, Old Xicheng District, the area within the 2nd Ring Road; the middle level taxi O-D temporal entropy in Wangjing; and low taxi O-D temporal entropy in Airport and Huilongguan.

3.5 Summary of This Chapter

Since big data creates good opportunities for a finer quantitative analysis and research on urban issues, some scholars began to evaluate the degree of urban function mix at a micro-scale, yet their studies are usually based on a relatively singular data structure. Based on these concerns, this chapter conducts a case study on Beijing in accordance with the entropy model, the 300m * 300m grid cell, the POIs data and taxi O-D, and the corresponding POIs-spatial entropy and the taxi O-D temporal entropy, to identify the degree of urban function mix, for the purpose of largely improving the accuracy and reasonableness in identification of mixed urban functions.

Firstly, based on the POIs data, this chapter analyzes the urban functional distribution patterns in the 300m * 300m grid cells and comes to the conclusion that different types of functions form different aggregation or discrete modes. In general, the profit-oriented functions are more likely to aggregate in a heterogeneous way, of which the large-scale functions tend to be heterogeneous, while the small-scale functions are more likely to be more homogenous and hierarchical; public-service-oriented functions incline to present homogeneous and discrete pattern, of which the partial profit-oriented functions are gathered in a hierarchical and heterogeneous way and in small scales. However, the interaction of the profit-oriented and public-service-oriented functions contributes to the functional mix. In addition, catering plays an important role in promoting the mixing and the diversification of urban functions; and some the profit-oriented and public-service-oriented functions in the same place can also increase the degree of urban function mix in Beijing.

Secondly, based on the issues mentioned here, and in combination with the taxi O-D of one week, we can directly identify the Zhongguancun, the financial street of the Xicheng District, CBD, Wangjing, etc. where are higher the degree of urban function mix than other places.

Finally, an analysis of urban function mix is conducted based on the POIs-spatial entropy and the taxi O-D temporal entropy. In general, the POIs-spatial entropy is of significant correlation with both the taxi-Os temporal entropy and the taxi-Ds temporal entropy, proving that it is reasonable to assess the degree of urban function mix by the two kinds of entropy. Beijing’s the degree of urban function mix are featured by gradually decreasing from the 3rd and 4th Ring Roads out to the periphery, among which traditional core areas including the Old Dongcheng District, Old Xicheng District, the area within the 2nd Ring Road, and CBD with the highest mixing degree; followed by Zhongguancun and Wangjing; then Tiantongyuan and Huilongguan communities with the lowest mixing degree because of their dominant residential functions. In terms of the key research sites, the POIs-spatial entropy is highly correlated with the taxi O-D temporal entropy, the traditional core areas such as Zhongguancun, CBD, the Old Dongcheng District, the Old Xicheng District, and the areas within the 2nd Ring Road are marked with the highest mixing degree and mature urban functions; Wangjing also holds a high mixing degree and strong consumer purchasing power, while its urban function needs to be further improved; Tiantongyuan and Huilongguan communities are represented by a lower mixing degree because of their largely residential function and attention should be paid to improving supporting facilities and other sectors for their future development; Although the Airport Terminal is closely associated with the taxi O-D, the Airport area has only a single functional structure and poor mixing degree.

The analysis conclusions represent the actual situation in Beijing, confirming the accuracy and reasonableness of the quantitative analysis on the degree of urban function mix that is conducted based on an entropy model built on the big data of POIs and taxi O-D. The research method can be applied in other areas and support for enhancing urban comprehensive capabilities in urban planning.

Chapter4

4. Uncovering the Relationship between Spatio-Temporal Distribution of Population and Urban Function with Location-Based Service Data

4.1 Introduction

With the increasingly sprawled urban boundaries and social networks, individuals have more opportunities to take part in social activities within an urban area. As a result, populations in different inner city areas change with time in accordance with different rules. For example, only high-end services, such as finance, business, trade, exhibitions, and foreign affairs, are located in the CBD of Beijing, which embraced a vibrate daytime, yet a silent nighttime. In contrast, a large number of people leave the mega-community of Tiantongyuan for the city center in the morning and return in the evening. Such variations reflect the dynamics of an urban space, and the variations in population distribution over time reflect the activities of urban residents behind static physical structures. Daily activities of residents are closely related to urban space and are determined by different urban functions. Therefore, the knowledge of social behaviors within a city and their interaction with physical space is crucial to understanding the urban environment.

Many research papers address urban space, which is a constant topic in urban geography. Traditional research mainly focuses on static urban space and is based on population census data or questionnaire surveys at the community or sub-district level. Time geography introduced time to study dynamic urban space using travel log surveys of residents. For example, Zhou (Suhong Z., 2015) proposed the concept of "urban rhythm" and studied the dynamic spatial structure of the population in Guangzhou with a twenty-four-hour period using travel log surveys. The traditional data sources (questionnaire, survey etc.) have advantages in collecting abundant information about profession, income, age, residents' willingness, etc., precisely targeting specific groups such as the disabled or the elderly and explaining further reasons. However, when it comes to intensive urban management, knowing urban running states quickly and comprehensively is becoming a necessity. Traditional data sources may not be sufficient due to their small sample size and high time, financial, and human costs of data collection.

Big data, comparatively, has advantages in its high volume, variety, and celerity. Many people thus use this new data source to discover perspectives on spatio-temporal dynamics of cities. Different types of big data, ranging in a wide spectrum from the logs of handheld GPS devices, to mobile phone data (MPD) (Kitchin R., 2014; Kang C. et al., 2012; Xia M. et al., 2010), to smart card data (SCD) (Wang et al., 2016), to GPS floating-car (taxi) data, to Weibo, to social media check-ins have been used in the studies. In 2010, Xia Mao (Xia M. et al., 2010) studied population distribution dynamics of Shenzhen at day and hour time scales separately using data from mobile base stations at a spatial resolution of 1 km. This is the earliest paper examining population distribution using big data in China, but it is only descriptive and does not explore the underlying population dynamics and their interactions with urban functions. Similarly, Pierre Deville's research on mapping population dynamics using mobile phone data (Deville P. et al., 2014) demonstrates the possibility of dynamically examining urban space in a timely and precise way.

Moreover, many papers attempt to extend the study of dynamic urban space to job-housing commuting (Long Y. et al., 2015; Ning X. et al., 2014; Long Y. et al., 2012), transportation structure (Liu L. et al., 2009; Steenbruggen J. et al., 2013), urban spatial structure (Xinyi N. et al.,2014; Louail T. et al., 2014), and urban function zoning (Bo W. et al.,2015; Pei T. et al., 2014; Y L. et al., 2014; Yuan J. et al., 2012). For example, Niu

Xinyi (Xinyi N. et al., 2014) assumes that changes in the population distribution at certain hours, such as 10:00 and 23:00 on weekdays, represents changes in activity status from working to staying at home. He uses this assumption to identifying the city center and urban function zoning. Thomas Louail (Louail T. et al., 2014) also uses mobile phone data to study the spatial structure of cities, yet he examines dynamics at a very large scale, which is insufficient for precise detection of urban space. In terms of urban function zoning, the most recent study in China concerns discovering urban functional zones based on human activity using latent activity trajectories (Yuan N.J. et al., 2015). A topic modeling-based approach is developed to cluster segmented regions into functional zones to leverage mobility and location semantics mined from LAT and to identify the intensity of each functional zone using kernel density estimation. By measuring mobile phone data from a 500 m by 500 m ‘pixel’ grid in Rome (Italy), Reades et al. (Reades J. et al., 2009) employed a variant of principal component analysis to cluster the pixels into regions with similar usage patterns and further made a qualitative link between the identified spatial patterns and the number of businesses in the corresponding areas. Mobile phone data used in these examples cover the broadest range of different ages and income groups compared to other big data. For example, smart card records mostly represent the low-income group whereas floating-car data is not representative of populations with lowest or highest incomes. However, mobile phone data are insufficient in terms of precision or scale.

In summary, traditional studies about urban space mainly focus on static urban space, and thus are inefficient in capturing the dynamics of urban space. Studies about dynamic urban space using big data mainly focus on spatio-temporal population distribution from single aspects such as job-housing commuting, urban spatial structure, urban function zoning, etc. Basically, those studies are conducted in a relatively narrow scale and limited research scope due to the difficulty in data acquisition. Besides, the relationship between urban functions and human vitality is still missing, which is important in terms of knowing the demand for people in different urban functions and the current situation of supply that urban space provides.

How do human beings move within a city? What kind of temporal characteristics exist in the mobility within a city? Is the urban function more mixed the more vital the urban space is? Are there preferences that human beings have that attract them to different urban functions? In this chapter, we try to reveal the relationship between spatio-temporal distribution and urban functions using an unprecedented high-resolution and broad-coverage-crowd LBS data from Tencent, one of the biggest internet companies in China, recording population with an hour as a time scale. We then discuss how our work can help understand the dynamic nature of our city and how our work can help urban planning decision-making.

4.2 Data Source

Using Beijing as a case study, this chapter is based on a multivariate dataset composed of POIs obtained via data-mining techniques and LBS spatio-temporal data from Tencent.

The data source – POIs

While maps, mainly consisting of streets and buildings, are basic to describe the form of urban areas, the POIs data can describe the basic function of each building. Therefore, the POIs may, to a certain extent, reflect the distribution pattern of urban functional zones. The POIs used in this chapter are obtained through Place API provided by Baidu, which is a free HTTP interface. Users could get POI data in searched areas through programming language like Python, etc. The POIs data is in json format and contains information about name, type, longitude, and latitude of the POI. Through the attribute of longitude and latitude, it is easy to locate the POI on the map. The POIs dataset used in this chapter represents the existing POIs in 2014, and its number is up to 649,359.

The data source – Spatio-Temporal LBS data from Tencent

LBS spatio-temporal data from Tencent is obtained by large-scale crawler technology through the Internet. The data is produced through one of Tencent's social networking applications—Tencent QQ—including both its mobile and desktop applications. Therefore, the data mainly represents Tencent QQ users, who are primarily young and middle-aged people. Due to different client preferences, it is almost impossible to expect a single type of social media data to cover age and income levels, the LBS spatio-temporal data from Tencent achieves an absolute advantage over other available big data, such as taxi GPS data, bus card data, questionnaire data, etc.

The LBS spatio-temporal data from Tencent are stored as points, with each point being the centroid of a 25m by 25m grid and representing the population within the grid. The data are collected every hour, which means that data collection occurs 24 times a day. Approximately 1.2 million points are collected every hour. In this chapter, the dataset is mainly distributed within the Sixth Ring Road (Fig 4.1) and was collected from July 29 to August 2 in 2015. In particular, July 29 and 30 are weekdays, and August 1 and 2 are weekend days. This research is based on the average hourly population on the weekdays and weekend days.

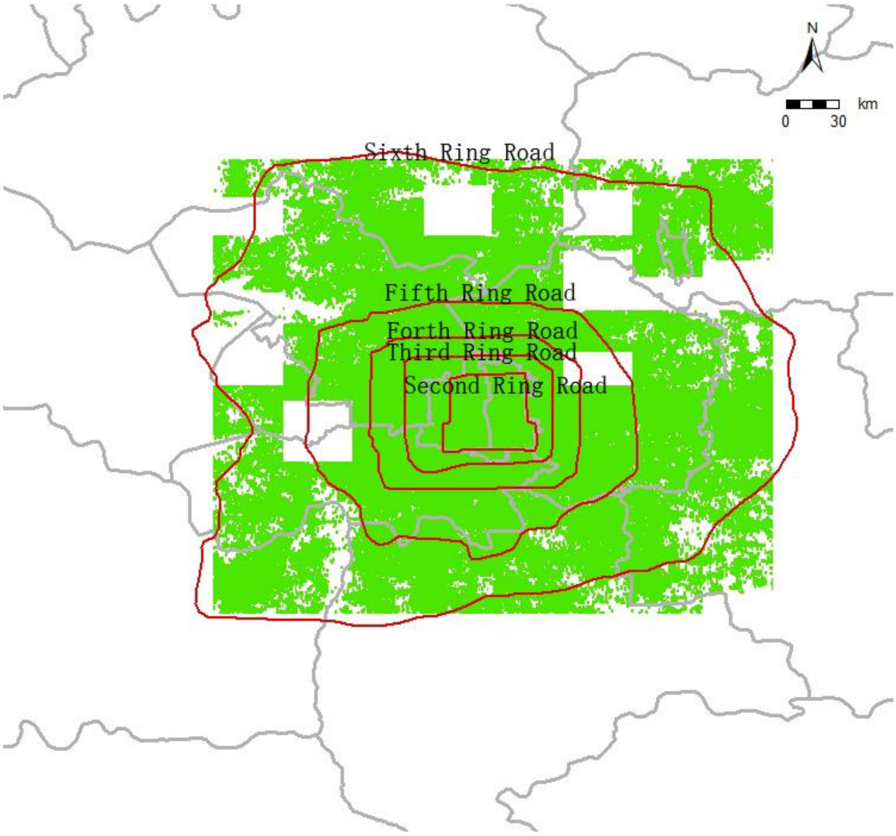


Fig 4. 1 Population distribution data

4.3 Methods

4.3.1 Urban Spatial Entropy and Temporal Entropy

In this chapter, we also create the spatial entropy model based on POIs data according to chapter3 section 3.3.1(Miaoyi L. et al., 2015). In this section, we will introduce how to use the entropy model and extend it to examine the temporal repartition of the population by hours on weekdays and weekends.

Basically, a temporal entropy model is first established according to the 24-hour grid population data for one weekday. We can calculate the total population of the day based on the size of the population in each hour. If the average population of each hour is $A_{1,}$ A_2 ... $A_{24,}$ then $A = A_1 + A_2 + ... + An = \Sigma_i A_i$ (i = 1, 2 ... 24), where A is the total population of the day. The probable population for each hour can thus be defined as:

$$Pi = A_i / A = A_i / \sum_i^n Ai \quad (4.\ 1)$$

Obviously $\Sigma i\ P_i = 1$, so the temporal entropy of the population is:

$$H = -\sum_{i=1}^{N} Pi * logPi \quad (4.\ 2)$$

where H ($H \geq 0$) is the temporal entropy. According to Miaoyi L. et al., (2015), when $A_1 = A_2 = ... = A_{24}$, $P_1 = P_2 = \ldots = P_{24} = 1 / N$, H reaches the maximum, H_m, when:

$$H_m = logN \quad (4.\ 3)$$

As shown, if $P_e = 1 / N$, the population in each hour remains constant.

Formula 4.2 shows the way to get the temporal entropy. However, when it comes to spatial-temporal entropy of a city, distributing temporal entropy in an urban space is needed. To further explore the spatio-temporal entropy, we will divide the urban space into grid cells. Urbanized area is divided into M *M grid cells (M = 300 meters). The spatial distribution pattern is then observed.

