EXPLORING PEDESTRIAN MOVEMENT PATTERNS WITH URBAN ENVIRONMENTAL FACTORS IN BEIJING

A Thesis in
Landscape Architecture

by

Qixiang Qin

© 2016 Qixiang Qin

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

August 2016
The thesis of Qixiang Qin was reviewed and approved* by the following:

Neil Korostoff  
Associate Professor of Landscape Architecture  
Thesis Advisor

Clio Andris  
Assistant Professor of Geography

Mallika Bose  
Associate Professor of Landscape Architecture

Frederico Fonseca  
Associate Professor of Information Sciences and Technology  
Director, Center for Innovative Online Learning  
Affiliate Assistant Professor of Geography

Eliza Pennypacker  
Professor of Landscape Architecture  
Head of the Department of Landscape Architecture

*Signatures are on file in the Graduate School
ABSTRACT

This research is an exploratory investigation to understand Beijing citizens’ travel modes by looking into their daily movement patterns. This research also intends to study pedestrian route selection criteria and the properties of their walking trajectories with different environment attributes (the mean distances to different facilities, mixed land use, population density, and junction density). People’s travel modes and wayfinding behaviors are affected by both internal personal characteristics and external environmental factors. In other worlds, the spatial distributions of travel destinations and the spatial patterns of walking trajectories reflect people’s lifestyles and preferences about the walking environment. As for the travel mode analysis, this thesis studies the distributions of respondents travel origins and destinations with their activity radiiuses, and the trajectory study here focuses on typical segments (e.g., stops) within trajectories and different trajectory types. In contrast to the experimental approaches, interviews and surveys, this research project takes more advantage of social media data and GPS data. In order to discover pedestrian’s route selection mechanism, travel mode and urban characteristics that are reflected in these typical trajectory segments, a method using mapping, GIS spatial statistics, and statistical tests is proposed. Taking Beijing’s urban area as a sample research area, this research finds out that people’s motivations of exploratory movements are correlated with their routine activity radiiuses and people prefer the walking environment to be less complicated. Moreover, through studying different surrounding environment along pedestrian trajectories, to some extent, Beijing still maintains its traditional “introverted” urban character that the land use patterns are more diverse inside blocks than along streets. By discovering pedestrian preferences in walking environment and urban characteristic which are reflected in people’s movements, the thesis also provides suggestions for building a more workable Beijing in the perspective of urban design and urban planning.
TABLE OF CONTENTS

List of Figures .......................................................................................................................... v
List of Tables ........................................................................................................................... vi
Acknowledgements .................................................................................................................. vii
Chapter 1 Introduction and Literature Review ........................................................................ 1
  1.1 Introduction ................................................................................................................ 1
  1.2 A Review of Research on Environmental Influences on People Movements .......... 6
  1.3 A Note on New Data Sources and Data Types .......................................................... 12
  1.4 Research Overview ............................................................................................... 14
  1.5 Significance .......................................................................................................... 15
  1.6 Definitions of Terms ........................................................................................... 17
Chapter 2 Introduction to Beijing and Research Questions ..................................................... 19
  2.1 An Introduction of Beijing and Its Urban Life .......................................................... 19
  2.2 Problem Statements ............................................................................................ 27
Chapter 3 Data and Methodology ............................................................................................ 29
  3.1 Data Collecting Strategy .......................................................................................... 29
    3.1.1 Trajectory Data ................................................................................................ 32
    3.1.2 Geo-Tagged / Check-in Data for Population Density ..................................... 38
    3.1.3 POI Data ........................................................................................................ 41
    3.1.4 Data for Mixed Land Use Map ..................................................................... 42
    3.1.5 Data of Green Spaces and Junctions ............................................................... 42
  3.2 Data Processing .......................................................................................................... 45
    3.2.1 Trajectory Data Processing ............................................................................. 45
    3.2.2 POI Data Processing ........................................................................................ 48
    3.2.3 Mixed Land Use Data Processing ................................................................ 50
    3.2.4 Junction Density Data Processing ................................................................ 51
    3.2.5 Population Density Data Processing ............................................................... 53
  3.3 Methodology .............................................................................................................. 54
    3.3.1 Research Question 1 about Travel Mode ........................................................ 54
    3.3.2 Method for Research Question 2 about Low Speed Movement (LSM) ........ 57
    3.3.3 Method for Research Question 3 about Trajectory Typology ......................... 57
  3.4 Limitation .................................................................................................................. 60
  3.5 Validation ................................................................................................................... 60
Chapter 4 Analysis and Results ............................................................................................... 61
  4.1 The Travel Mode of Beijing Citizens ......................................................................... 62
  4.2 The Spatial Structure of LSMs in Trajectories ........................................................... 65
  4.3 Typologies of Pedestrian Trajectory .......................................................................... 67
LIST OF FIGURES

Figure 1-1. A Sample of An original GPS Trajectory. ............................................................ 2
Figure 1-2. Different Pattern of Explorer and Returner.......................................................... 18
Figure 2-1. Map of Beijing. ..................................................................................................... 20
Figure 2-2. The Diagram of Ancient Beijiang’s Wall System.................................................. 22
Figure 2-3. Map of Ancient Beijing ........................................................................................ 23
Figure 2-4. Average Time for Leisure Walk per Day for Beijing Citizens......................... 25
Figure 2-5. Spatial Distribution Map of Beijing’s Building Density......................................... 25
Figure 3-1. Map of Research Area.......................................................................................... 30
Figure 3-2. A Sample Trajectory Which did not Meet the Selection Criteria..................... 35
Figure 3-3. Map of Trajectories .............................................................................................. 36
Figure 3-4. Distribution of Trajectory Durations ................................................................. 36
Figure 3-5. The Working Flow of Trajectory Data Processing ............................................. 37
Figure 3-6. Map of Flickr Photo Density .............................................................................. 40
Figure 3-7. Map of Weibo Check-in Density ......................................................................... 40
Figure 3-8. Map of Jiepang Check-in Density ....................................................................... 41
Figure 3-9. Map of Green Space ............................................................................................ 44
Figure 3-10. Sample Scatterplot of a trajectory .................................................................... 47
Figure 3-11. Map of Distance to Commercial Facility (meter) ............................................ 49
Figure 3-12. Map of Distance to Dining Space (meter) ....................................................... 49
Figure 3-13. Map of Distance to Green Space (meter) ........................................................ 50
Figure 3-14. Map of Mixed Land Use ................................................................................... 51
Figure 3-15. Map of Junction Density .................................................................................... 52
Figure 3-16. Point Density Algorithm .................................................................................... 53
Figure 3-17. Map of Population Density .............................................................................. 54
Figure 3-18. Boxplot Algorithm.................................................................59

Figure 4-1. Diagram of Travel Mode Analysis........................................62

Figure 4-2. Boxplots of Total Activity Radiuses on Weekdays and at Weekends. .........63

Figure 4-3. Breakdown of Trajectory Types..............................................68

Figure 4-4. A Sample of Type A Trajectory............................................68

Figure 4-5. A Sample of Type B Trajectory...............................................69

Figure 4-6. A Sample of Type C Trajectory...............................................69

Figure 4-7. A Sample of Type D Trajectory...............................................70

Figure 4-8. A Sample of Type E Trajectory...............................................70

Figure 4-9. A Sample of Type F Trajectory...............................................71

Figure 4-10. Boxplots of Durations of Trajectories (separated by types).........71

Figure 4-11. Boxplots of Junction Density (separated by types)...................72

Figure 4-12. Boxplots of Mixed Land Use (separated by types)...................72

Figure 4-13. Boxplots of Population Density (separated by types).................73

Figure 4-14. Boxplots of Distances to Different Types of POIs (separated by types)........73

Figure 4-15. A Mapping of the Least Cost Paths and Redundant Routes........75
LIST OF TABLES

Table 1-1. Summary of Environmental Factors Related to Walking........................................11
Table 3-1. Data Overview.......................................................................................................31
Table 3-2. Classification of Movement Vectors According to Movement Speed, COType, and Z score using a 5% Significance Level ...............................................................48
Table 4-1. Routine Activity Radius Comparison......................................................................63
Table 4-2. Weekday Activity Radiuses and Weekend Activity Radiuses Comparison...........64
Table 4-3. Shapiro-Wilk normality test ..................................................................................66
Table 4-4. Spearman's Correlation between LSM Duration and Environment Factors........66
Table 4-5. Shapiro-Wilk normality test ..................................................................................76
Table 4-6. Comparison between Redundant Routes and the Least Cost Paths.......................76
Table 4-7. Shapiro-Wilk normality test. ................................................................................80
Table 4-8. Comparison between Redundant Routes and the Shortest Paths .........................80
Table 4-9. Correlations between the Tortuosity of the Trajectory (Independent Variable) and the Ratio of the Surrounding Environmental Factors of RT to those of ST..............81
ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my thesis advisor, Professor Korostoff, who have offered valuable guidance and advice throughout the process of this research. I would also like to thank Professor Andris, Professor Bose and Professor Fonseca for serving as my committee members and provided me inspired advice with their expertise in different professional fields.