Providing that the total size of population over a whole day in the x row and y column is A_{xy} and the population of the i^{th} hour is A_{ixy}, then

$$P_{ixy} = A_{kxy} / A_{xy} \quad (4.\ 4)$$

Apparently, $\Sigma_i P_{xy} = 1$, so the temporal entropy of the population in the grid of x row and y column is:

$$H_{xy} = -\sum_i^M \sum_j^N Pixy * logPixy \quad (4.5)$$

Where H_{xy} is the temporal entropy in the grid of x row and y column. By dividing the urban space into grid cells (using software such as ArcGIS), it is not difficult to obtain a value for P_{ixy} based on each hour's average population data (sequentially using the intersect, summarize, and join tools in ArcGIS, and it is worth mentioning that when using the intersect tool, it is point representing the population intersecting with a grid with an edge of 300m, and the sum of the values of points located within the 300*300meter grid will be the population of the grid with an edge of 300m), so it is easy to obtain the value of H_{xy}. From the calculation of one grid (formula 4.5), it is possible to get the temporal entropy of every grid. Consequently, the distribution of the populations for different urban areas over an hour is obtained.

Generally speaking, the spatial and temporal distribution pattern of the population can, to some extent, reveal the behavior of people from which large populations and even more distributions can be observed in certain areas. Higher temporal entropy indicates that people appear more frequently in certain areas during different time periods; lower temporal entropy signifies that people appear more frequently in a certain area in only certain time periods; a larger population and higher temporal entropy indicate an area with 24-hour activity.

4.3.2 Calculating Mixture of Urban Functions based on POI Data

As stated above, entropy has widespread applications in studies of diversity and based on it, Chapter3 (Miaoyi Li. et al. (2015)) have developed spatial entropy of POIs to measure the mixture of urban functions. In the model, the closer the quantity of every type of POI stays, the more mixed the urban function is. The model has its drawbacks because it has not considered the weight of each type of POI. It is known that even 100 POIs of shopping could only show a highly clustered single function and one POI of a hospital is much more important than one POI of a catering shop.

TF-IDF (term frequency-inverse document frequency) is one of the most widely used techniques to weigh variables in information retrieval and data mining and assess the degree of importance of one word or phrase in one file or in one corpus. The main idea is that the importance of a word to a document increases as its frequency occurring in this document increases, but decreases as the number of documents that contain this word within the corpus increases. Jing Yuan et al. (2012) have applied the method to discover urban functions of land to eliminate the effect of different manners of the importance of different POI types. In this chapter, a method that combines TF-IDF and entropy measurements is introduced to measure the mixture of urban functions.

As shown in a fellowing corresponding relationship, an analogy between discovering the importance of one type of POI of a region in Beijing and the importance of one word in a file is drawn. Specifically, a formal region is regarded as a file and one type of POI is regarded as one word, and the whole region is regarded as a corpus. The importance of one type of POI to a region is directly proportional to its frequency in a region and inversely proportional to its frequency in the whole region.

One type of POI ->word

Region ->file

Whole region ->corpus

Therefore, the importance Vi of the i^{th} POI category in region r is calculated by:

$$\mathrm{Vi} = \frac{Ai}{Ni} \quad (4.\ 6)$$

where Ai is the number of the i^{th} POI category in region r and Ni is the number of the i^{th} POI category in the whole region. In this chapter, the Ni is the number of the i^{th} POI category in Beijing.

As to combining TF-IDF and entropy to calculate the mixture of urban functions, the importance of one type of POI – V is to be used as the input of the entropy model shown in Miaoyi Li. et al.'s (2015) paper. This means that the input of the number of one type of POI is replaced by the importance of one type of POI, and the calculation process is the same.

4.4 The Analysis of Spatio-Temporal Population Distribution and Urban Function

4.4.1 Spatio-Temporal Population Distribution based on LBS Data

4.4.1.1 Spatial Distribution of Population on Weekdays and Weekends

Based on the LBS spatio-temporal data from Tencent, the population distributions on weekdays and weekends are examined and used as starting points to determine the relationship between spatio-temporal distribution of population and urban functions. The data processing procedures are as follows: (1) average the population data by day, weekdays, and weekends; (2) create a fishnet with 300*300-meter grids and intersect the grids with the population data; (3) summarize the population size in each grid; (4) use the natural break method to rate the grid data by the 24-hour population size for the weekdays and weekends.

Fig 4.2(a) displays the distribution of the weekday population, and most people are distributed within the Forth Ring Road, the areas north of the Forth Ring Road and the Fifth Ring Road, and the area northeast of the Fifth Ring Road. This is because most employment posts are located within these areas, and some stand out, including Tian'anmen, Qianmen, Xidan, Wangfujing, Sanlitun, Chaoyangmen, Hujialou, CBD, Shuangjing, Zhongguancun, Zoo Clothing Wholesale Market, etc. Among these places, Xidan and Sanlitun are commercial centers; Zoo Clothing Wholesale Market is commercial-dominated; Wangfujing is a commercial and business center, and Chaoyangmen, Hujialou, CBD, and Shuangjing are business-dominated centers.

Zhongguancun serves residential, employment, and commercial functions. Moreover, Shangdi, Majuqiao Town, Outlets and Longde Square in Tiantongyuan, etc. are hotspots. Shangdi is a typical working area, Majuqiao Town is a predominantly residential area, and the Outlets and Longde Square in Tiantongyuan are district-grade commercial centers. Moreover, some traffic hubs, such as Beijing Station, Beijing West Station, some highways, such as the northeast Forth Ring Road, the east Third Ring Road, etc. are highly concentrated with people.

Figure 4.2(b) shows the weekend population distribution, and compared with that on the weekdays, the weekend population persists in some residential areas, such as Majuqiao Town, with most people still gathering around Tian’anmen. Commercial centers, such as Xidan and Sanlitun, commercial-dominated areas, such as the Zoo Clothing Wholesale Market, commercial and business centers, such as Wangfujing, and mixed-use areas, such as Zhongguancun, continue to be hotspots. Aeon International Mall becomes hot on the weekend, and as for the hotspot of Kehuiqiao, it is mainly due to its entertainment town and its relatively abundant commercials. Not surprisingly, some traffic hubs, including Beijing Station, Beijing West Station, etc., and highways, such as the east Third Ring Road, etc. continuously concentrate many people. The difference is that fewer people are observed in business centers, such as Chaoyangmen, Financial Street, etc.

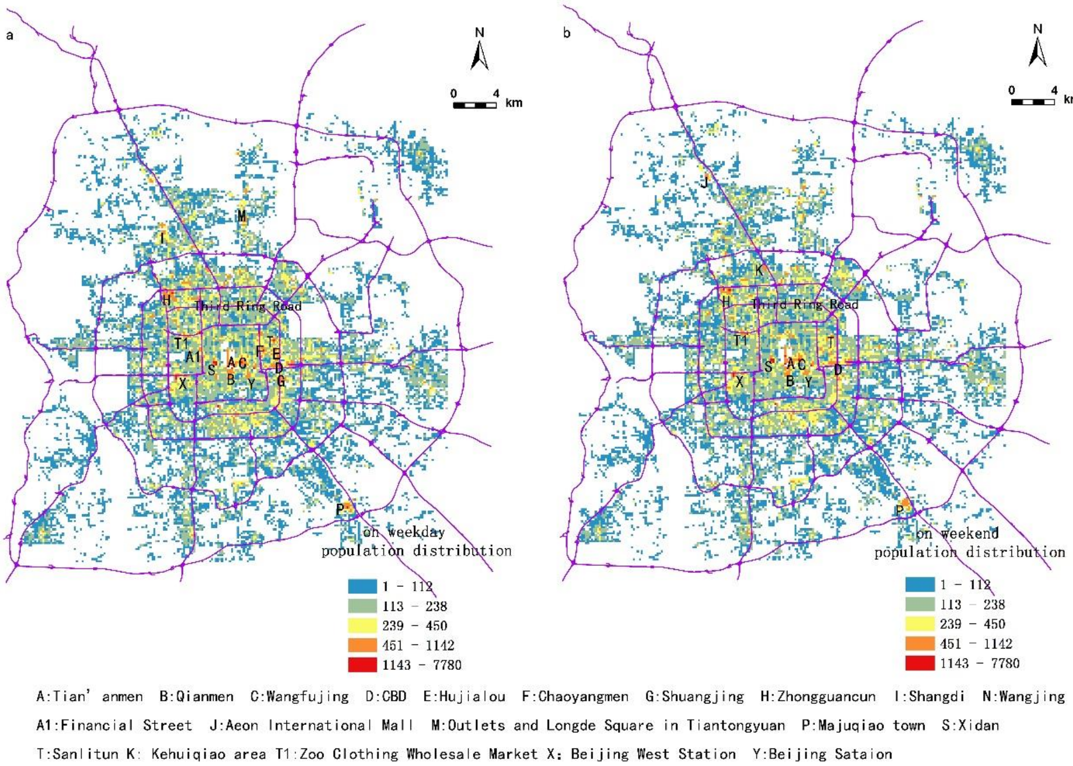


Fig 4. 2 Population distribution on weekdays and weekends

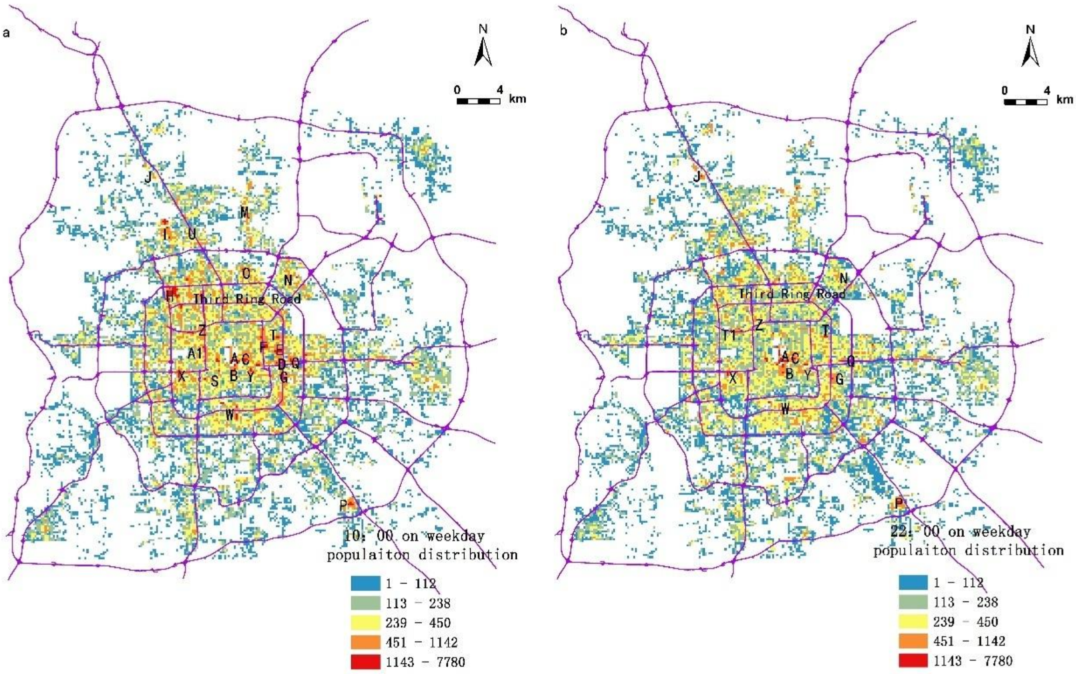
4.4.1.2 Spatial Distribution of Population at Several Certain Moments

In general, 10:00 and 22:00 on weekdays represent the statuses of people being on the job and staying at home respectively, whereas 15:00 on the weekends can represent the status of people seeking entertainment. The comparison between10:00 on weekends and on weekdays reveals the status of working overtime on weekends. Given these typical urban morphologies, this section chooses 10:00 and 22:00 on weekdays and 10:00 and 15:00 on weekends respectively to illustrate the different urban morphologies based on the population distribution at different time periods.

Figure 4.3(a) shows the weekday population distribution at 10:00, when most people are distributed within the Forth Ring Road, the area around the north Forth Ring Road and the Fifth Ring Road, and the area around the northeast Forth Ring Road and the

Fifth Ring Road. This pattern represents the weekday average. Moreover, most hotspots correspond with areas that have many employment opportunities, including Financial Street, Chaoyangmen, Hujialou, CBD, Shuangjing, Wangfujing, Shangdi, etc. Fengtai Headquarter Base and Shangdi, two typical employment areas, do not stand out mainly because they are of low building density and there are barely any commercial areas. Xidan and Sanlitun are small hotspots, and Majuqiao town stands out due to its high building density. Moreover, some traffic hubs, such as Beijing Station, Beijing West Station, Beijing North Station etc., some highways, such as the northeast Forth Ring Road, the east Third Ring Road, etc., are also hotspots.

Figure 4.3(b) shows the weekend population distribution at 22:00. Compared with the weekday population distribution at 10:00, the areas with many employment opportunities, such as Financial Street, Chaoyangmen, Jianguomen, Hujialou, CBD, Shuangjing, Shangdi, Zhongguancun, etc., do not stand out. A residential area with abundant commercials and entertainment facilities, such as Kehuiqiao, is hot at night. Among the commercial areas, Xidan, Sanlitun and Wangfujing are hot, and Aeon International Mall becomes more popular at night. Among the residential areas, only Majuqiao Town stands out due to its high density. Some large living communities such as Huilongguan, Tiantongyuan, and Tongzhou show a silent status at night mainly because of their low building density. Not surprisingly, some traffic hubs, such as Beijing Station, Beijing West Station, etc., and the subway stations in the north part of subway Line 5 remain hot.



A:Tian' anmen B:Qianmen C:Wangfujing D:Guomao E:Hujialou F:Chaoyangmen G:Shuangjing H:Zhongguancun I:Shangdi N:Wangjing
A1:Financial Stret J:Aeon International Mall M:subway stations in north part of subway line 5
O:Asian Games Village P:Majuqiao town Q:CBD S:Xidan T:Sanlitun T1:Zoo Clothing Wholesale Market
U: Qinghe V:Tongzhou X: Beijing West Station Y:Beijing Sataion W:Beijing South Station Z:Beijing North Station

Fig 4. 3 Weekday population distribution at 10:00 and 22:00

Figure 4.4(a) shows the weekend population distribution at 10:00. Compared to the population distribution at 10:00 on weekdays, the areas with many employment opportunities, such as Financial Street, Chaoyangmen, Jianguomen, Hujialou, Panjiayuan, Shangdi, etc., do not stand out. Among the high employment areas, CBD and Guomao attract many people. At the same time, mixed areas, such as Zhongguancun, residential areas with abundant commercial areas and entertainment facilities consistently concentrate many people and are the same size as the commercial areas such as Xidan and Aeon International Mall. Moreover, Majuqiao Town continues to stand out due to its high building density. The traffic hubs, such as Beijing Station, Beijing West Station, etc., are consistently hot.

Figure 4.4(b) shows the weekend population at 15:00, and it is similar to the weekend population distribution at 10:00, except that people seem to concentrate more at some places at 15:00, especially commercial areas such as Xidan and Zhongguancun. The hotspots include CBD and Guomao, Zoo Clothing Wholesale Market, Zhongguancun, Xidan, etc.