A special thanks to Professor Henderson, who is my previous thesis advisor. His stimulating motivation and valuable comments helped me to complete this work.

I am grateful to my friends and my family for their supports during the work.
Chapter 1
Introduction and Literature Review

1.1 Introduction

Although walking is the most common and familiar form of movement, in fact, there are more challenges in studying pedestrian behaviors than vehicle transportation and animal movement. This is because pedestrian movements are motivated by much more complicated inner process and affected by many heterogeneous external environmental factors. Pedestrian movement research usually focuses on pedestrian wayfinding mechanisms and the navigation process which consists of a sequence of decisions on routes and destinations. The decision making process is an interactive process between external information and internal preferences. Stern and Portugali (1999) have identified four kinds of factors that affect this decision making process. These four kinds of factors are: (1) the trip purpose and the destination; (2) the personal characteristics (age, sociodemographic, spatial knowledge, personal preference, etc.) of the navigator; (3) physical restrictions; and (4) events occurring on the way (Stern and Portugali 1999). Similarly, Golledge (1995) points out that destination selection and route selection, which are affected by the same factors (2), (3), and (4) that Stern and Portugali mentioned (1999), are two critical research perspectives in the pedestrian movement study. Studies, which focus on pedestrians’ internal factors of wayfinding process, have provided the basis for studies about effective external environmental factors of pedestrian movement. Allen (1999) reviewed previous research on human spatial ability and wayfinding behavior and pointed out that the wayfinding means are generated by personal spatial perception, motor capabilities, information processing abilities and spatial knowledge. He also explained that “perception, knowledge, and information-processing capabilities are portrayed as fully interactive
resources, as are perception and motor capabilities (Allen 1999, 78)”. Allen’s work shows the psychological decision making process of pedestrian wayfinding behavior systematically. Instead of providing a conceptual model of pedestrian’s internal decision making process, Gärling (1999) identified attributes that were traded off through the decision making process and these attributes were distance, spatial configuration, time and priority.

In additional to studying pedestrian movement from the perspective of psychology, researchers are also interested in properties (e.g. Speed) and spatial pattern of pedestrian movement. Spatially, pedestrian movement studies usually involve one or more following components: origin-destination (O-D) network, trajectories and flows. A user’s origin-destination (O-D) network is a directed geospatial network whose nodes are the locations of origins and destinations. Links within a user’s origin-destination (O-D) network come from origins to destinations. A pedestrian trajectory stands for the trace of a pedestrian’s movement during a period of time, while a travel flow of pedestrians means the collective trend of pedestrian trajectories. Different approaches of presenting pedestrian movements indicate research on different aspects of human movement.

Figure 1-1. A Sample of An original GPS Trajectory.

The origin-destination (O-D) network as a network model weakens the effects of actual spatial patterns of trips (trajectories) while dealing with people’s destination choices and
destination distributions. Guy Debord (1958, 1956) argues that walking is an exploratory movement of pedestrians from one attractive location to another and those locations are motivations of movements. Therefore, there is a considerable amount of research that uses the O-D network or the destination network to study people’s preferences on destination selections. One of the most classic traffic models with the origin-destination (O-D) network is the four step model. It is called the four step model, because it consists of four steps: (1) trip generation (gives a rating schedule for origins and destinations based on land uses or socio-economic factors); (2) trip distribution (link origins to destinations); (3) Mode choice (computes the proportion of trips by different transportation modes); (4) Route assignment. The simulation of the four step model not only depends on physical restrictions of transport infrastructures (capability, speed limit, distance, etc.), but it also generate the motivations of traffic and the distribution of trips based on O-D network and properties of nodes (origins and destinations), for example the employment rates of different areas (McNally 2008). Other kinds of network models for transportation analysis also use origin-destination (O-D) network. Take the World’s Eyes project as an example, Fabien Girardin and his team obtained data about frequently visited sites in various cities through geo-tagged photos on Flickr. They were able to generate the network of visitors’ destinations and weighted those locations by their visiting frequencies. By studying the spatial pattern and network properties of this network, the research figures out the need of transportations among these locations and people’s preferences on destination selections (Girardin et al. 2009). De Wang et al. (2009) carry out a network simulation model that predicts the spatio-temporal distribution of visitors, visitor flows and dining demands through collecting O-D patterns of respondents’ virtual trips and related temporal factors. In addition to studying people’s destination selections, some research also uses O-D network to study the travel models of people. Pappalardo et al. (2015) tried to identify two different types of movement patterns, which related to two different kinds of personal characteristics, through looking into the ratio of people’s routine activity radiuses to their total
activity radiuses. Pappalardo et al. (2015) call these two types of movement patterns the explorer and the returner. The explorer is the pattern of movement of wandering in a large network of various locations at a large scale, while the returner is the pattern of movement of moving between a few routine places (Pappalardo et al. 2015). In conclusion, research using O-D networks is macroscopic research that focuses more on the destination selection and the travel mode aspects of people’s movements.

Trajectory studies have two primary subjects, one is parameters of pedestrian movements, and the other is spatial patterns of trajectory (This kind of trajectory will be introducing more in Section 1.2). Usually, the speed of pedestrian movement is the most common parameter in pedestrian trajectory research. Zheng et al. (2010) point out that parameters related to speed (average speed, acceleration, etc.) are widely used indicators in identifying trajectories of different transportation modes. Walking speed is also a sensitive indicator of the influence of demographic, environmental and individual factors (Finnis and Walton 2008). There are also some other parameters mentioned in previous trajectory research, including spatial layout, heading change rate (heading change rate is the frequency of the direction changes during the movement.), stop rate, and velocity change rate, etc. (Zheng et al. 2010). Gjersoe et al. (2004) use a video-based observational method to study (1) pedestrian trajectories within deserted environments, in particular: (a) walking speed, (b) microscopic position preferences, and (c) interpersonal distances; and (2) how these variables might be influenced by the various personal, situational, and environmental factors. In addition to concentrating on parameters of trajectories, researchers also try to study pedestrians’ preferences in route selections through looking into the spatial patterns of trajectories. Golledge (1995) studied this issue through analyzing the characters of trajectories (fewest turns, longest segment first, preference for curves, preference for diagonals, etc.) with observations and experiments. Rhee et al. (2011) argues that outdoor pedestrian trajectories within 10 km presented a statistical similarity with a Levy walk model which consisted of many short
segments towards different directions and occasionally long straight ones. All in all, trajectory studies are usually microscopic research, considering the continued spatio-temporal transitions within trajectories.

Research about pedestrian flow use different scales. As for macro scale, researchers study the overall traffic patterns and volumes of different locations to show (1) pedestrians’ preferences in destination choices and route selection, and (2) the impact of external environmental factors, including socio-economic factors, demographic factors and spatial quality factors (Kurose, Deguchi, and Zhao 2009, Spek 2008, Nakamura 2015, Mohamed 2016). In terms of the micro scale, it is common to study the influences of environmental factors on parameters (such as speed, density, etc.) and patterns of pedestrian flow under particular scenarios. Usually these scenarios are related to issues about safety, capacity and evacuation. For example, Daamen and Hoogendoorn (2003) study the transitions of traffic flows while passing through bottlenecks which are narrow corridors between two spaces. In the same spirit, Rupprecht, Klingsch, and Seyfried (2011) study the influence of different shapes of bottlenecks on parameters of pedestrian flows.

Compared to vehicle movement, pedestrian movement receives less attention in research. However, with the growing interests in urban non-motorized traffic systems, there is a growing need of pedestrian movement research. Previous health research (Satcher 2001, Williamson 1999) has identified walking (active living) as an effective way to improve national health and reduce related medical care costs. Also, it is evident that the development of the urban non-motorized traffic system benefits cities’ sustainable developments and urban environment goals. These benefits lead to campaigns, policies, planning and design project to accommodate pedestrians in urban environments. On 26th April, 2007, a dressed-up campaign called “Walking Works” was carried out, aiming to promote the idea of making London one of the world’s most walking friendly cities (Middleton 2009). Over the past decades, regulations like the Federal Highway Program’s Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) and the 1998 Transportation
Equity Act for the 21st Century (TEA-21) have been developed to deal with the major shift in policy away from auto-centric planning, to non-motorized traffic system in federally supported transportation projects in the U.S. (Southworth 2005, 246).” The greenway systems, which are multifunctional corridors built on the linear transportation routes for sustainable development, nature preserve, water resource management, education and recreation, are widely built to accommodate pedestrians in the built environment. Recently, there is an increasing amount of greenway projects in most of the major cities in China and most of the greenway projects are focusing on improving walkability (a measure of the comfort extent for walking) in cities and their suburbs (Yu et al. 2013). Since promoting non-motorized transportation has become a worldwide trend, research on pedestrian movement plays a more important role in urban planning and design. Effective planning for non-motorized transportation requires an understanding of two broad issues: what constraints people’s decisions to walk, and what variables favor pedestrians (Willis et al. 2000).

1.2 A Review of Research on Environmental Influences on People Movements

In terms of discovering the variables that favor pedestrian, psychological research has illustrated basic rules and the theoretical basis for research about how environmental factors impact people’s walking behavior. Ecopsychologists like Barker and Wright (1949) argued environment and human behavior are interactive and coexisted through their research using behavior settings approach. Berlyne defined two types of human exploration behaviors. One was the diversive exploration that usually referred to aimless behaviors and the other was the specific exploration which was motivated by stimulation or curiosity (Berlyne 1960). The behavior setting theory argues that particular spatial configuration supports particular behavior pattern. Regardless of individual differences (Lin and Hu 2006). For example, in the villages, activities such as chatting, storytelling,
outdoor markets, and performance are often associated with particular environments like pavilion, big tree, or open space surrounded by woods. In Berlyne’s following study (Berlyne 1960, 1962, 1970), he called characters of exploration-related environments ‘collative stimulus properties’ that included four aspects: complexity, novelty, surprisingness, and incongruity, while in Kalplans’ preference matrix (Kaplan, Kaplan, and Ryan 1998), complexity and mystery (incongruity and surprisingness) are also related to exploratory behavior. Through quasi-experiments with video and slides, Schwarz and Werbik (1971) finds out that landscape with moderated complexity usually has a higher aesthetic value to people. Similarly, Lin and Hu argued that regarding natural landscape, higher complexity led to higher degree of preference (Lin and Hu 2006). However, in terms of the artificial environment, Wohlwill (1976) discovers a negative correlation between mystery and incongruity of spatial experience of an artificial environment. Furthermore, as for the mixed landscape condition of natural elements and artifacts, layouts of artifacts and the percentage of natural elements have primary effects on the harmony of the landscape (Lin and Hu 2006). Although there seems to be an agreement that environmental factors are able to influence human behaviors, in terms of how environmental factors can affect human behaviors and to what extent environmental factors can affect human behaviors, different researchers hold different opinions. Experimental evidences from several behavioral domains identified circumstances in which direct environmental influence could be a stronger determinant of behavioral choice than were cognitively mediated influences (Bargh and Chartrand 1999, Bargh and Ferguson 2000). However, in contrast to that, Bourassa argued that (1) the way people perceive landscapes was influenced but not determined by physical landscape attributes, (2) a complex mental process of information reception and processing between the physical landscape and the mental landscape, and (3) various factors, including biological, cultural and individual factors, could influence this mental process (Bourassa 1992, 1990).
According to previous research on pedestrian route choice, it is still unclear to what extent environment can affect the pedestrian route choice. Many route-choice studies or related studies usually assume that pedestrian prefer to choose the shortest route or easiest route (fewest turns) to minimize the cost of moving (Borgers and Timmermans 2005, Hillier 2007). This point of view is supported by sufficient empirical evidences. For instance, Seneviratne and Morrall (1985) interviewed 2685 pedestrians in downtown Calgary, Canada and found that 50.7% of the respondents chose the quickest path. The second largest amount of respondents (21.7%) chosen route was their routine ones, presumably originally chosen because it was the shortest. Therefore, nearly three-quarters (72.4%) of the respondents chose the shortest route (Seneviratne and Morrall 1985). A similar result was found by Guy (1987) in Jerusalem, where two thirds of the survey participants chose the shortest route.

However, a considerable number of studies have also found that the distance or the cost of time is not always the dominant factor. Borst et al. (2009) examined 1294 walking routes of 364 elderly individuals in the Netherlands and found that only 20% of the chosen routes were the shortest. Rhee et al. (2011) find that human intentions instead of geographical artifacts lead to the variety of pedestrian route choice in outdoor spaces. They also conjecture that this is caused by the power-law tendency of human interests or popularity of locations people visit. (Rhee et al. 2011). Thus, except the distance and the cost of time, there must be some other environmental factors that affect the pedestrian route choice. Rodriguez et al. (2009) associated the spatial quality factors (i.e., land use, sidewalk width, sidewalk continuity, trash bins, crossing aids, density, and road density) of 338 street segments with pedestrian flows during a 10-min interval and found a significant correlation between these two.

Since environmental factors and pedestrian trajectory pattern (route choice) are related to each other mutually, some researchers study pedestrian route choice through analyzing the spatial cognitive mechanism of pedestrian, while others do the research in an inverse direction in that they
try to identify pedestrian’s criteria in route choice by studying different pedestrian trajectories which are under different situations. The former type of research mainly focuses on the ‘cognitive maps’ in people’s mind and its typical methods are surveys and interviews. The latter type of research usually carries out experiments or quasi-experiments under different situations, or correlates trajectory data, which is collected by tracking pedestrians, with different kinds of variables, often GIS data.

The word ‘cognitive map’ was created by Gestalt psychologist, Edward C. Tolman (Lin and Hu 2006). This concept was generated from a rat experiment. Cognitive map indicated spatial knowledge about structure of external environment, which finally determines route choices (Tolman 1948, 192). According to Golledge’s argument (2002), even though there are individual differences in travel behaviors and decision making process, it is unlikely to understand and predict every aspect of trajectories. However, combing the contextual information of trips, it is possible to identify environmental factors related to route selection (Golledge and Garling 2002). Interviews and questionnaires are effective methods in cognitive study to get contextual information on trips and pedestrians. For example, while studying the pedestrian movement in commercial streets, in addition to characteristics about trajectories (i.e. Duration, stop points, direction, and length), researchers were able to get other information, including socio-demographic data, trip purposes, previous plan of activities etc. (Kurose, Deguchi, and Zhao 2009). Other than interviews and surveys, Lynch asked respondents to map their images (impression) of cities to gain information about their cognitive process. Lynch (1960) reports that citizens understand their surrounding environment and create their cognitive maps based on five environmental elements: Paths, Edges, Districts, Nodes and Landmarks. Through statistical calculation of the frequencies of various elements in maps, a collective cognitive map which reflected the human spatial experience related characters of a city could be drawn, too (Lynch 1960). Seneviratne and Morrall (1985) found six additional factors to be influential: route attraction, junction amount, degree of crowding, shading
condition, noise, and security through interviews. Guo and Ferreira (2008) and Guo (2009) studied pedestrians’ route choices in downtown Boston and found that the presence of retail and open space, sidewalk width, junctions, and topography affected the selections. There are additional studies that identify environmental factors related to pedestrian movements and I summarize these factors in Table 1-1.

As I mentioned earlier in this section, particular spatial configuration supports particular behavior patterns and Kim (2001) points out that spatial configuration and spatial cognition are correlated. As for study of pedestrian movement, the spatial pattern of pedestrian movement is able to inform us about route choice criteria, preference of locations and wayfinding process of pedestrians. Golledge (1995) carried out experiments in virtual environment settings by asking respondents to draw their chosen routes on maps, and experiments in real world settings with tracing respondents’ trajectories. Through comparing results of these two kinds of experiments, Golledge figured out the common route choice criteria and importances of different criteria used by respondents in both situations (Golledge 1995). Brown et al. (2007) identified the social milieu, sidewalk amenities and events, building attractiveness, and safety as important route attributes in attracting pedestrians through an experiment in downtown Salt Lake City with 76 subjects. Furthermore, some researchers pay more attention to some ‘special’ patterns (such as stops and crossings) within a trajectory, since these ‘special’ patterns are more efficient indicators of pedestrian behaviors and preferences (Lassarre et al. 2012), Zacharias (1997), and Orellana and Wachowicz (2011) chose stops within pedestrian trajectories as indicators to study the pedestrian’s preferences of the external environment. As for this research project, I also intend to study the route choice criteria, preference of locations and wayfinding process of pedestrian through analyzing spatial patterns of pedestrian trajectories.
<table>
<thead>
<tr>
<th>Sources</th>
<th>Factors</th>
<th>Constrain Factors (For This Research)</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Moudon et al. 2006, Lin and Hu 2006, Saelens, Sallis, and Frank 2003, Humpel, Owen, and Leslie 2002)</td>
<td>Proximity/Accessibility to Amenities</td>
<td>Dining Space, Commercial Facilities, and Green Space (Three types of amenities that are widely mentioned.)</td>
<td>Amenities provide supplies, activities, and space for daily life, so they are common destinations and motivations for pedestrian movements.</td>
</tr>
<tr>
<td>(Lin and Hu 2006, Lee and Moudon 2004, Humpel, Owen, and Leslie 2002)</td>
<td>Aesthetics</td>
<td>Green Space (especially water bodies)</td>
<td>Green space have both physically and mentally attractiveness to people and they are also traditional social space.</td>
</tr>
</tbody>
</table>
1.3 A Note on New Data Sources and Data Types

The development of trajectory research has been greatly benefited from the evolution of computing science and sensing technology. Nowadays, GPS recording programs have been widely deployed in mobile phones and different kinds of wearable devices in order to record personal mobility information, including tracks, distances, speeds, durations, and timestamps for each node. “The high accuracy of GPS data enables the micro-analysis of environments with various approaches to achieve a better understanding of how people consume space” (Shoval 2008, 19). Through studying the actual spatial pattern of pedestrian trajectories, researchers are able to identify spatial characters of pedestrian trajectories and correlate them with different variables in statistical analysis more directly and more easily.