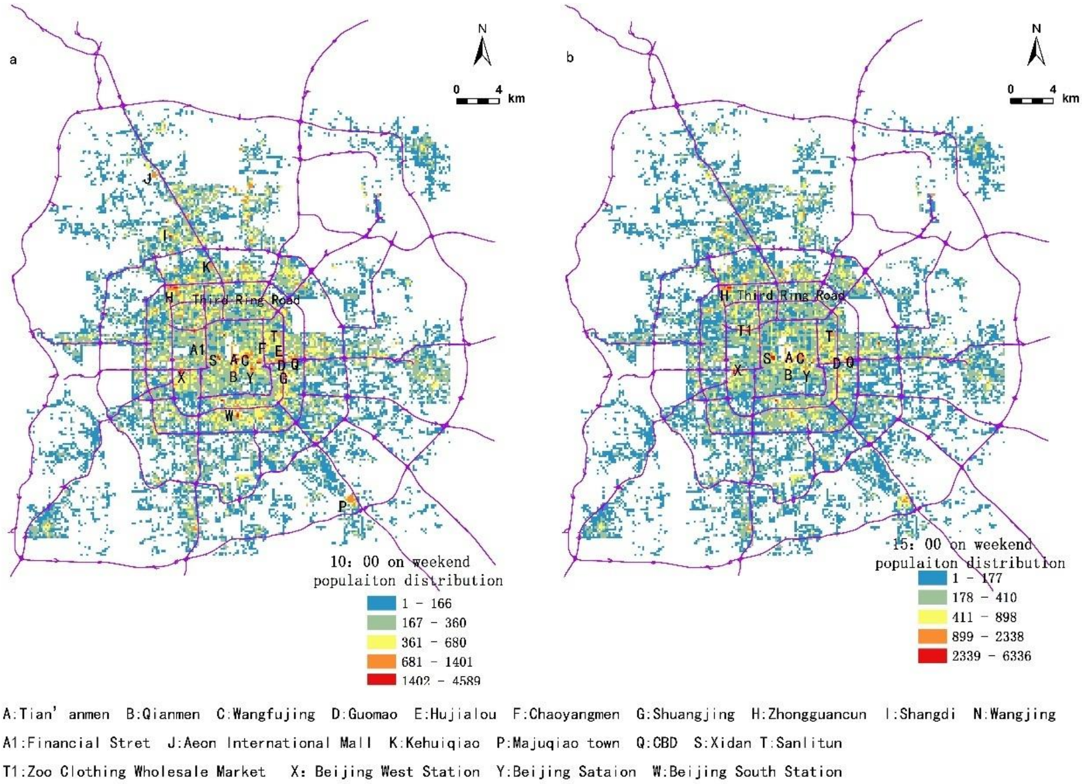


Fig 4. 4 Weekend population distribution at 10:00 and 15:00

From the specific and average weekday and weekend population distributions, it is obvious that there are different urban morphologies at different times. Some places with mixed functions attract many people from morning to night, while some places with single functions only concentrate people during specific time periods. For example, Zhongguancun, which integrates commercial, residential, and employment functions, is always hot, but Financial Street, a pure employment area, is only hot during the day and silent at night. People gathering at different places and times reflects urban vitality, forming a city spectrum, as well as daily human activity. To observe the spatio-temporal distribution in homogeneity and heterogeneity of population dynamics, temporal entropy must be used to measure it.

4.4.2 Temporal Distribution of Population in Different Urban Function Areas

4.4.2.1 Temporal Distribution of Population in Different Urban Function Areas

Based on the spatio-temporal entropy of the population on weekdays and weekends, we can identify places with uneven population distribution. According to section 4.3, the population distribution remains the same between hours when the temporal entropy is equal to 3.18. Most places have high entropy values with some exceptions (Fig 4.5), suggesting that the spatial distribution of temporal entropy is not correlated with location. Figure 4.5(a) shows that places with the lowest entropy (the blue area in Fig 4.5(a)) are employment-dominated areas including Shangdi, Zhongguancun Software Plaza, Wangjing, and Zhongguancun's enterprise clustering area, the Yizhuang Development Zone, the Fengtai Headquarter Base, Financial Street, Hujialou, Chaoyangmen, CBD, etc.; scenic spots, such as the 798 Arts District; schools, such as Tsinghua University, Beijing University of Technology, and Beijing City School; markets, such as the Yuquanying Decorative Building Materials Market and the Zoo Clothing Wholesale Market. Places with the second lowest entropy (green area shown in Fig 4.5(a)) are mainly distributed in scenic spots, including the Forbidden Palace, Temple of Heaven Park, Yuyuantan Park, the Beijing Zoo, Yuanmingyuan, the Summer Palace, the National Olympic Sports Center as well as residential areas, such as Sihui.

Figure 4.5(b) shows the spatial distribution of the weekend temporal entropy of the population. Compared with the spatial distribution on weekdays, scenic spots, such as the 798 Arts District, the Forbidden Palace, Temple of Heaven Park, Yuyuantan Park, the Beijing Zoo, Yuanmingyuan, the Summer Palace, the National Olympic Sports Center, etc.; schools, such as Tsinghua University and Beijing University of Technology;

and business centers, such as CBD, remain areas of low entropy. A small part of the Yizhuang Development Zone, Shangdi and the Fengtai Headquarter Base also continue to exhibit low temporal entropy. Moreover, Lanlishilu, which is close to Xidan, and Yongdingmen Park, which is close to Temple of Heaven Park, have low entropy on the weekends. Moreover, the commercial and business center of Wangfujing has low entropy, because it primarily performs a commercial function on the weekends. The temporal entropies of the populations in Tsinghua Science Park and the shopping mall area in Zhongguancun area are low.

It can be concluded that employment areas, commercial areas, scenic spots, and markets, especially wholesale markets, have the most unevenly temporally distributed populations on both weekdays and weekends, meaning that gatherings of people in these areas occur during certain time periods. Residential and mixed areas display high entropy on the weekdays and weekends. That is, these areas have evenly distributed populations over time on weekdays. It is worth mentioning that commercial areas and the commercial sections of mixed areas have more time periods during which people concentrate.

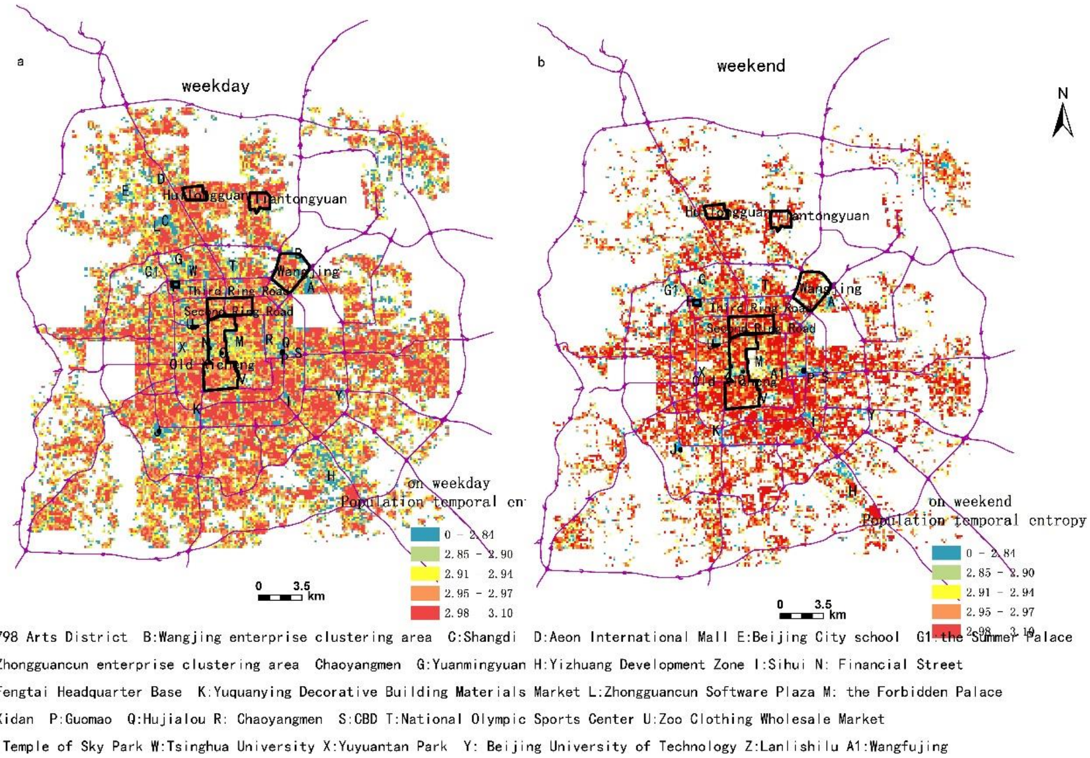


Fig 4. 5 Temporal entropy on weekdays and on weekends

4.4.2.2 Comparison with Population Numbers

To identify spatial distribution and temporal entropy of different population groups on weekdays and weekends, this part have to follow a particular process. Firstly, we classified the average population and temporal entropy on weekdays and weekends separately into three grades using the natural break method. The three classes are assigned to 1, 2, 3 indicating low, medium, and high values, respectively. Secondly, we grouped the temporal entropy and population on weekdays and weekends separately using the formula of "10*population + temporal entropy". In this way we can get 3*3 kinds of groups. The processes mentioned are implemented by the field calculator tool of ArcGIS software. To emphasize the circumstances of low or high, we highlight four groups, which are less population, high temporal entropy; less population, low temporal entropy; high population, high temporal entropy; high population, low temporal entropy.

Figure 4.6 shows the distribution of four scenarios. Most areas fall into the category of low population and high temporal entropy both on weekdays and weekends, suggesting that most areas are less-populated and have relatively even population distribution over time. It is easy to understand that the area with less population will have less opportunity to let people get in and out. These areas may be of low building density and mixed functions with residential, commercial etc.

Figure 4.6 (a), the areas with less population and low temporal entropy are Fengtai Headquarter Base, YuQaunYing Building Materials Market, Yizhuang Economic Development Zone, office area of Wangjing, 798 Arts District and Yuanquan village, etc. Among them, Fengtai Headquarter Base, Yizhuang Economic Development Zone and office area of Wangjing are typical working areas with less public services facilitating and of low build density; 798 Arts District is scenic spot; YuQaunYing Building Materials Market is specialized market which is not that hot like clothing wholesale market; Yuquan village is of low population density, which is no doubt, as for its low temporal entropy, it need be distracted by transit traffic. Among those of high population, Xidan, Zhongguancun, Shangdi, Guomao, Hujialou, CBD are of low temporal entropy, while Tian'anmen, Qianmen, Beijing station, Shuangjing, Aeon International Mall, Outlets and Wande Square in Tiantongyuan and Majuqiao Town are of high temporal entropy. It can be concluded that employment-dominated area with less commercial and entertainment facilities with high building density like CBD, Shangdi etc., big commercial centers like Xidan and Zhongguancun tend to be of the high population with even population distribution over time. As for district-grade commercial centers like Aeon International Mall and Outlets and Wande Square in Tiantongyuan, high population and high temporal entropy may reflect human activities' rule in district-grade commercial centers. Tian'anmen is hot all the time. Meanwhile, Qianmen is also hot all the time due to its commercial, residential and tourism functions. Besides, Majuqiao town is a residential area with high building density and is hot all the time, while Beijing station is a traffic hub.

Figure 4.6 (b) shows similar spatial distribution on weekend compared with a weekday. The differences are that employment-dominated areas have less population and high entropy, while Kehuiqiao shows high population and low entropy, possibly resulted from its entertainment function.

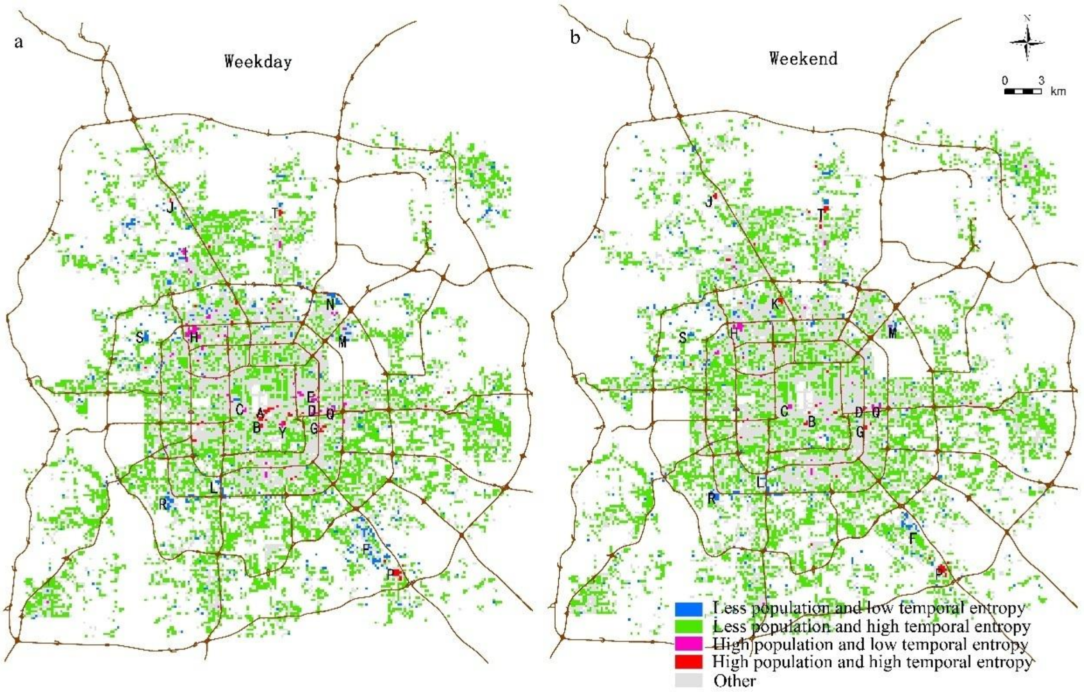


Fig 4. 6 Comparison between temporal entropy and population distribution

4.4.3 Correlation between Spatio-Temporal Distribution of Population and Urban Function

Above description shows that some places have large temporal variations in the distribution of the population while others have relatively stable population size overtime. These differences reflect daily human activities that are related to urban functions. This part of the chapter examines the correlation between urban functions and spatio-temporal distribution of the population from three perspectives. The first one is based on 300*300-meter grids, which are used to examine how the degree to which urban functions are mixed affect the temporal population distribution. The second is based on the analysis of key areas and the type of urban functions that are important. The third is based on examining the temporal population distribution curve in the key areas to determine the characteristics of each function that attract people.

4.4.3.1 Correlation between Spatio-Temporal Distribution of Population and Urban Function Mix

This chapter uses the method introduced in part 3.3.1 of Chapter3 to calculate the mixture of urban functions. To examine whether there is a correlation between the spatial entropy of the POIs and the temporal entropy of the population, the weekday and weekend temporal entropies are calculated using the method mentioned in section 3.3.2 of Chapter3, and the three types of entropy are incorporated into a table. To calculate POI-spatial entropy, the weekday and weekend temporal population entropy values are used as variables, and it is not difficult to detect the correlations between each pair of variables using the bivariate correlation tools in SPSS software.

As shown in Table 4.1, this approach determines the correlation between any two entropies; and the correlation coefficients between the spatial entropy of the POIs and the weekday and weekend temporal entropies of the populations are 0.046 and 0.067, respectively, a lower degree of correlation. Additionally, there is a correlation between each pair of variables at a confidence level of 0.01. The significant correlations prove, to some extent, that mix-used areas are more likely to have temporally even-distributed populations.

Table 4. 1 Pearson correlation between the spatial entropy of the POIs and the temporal entropy of the population

	Temporal entropy on weekday	Temporal entropy on weekend
POIs entropy	.046**	.067**

Note: **. Correlation is significant at the 0.01 level (2-tailed)

4.4.3.2 Correlations between Spatio-Temporal Distribution of Population and Urban Function in Key Areas

Based on the distribution of the temporal entropy of the weekday and weekend populations and considering different types of function, such as residential, employment, etc., this chapter chooses key areas (Fig 4.6), among which Financial Street, CBD, the Fengtai Headquarter Base and Shangdi occupy a single land parcel within the core district and are predominantly employment areas. Xidan and the Zoo Clothing Wholesale Market also form a land parcel within the core district and are commercial areas. Zhongguancun is a land parcel within its core district that is a both employment and commercial area, and the areas around the north Second Ring Road and the Third Ring Road, Wangjing, and Old Xicheng are all large mixed areas.

The data processing procedures are as follows: (1) create a Thiessen polygon with the average daily weekday population data as an input to obtain the range represented by one population point; (2) select those Thiessen polygons that are completely within a land parcel to remove the influence of roads and subways and then intersect the polygons with the population data; (3) calculate the population of each key area; (4) obtain the temporal weekday and weekend entropy using the method mentioned in section 3.3.2 of Chapter3 and then make a table and scatter plot of the two types of entropy.

The results of the analysis of which functions play an important role in spatio-temporal distribution of population and the scattered plot analysis of the key areas are as follows (see Fig 4.7):

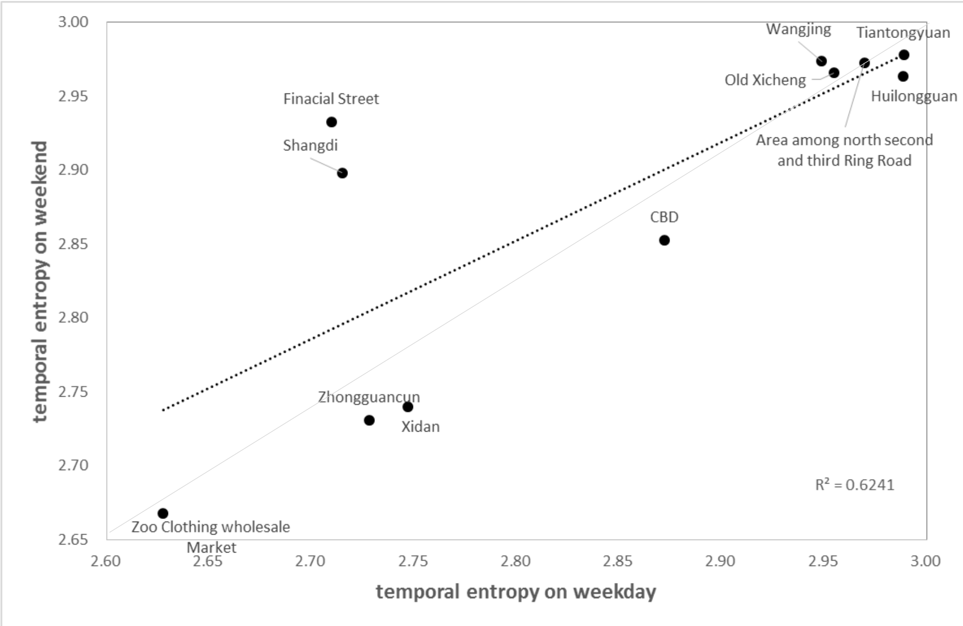


Fig 4. 7 weekday and weekend temporal entropy in the key areas

(1) Both the spatial entropy of the POIs and the weekday and weekend temporal entropies are high in the area around the north Second Ring Road and the Third Ring Road and in Old Xicheng, which indicates that these areas have a high mix of urban functions, relatively mature urban amenities, and functional development; concentrated residences, business offices, and recreation; high consumption and high frequency. Wangjing shows high temporal entropy on both weekdays and weekends because it is a relatively newly developed urban area that is focused on residential and business functions, which have two of the highest frequencies of occurrence in the population. Huilongguan and Tiantongyuan also exhibit high temporal entropy on weekdays and weekend, similar to that in employment areas. They are large-scale living communities, but as people gather, some commercial enterprises and some other facilities develop as predicted by market law. That is to say, they are residential-dominated areas with relatively abundant community services and recreational facilities, which causes some population to stay in the residential areas evenly in the daytime. Furthermore, people

primarily stay in residential areas at night, the time when people sleep, making the population appear to be lower than it actually is. This means that the population gap between the day and night appears smaller than the true status.

Financial Street and Shangdi show relatively high entropy on the weekends and low entropy on the weekdays because they serve a single function. Financial Street mainly provides offices for financial businesses while Shangdi mainly focuses on IT enterprises, most of whose employees are not working on weekends. The high entropy on the weekend is resulted from few people working all day on weekdays, but would come to these places for work during the day time and leave after work. Therefore, the gap between day and night will be large. This also shows that the population of Financial Street is a little higher than Shangdi on the weekends because working extra shifts is very common in the IT industry but not in the financial industry.

CBD shows a relatively moderate entropy on the weekdays and weekends as it is a business center that also supports commercial enterprises. It is located in the city center, and most of the companies are foreign. Although its main function is business, the temporal distribution of its population is more even than other employment-dominated areas on weekdays because of its location and surrounding commercial center.

Both Zhongguancun and Xidan have relatively low entropy on weekdays and weekends. Zhongguancun is a land parcel within the core area that performs commercial and employment functions at the district level but also has a clustered IT industry. The underground commercial activity at the place also attracts many people. Xidan is one of the biggest commercial centers in Beijing. It is thus vital all the time due to many shopping and leisure activities. Compared with the employment areas where people stay at particular times, people may gather anytime in commercial areas. The diversities of people’s activities contribute to the uneven temporal distribution of the population.

The Zoo Clothing Wholesale Market has the lowest entropy on weekdays and weekends. Most of the people at these places are clothing merchants, salesmen, porters, or couriers, who have rapid paces of life. The environment is usually noisy and disorderly with many people coming in and out. The enormous flow of people thus makes the population very unstable over time.

(2) With an R^2 coefficient reaching 0.6241, the temporal entropies on the weekdays and weekends are strongly correlated; when entropy is high on weekdays, it is also high on the weekend. Areas such as Huilongguan, Tiantongyuan, Old Xicheng, Wangjing, and the area around the north Second Ring Road and the Third Ring Road all have similar values on weekdays and weekends, possibly related to the high mixture of urban functions. At the same time, the Zoo Clothing Wholesale Market also shows the same approximate entropy value on weekdays and weekends because those that gather here do not distinguish between weekdays and weekends; they have free time and do not need to go to an office. Employment areas, such as Financial Street and Shangdi, have weekend population entropy values that are higher than those on weekdays, primarily because they only provide a single employment function. Moreover, Zhongguancun, Xidan, and CBD show lower population entropy on weekends than that on weekdays, which is related to people’s behaviors, such as shopping, entertainment, etc. On weekends, people have more free time to go to these places and engage in these activities.

It can be concluded that more mixed urban functions drive more even population distributions over time in general, but they are of weak correlation. Besides the mixture of urban functions, the type of function is also correlated. For example, residential-dominated areas with abundant community service and recreational facilities will be stable, while the commercial function is associated with pretty uneven temporal population distributions due to variations in human behavior. As for the comparison of weekday and weekend, more mix areas tend to be similar. At the same time, the difference between them is also correlated with the type of urban function. Employment-dominated functions, as in places that attract the IT and financial industries, show uneven population distributions on weekdays but more even distributions on weekends, while the commercial-dominated areas show the opposite.

4.4.3.3 Exploring Temporal Patterns of Population in Different Urban Function Areas

Temporal entropy indicates whether the size of a population is evenly distributed in each hour; but cannot tell at what time there are more people or less. To further explore the hourly population distribution in more details, a population temporal distribution curve is constructed. To eliminate the influence of size, the population is normalized by equation 4.7.

$$Y = (X - X_{min})/(X_{max} - X_{min}) \qquad (4.7)$$

Where X is the population value of a given hour; X_{min} is the minimum population value in 24 hours; X_{max} is the maximum population value in 24 hours; Y is the normalized value.

Fig 4. 8 Population curve in the key areas at different times

Based on the curve of the three employment areas (Fig 4.8), Financial Street and Shangdi show similar characteristics. There are much fewer people on weekends, most gathering from 10:00 – 18:00 with little fluctuation. The populations of both rise sharply before noon and descend at a relatively slower speed after 18:00, which coincides with people's commutes. That is, most people arrive at work before a certain time and leave work after a certain time, such as 18:00. As for CBD, the population on weekend is slightly less than that on a weekday, which is due to the commercial function of the area. The curve also rises sharply before noon and descends at a slower speed after 18:00. However, the curve stops descending at 21:00 because people gather for a late snake.

Two commercial areas show completely different curves. People coming to the Zoo Clothing Wholesale Market arrive several hours earlier than those coming to Xidan, which is due to the different operating hours of the businesses. This difference in business times results from the purposes of the people coming to these places. Many people buy goods for sale at Zoo Clothing Wholesale Market. Therefore, they go to the place earlier so that they can obtain their goods and get prepared before customers arrive. As for Xidan, people come for leisure or to buy goods as well, but they are not hurrying to meet a deadline. Moreover, the numbers of people who come to Zoo Clothing Wholesale reach high points at 9:00 and 14:00, rising and descending quickly before 9:00 and after 14:00, which coincides with the start of the morning and afternoon. The number of people coming to Xidan peaks at 12:00 and 15:00, rising quickly and descending slowly before 12:00 and after 15:00, implying that people prefer to shop at 15:00. Another difference is that the population of Zoo Clothing Wholesale remains the same on the weekend and weekdays, while the population of Xidan is lower on weekdays than on the weekend. This is because people have more free time to go shopping on the weekend, but there is no difference for people who have their own business.

Two curves are very similar for residential areas. They both peak at 9:00 and 22:00 with a depression in the middle of the week and peaks at 10:00 and 22:00 with a shallow depression in the middle on the weekend. This is easy to understand, considering that most people are not going to work on the weekend. As for being almost no one active after midnight, this is caused by the powering off of cellphones when sleeping.

The curves of the areas along the north Second Ring Road and the Third Ring Road, Wangjing and Old Xicheng are similar, and people come to Zhongguancun a little bit later than to the other three. Moreover, there are fewer people at night in Zhongguancun, but many people come to these areas at 22:00. This is because Zhongguancun undertakes a large proportion of commercial function without performing a residential function. The peaks of the Zhongguancun curve occur at 12:00 and 15:00 on weekdays and weekends, indicating its commercial characteristics, while the curves of the other three areas peak at 10:00, 15:00, and 22:00 on both weekdays and weekends, which coincides with commuting, shopping and returning home, respectively. Moreover, the populations of areas along the north Second Ring Road and Third Ring Road, Wangjing and Old Xicheng fluctuate over a small range at a relatively high level from 10:00 to 22:00.

From the curve of temporal population distributions of the key areas, it is clear that the population of each key area varies over time, and this is no doubt closely related to human behavior. From the above descriptions, it is clear that the time at which people appear in a commercial area is highly concentrated at several specific times, but people gather at mixed areas at any time. As for employment-dominated areas and large residential areas, population sizes vary during the day and at night. However, compared to employment areas, the gap in the residential area between day and night is much less.

4.5 Summary of This Chapter

Due to the limitations of traditional research data, most of the previous studies of urban space are static. Big data provides opportunities for scholars to study dynamics of urban space, such as urban structure, job-housing commute and transportation structure, and urban function zoning, at a micro scale. A general understanding of spatio-temporal distribution of the population can help us examine cities with a renewed view, and this chapter explores spatio-temporal distribution of population and its correlation with urban functions.

First, it is obvious that there are different urban morphologies based on the distribution of the population at different times. Places with mixed functions attract people from morning to night, whereas places with the single function may only concentrate people during a specific time period. For example, Zhongguancun, which integrates commercial, residential and employment functions, is always busy, while Shangdi, an employment-dominated area, is only popular during the daytime.

Second, employment-dominated areas, such as Shangdi and Financial Street; commercial areas, such as Xidan; scenic spots such as the 798 art district; and markets, especially wholesale markets, such as the Zoo Clothing Wholesale Market, have the most unevenly temporally distributed populations on both weekdays and weekends. Residential and mixed areas are more evenly temporally distributed on weekdays and weekends. Grouping temporal entropy and population, most areas are of less population and having evenly temporally distributed population both on weekdays and weekends. Working areas with fewer facilities and low building density tend to show high population and unevenly temporally distributed population. Large commercial centers show high population with unevenly temporally distributed population, while district-grade commercial centers show high population with the evenly temporally distributed population.

Third, the spatio-temporal distribution of the population is closely related to urban functions. In general, more mix urban functions drive a more even temporal distribution of the population. Moreover, the appearance of people at certain places is in accordance with behavioral habits. Employment areas show large population fluctuations on weekdays, but their numbers are distributed relatively evenly on weekends. Commercial areas only peak for several hours both on weekdays and weekends, while mixed areas and large-scale living communities accommodate a stable number of people all of the time.

Chapter5

5. Examining Taxi Ridership Impacts from Newly Opening Subway Line with Taxi Trip Data

5.1 Introduction

Fast-changing urbanization contexts demand knowledge to provide strategies for sustainable transportation development in the context of ever-increasing volumes of geospatial data. Transit ridership has long been a hot research topic among scholars and policy-makers in the fields of transport geography and urban planning (Markovich J. et al., 2011; Cervero R. et al., 1996; Liu X. et al., 2015). Much work has been conducted to identify the interdependencies of transportation systems and their ridership (Kang C. et al., 2013). Few studies focus on the spatial dependency of transit ridership in comparing subway and taxi, which is one of the vital studies in urban sustainability.

The relationship between subway and ground systems is interrelated, as they mutually affect each other. Understanding the connection of subway with taxi is essential for society-wide policy issues. Subway, as one of the popular transportation modes in urban areas for most passengers, is an important means of sustainable transportation development in reducing private travel demand (Liu X. et al., 2015) . Lin et al. (Lin D. et al., 2016) suggest that public transport networks, the location of housing and employment have played significant roles in commuting. They indicate that subway significantly and negatively influenced the commuting times of low- and middle-income workers, but neither subway nor taxi had a significant influence on the commuting times of high-income workers. On the other hand, taxi meets a large amount of citizens' travel demands and covers a wide range of urban areas with high accessibility and flexibility. Taxi, as an indispensable mode of transportation in large cities, complements other public transport modes in terms of flexible door-to-door service and 24/7 operations. This has drawn many researchers to examine a new subway system (Kain J.F. et al., 1999; Yuan Y. et al., 2012) , while less attention is paid to a subway's influence on taxi ridership (Barthélemy M. et al., 2011; Kivelä M. et al., 2014) .

Understanding the association between the travel patterns of subway users and the choice of taxi trip is an essential part of any transportation plan to enhance sustainable urban development (Kivelä M. et al., 2014) . A newly opened subway would have both positive and negative effect on taxi ridership in the nearby region (Liu Y. et al., 2012). For example, Knowles (Knowles R.D. et al., 1996) observed a much larger shift from vehicle travel to transit travel because of the opening of a new subway line (Kim S. et al., 2007) , which encouraged people to use transit and more people who like to use transit would choose to live near transit stations (Cervero R. et al., 1996) . Previous literature focuses on examining different variables, such as travel time, costs and the level of service that are closely associated with subway and taxi ridership; however, few studies have quantitatively analyzed the impact of opening a new subway line on spatial variation of urban taxi trajectories and evaluated the changes of taxi passengers' transportation modes that reveal the critical locations in people's mobilities, because of the absence of digitalized data both in subway and in ground transportation.