Furthermore, the rise of social media and self-reported data also benefits research. In the last decade, there was a huge growth of user-generated contents on the Internet through the use of a series of sharing tools that have generally been defined as online social media (Sigala and Marinidis 2009). This was one of the fastest growing online segments. Social media have various forms, they include opinion sharing blogs and microblogs (i.e. Blogger and Twitter), photo and video sharing media (i.e. Flickr and YouTube), sharing knowledge bases (i.e. Wikipedia), social bookmarking (i.e. Delicious) and many other forms (Parra-López et al. 2011, 640). Moreover, the contents of some social media are geo-tagged and many social media provide geo-location information through their APIs (Application Programming Interface) so that these can be new data sources for human movement research. Usually social media data is highly accessible and sometimes free so that it benefits individual researchers a lot. However, one major disadvantage of social media data is that they are mostly anonymous. Such kind of anonymous data usually lacks contextual information. For example, when we extract the social network data from Facebook, we
may not be able to get the personal information like age, job and income for each one. But this may not be problematic, if the research is less focused on demographic information, for example, research on the spatial distribution of popular locations. Another major disadvantage is that social media data contain unavoidable and uncontrollable bias and errors, because these data are user-generated contents. However, data from other sources have such problem as well.

Regarding research related to urban design and planning, social media can provide alternative data for information that is hard to access through other approach (i.e. the actual land use data). For example, in China, land use data are not open to the public; even research institutes are not able to easily access this data. Yet, there are related data online like POI (points of interest) data from online maps (i.e. OpenStreetMap) and the geo-tag of social media (i.e. Flickr and Twitter). The Beijing City Lab proposes “a method for automatic identification and characterization of parcels (AICP), based on ubiquitous OpenStreetMap (OSM) and Points-of-Interest (POI) data. This proposed method could (1) provide quick and robust delineation of land parcels; and (2) generate a variety of parcel level attributes, allowing for the examination of urban functions, development density and mixed land use (Long and Liu 2013b)”.

The importance of the use of social media in the context of tourist preference has been widely studied in recent years, many studies related to various types of social media (i.e. Blogs or photo sharing tool) or different online communities (i.e. Trip Advisor) have been done (Hsu and Lin 2008, Wang and Fesenmaier 2004, Chung and Buhalis 2008). APIs of the photo-sharing platform like Flickr allows researchers to map and visualize the active footprint of tourists. Through the temporal and geographic distribution of tourists, preference of tourists and attractiveness of locations can be identified (Girardin et al. 2009). Moreover, social media data usually not only provide the geographic network of movements or users, but also users’ social networks, researchers can also look into the social connection among online community actors with geographic distributions (Cho, Myers, and Leskovec 2011, Lin, Halavais, and Zhang 2007).
Guo and Loo point out the difficulty that “density, land-use mix, street network, and accessibility are often readily available in GIS data formats. However, other environmental features, such as design features, traffic, and pedestrian infrastructure, may not be available (Guo and Loo 2013, 125).” However, by using the social media data, we can generate additional needed GIS data layers. Besides, few researchers have paid enough attention to online trajectory data from trip information sharing communities like Wikiloc and EveryTrail. Therefore, this thesis intends to fill this research gap through exploring the processing and analysis of social media data and GPS data from online communities.

1.4 Research Overview

A lot of research (Mohamed 2016, Kim 2001, Hillier 2007) has studied the relationship between connectivity and people’s movement through spatial configuration analysis. However, Nakamura (2015) argues that “configurational accessibility analysis does not consider the quality attributes of routes” (Nakamura 2015, 1). According to the literature, I found that it is still unclear whether environmental factors, which relate to the spatial qualities of surrounding environment, impact pedestrian route choice and which environmental factor or factors affect pedestrian route choice and movement. In order to fill the gap in the knowledge, this research intends to explore the relations between environmental attributes (distance to amenities, mixed land use, population density, and aesthetics) and pedestrian route choice through comparing different types of trajectories and their surrounding environment. Regarding factors like distances to different types of POIs (point of interest), mixed land use, density, and aesthetics, the most widely mentioned constrain factors in previous research were selected to be examined in this thesis (See Table 1-1). Because of the lack of data, neighborhood type issues were not involved in this project, too. Instead, this project intends to examine the impact of environmental factors on pedestrian behavior by using
publicly available social media data and GPS data. By doing so, a new approach that combines social media data, programming and GIS platform is proposed for future trajectory studies. In addition, this research will examine whether Beijing citizens’ travel mode is closer to the mode of explorer or the one of returner through studying the differences between their routine activity radiuses and their total activity radiuses.

Facing the growing need of enhancing the ‘intelligence’ of urban planning and design approaches in this information era (Long and Shen 2015), this research studies the correlation between pedestrian movements and external environmental factors to provide scientific suggestions and design guidelines for Beijing’s future planning practices. Specifically, through mapping, cluster analysis, spatial statistics and statistics, this research is able to answer the following two questions related to China’s ongoing planning practices: How dense should the green space network of Beijing be? And what kinds of outdoor spaces are desired by pedestrians?

1.5 Significance

Researchers in both public health and urban planning have highlighted the importance of using quantitative measures to better understand the relationships between built environment features and human behavior (Saelens, Sallis, and Frank 2003, Sallis et al. 2004, Owen et al. 2004). This project studied pedestrian trajectory based on self-reported GPS trajectory data, GIS data and social media data, in particular to identify effective urban environmental factors, and provide suggestions on how urban planning can be improved for walkability in Beijing. Based on a review of the current data sources and methodologies used by designers, planners and researchers who are working on the built environment and human behavior, it is evident that few researchers have explored and analyzed the rich geo-location data and trajectory data from online sharing communities. The study explores the possibility and methodology for analyzing human behavior
and urban environment through social media data from a number of perspectives. Also, the methodology of this research is a combination of traditional statistics and spatial statistics through an integration platform of Rstudio and ArcGIS. GIS has served as a planning and analytics tool and is familiar to landscape architects. Through integrating Rstudio with GIS, the project also intends to introduce a new tool to landscape architects and improve the ability of dealing with big data in landscape architecture. Furthermore, previous research has usually examined pedestrian trajectories in a general manner while this research project concentrates on comparing different typologies of pedestrian trajectories in terms of various environmental factors. It establishes strategies and new knowledges with which landscape architects and researchers can contribute to the work of constructing multi-functional landscape corridors and walkable cities.
1.6 Definitions of Terms

Based on literature review, ground truth, and life experiences, important concepts and terms in the research are defined as following:

1. Trajectory: A trajectory is the route (curve) described by a person when he/she moves.

2. GPS Log (GPS trajectory): “Typically a GPS trajectory consists of a sequence of points with latitude, longitude, and timestamp information (Lou et al. 2009).”

3. Easiest Route: the route which has fewest turns and follows the major road grid from the start point to the destination.

4. Major Road Grid: The road network consists of roads whose levels are higher or equal to “county road” in the road hierarchy system of Beijing. In other words, it is formed by highways, urban expressways, national roads, provincial roads and county roads.

5. Redundant route: Rather than following shortest path, human’s wayfinding also leads to ‘preferred paths’ which are the results of balancing various variables including ‘short distance’, ‘security’, ‘minimal work’, ‘maximum experience’ etc. (Lewin 1934). If an unclosed route’s length is longer than the Manhattan distance between the origin and destination and the unclosed route has more turns than necessary, this unclosed route is a redundant route.

6. Walkability: “Walkability is the extent to which the built environment supports and encourages walking by providing for pedestrian comfort and safety, connecting people with varied destinations within a reasonable amount of time and effort, and offering visual interest in journeys throughout the network (Southworth 2005, 247-248).” Walkability depends on various environmental factors that relate to pedestrians’ preferences and spatial experiences. These factors include, but are not limited to, quality of pedestrian
facilities, road safety, land use patterns, spatial accessibility and aesthetics.

7. Low Speed Movement (LSM): Since it is quite unlikely to always have GPS records in the same location or a zero movement (speed = 0 or near to zero), low speed movement is the low value (low speed) cluster which is defined by Anselin Local Moran's I approach in GIS (Orellana and Wachowicz 2011).

8. Explorer (Figure 1-1): “Explorer” who prefer to travel to many different locations that further than their routine locations (Pappalardo et al. 2015).

9. Returner (Figure 1-1): “Returner” is the kind of people who have frequent movements between only a few preferred locations (Pappalardo et al. 2015).

![Figure 1-2. Different Pattern of Explorer and Returner. Source: Adapt from Pappalardo et al., 2015, “Returners and explorers dichotomy in human mobility”, Nature Communication 6, page 5.](image-url)
Chapter 2

Introduction to Beijing and Research Questions

I choose Beijing as the case study site because of my personal interest and the availability of data. Beijing is a historic city and its urban environment has influenced the formation of customs and its residents’ lifestyle. Thus, in this chapter, I will briefly introduce Beijing and its urban life to provide some contextual information. Since the basic spatial structure of Beijing has been maintained for a long time, I assume that modern Beijing city will still have an urban character similar to the ancient Beijing city to some extent and I will examine this hypothesis in later analysis.

2.1 An Introduction of Beijing and Its Urban Life

Beijing is China’s capital city which is also the political and cultural center of China. Beijing was first designed as China’s capital city in the Mongol-led Yuan Dynasty of the 13th century, and is the city that has the most number of World Cultural Heritage sites. Beijing is located in the northern part of the North China Plain, which is in the northeast of China, and near the Bohai Bay. The total area of Beijing is 16,410.54 km² and 62% of it is mountainous. Beijing’s urban area is located in a plain area which is in the southeastern part of Beijing. The administrative area of Beijing consists of 16 districts. Dongcheng District, Xicheng District, Haidian District, Chaoyang District, Fengtai District, and Shijingshan District are the six major urban districts of Beijing. The area encompassed by the Fifth Ring Road is regarded as the urban area of Beijing by Chinese people (Figure 2-1). It is centered on the Forbidden City and the city spreads out in concentric ring roads. The First Ring Road overlaps the boundary of the Forbidden City and the Second Ring Road.
traces the original city walls of Beijing. The urban area of Beijing is about 1,381 km² and its population was about 12,763,000 according to the 2014 data.

Figure 2-1. Map of Beijing.
The concentric structure of modern Beijing city is based on its ancient urban fabric. In ancient times, the walls served as the boundaries of the territories as well as the administrative power. Furthermore, the hierarchy of different urban spaces was also defined by different types of walls. There were five types of walls from the top to the bottom of the hierarchy (Figure 2-2). These were: city walls that defined the boundary of a city, inner city walls that defined the boundary between the inner city and the outer city, palace walls, walls of Li-Fang (blocks), and walls of Siheyuan (Courtyard houses). Li-Fang were enclosed walled-wards that were similar to the gated communities of today (Liu and Lai 2008). Ancient Beijing had an introverted urban character with facilities built inside the blocks while the streets were basically only used for transportation. The inner Li-Fang paradigm has lasted for over 3000 years. Because of this, it gives people an impression that China’s traditional cities are introverted in comparison to the ancient western cities. However, due to the spatial revolution, Beijing changed from being a Fang (block) -based city to a street-based city, between Tang Dynasty and Song Dynasty (between 618 AD and 1279 AD), with the citizens’ daily social activities spread from Fang and Hutong to market and street (described below).
Figure 2-2. The Diagram of Ancient Beijing’s Wall System.
Source: The Library of Congress, U.S (base map)
The layout of the ancient Beijing was a chessboard-shaped pattern with uniform blocks and a strong central axis (Figure 2-3). There were three levels in the street structure of the ancient Beijing city. They were boulevards along the central axis, major streets which connected city gates, and Hutongs which were alleys connecting courtyard houses. Hutongs began to refer to residential blocks after the spatial revolution. The widest part of the axis boulevard was about 28 meters wide while other major streets were 25 meters wide. Hutongs are usually 5 ~ 6 meters wide. Because of the water bodies in the city and the asymmetric layout of the city gates, many junctions of the major streets were T-shaped. On both sides of the north-south major streets, many east-west Hutongs were distributed equidistantly (Dong 2004, 113). Even in today’s Beijing City, we can still recognize such chessboard-shaped layout and street-block (Hutong) system easily.

Figure 2-3. Map of Ancient Beijing. 
The structures of ancient Beijing’s urban spaces and residential spaces (Hutongs and Courtyard Houses) are characterized as ‘introverted’ structures. To some extent, Beijing nowadays still maintains this character. The life style of Beijing citizens is also ‘introverted’ in that Beijing citizens prefer indoor activities more than outdoor ones. Less than 2% of Beijing citizens take exercises (other than running) routinely for more than 100 days annually and less than 3% of Beijing citizens exercises (other than running) at least one time a week on average annually (Wang and Ma 2015). Regarding recreation activities and exercises, walking and running were the most popular activities. 39% of Beijing citizens have leisure walks for more than 100 days per year (Wang and Ma 2015). Beijing citizens spent 24 minutes on leisure walks during weekdays on average, while they usually spent 40 minutes on leisure walks at weekends (Wang 2002). According to Yanqi Wang (2002), older people usually intend to have a longer walk per day (Figure 2-4). Because other age groups had bigger pressures at work and they also had to spend a lot of time on commuting on weekdays for the congested traffic conditions in Beijing (an average of 74 minutes on weekdays) (Wang, Lei, and Shi 2002). This makes it difficult for the working class to engage in recreation activities on weekdays. The core urban area has high building density (\(\frac{\text{Footprint area}}{\text{land lot area}}\)) that is between 20% - 30% on average (Figure 2-5). Because of the rapid development and high density, current green spaces and the public recreation services do not meet citizens’ need and the diversity and spatial distribution of recreation facilities needed to be improved.
Figure 2-4. Average Time for Leisure Walk per Day for Beijing Citizens.

Figure 2-5. Spatial Distribution Map of Beijing’s Building Density.

Because of the needs of healthy living, there is a growing concern for non-motorized transportation (active transportation) planning and green space network development to provide a better urban environment for pedestrians. TOD (Transit Oriented Development) is one of the planning strategies for sustainable development and TOD is one of the most popular urban planning strategies for current China’s urban practices. TOD intends to improve the transportation system to encourage active transportation and build a safer and more comfortable walking environment (Song and Li 2012). Greenway is an important component in TOD and greenways have been or are being constructed in many cities in China. Greenways are multi-functional landscape corridors which are built on the linear connections in cities, focusing on not only the needs of urban beautification and recreation, but also the needs of wildlife, flood damage reduction, water quality, education and other infrastructure (Seams 1995, 66). The Beijing government hopes to increase citizens’ satisfaction with the overall quality of life by improving Beijing’s Green Space System. In 2004, Beijing New Master Plan (2004–2020) proposed an overarching agenda to improve the walkability of the city and build multifunctional corridors based on the existing street network and the green space system. Furthermore, in Beijing Green Space System Planning (2010), the planners also suggested to set up a multi-layered green space network with rational service radiuses and this needs to be supported by further behavioral research.

Although the Beijing government is trying to improve Beijing’s Green Space System, the rapid urbanization process still has greater impacts than the government’s efforts on Beijing’s green spaces. In recent years, the development of the ring road system has greatly changed Beijing’s spatial structure and accelerated the developments of adjacent spaces (Li, Xiao, et al. 2015). Although the green space coverage was increasing, the majority of these green spaces are small dispersive fragments with low diversity (Yang 2015). In the built environment, grasslands have been the domain type of green space (Li, Xiao, et al. 2015). However, in China, people are prevented to get into grassy areas in parks and gardens. That grassland has become the domain type
of green spaces in fact makes urban green space less accessible to the citizens of Beijing. Besides, water bodies are one of the most attractive landscape elements in Beijing. But, with rapid urbanization, urban water bodies all show a trend of shrinking (Yang 2015). There’s a mismatch between the needs of citizens and the actual urban development. Thus, more research on urban characters and human behavior needs to be carried out to direct future planning. This research identifies the urban characters of Beijing through studying pedestrian route choice and GPS trajectories, and in the end, it gives suggestions for the planning and design of Beijing’s open space structure and walking environment.

### 2.2 Problem Statements

This thesis is an analytic investigation about human movement, trajectory, and urban environment. This research focuses on three sets of specific research questions that cover both macroscopic and microscopic study:

1. Does Beijing citizens’ movement pattern more closely approximates that of an “Explorer” or a “Returner”?
   1.1 What are the differences between the explorers’ and the returners’ routine activity radiuses?
   1.2 What are the differences between weekday activity radiuses and weekend activity radiuses of Beijing citizens?

2. Are the people’s stops (low speed movements / LSMs) influenced by the junction (road intersections) density, the distance to dining place, the distance to commercial facilities, the distance to green space, or the population density (popularity degree of locations)?