The majority of existing evaluation studies on ridership reply on traditional data, including questionnaire and travel surveys, in which it is difficult and time-consuming to capture the information of origins and destinations, leading to small sample sizes, inaccurate location formation, and shorter periods of time coverage (Stopher P.R. et al., 2007) . These studies seldom scrutinize the ridership and transport mode choice spatially, which may include evaluations of the expected and realized benefits from the construction of new public transport systems through questionnaires (Mackett R.L. et al., 1998) , studies on taxi ridership with data collected by counting the number of taxis passing roadside checkpoints (Yang H. et al., 2000) , ridership studies on origin–destination (O–D) estimation from travel diary surveys, e.g. face to-face interviews (Brög

A.R.G., 1983) , place-based trip or activity diary (Harvey D., 2003; Battellino, H. et al.; 2003) , and traffic count data, e.g. methods reviewed for estimating an O–D matrix (Willumsen L.G., 1978) , dynamic O–D estimation to derive day-to-day demands (Zhou X. et al., 2006) , and decomposition framework for estimating dynamic O–D flows (Lou Y. et al., 2010) .

On the other hand, the distinct techniques that capture large volumes of mobility data collected from GPS-equipped vehicles, mobile phones, and smart cards lead to the discovery of spatial-temporal features of human mobility from new perspectives and employing new tools (Yuan Y. et al., 2012; Calabrese A. et al., 2013; Chen C. et al., 2014). In the environment of such new data (Long Y. et al., 2016), the new emerging data are more accurate, objective, plentiful, and cost-effective in describing human mobility in a spatially embedded flow network (Batty M., 2013). For example, researchers have adopted taxi trajectory data for understanding spatio-temporal characteristics of human mobility within a city (Chowdhury D. et al., 2000; Cichocki A. et al., 2009; Ciscal-Terry W. et al., 2006; Fan Z. et al., 2014; Gonzalez M.C. et al., 2008), specifically exploring human travel patterns (Gonzalez M.C. et al., 2008; Schneider C.M. et al., 2013; Ratti, C. et al., 2006), observing mobility pattern at the individual level (Kwan, M.P., 2000), and deriving O–D information and trip purpose (Wolf J. et al., 2013; Bohte W. et al., 2009). Thus, these works inspire the pioneering work on characterizing spatial variation among alternative travel modes.

This study attempts to examine transit ridership among alternative travel modes from O–D pairs by using GPS tracking data of taxis and subway transaction data from the new subway line to analyze human mobility in the network analysis by constructing a 1km by 1km cell grid at the city level in Wuxi, China. Smaller and medium-sized developing cities have great potential to develop sustainable transport systems, which have low-cost investments and the imposition of modest fees to promote more sustainable urban transport, therefore different measures may be more appropriate for smaller and medium-sized developing cities than megacities (Pojani D. et al., 2015). This study apply the spatio-temporal network analysis method for the comparative analysis to deploy more accurate and complete information on the interaction between subway and taxi ridership, which has not yet appeared in the existing transit-ridership literature. The three-fold research objectives are: (1) to present a novel quantitative method to observe the divergent behavior of human mobility by developing a geospatial citywide database of floating car trajectories and subway transaction records; (2) to explore the impacts of opening a new subway line on taxi trips in terms of spatial and temporal dimensions by taking advantage of the longitudinal large-scale trip data; and (3) to investigate the interaction between taxi trips and subway trips at the city level.

5.2 Materials and Methods

5.2.1 Study Area

Wuxi, located in eastern China, is one of the cities representative of fast urbanizing China and also has great potential to develop sustainable transport systems. We selected Wuxi city as the study area for two main reasons: Firstly, as one of the largest cities in the Yangtze River Delta region of China, from 2000 to 2015, Wuxi has approximately tripled its urban built-up area from 110 km^2 to 320 km^2 and increased fivefold its urban population from 0.6 to 3 million, depicting an ideal representation of Chinese cities in the fast urbanization process. Secondly and the most importantly, the first subway line (Line 1) of Wuxi just opened in July 1, 2014, making it possible to measure the impact of subway to urban transportation networks with data collected via Metro Card and taxi GPS. Figure 5.1 shows the boundary of Wuxi and the 24 subway stations (green dots) of Line 1.

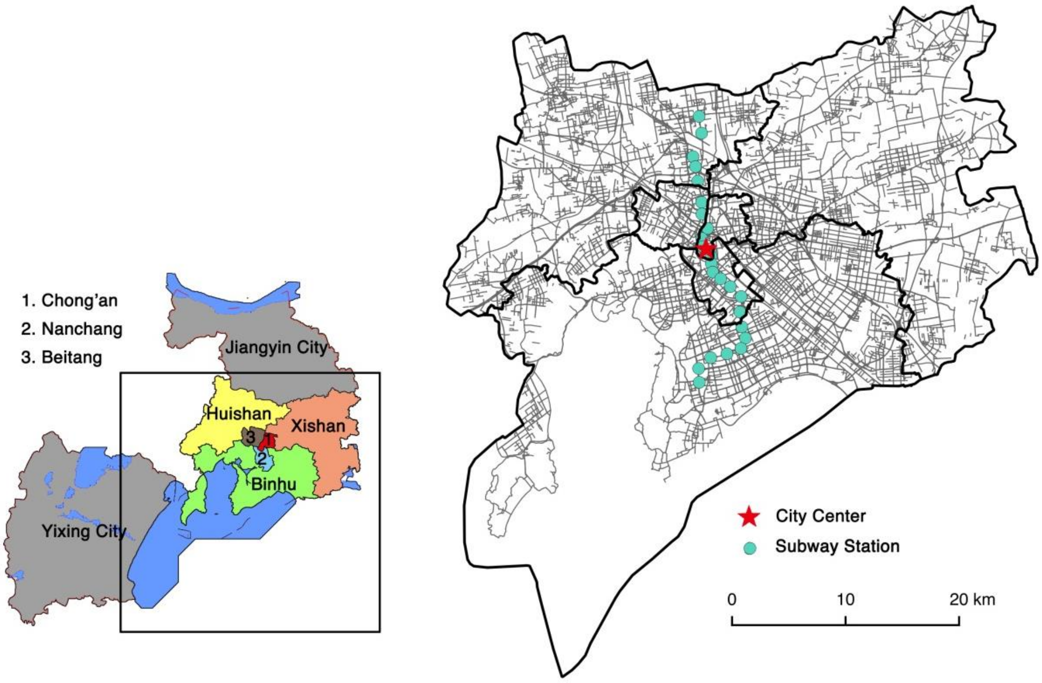


Fig 5. 1 The Wuxi study area and the subway stations (green dots)

5.2.2 Dataset

Two datasets are used in this study. The first one is subway O–D data that covers four weeks since the opening of Line 1. This dataset is extracted from the transaction records of each Metro Card, including anonymized card ID, transaction time, transaction type (enter or exit), and station ID.

The second dataset is the trajectories of 1500 taxis in Wuxi, which covers four weeks before and four weeks after the opening of the new subway. Taxi trajectory data includes drive routes, pick-up and drop-off locations, speed, and time slots, which are collected by the widely used floating car technology (Probe car) with a GPS device on board. Compared to traditional surveys, taxi trajectory data reflects more dynamical and precise travel behaviors in cities, and therefore is employed to represent the ground transportation. For the validity of this study, we only extract trajectories with passengers on board, and then keep the origin and destination points from each trajectory to construct a network (detailed in Methods). Since people may have different travel behaviors during weekdays and weekends, we analyze the travel flows on weekdays and weekend separately. Table 5.1 gives a short summary of the daily number of taxi trips in different time periods.

Table 5. 1 Daily number of taxi trips in different time periods.

Date	June Weekday	June Weekend	July Weekday	July Weekend
Count	136,135	144,036	130,839	134,107

5.2.3 Methods

The primary goal of this study is to explore the impact of a new subway on ground transportation in a city, and we assume that in a relatively short period (four weeks, in our case) before and after the subway's opening, the average variation of ground transportation (e.g. taxi trips) is mainly caused by the subway's influence. Therefore, we construct two networks (subway network and taxi network) and evaluate the statistical results before and after the opening of a new subway. Firstly, we build the taxi network in 1km by 1km grids, which are commonly used for transportation network studies (Liu X. et al., 2015). In this network, the 1km^2 cells are nodes and the O–D vectors between nodes are inbound and outbound edges; the total volume of each edge is its weight, it also belongs to the one type of spatio-temporal network.

Here, we average the daily O–D volume between cells for weekdays, weekends, in June and July respectively, and remove edges with a daily traffic volume of less than 1. Thus, we construct four direct networks (June weekday, June weekend, July weekday, and July weekend), two of which are before the opening of the subway, and two of which are after. For this directed network, the degree of node i can be divided into in-degree and out-degree. The in-degree is the sum of connections onto node i, $k_i^{in} = \sum_j a_{ij}$, and the out-degree is the sum of connections coming from node i, $k_i^{out} = \sum_j a_{ji}$, here a_{ij} is the element of adjacency matrix A. As shown in Figure 5.2, the in-degree of Node 1 is 1, and the out-degree is 1, while for Node 3, the in-degree is 1, and out-degree is 2. Table 5.2

shows the basic network statistics. We can see that the number of edges drops nearly 20% for both weekday and weekend after the opening of the subway.

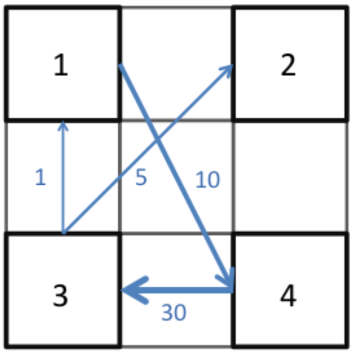


Fig 5. 2 Illustration of Network construction. Boxes 1-4 represent four nodes, and gray arrows represent four direct edges, number on each arrow represents the traffic volume.

Table 5. 2 Basic Network statistics based on taxi GPS data in Wuxi.

	Number of Nodes	**Number of Edges**	**Average Edge Volume**
June weekday	404	14,902	7.02
June weekend	407	15,910	7.16
July weekday	405	12,168	7.77
July weekend	416	12,929	7.86

5.3 Results

5.3.1 Subway Influence on Taxi Ridership

Firstly, we plot the subway O–D matrix to show the subway network structure. Figure 5.3 depicts each station's ridership volume with the weekdays' result on the left and the weekends' on the right. In Figure 5.3, we can easily identify that *Sanyang Square* and *Nanchansi* are the stations with the darkest red as they are the central area of the city, and the origin-destination area is more concentrated near the city center. Additionally, the overall subway ridership intensity on weekends is higher than weekdays, which indicates that people are more likely to travel by subway Line 1 on the weekend.

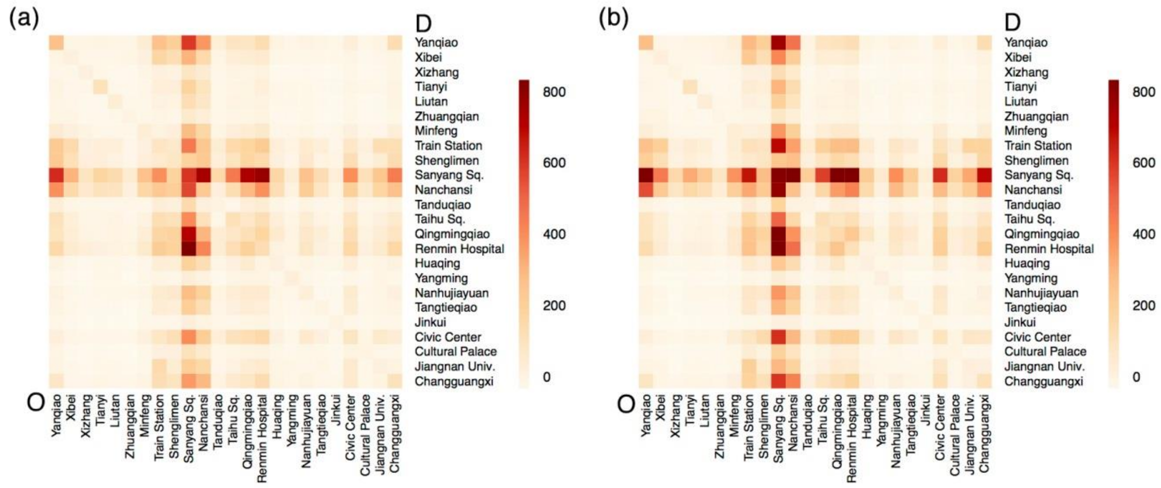


Fig 5. 3 Subway O–D matrix; the order of stations is organized as the spatial sequence along Line 1. (a) Weekday; (b) Weekend.

Secondly, we analyze the daily taxi O–D data within a 1km radius of the subway stations. Figure 5.4 and Table 5.3 present us with more detailed evidence of the ridership change in taxi and subway along the new subway's (Line 1) service corridor before and after its opening. Within the observed period, we can see that: (1) The daily volume of taxi ridership near the service corridor of the subway line shows a decrease since the subway opening, indicating a sign of substitution effect, which means people would choose subway as their means of transportation rather than taxi once the subway was in operation. (2) There is a fluctuation in the ridership of the subway during the weekends, while the subway's ridership on weekdays remains steady (Figure 5.4). Specifically, there is an obvious drop in the volume of subway ridership in the second weekend compared to the first weekend. One explanation is that people took the subway during weekends as a curiosity or trial to test if the subway is convenient since the subway had just opened. After that, people would choose their preferred way of traveling, thus there were not as many passengers the second weekend as there were in the first weekend.

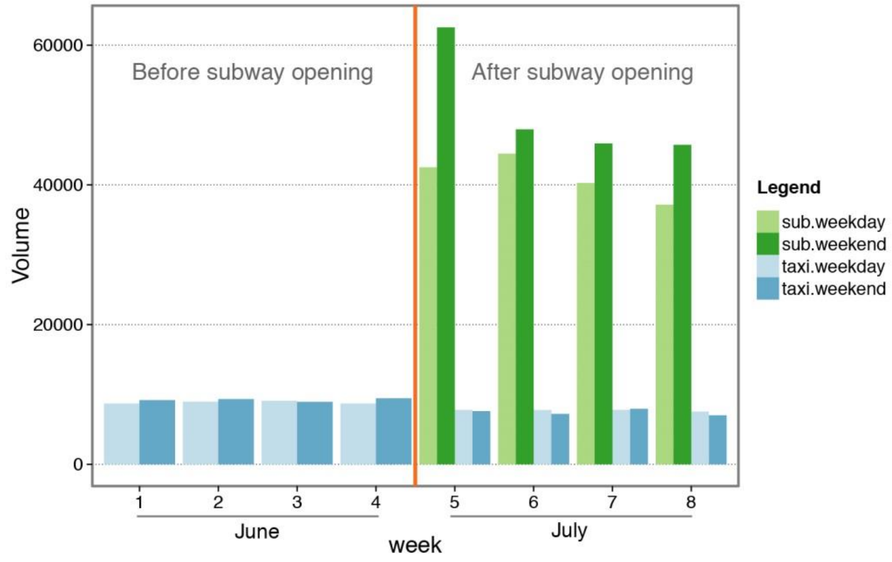


Fig 5. 4 Comparison of taxi and subway ridership.