3. What are the different types of trajectories followed by Beijing pedestrians?
3.1 What does the pedestrian route choice of Beijing citizens reveal about the urban character of Beijing?

3.2 What are the differences among different types of trajectories in the factors (the distance to dining place, the distance to commercial facility, the distance to green space, the degree of mixed land use, and the population density) of the surrounding environment?

The first set of the questions focuses on the overall travel mode of respondents. The second and third sets of research problems concentrate on the relationships between environmental factors and particular behavior patterns of Beijing citizens.
Chapter 3

Data and Methodology

3.1 Data Collecting Strategy

The first research question is about the travel modes (patterns) of Beijing citizens, whether they are explorers or returners. Since the trajectory data used in the analysis of research question 1 include both walking trajectories and long distance movements, the research site of research question 1 is the entire Beijing city. However, research question 2 and 3 are about walking trajectories, so the research site of these two questions are smaller than that of research question 1. The research site of research question 2 and 3 is the urban area of Beijing, which is about 1,381 km². This research site covers the areas of: 1) Haidian District, 2) Chaoyang District, 3) Shijingshan District, 4) Fengtai District, 5) Xicheng District, and 6) Dongcheng District. Within the 6 district area, the core research area is basically the core urban area encompassed by the Fifth Ring Road (Figure 3-1). As I mentioned in Chapter 2, there are seven factors in this research: i) pedestrian GPS trajectories, ii) dining places (including restaurants, bars, and coffee shops), iii) commercial space, iv) green space (including water bodies), v) junction density, vi) population density (popularity of location), and vii) degree of mixed land use. All the related data are from online open sources (secondary data), and they are integrated into a GIS platform for further analysis.
Figure 3-1. Map of Research Area.
<table>
<thead>
<tr>
<th>Factor</th>
<th>Input Data</th>
<th>Source</th>
<th>Date</th>
<th>Link</th>
<th>Type</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>included in all research</td>
<td></td>
<td></td>
<td></td>
<td>daa38f2b2e13/)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>questions, while other 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>datasets are only included</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>in research question 2 &amp; 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trajectory Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12 users</td>
</tr>
<tr>
<td>Trajectory Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12 users</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8 users</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2013a, Thomeee et al. 2015)</td>
<td></td>
<td></td>
<td>Shapefile</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shapefile</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance to Dining Space</td>
<td>Beijing City Lab</td>
<td>2013</td>
<td><a href="http://www.beijingcitylab.com/">http://www.beijingcitylab.com/</a></td>
<td>Vector in</td>
<td>35,143 nodes</td>
</tr>
<tr>
<td></td>
<td>Mixed Land Use</td>
<td>Beijing City Lab / Dr. Dong Li</td>
<td>2011</td>
<td><a href="http://www.beijingcitylab.com/">http://www.beijingcitylab.com/</a></td>
<td>Vector in</td>
<td>211,723 POI</td>
</tr>
<tr>
<td></td>
<td>Junction Density</td>
<td></td>
<td></td>
<td></td>
<td>Shapefile</td>
<td>62,258 junctions</td>
</tr>
<tr>
<td></td>
<td>Distance to Commercial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16,337 nodes</td>
</tr>
<tr>
<td></td>
<td>Facility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance to Green Space</td>
<td>FROM-GLC Land Cover Data</td>
<td>2013</td>
<td><a href="http://data.ess.tsinghua.edu.cn/">http://data.ess.tsinghua.edu.cn/</a></td>
<td>Raster</td>
<td>21,824 features</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tsinghua University (Gong et al. 2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shapefile</td>
<td></td>
</tr>
</tbody>
</table>
3.1.1 Trajectory Data

Trajectory data used in this research project are GPS logs that basically contain sequences of points with their longitudes, latitudes, elevations, and timestamps. They are from four sources, including Microsoft Research Asia’s GeoLife Project, Run.GPS, Wikiloc, and EveryTrail. These trajectories were recorded by different devices which included various types of cell phones and professional GPS devices. However, most of them are either recorded by distance, for example, every 10 meters, or recorded by time period, for example, every 5 seconds. Trajectories whose durations are less than 10 minutes were excluded since they may not be long enough to have variety in their surrounding environment.

The GeoLife Trajectory Dataset v1.3 (Zheng, Xie, and Ma 2010, Zheng et al. 2009, Zheng et al. 2008) of Microsoft Research Asia’s GeoLife Project records 182 users’ discontinuous movement segments within a period of over six years (from April 2007 to August 2012). This dataset contains 17,621 trajectories with a total distance of 1,292,951 kilometers and a total duration of 50,176 hours. Although these trajectories are widely distributed, the majority was located in Beijing. Because of privacy protection, user names are coded into 3 digit numbers and they do not provide the key for deciphering. There is no personal demographic information included in this public dataset. But according to the related publications, most of the respondents were college students who were under 30 years old, while others are companies’ employees and government staffs. Most respondents were living and working in or near universities and Zhongguancun area which is a technology hub in Haidian District (Zheng et al. 2009). Seventy-three respondents have labeled their trajectories with transportation mode, such as driving, taking a bus, riding a bike, running, and walking. A GPS trajectory here is represented by a sequence of time-stamped points, each of which contains information of latitude (in decimal degree), longitude (in decimal degree),
altitude (in feet), date and time. A GPS trajectory of this dataset usually contains several different transportation modes. Due to the low accuracy and complex process of automatic identification of travel modes (Zheng et al. 2010), in order to extract walking trajectories for the analysis of research question 2 and 3, I only used the labeled data of these 73 users. Furthermore, the travel mode labels were based on users’ memory and GPS devices sometimes were influenced by some interference. Therefore the sampling process for GeoLife Trajectory Data contains two steps. In Step 1, I programmed to automatically pick out data whose labels are “walking” with Rstudio (a programming tool which uses a programming language called R) and its SQL package. Step 2 is a graph-based post-processing algorithm which can remove error data and data with strong interference from the subset extracted from Step 1. Particularly, Step 2 involves two basic identification rules for inappropriate data: (1) The average speed of an inappropriate trajectory data is faster than 4.44 km/h (Cheung and Lam 2000). (2) The pattern of an inappropriate trajectory data is chaotic and has overlaps with unwalkable locations like water body, railway and buildings. In addition, the trajectories with durations under 10 minutes are excluded in this research, because they are too short to present enough information. After cleaning the GeoLife Trajectory Data, I selected totally 104 trajectories of 34 users for use in this research. In addition to walking trajectory analysis, the movement pattern analysis (Research Question 1) used the data from 50 respondents who have most data. The detailed processing approach for these 50 respondents’ O-D networks will be introduced in Section 3.3.1.

Run.GPS, Wikiloc, and EveryTrail are three other data sources and they are all free online activity log and GPS navigation data sharing communities. Wikiloc, which was founded in 2006, described itself as a volunteered geographic information platform for people to discover and share outdoor experiences and trips. Run.GPS is a GPS community for training and competition. Users can upload and download GPS training profiles of each other. Similarly, EveryTrail is a global web2.0 platform for users to share their self-generated geo-tagged travel information to others who
share similar interests in travel and outdoor activities. All of them have released their own outdoor activity apps for users to record their daily activities. Uploaded trajectories on these websites are categorized with different kinds of labels including road trips, sightseeing tours, sailing trips, hiking, cycling, flying, hang gliding, geocaching, skiing, kayaking trips and more. No matter what kinds of GPS devices these online communities’ users use, their trajectories are often recorded as sequences of time-stamped points, each of which contains information of latitude (in decimal degree), longitude (in decimal degree), date and time. Besides, some users also attached some contextual information about themselves or their activities. For example, some users presented short descriptions about their activity purposes and destinations, while some others might provide information which can be used to infer their jobs or whether they were Beijing citizens or not.

Regarding the trajectory data collecting, I went over trajectories under the labels of “walking” and “hiking” with three selection rules: (1) more than 75% of a trajectory should be located in an urban area instead of natural dominated environment, because this research focuses on studying pedestrian movements in an urban environment; (2) a selected trajectory shouldn’t be greatly mismatched with the urban fabric. The sample trajectory showed in Figure 3-2 mostly did not follow the street network because it had severe interferences in GPS record. This kind of trajectory (like the one in Figure 3-2) will be excluded from the data, and (3) if there are more than five trajectories of one user, five of those trajectories were randomly selected.
Finally, I got 18 trajectories of 12 users from Wikiloc and 36 trajectories of 12 users from Run.GPS and 14 trajectories of 8 users from EveryTrail. In the end, totally 172 trajectories of 66 different people were collected from these four different data sources (Figure 3-3). The longest trajectory is 30,120.5 meters long while the shortest one is 468.31 meters long. In terms of duration, the GeoLife Trajectory Datasets contain more short-duration commuting trajectories than the online community data. In terms of the whole dataset, short duration trajectories from 10 minutes to 20 minutes occupy the majority while the numbers of other duration categories’ trajectories are not so different from each other. The most time-consuming trajectory is a sightseeing activity which lasts about 12 hours. The working flow of trajectory data analysis is shown in Figure 3-5 and the methods of it will be introduced further in Section 3.2.1, 3.3.2 and 3.3.3.
Figure 3-3. Map of Trajectories.