Table 5. 3 Total daily taxi and subway volume.

	Before Subway Open		**After Subway Open**	
	Weekday	Weekend	Weekday	Weekend
Taxi	8,860 (202) *	9,245 (182)	7,680 (116)	7,435 (420)
Subway	-	-	41,000 (3,066)	50,528 (8,086)

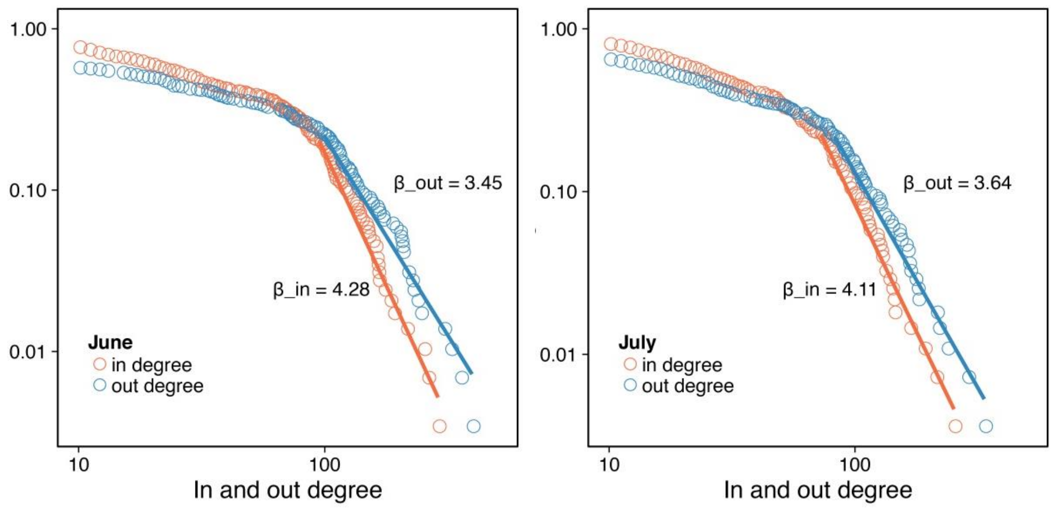
* Note: Number within brackets is standard error.

In general, if we compare the change in total volume (see Table 5.3), the subway volume is around 40,000–50,000 persons/day, which is much larger than the reduction in taxi volume, being only 1,200 on weekdays and 1,800 on weekends (note that we only use 1500 taxis, accounting for about 40% of the total number of taxis). Therefore, the operation of the subway may have more impact or substitution effect on other transportation means such as buses or private cars. In other words, a number of people who took a bus or private car before are more likely to choose the subway after its opening. At the same time, people's travel demands are very likely stimulated by the new subway, which led to the huge differences between the ridership increase in subway and decrease in taxi.

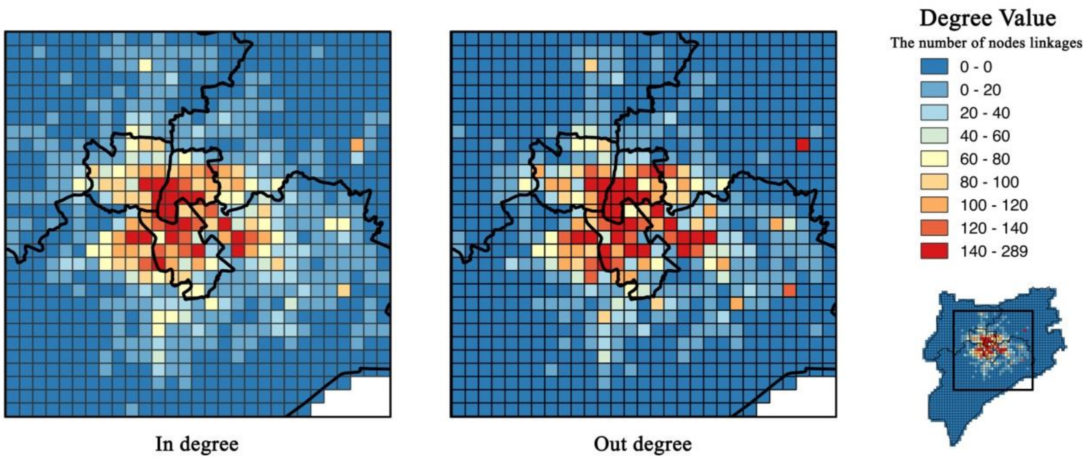
To add a more robust check, we also perform the above-mentioned analysis to the datasets of April and September, two typical months without a long public holiday. April is two months before the subway's opening, and September is two months after. The mean daily taxi volume of April is very close to that of June, and September's is very close to that of July (the difference is less than 5%), showing that the proposed method is quite promising.

5.3.2 Subway Influence on Taxi Network Structure

We move our scope onto the taxi network in the entire urban area and examine the influence of subway on taxi network structure. Figure 5.5(a) shows the cumulative in- and out-degrees of the taxi network distributions before and after the subway was in use, and Figure 5.5(b) shows the geospatial distribution of in- and out-degrees values of nodes linkages of the taxi networks distributions before the subway was in use. As shown in previous research, degree distribution is one of the most important indicators for network structure (Barthélemy M., 2011). In our case, the cumulative degree distributions of the taxi network are well fitted by power-law distribution, $P(d) \sim d^{-\beta}$, with $\beta \approx 4.28$ and 3.45 for June in-degree and June out-degree respectively (the fitting method is the maximum likelihood estimation). We find that the July out-degree distribution increases to 3.64 after the subway's opening, while the July in-degree decreases to 4.11 (Fig 5.5a), indicating that given the total volume of taxi ridership decreases after the subway's opening, and the effects of the newly opened subway are also rippling through the travel patterns of taxis. In addition, spatial distributions of the in-degree and out-degree (Fig 5.5b) identifies that the intensity of the distribution of in- and out-degrees value decreases from the city center to the suburbs, which implies the mono-centric nature of Wuxi's urban structure. We also identify hot spots from Figure 5.5b, such as *Sanyang Square*, *Nanchansi* and the train station. In addition, *Sanyang Square* was chosen as the interchange station with Line 2 in the following transportation construction, which may partially verify our conclusion.



(a)



(b)

Fig 5. 5 (a) Cumulative degree distribution of spatial connections in nodes linkages of taxi networks before and after the opening of the subway. Left: June cumulative in- and out-degree distribution (log-log); right: July cumulative in- and out-degree distribution (log-log). The straight line in each graph is the power-law fitting line; (b) Geographical distribution of in-degree (left) and out-degree (right) before the opening of the subway. Left: June in-degree distribution; right: June out-degree distribution. (All colors are in the same scale). The degree value is the number of nodes (not edge) linkages rather than O-D counts numbers, which denotes the degree of spatial connection.

When examining the subway's impact on taxi trips, we also need to consider the changing volume of taxi pick-ups and drop-offs before and after the subway's opening. In order to evaluate the balance between taxi pick-ups and drop-offs in each cell, we created an index, η, to represent the degree of balance for taxi pick-ups and drop-offs. η is equal to drop-off divided by pick-up:

$$\eta_i = \frac{Drop_i}{Pick_i}, \text{(5. 1)}$$

where $Drop_i$ equals the number of drop-offs in cell i, and $Pick_i$ equals the number of pick-ups in cell i. If pick-ups are equal to drop-offs, η will be equal to 1, which implies that the pick-up and drop-off of taxi trips are balanced in the selected area. Table 5.4 shows the statistical result of η based on Equation 5.1. However, we cannot directly compare numbers in Table 5.4, since cells with few trips may have very large or small η_i factor, which may bias the result. In order to remove this effect and get the overall balance index, we calculate the overall balance factor η_{total} by adding each cell's trip percentage as weight:

$$\eta_{total} = \frac{\sum_i (Drop_i + Pick_i) \times \eta_i}{\sum_i (Drop_i + Pick_i)}. \quad \text{(5. 2)}$$

Based on Equation 5.2, the value of η_{total} increases from 1.09 to 1.18 after the opening of subway Line 1, straying away from 1, inferring the increasingly unbalanced taxi pick-ups and drop-offs. These findings show that the opening of a new subway line may produce a certain decremental effect on the previous taxi trip pattern, e.g., spatial distribution of pick-ups and drop-offs.

Table 5. 4 Taxi pick-ups and drop-offs (weekday) before and after the subway's opening.

	Minimum	**Median**	**Mean**	**Maximum**
Before	0.37	1.63	2.48	16.00
After	0.25	1.51	2.87	81.00

5.3.3 Subway Influence on Travel Patterns

After looking into the variation of taxi volume and pick-ups/drop-offs, we zoom into the variation of taxi travel patterns since the subway was put into use. Figure 5.6 summarizes the change rate of taxi ridership after the opening of subway Line 1. Figure 5.6(a) depicts the taxi O–D diagram. There are four modes: (i) A trip from other areas to a subway station (note as $O2A$); (ii) A trip from a subway station to other areas ($A2O$); (iii) A trip from station A to all other stations ($A2\bar{A}$, where $\bar{A}$ represents the sum of all other stations except A); and (iv) A trip from all other stations to station A ($\bar{A}2A$). The ridership change rate is shown in a heat map, Figure 6c, where the x-axis represents eight different travel modes (four modes times two time periods), the y-axis represents 24 stations, and the red and blue grids represent the rate of increase and decrease in ridership respectively. For example, the first column "A2O Weekday" is the ridership change rate of Mode i during the weekday.

Figure 5.6(b) shows that most of the taxi trip volume between stations drops dramatically, and the farther from the city center (*Sanyang Square* and *Nanchansi*), the greater the substitution effect of the subway (the color ranges from light blue to dark blue). Almost 40%–50% of the taxi volume between stations decreases in suburban areas, such as *Xibei* and *Jiangnan University* areas. Another important finding is that the closer to the station, the greater the substitution effect of the subway. The first four columns of Figure 5.6(b) represent ridership change rate between one station and other places, while the last four represent ridership change rate between stations. It is obvious that taxi ridership between subway stations declines much more than that of the first four columns. Interestingly, there is a general increase in the taxi trips between other areas and a subway station in the urban fringe (red box in Fig 5.6(b)), indicating that people may tend to use taxis as a connection between stations and other areas without subway service, covering the last mile of their trips. Figure 5.6(a) depicts a possible scenario to explain this travel behavior, showing the preference that people will cover most of their travel demand by subway while only using taxi for the last mile connection.

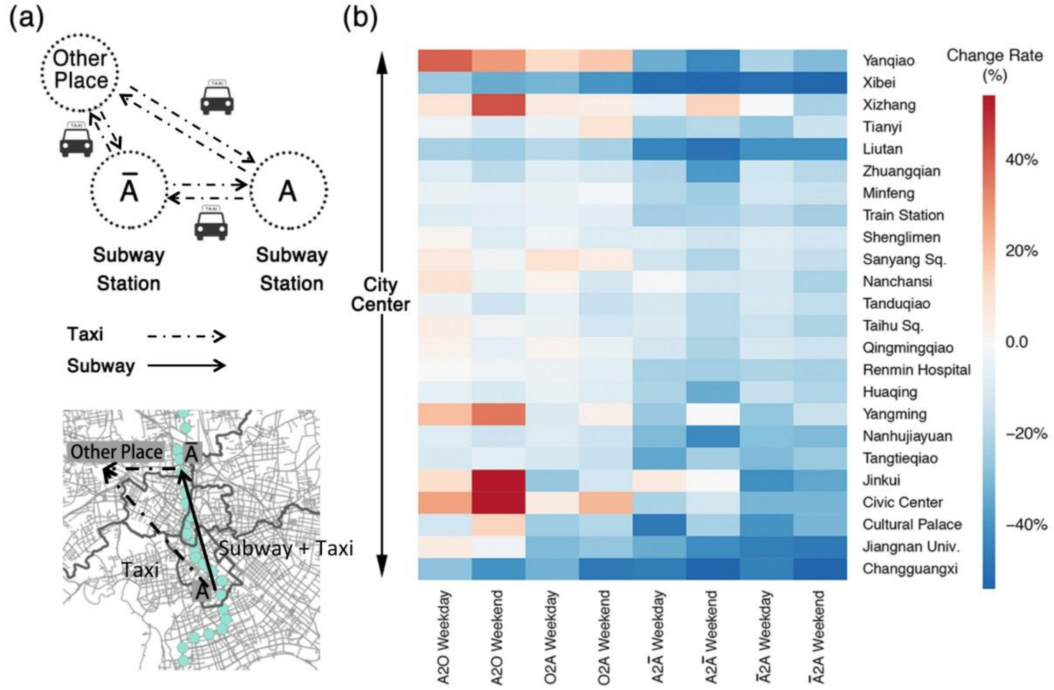


Fig 5. 6 Taxi ridership patterns. (a) Taxi O–D diagram. A to O(A2O) means from station A to Other place by taxi, O to A(O2A) means from other place to station A by taxi, A to A¯(A2A¯) means from station A to A¯by taxi, A¯to A (A¯2A) means from station A¯to A by taxi, where A¯represents the sum of all other stations except A; People may choose subway + taxi for longer distance trips after the opening of subway; (b) The heat map of taxi volume variations. X-axis represents eight different travel modes (four modes mentioned above (a), two time periods are Weekday and Weekend). Y-axis represents 24 stations, with Sanyang Square and Nanchansi being the urban central areas. The red and blue indicate increasing and decreasing rates respectively.

5.4 Summary of This Chapter

This study quantitatively examines the impact of the first and newly opened subway, using spatio-temporal network analysis based on the combination of emerging taxi trips and subway transaction records, which can lead to a deeper understanding of human mobility and sustainable urban development as well as provide informative insights in population mobility and urban configuration. We calculate the subway network and evaluate its impact on the surrounding taxi volume and analyze the influence of the subway on the taxi network structure (pick-ups and drop-offs) by defining an index to analyze the taxi imbalance issue within the city. Further, we divide the taxi trips into four modes to see where the subway impacted taxi ridership. Finally, from our findings we conclude that the farther away from the city center, the greater the impact; the closer to the subway station, the greater the impact; and we find that not all taxi ridership decreased but in fact increased in some locations, which was explained with analysis.

We identify that there is more subway ridership on weekends than on weekdays. One reason, from the demand side, is that people are more flexible and willing to take the subway over a taxi during weekends, since punctuality on weekends is not as crucial as it is for fixed business hours on weekdays. The other reason, from the service side, is that the new subway Line 1 serves a corridor mainly along more consumption-oriented places rather than job-oriented places, from the city center extending to the northern and southern parts of the suburbs. Therefore, this ridership pattern may demonstrate the variability in travel behavior within a week resulting from work–home separation.

When examining the interactions between subway and taxi in a spatial dimension, we notice the farther from the city center, the greater the substitution effect of the subway. In addition, the taxi trip volume decreases most drastically between each O–D pair if they are both close to subway stations, which could be a reduction of 40%–50% in most suburb areas within 1 km of the subway line corridor. There is a substantial shift in the taxi volume and trip pattern within a 1 km radius of the subway line along the whole subway line. Thus, the introduction of a new subway is expected to considerably upgrade public transport within the influence area it serves. On the other hand, the connection between subway stations in the suburbs and other areas of the city without subway services shows an inverse result. People who used to take a taxi for their daily commute now choose to take the combination of subway and taxi, by using a taxi to cover the last mile to their destination, which is a way of maximizing travel efficiency and cost effectiveness. However, the volume of subway ridership is much larger than the decrease of the volume of taxi ridership, indicating a latent influence upon other means of transportation from the subway.