Figure 3-4. Distribution of Trajectory Durations.
Figure 3-5. The Working Flow of Trajectory Data Processing.
3.1.2 Geo-Tagged / Check-in Data for Population Density

In previous active transportation research, population density usually referred to resident density. However, by using the check-in data of social media, the temporal population density is able to be described through the popularity of locations. It describes the density pattern of people who are motivated by events or certain kinds of activities, for example, shopping, eating in restaurants, or sightseeing. In this research, the temporal population density was generated from check-in data of three datasets. These three datasets are Weibo check-in data, Flickr geo-tagged photo data, and Jiepang check-in data.

The Flickr photo data were extracted from the original Yahoo Flickr Creative Commons 100M dataset released by Yahoo (Thomee et al. 2015). The Flickr geo-tagged photo dataset of China in this research was prepared by Dr. Dong Li (Senior Engineer of China Academy of Urban Planning & Design) in shapefile format for GIS platform. It was released to the public on the Beijing City Lab website (http://www.beijingcitylab.com/). It represents the distribution of the Flickr photos of China by location points in GIS. It contains totally 2,171,162 tagged locations in China, where Flickr users have ever taken photos and uploaded them to Flickr before March, 2014. The extracted data, which are located in the research area, contains 126,326 location nodes. Every location node in this dataset contains information about place ids, user ids, the models of cameras, photo dates, latitudes, and longitudes.

The Beijing Weibo check-in data were collected from Sina Weibo (the Chinese version of Twitter) by Beijing City Lab (Long and Liu 2013a) in 2013. Similar to the Flickr data, it represents the distribution of the Weibo check-in geo-locations in Beijing and these locations were categorized into different land use types. The dataset has totally 8,678,743 check-ins which are located in 143,576 different places in Beijing. This information was edited into data points in shapefile format.
for GIS platform. The extracted data, which are located in the research area, contains 113,441 location nodes. Every check-in location point contains information about POI ids, total check-in amounts, the types and categories of locations, location names, postcode, phones, addresses, latitudes, and longitudes.

Compared to the other two social media datasets, the typical character of Jiepang was that Jiepang was a geo-tag-based social media like Foursquare. Users shared the GPS information of places, which they visited, and their comments on these places to inform and get to know others who share the same interests as them. Jiepang shut down its service and released a check-in dataset in 2012. This dataset is shared on Shujutang (a data sharing platform) and originally modified by MIT’s Civic Data Design Project (Williams et al. 2012). The data contains information about 5,597 places in Beijing’s core urban area with a total of 2,678,232 check-ins. The extracted data, which are located in the research area, contains 5,237 location nodes. Every check-in location point contains information about total check-in amounts, the categories of locations, location names, addresses, latitudes, and longitudes.

According to the visualization of these datasets (Figure 3-6, Figure 3-7, and Figure 3-8), even though those photo-taken locations and check-in locations were widely distributed in the research area, the photo-taken locations and check-ins still show strong spatial clustered patterns which indicate the popular places and areas in the urban area of Beijing and the distribution of the temporal population density.
Figure 3-6. Map of Flickr Photo Density.

Figure 3-7. Map of Weibo Check-in Density.
3.1.3 POI Data

A point of interest (POI) is a particular place that has a specific function that someone may find interesting. In terms of POI data, POIs usually refers to the location signs with coordinating information on a map. For example, the location sign of the Stuckeman Family Building at Penn State on the Google Map is a POI. There are three types of POI involved in this research: they are dining places (including restaurants, bars, and coffee shops), commercial facilities, and green spaces (the green space data is introduced in Section 3.1.5). The dining place data was extracted from The Beijing Weibo check-in data that was introduced in section 3.1.2. While the other two were extracted from the China’s 2011 POI data which was prepared and provided by Beijing City Lab and Dr. Dong Li. This data contains two parts, one is a series of POI data which were separated
by categories of functions, and the other is a set of environmental elements including roads, islands, green space, mountains, etc. The POI data of commercial facilities is one of those POI datasets.

### 3.1.4 Data for Mixed Land Use Map

The metadata of the mixed land use map are selected POI datasets from the original China’s 2011 POI data (see section 3.1.3 and the calculations of the mixed land use map will be introduced in section 3.2). The selection criteria of the POI datasets is that the functions / land uses of these locations must relate to most people’s daily life. The selected types of POI datasets include government institutes, banks, hotels, office buildings, companies, residential communities, green spaces, hospitals, academic and research facilities, commercial facilities, dining and entertainment places, train stations, subway entrances, and bus stops.

### 3.1.5 Data of Green Spaces and Junctions

In this research, green spaces refer to the major chunks of vegetation, recreation facilities and water bodies. Green spaces include parks, gardens, water bodies, riverbanks, forests, farmlands, grasslands, and avenue with trees. There are three data sources: OpenStreetMap data, China’s 2011 POI data and FROM-GLC data (Finer Resolution Observation and Monitoring of Global Land Cover).

Green spaces with large footprints, especially waterbodies and major parks, were extracted from the base maps provided by OpenStreetMap. OpenStreetMap is a free editable worldwide map database supported by the collaborative effort of different groups of people. Because of the collaborative work, the OpenStreetMap has updated information and reliable accuracy for various types of applications. Besides, it also provides free access to the data of different kinds of map
features separately in vector format (shapefile) for GIS. Because usually it only provides most updated data, I used the 2015 OpenStreetMap data in this research. Similar to the data from OpenStreetMap, the China’s 2011 POI dataset also contains data of different kinds of environmental features. But the China’s 2011 POI dataset has more data of medium scale green spaces than the OpenStreetMap. Through combining the OpenStreetMap data and China’s 2011 POI data, the general structure of green spaces in Beijing was developed.

FROM-GLC (Finer Resolution Observation and Monitoring of Global Land Cover) is the first 30m resolution global land cover maps produced using Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) data (Gong et al. 2013). It is a kind of land cover data extracted from satellite images. Thus, the FROM-GLC data covers much more green spaces than the OpenStreetMap data and China’s 2011 POI data. However, because of its 30m resolution and raster format, the form and boundaries of green spaces are not as accurate as the other two types of green space data. In order to generate a more accurate and comprehensive map of green spaces in Beijing, these three datasets (the OpenStreetMap data, China’s 2011 POI data and FROM-GLC data) were integrated into one.
Figure 3-9. Map of Green Space.

Regarding the road map of Beijing, it was also provided by China’s 2011 POI dataset. The road data not only contains the major roads and highways, it also contains streets and alleys. The data are categorized based on road hierarchy and presented as polylines. If a road contains several lanes, the road is presented as a group of polylines and one polyline indicates one lane. The intersections of polylines indicate the junctions of roads and streets, so the junction of a multilane road counted as several intersections.
3.2 Data Processing

3.2.1 Trajectory Data Processing

The trajectory data was recorded by different GPS devices and collected from four different sources. Firstly, I converted all GPS trajectories into CSV (Comma Separated Values) format. By programming within R studio, in every trajectory, several new fields are added, for example, average speeds of the distances between every two points were calculated. CSV files of trajectories were normalized in the following format:

Field 1: Latitude in decimal degrees.
Field 2: Longitude in decimal degrees.
Field 3: User code.
Field 4: Trajectory code.
Field 5: Date and time as a string.
Field 6: Date as a string.
Field 7: Time as a string.
Field 8: Average Speed between two points.
Field 9: Distance between two points.
Field 10: Duration between two points.
Field 11: On Weekday or on Weekends (if the trajectory is a movement on a weekday, it will be coded with 1, otherwise, it will be coded with 0).