From the result of taxi trips' O–D density in the travel distribution and the distance effect of distribution, we come up with the conclusion that Wuxi is a mono-centric city. The spatial distribution of taxi trip density coincides with the location of subway stations, revealing the most popular destinations in the central city. The city center shows a very high concentration of taxi trips compared to the urban fringe. People who live in places far from the city center generate large amounts of travel flow and they come to the nearest subway station by taxi for cheaper long-distance travel. Suburban areas that contain transfer facilities such as subway stations have the potential to be developed into sub-commercial centers. Thus, some long-distance trips could be turned into short local trips by encouraging densification and diversification of station neighborhoods in the suburban areas. Nevertheless, the distance decay effect makes the spatial distribution of the trips more concentrated. The farther the cell is from the city center, the lower the density of its taxi trips. This also suggests that although the subway might have impacts on the taxi ridership in the region close to subway line, taxi is still a crucial alternative for people to satisfy their travel demands properly. Transport connectivity is critical over wider spatial ranges in determining subway ridership.

From a policy perspective, this study may suggest several policy implications in aspects of urban planning for sustainable urbanization. Promoting more sustainable patterns of urban development is also crucial for improving the subway ridership of cities but the appropriateness of different forms of development is context-dependent. There is increasing recognition that combinations (or packages) of measures are necessary. Certain combinations of policies can work together and give rise to synergies, leading to more sustainable urban transport. Finally, caution is advised both in terms of the appropriateness and effectiveness of policy solutions being transferred. This study also has practical implications for urban planning and management, which contributes to a better understanding of people's travel behavior and ways to balance the demand between subway ridership and taxi trips. Moderate interventions by spatial planning could improve subway ridership and efficiency for sustainable urbanization. The future development of subway systems should include new subway lines that have a greater focus on the outer suburbs where public transport dependent people are concentrated. This would be beneficial for peripheral residents who have less capacity to adjust their housing locations to secure connectivity to the city center through the subway.

Moreover, a promising direction is to utilize spatio-temporal big data with regard to human mobility associating subway and taxi usages. This can broaden the literature of human mobility, origin–destination estimation, emerging data and public transit analysis (Huang X. et al., 2016; Chong Z. et al., 2016; Wang Y. et al., 2016; Zhang F. et al., 2016; Al-Dohuki S. et al., 2017; Ye X. et al., 2016). The findings would provide an objective bottom-up view to depict human mobility as well as new insights for traffic optimization and urban transport planning policy. In future research, it is reasonable to investigate the urban form according to land use and expand the data source to include private car and bus to explore the pathways underlying the effects of the subway line. By comparing the change in points of interest composition along the subway line, we will receive a more detailed and explicit picture of a new subway's impact on human mobility.

Chapter6

6. Conclusions

6.1 Summary

After development, planner and designer in Chinese cities have to pay more attention to find out the urban problem, understanding the reason of urban problem in current urban form is the start point for making the future plan or design of a city. For this, my PHD research work developed the theoretical framework of spatio-temporal analytics, according to it, we employ some specific methods based on spatio-temporal big datasets for identifying urban problems hidden in current urban form through analyzing human behaviors and their spatial patterns.

Firstly, we develop a comparative spatio-temporal framework by combining spatial, temporal, and statistical distributions (three dimensions) with individual, local, meso, and global levels (four scales). By customizing the combination of analysis units, the researcher can measure human mobility pattern reflected by trajectory data dynamically and comparatively. Two case studies of Chinese cities are carried out to evaluate the usefulness of proposed conceptual framework. Our results suggest that the proposed framework can comprehensively quantify the variation of human mobility across various scales and dimensions.

Meanwhile, under this framework thinking, we continue to use Taxi O-D data, combining with POIs data, to create the spatial and temporal entropy model to identify the spatial distribution pattern of urban function and its mixture. In this part, we considered individual trip behaviors in different functional zones when we had evaluated urban function mix. Furthermore, we discuss the relationship between the spatio-temporal distribution of population and urban function, that help me to deep understand different urban functions and their mix of different attraction to people.

Finally, we apply the spatio-temporal network analysis method for the comparative analysis of subway and taxi ridership and their interactions in urban area, this chapter aims to examine the spatial variation of urban taxi ridership due to the impacts of a new subway line. We examine the spatio-temporal patterns and interactions of ridership in Wuxi by integrating taxi O-D trips from taxi GPS data and subway smart card data from continuously collected fare transactions.

However, in this stage, we just proposed the spatio-temporal analytics framework and from the perspective of human behaviors according to this analytics framework to explain spatial distribution pattern of urban function and its mixture, spatio-temporal distribution pattern of population and correlation with urban function, spatio-temporal patterns and interactions between subway and taxi ridership. In this PhD dissertation, we have not done the comprehensive analysis of the urban problem and also haven't found solutions for urban problem. But this framework and analyzing methods can provide the good support for the urban planner and designer to consider the future plan of a city based on the findings of spatio-temporal analytics.

6.2 Future Work

For future work, we should design the complete analyzing indicator putting into this framework to comprehensive analyze urban problem. And then, we could based on open source spatial analysis software such as Open Source Package STARS (Spatio-temporal Analysis of Regional Systems) and PySAL (Opento develop the Source Python Library for Spatial Analytical Functions) (Rey and Janikas, 2006; Rey and Anselin, 2007) to develop the new tools for analyzing and evaluating urban problem. It's easy to use for urban planning and urban design practices.

Publications

- Miaoyi Li, Xinyue Ye*, Shanqi Zhang, Xiaotong Tang, Zhenjiang Shen. A Framework of Comparative Urban Trajectory Analysis[J]. Environment Planning Part B: Urban Analytics and City Science.**(SSCI IF:1.527)**
- Miaoyi Li, Zhenjiang Shen* et al. Application of Spatial and Temporal Entropy Based on Multivariate Data for Measuring the Degree of Urban Function Mix[J]. China City Planning Review, 2015, Vol.24, No.1, 40-48. **(CSCD)**
- Miaoyi Li，Zhengjiang Shen*，Xinhua Hao. Revealing the relationship between spatio-temporal distribution of population and urban function with social media data[J]. GeoJournal, 2016, Volume 81, Issue 6, 919-935. **(EI/SCImago)**
- Miaoyi Li, Lei Dong, Zhenjiang Shen*,Wei Lang, Xinyue Ye*. Examining the interaction of taxi ridership and subway for sustainable urbanization[J]. Sustainability. 2017; 9(2):242. **(SSCI/SCI IF:1.789)**

Conference:

- Miaoyi Li, Dong Lei, Zhenjiang Shen*, Wang Jingyuan, Huang Ling. Estimating the Influence of a New Subway Line with Taxi Trip Data. The 9th IACP Conference: Smart Growth and Sustainable Development(IACP, Chongqing)(2015.6)

References

Al-Dohuki, S., Kamw, F., Zhao, Y., Ma, C., Wu, Y., Yang, J., Ye, X., Wang, F., Li, X. & Chen, W. (2017) SemanticTraj: A New Approach to Interacting with Massive Taxi Trajectories, IEEE Transactions on Visualization and Computer Graphics. 23(1), 11-20.

Agryzkov, T., Martí, P., Tortosa, L., & Vicent, J. F. (2016). Measuring urban activities using Foursquare data and network analysis: a case study of Murcia (Spain). International Journal of Geographical Information Science, 8816(May), 1–22. http://doi.org/10.1080/13658816.2016.1188931

Andrienko G, Andrienko N, Demsar U et al., 2010. Spatio-temporal and visual analytics. International Journal of Geographical Information Science, 24(10): 1577–1600. doi: 10.1080/13658816.2010.508043

Andrienko G, Andrienko N, Keim D et al., 2011. Editorial: Challenging problems of geospatial visual analytics. Journal of Visual Languages and Computing, 22(4): 251–256. doi: 10.1016/j.jvlc.2011.04.001

Andris, C. (2016). Integrating social network data into GISystems. International Journal of Geographical Information Science, 8816(March), 1–23. http://doi.org/10.1080/13658816.2016.1153103

Barthélemy, M. Spatial networks. Phys. Rep. 2011, 499, 1–101.

Battellino, H.; Peachman, J. The joys and tribulations of a continuous survey. In Transport Survey Quality and Innovation; Emerald: Kruger National Park, South Africa, 2003; pp. 49–68.

Batty, M. The New Science of Cities; Mit Press: Cambridge, MA, UK, 2013.

Bohte, W.; Maat, K. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in The Netherlands. Transp. Res. Part C: Emerg. Technol. 2009, 17, 285–297.

Bonnell T R, Dutilleul P, Chapman C A et al., 2013. Analysing small-scale aggregation in animal visits in space and time: The ST-BBD method. Animal Behaviour, 85(2): 483–492. doi: 10.1016/j.anbehav.2012.12.014

Bo, W., Feng, Z., & Hao, Z. (2015). The dynamic changes of urban spatio-temporal activity and activity zoning based on check-in data in Sina web. Scientia Geographica Sinica, 35(2), 151–160.

Brög, A.R.G. Old State, new society: Indonesia's new order in comparative historical perspective. J. Asian Stud. 1983, 42, 477–496.

Calabrese, A.; Costa, R.; Menichini, T. Using Fuzzy AHP to manage Intellectual Capital assets: An application to the ICT service industry. Expert Syst. Appl. 2013, 40, 3747–3755.

Cervero, R. Mixed land-uses and commuting: Evidence from the American Housing Survey. Transp. Res. Part A: Policy Pract. 1996, 30, 361–377.

Chen, C.; Bian, L.; Ma, J. From traces to trajectories: How well can we guess activity locations from mobile phone traces? Transp. Res. Part C: Emerg. Technol. 2014, 46, 326–337.

Chen X, Kwan M P, 2012. Choice set formation with multiple flexible activities under spatio-temporal constraints. International Journal of Geographical Information Science, 26(5): 941–961. doi: 10.1080/13658816.2011.624520

Chong, Z.; Qin, C.; Ye, X. Environmental regulation, economic network and sustainable growth of urban agglomerations in China. Sustainability 2016, 8, 467.

Chowdhury, D.; Santen, L.; Schadschneider, A. Statistical physics of vehicular traffic and some related systems. Phys. Rep. 2000, 329, 199–329.

Cichocki, A.; Zdunek, R.; Phan, A.H.; Amari, S.I. Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation; John Wiley & Sons: West Sussex, UK, 2009.

Ciscal-Terry, W.; Dell'Amico, M.; Hadjidimitriou, N.S.; Iori, M. An analysis of drivers route choice behavior using GPS data and optimal alternatives. J. Trans. Geogr. 2016, 51, 119–129.

Cranshaw, J., Hong, J. I., & Sadeh, N. (2012). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (pp. 58–65).

Dempwolf, C. S., & Lyles, L. W. (2012). The Uses of Social Network Analysis in Planning: A Review of the Literature. Journal of Planning Literature, 27(1), 3–21. http://doi.org/10.1177/0885412211411092

Deville, P., Linard, C., Martin, S., et al. (2014). Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences, 111(45), 15888–15893.

Downs J A, Horner M W, 2012. Probabilistic potential path trees for visualizing and analyzing vehicle tracking data. Journal of Transport Geography, doi: 10.1016/j.jtrangeo.2012.03.017

Fan, Z.; Song, X.; Shibasaki, R. CitySpectrum: A non-negative tensor factorization approach. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, NY, USA, 13 September 2014.

Farber S, Neutens T, Miller H J et al, 2012. The social interaction potential of metropolitan regions: A time-geographic measurement approach using joint accessibility. Annals of the Association of American Geographers, 103(3): 483–504. doi: 10.1080/00045608.2012.689238

Frank, L. D., & Pivo, G. (1994). Impacts of mixed use and density utilization of three modes of travel: Single-occupant vehicle, transit, and walking. Transportation Research Record, 1466, 44–52.

Gao S, Liu Y, Wang Y et al., 2013. Discovering spatial interaction communities from mobile phone data. Transactions in GIS, 17(3): 463–481. doi: 10.1111/tgis.12042

Girardin, F., et al. 2008. Digital footprinting: uncovering tourists with user-generated content. IEEE Pervasive Computing, 7 (4), 36–43. doi:10.1109/MPRV.2008.71

Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. Nature 2008, 453, 779–782.

González, M. C., Hidalgo, C. A., Albert-László B(2008). “Understanding individual human mobility patterns”.Nature, 453(7196),779-782 .

Guo D, Zhu X, Jin H et al., 2012a. Discovering spatial patterns in origin-destination mobility data.

Transactions in GIS, 16(3): 411–429. doi: 10.1111/j.1467-9671.2012.01344.x

Guo J, Chen H, Xu C Y et al., 2012b. Prediction of variability of precipitation in the Yangtze River Basin under the climate change conditions based on automated statistical downscaling. Stochastic Environmental Research and Risk Assessment, 26(2): 157–176. doi: 10.1007/s00477-011-0464-x

Hägerstrand T, 1970. What about people in regional science? Papers of the Regional Science Association, 24(1): 6–21. doi: 10.1007/BF01936872

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: Views from urban planning. American Journal of Preventive Medicine, 23(2 SUPPL. 1), 64–73. http://doi.org/10.1016/S0749-3797(02)00475-0

Hao, T., Ma, X.J., et al.(2010). "A Novel Approach to Estimate Human Spatio-temporal Path Based on Mobile Phone Call Records",in: 18th International Conference on Geoinformatics, Beijing,China, IEEE, 1-6.

Harvey, D. The New Imperialism; Oxford University Press: New York, NY, USA, 2003.

Horner M W, Zook B, Downs J A, 2012. Where were you? Development of a time-geographic approach for activity destination re-construction. Computers, Environment and Urban Systems, 36(6): 488–499. doi: 10.1016/j.compenvurbsys.2012.06.002

Harvey, C., & Aultman-Hall, L. (2015). Measuring Urban Streetscapes for Livability: A Review of Approaches. The Professional Geographer, (August), 1–10. http://doi.org/10.1080/00330124.2015.1065546

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr to describe city cores. Journal of Spatial Information Science, 1(1), 21–48. http://doi.org/10.5311/JOSIS.2010.1.3

Huang, Q., & Wong, D. W. S. (2016). Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? International Journal of Geographical Information Science, 8816(March), 1–26. http://doi.org/10.1080/13658816.2016.1145225

Huang, X.; Zhao, Y.; Yang, J.; Zhang, C.; Ma, C.; Ye, X. TrajGraph: A graph-based visual analytics approach to studying urban network centralities using taxi trajectory data. Vis. Comput. Graph. 2016, 22, 160–169.

Jones O, Cloke P, 2008. Non-human agencies: Trees in place and time. In: Malafouris L et al. (eds.). Material Agency. Springer US, 79–96.

Kain, J.F.; Liu, Z. Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. Trans. Res. Part A: Policy Pract. 1999, 33, 601–624.

Kang, C.; Sobolevsky, S.; Liu, Y.; Ratti, C. Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages. In Proceedings of the 2nd ACMSIGKDD InternationalWorkshop on Urban Computing, New York, NY, USA, 11 August 2013.

Kang, C., Liu, Y., Ma, X., et al. (2012). Towards estimating urban population distributions from mobile call data. Journal of Urban Technology, 19(4), 3–21.

Kim, S.; Ulfarsson, G.F.; Hennessy, J.T. Analysis of light rail rider travel behavior: impacts of individual, built environment, and crime characteristics on transit access. Trans. Res. Part A: Policy Pract. 2007, 41, 511–522.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. GeoJournal, 79(1), 1–14.

Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer networks. J. Complex Netw. 2014, 2, 203–271.

Knowles, R.D. Transport impacts of Greater Manchester's Metrolink light rail system. J. Trans. Geogr. 1996, 4, 1–4.

Kuijpers B, Grimson R, Othman W, 2011. An analytic solution to the alibi query in the spatio-temporal prisms model for moving object data. International Journal of Geographical Information Science, 25(2): 293–322. doi: 10.1080/13658810902967397

Kuijpers B, Miller H J, Neutens T et al., 2010. Anchor uncertainty and spatio-temporal prisms on road networks. International Journal of Geographical Information Science, 24(8): 1223–1248. doi: 10.1080/13658810903321339

Kuijpers B, Othman W, 2009. Modeling uncertainty of moving objects on road networks via spatio-temporal prisms. International Journal of Geographical Information Science, 23(9): 1095–1117. doi: 10.1080/13658810802097485

Kwan, M.P. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. Transp. Res. Part C: Emerg. Technol. 2000, 8, 185–203.

Kwan M P, 2000a. Human extensibility and individual hybrid-accessibility in spatio-temporal: A multi-scale representation using GIS. In: Janelle D et al. (eds.). Information, Place, and Cyberspace. Springer Berlin Heidelberg, 241–256.

Kwan M P, 2000b. Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set. Transportation Research Part C: Emerging Technologies, 8(1–6): 185–203. doi: 10.1016/S0968-090X(00)00017-6

Kwan M P, Lee J, 2003. Geovisualization of human activity patterns using 3D GIS: A time-geographic approach. In: Goodchild M et al. (eds.). Spatially Integrated Social Science: Examples in Best Practice. Oxford: Oxford University Press.

Huang, L., Zhu, X., Ye, X., Guo, W., & Wang, J. (2015). Characterizing street hierarchies through network analysis and large-scale taxi traffic flow: A case study of Wuhan, China. Environment and Planning B: Planning and Design, doi: 10.1177/0265813515614456

Huang, X., Zhao, Y., Yang, J., Zhang, C., Ma, C., & Ye, X. (2016). TrajGraph: A Graph-Based Visual Analytics Approach to Studying Urban Network Centralities Using Taxi Trajectory Data. IEEE Transactions on Visualization and Computer Graphics, 22(1), 160-169

Lathia, N. and Capra, L. 2011. How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives.Proceedings of the 13th international conference on ubiquitous computing, UbiComp’11, 17–21 September, Beijing, China, 291–300.

Lee, T. (1968). Urban neighourhood as a socio-spatial schema. Human Relations, 21(3), 241–267.

Li, M., Dong, L., Shen, Z., Lang, W., & Ye, X. (2017) Examining the interaction of taxi ridership and subway for sustainable urbanization. Sustainability. doi:10.3390/su9020242

Li, M., Shen, Z, & Tao, Y. (2015). Application of spatial and temporal entropy based on multivariate data for measuring the degree of urban function mix. China City Planning Review, 24(1), 40–48.

Lin, D.; Allan, A.; Cui, J. Exploring Differences in Commuting Behaviour among Various Income Groups during Polycentric Urban Development in China: New Evidence and Its Implications. Sustainability 2016, 8, 1188.

Liu, L., Hou, A., Biderman, A., et al. (2009). Understanding individual and collective mobility patterns from smart card records: A case study in Shenzhen. In 12th international IEEE conference on intelligent transportation systems, 2009. ITSC'09. IEEE 2009 (pp. 1–6).

Liu, L., Andris, C., and Ratti, C., 2010. Uncovering cabdrivers' behavior patterns from their digital traces. Computers, Environment and Urban Systems, 34 (6), 541–548. doi:10.1016/j.compenvurbsys.2010.07.004

Liu, X.; Gong, L.; Gong, Y.; Liu, Y. Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr. 2015, 43, 78–90.

Liu, Y., Wang, F., Xiao, Y., et al. (2014). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. Landscape Urban Plan, 106(1),73–87.

Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PloS One, 9(1), e86026. http://doi.org/10.1371/journal.pone.0086026

Liu, Y.;Wang, F.; Xiao, Y.; Gao, S. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. Landsc. Urban Plan. 2012, 106, 73–87.

Long J A, Nelson T A, 2012a. A review of quantitative methods for movement data. International Journal of Geographical Information Science, 27(2): 292–318. doi: 10.1080/13658816.2012.682578

Long J A, Nelson T A, 2012b. Time geography and wildlife home range delineation. The Journal of

Wildlife Management, 76(2): 407–413. doi: 10.1002/jwmg.259

Long, Y.; Lang, W. Fine-scale human settlement studies of China in the new data environment. J. Urban Reg. Plan. 2016, 8, 10–32.

Long Y and Liu X (2013). “How Mixed is Beijing, China? A Visual Exploration of Mixed Land Use”. Environment and Planning A . 45, pp. 2797 – 2798.

Long, Y., & Thill, J. (2015). Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. Computers, Environment and Urban Systems, 53, 19–35.

Long, Y., Zhang, Y., & Cui, C. Y. (2012). Identifying commuting pattern of Beijing using bus smart card data. Acta Geographica Sinica, 67(10), 1339–1352.

Louail, T., Lenormand, M., Ros, O. G. C., et al. (2014). From mobile phone data to the spatial structure of cities. SCI REP-UK, 4.

Luo, F., Cao, G., Mulligan, K., & Li, X. (2015). Explore Spatio-temporal and Demographic Characteristics of Human Mobility via Twitter: A Case Study of Chicago. Applied Geography, 70, 11–25. http://doi.org/10.1016/j.apgeog.2016.03.001

Lou, Y.; Yin, Y. A decomposition scheme for estimating dynamic origin–destination flows on actuationcontrolled signalized arterials. Transp. Res. Part C: Emerg. Technol. 2010, 18, 643–655.

Mackett, R.L.; Edwards, M. The impact of new urban public transport systems: Will the expectations be met? Trans. Res. Part A: Policy Pract. 1998, 32, 231–245.

Markovich, J.; Lucas, K. The Social and Distributional Impacts of Transport: A Literature Review; Working Paper; Transport Studies Unit School of Geography and the Environment, University of Oxford: Oxford, UK, 2011.

Miller H J, 2005. A Measurement Theory for Time Geography. Geographical Analysis, 37(1): 17–45. doi: 10.1111/j.1538-4632.2005.00575.x

Murray A, Liu Y, Rey S et al., 2012. Exploring movement object patterns. The Annals of Regional Science, 49(2): 471–484. doi: 10.1007/s00168-011-0459-z

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. Society for Industrial and Applied Mathematics (SIAM) Review, 45(2), pp. 167–256.

Ning, X., Ling, Y.,&Jinxing, H. (2014). Identifying home-work locations from short-term, largescale, and rgularly sampled mobile phone tracking data. Geomatics and Information Science of Wuhan Univers, 39(6), 750–756.

Pei, T., Sobolevsky, S., Ratti, C., et al. (2014). A new insight into land use classification based on aggregated mobile phone data. International Journal of Geographical Information Science, 28(9), 1988–2007.

Pojani, D.; Stead, D. Sustainable urban transport in the developing world: Beyond megacities. Sustainability. 2015, 7, 7784–7805.

Rajamani J, Bhat C R, et al. (2003). Assessing Impact of Urban Form Measures on Nonwork Trip Mode Choice After Controlling for Demographic and Level-of-Service Effects. Transportation Research Record: Journal of the Transportation Research Board . 1831, pp. 158 – 165.

Ratcliffe J H, 2006. A temporal constraint theory to explain opportunity-based spatial offending patterns. Journal of Research in Crime and Delinquency, 43(3): 261–291.

Ratti, C.; Frenchman, D.; Pulselli, R.M.; Williams, S. Mobile landscapes: using location data from cell phones for urban analysis. Environ. Plan. B: Plan. Des. 2006, 33, 727–748.

Reades, J., Calabrese, F., & Ratti, C. (2009). Eigenplaces: Analysing cities using the spatio-temporal structure of the mobile phone network. Environment and Planning B: Planning and Design, 36(5), 824–836.

Robertson, C., & Feick, R. (2015). Bumps and bruises in the digital skins of cities: unevenly distributed user-generated content across US urban areas. Cartography and Geographic Information Science, http://doi.org/10.1080/15230406.2015.1088801

Schneider, C.M.; Belik, V.; Couronné, T.; Smoreda, Z.; González, M.C. Unravelling daily human mobility motifs. J. R. Soc. Interface 2013, 10, 20130246.

Shaw S L, Yu H, 2009. A GIS-based time-geographic approach of studying individual activities and interactions in a hybrid physical–virtual space. Journal of Transport Geography, 17(2): 141–149.

She, B., Duque, J., & Ye, X. (2016) The Network-Max-P-Regions Model. International Journal of Geographical Information Science doi: 10.1080/13658816.2016.1252987

Song C, Qu Z, Blumm N et al., 2010. Limits of predictability in human mobility. Science, 327(5968): 1018–1021.

Spielman, & Thill. (2008). Social area analysis, data mining, and GIS. Computers Environment and Urban Systems, 32(2), 110–122.

Steenbruggen, J., Borzacchiello, M. T., Nijkamp, P., et al. (2013). Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: A review of applications and opportunities. GeoJournal, 78(2), 223–243.

Stopher, P.R.; Greaves, S.P. Household travel surveys: Where are we going? Trans. Res. Part A: Policy Pract. 2007, 41, 367–381.

Suhong, Z. (2015). Study on dynamic structure of cities based on questionnaire and floating cars' GPS data. In Symposium on big data and spatio-temporal behavior planning, Tongji University, Shanghai 2015-01-04 (in Chinese).

Sui D, 2012. Looking through Hägerstrand′s dual vistas: Towards a unifying framework for time geography. Journal of Transport Geography, doi: 10.1016/j.jtrangeo.2012.03.020

Timmermans H, Arentze T, Joh C H, 2002. Analysing spatio-temporal behaviour: New approaches to old problems. Progress in Human Geography, 26(2): 175–190.

Troped, P. J., Wilson, J. S., Matthews, C. E., Cromley, E. K., & Melly, S. J. (2010). The Built Environment and Location-Based Physical Activity. American Journal of Preventive Medicine, 38(4), 429–438.

Wang, M., Zhou, J., Long, Y., & Chen, F. (2016). Outside the ivory tower: Visualizing university students' top transittrip destinations and popular corridors. Regional Studies, Regional Science, 3(1), 202–206.

Wang, Y., Jiang, W., Liu, S., Ye, X, & Wang, T. (2016a) Evaluating trade areas using social media data with a calibrated Huff model. ISPRS International Journal of Geo-Information. doi:

10.3390/ijgi5070112

Wang, Y.;Wang, T.; Ye, X.; Zhu, J.; Lee, J. Using social media for emergency response and urban sustainability: A case study of the 2012 Beijing rainstorm. Sustainability 2016, 8, 25.

Wang, Z., Ye, X, & Tsou, M. (2016b) Spatial, temporal, and content analysis of Twitter for wildfire hazards. Natural Hazards doi: 10.1007/s11069-016-2329-6

Willumsen, L.G. Estimation of an OD Matrix from Traffic Counts–A Review; Institute of Transport Studies; University of Leeds: Leeds, UK, 1978; Working paper 99.

Winter S, Raubal M, 2006. Time geography for Ad Hoc shared-tide trip planning. Proceedings of the 7th International Conference on Mobile Data Management, Washington

Winter S, Yin Z C, 2011. The elements of probabilistic time geography. GeoInformatica, 15(3): 417–434.

Wolf, J.; Loechl, M.; Thompson, M.; Arce, C. Trip rate analysis in GPS-enhanced personal travel surveys. In Transport Survey Quality and Innovation; Elsevier Science: Oxford, Kruger Park, South Africa, 2013; Volume 28, pp. 483–498.

Woodard, D., Nogin, G., Koch, P., Racz, D., Goldszmidt, M., & Horvitz, E. (2017). Predicting travel time reliability using mobile phone GPS data. Transportation Research Part C: Emerging Technologies, 75, 30-44.

Xia, M., Rongrong, X., Xinshuo, L., et al. (2010). Fine grid dynamic features of population distribution in shenzhen. Acta Geographica Sinica, 65(4), 443–453.

Xinyi, N., Liang, D., & Xiaodong, S. (2014). Understanding urban spatial structure of shanghai central city based on mobile phone data. Urban Planning Forum, (6), 61–67.

Yang, H.; Lau, Y.W.;Wong, S.C.; Lo, H.K. A macroscopic taxi model for passenger demand, taxi utilization and level of services. Transportation 2000, 27, 317–340.

Yao, X., & Zhang, S. (2014). Social-spatial structure of Beijing: a spatial-temporal analysis. International Journal of Society Systems Science. http://doi.org/10.1504/IJSSS.2014.059923

Ye, X., Huang, Q., & Li, W. (2016) Integrating Big Social Data, Computing, and Modeling for

Spatial Social Science, Cartography and Geographic Information Science, doi: 10.1080/15230406.2016.1212302

Ye, X. and Rey, S. J. (2013) A Framework for Exploratory Spatio-temporal Analysis of Economic Data. Annals of Regional Science, 50(1): 315-339

Yuan, J., Zheng, Y., & Xie, X. (2012). Discovering regions of different functions in a city using human mobility and POIs. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12 (p. 186). Beijing, China. http://doi.org/10.1145/2339530.2339561

Yuan, N. J., Zheng, Y., Xie, X., et al. (2015). Discovering urban functional zones using latent activity trajectories. IEEE Transactions on Knowledge and Data Engineering (TKDE), 3(27), 1041–4347.

Yu H, Shaw S L, 2008. Exploring potential human activities in physical and virtual spaces: A spatio-temporal GIS approach. International Journal of Geographical Information Science, 22(4): 409–430. doi: 10.1080/13658810701427569

Yuan, Y.; Raubal, M.; Liu, Y. Correlating mobile phone usage and travel behavior–A case study of Harbin, China. Comput. Environ. Urban Syst. 2012, 36, 118–130.

Zhang, F.; Zheng, Y.; Xu, D.; Du, Z.; Wang, Y.; Liu, R.; Ye, X. Real-time spatial queries for moving objects using storm topology. ISPRS Int. J. Geo-Inf. 2016.

Zhang, F., Zhu X., Guo W., Ye, X, Hu T., & Huang, L. (2016). Analyzing Urban Human Mobility Patterns through Thematic Model at the Finer Scale. ISPRS International Journal of Geo-Information. 5(6), 78; doi: 10.3390/ijgi5060078

Zhao, P., Qin, K., Ye, X., Wang, Y, & Chen, Y. (2016) A Trajectory Clustering Approach Based on Decision Graph and Data Field for Detecting Hotspots. International Journal of Geographical Information Science doi: 10.1080/13658816.2016.1213845

Zhou, X.; Mahmassani, H.S. Dynamic origin-destination demand estimation using automatic vehicle identification data. IEEE Trans. Intell. Transp. Syst. 2006, 7, 105–114.