Since the trajectory data were collected by various devices, the units of elevation can hardly be identified so that the elevation information is not used in this research. Then, all processed data were imported into ArcMap. With satellite images, error points in the data were excluded. In addition, every trajectory was clipped by the research area and the exceeded portions excluded.
Another part of trajectory data processing was identifying and extracting the low speed movements or LSMs within trajectories. According to Orellana and Wachowicz (2011), usually GPS devices were less satisfactory for recording LSMs. “Consecutive GPS recordings of the position of a pedestrian standing still are quite unlikely to be in the same location, but rather will lie within an area defined by the GPS error circle (the circle inside of which the true horizontal coordinates of a position have a 50% probability of being located). Therefore, based on the commonly used tracking technology, no real gaps of zero movement (speed = 0 or near to zero) exist in GPS recordings (Orellana and Wachowicz 2011, 243).” Thus, in addition to extracting points with speeds near to zero (lower than 0.05 m/s), in the GIS environment, “Cluster and Outlier Analysis (Anselin Local Moran's I)” function was used to detect LSMs of pedestrian trajectories. In addition, the average speed of every segment between two record points was needed and these speeds were assigned to record points. The “Cluster and Outlier Analysis (Anselin Local Moran's I)” calculated the COType (clustering / outlier type), Z score (LMiZScore), and P value (LMiPValue) for every node of every trajectory. The COType contains 5 different types, they are “Not Significant (Random)”, “High-High Cluster (HH)”, “High-Low Outlier (HL)”, “Low-High Outlier (LH)”, and “Low-Low Cluster (LL)”. High-High Cluster means high value cluster while Low-Low Cluster mean low value cluster. If a point’s COType is High-Low Outlier, it means the point is a high value point in a low value cluster. Conversely, if a point’s COType is Low-High Outlier, it means the point is a low value point in a high value cluster. In this research, the values here mean the average speeds of segments that are assigned to points. P value is a variable that indicates the possibility of rejecting a null hypothesis. When the p value is less than or equal to the chosen significance level (usually 5% or 1%), the null hypothesis must be rejected. A Z-Score is a statistical measurement of a score's relationship to the mean in a group of scores. Based on Orellana and Wachowicz (2011), I used the following classifications (Table 3-2 and Figure 3-10) to identify LSMs in trajectories. As it was shown in Figure 3-7, the movement vectors in zone 4 are classified
as LSMs. However, the outcomes of this classification were clusters which were areas of possible LSM areas and they needed to be converted into location points. Firstly, “Aggregate Points” function was used to convert these clusters into polygons. According to previous research on GPS device (Zhang et al. 2006), the error bounds of GPS devices were generally between 15 meters to 30 meters. Thus, the aggregation distance was 30 meters and if the distance between two points was smaller than the aggregation distance, these two points were in the same cluster. Then, centroids of the polygons (clusters were used as LSM areas. After the classification and data revision, the map of LSM was drawn.

![LMI Z Score - Movement Speed](image)

Figure 3-10. Sample Scatterplot of a trajectory.

Table 3-2. Classification of Movement Vectors According to Movement Speed, COType, and Z score using a 5% Significance Level.

<table>
<thead>
<tr>
<th>Speed</th>
<th>COType</th>
<th>LMi Z Score</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; Avg</td>
<td>HH</td>
<td>&gt;1.96</td>
<td>High-speed vector surrounded by other high-speed vectors (high-speed cluster). See zone 1 in Figure 3-9.</td>
</tr>
<tr>
<td>&gt; Avg</td>
<td>HL</td>
<td>&lt;1.96</td>
<td>High-speed vector surrounded by low-speed vectors (high-speed outlier). See zone 2 in Figure 3-9.</td>
</tr>
<tr>
<td>&lt; Avg</td>
<td>LH</td>
<td>&lt;1.96</td>
<td>Low-speed vector surrounded by high-speed vectors (low-speed outlier). See zone 3 in Figure 3-9.</td>
</tr>
<tr>
<td>&lt; Avg</td>
<td>LL</td>
<td>&gt;1.96</td>
<td>Low-speed vector surrounded by other low-speed vectors (low-speed cluster). See zone 4 in Figure 3-9.</td>
</tr>
<tr>
<td>&gt; Avg or &lt; Avg Random</td>
<td>&gt;1.96 and &lt;1.96</td>
<td></td>
<td>All vectors with nonsignificant spatial association. See zone 5 in Figure 3-9.</td>
</tr>
</tbody>
</table>

3.2.2 POI Data Processing

The POI data are used to examine the correlation between the distance to specific kinds of POIs and pedestrian route choice or pedestrians’ LSMS. The distance map of specific kinds of POIs (dining places, commercial facilities, and green spaces) is generated by the “Euclidean distance” function in the ArcMap. The trajectory analysis is a micro scale analysis that requires a high resolution, however, calculating an output raster with a too high resolution can be very time consuming and may lead to a system crash, so I had to find a balance between output resolution and efficiency. After experimenting with different output cell size, I choose the cell size of 5m as the output cell size for these maps. Figure 3-11, 3-12, and 3-13 are outcome maps and on the maps, the colder areas are closer to those facilities.
Figure 3-11. Map of Distance to Commercial Facility (meter).

Figure 3-12. Map of Distance to Dining Space (meter).
3.2.3 Mixed Land Use Data Processing

According to Long and Liu (2013b), the mixed land use degree of a parcel (M) can be calculated with the following equation:

\[ M = - \sum_{i=1}^{n} (p_i \times \ln p_i) \]  \hspace{0.5cm} (3-1)

Where \( n \) denotes the number of POI categories, and \( p_i \) is the proportion of POIs of a particular category \( i \) among all POIs in the parcel (Long and Liu 2013b). The calculations were based on a 300m \( \times \) 300m (which is the common scale of city blocks in China) grid. The green space data needed to be converted into points from polygons. The green space data was first converted into a raster map with 300m \( \times \) 300m cell size, and then centroids of cells were generated for later calculations. After that, I used the “spatial join” function of ArcMap to join the different categories of POI data (government institutes, banks, hotels, office buildings, companies, residential...
communities, hospitals, academic and research facilities, commercial facilities, dining and entertainment places, train stations, subway entrances, bus stops, and green spaces) into the cells of the grid. As a result, the numbers of POIS of different categories within cells and the total amounts of all POIs within cells were calculated separately. Then, the mixed land use degree of each cell was calculated with the equation.

Figure 3-14. Map of Mixed Land Use.

3.2.4 Junction Density Data Processing

The junction density was used as an indicator of the complexity of traffic condition (Wegman, Lynam, and Nilsson 2002). Since a person chooses his / her path based on the
environment within a particular surrounding area, every junction density value was calculated with junctions within an area with a 100m radius. The junction points were generated by “Create Junction Connectivity Feature” tool and the tool created junctions with counts of roads connected to them. Then, the junction density was calculated by the “Point Density” tool. “Point Density” is a neighborhood level function that calculates a cell’s density value based on the values of all points within the moving search window whose center is the cell. Through moving the search window all over the analysis area, all cells are assigned with density values and generate a smooth surface of density values (Figure 3-16). The output cell size is 5m × 5m and the search radius is 100m. The count of connected roads was used as a weighted factor in the calculation. A higher junction density value means a more complicated traffic condition and a less safe environment for pedestrians.

Figure 3-15. Map of Junction Density.
3.2.5 Population Density Data Processing

The check-in densities of locations can help us to understand the popularities of locations and the population clustering pattern related to mobility (Li, Yang, et al. 2015). Since the social media data only present an incomplete check-in pattern with disconnected points so that the metadata needs to be interpolated to generate a complete surface of population (check-in) distribution with the “Point Density” function. Every 5m × 5m cell of the outcome was assigned the population density of its surrounding area (a 300m × 300m area). By using the “Point Density” function to interpolate, the outcome density values of in-between spaces of data points sometimes may be higher than the density values of data points around them. This actually describes the population densities of “transportation hubs” which are surrounded by several popular places. Then, the point density maps were normalized in to 0-100 scale so that they could be integrated into one population density map through raster calculation. The combination of maps uses following formula and the rating schedule is based on the popularity of social media:

\[
\text{Population Density} = (\text{Point Density of Flickr}) \times 0.2 + (\text{Point Density of Jiepang}) \times 0.35 + (\text{Point Density of Weibo}) \times 0.45
\]

Finally, after converting the result into 0-100 scale, the map of population density was generated.
3.3 Methodology

3.3.1 Research Question 1 about Travel Mode

According to Pappalardo et al. (2015), the ratio of the radius of a person’s regular activities ($R_s$) to the radius of all this person’s activities ($R_a$) can be used to measure an individual’s travel mode through studying the dispersion of his/her recurrent O-D (Origin-Destination) pairs. If a person’s $\frac{R_s}{R_a}$ is larger than 0.5, this person’s movement pattern approximates that of returner’s. Otherwise, it’s closer to an explorer’s. In order to identify whether a person was a returner or explorer, the ratio of $R_s$ to $R_a$ needs to be calculated. Pappalardo et al. (2015) used a
weighted “Standard Distance” method to calculate the activity radiuses of respondents. “Standard Distance” measures how dispersed a group of features is around its mean center. The “Standard Distance” tool in GIS generates a circle polygon feature centered on mean center of input features and the radius of polygon is equal to the standard deviation. The standard deviation (σ) is calculated by the following equation:

\[ \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2} \]  \hspace{1cm} (3-2)

Where \( N \) denotes the total number of features, and \( x_i \) is the \( i \)-th feature, \( \mu \) is the mean of features.

Pappalardo et al. (2015) replace the mean of features (\( \mu \)) by the most frequently visited location of all \( k \) locations (\( k \leq \) the total amount of locations) and use the count of visits as the weight. All the locations in the calculation need to be sorted by visit counts of locations and the location \( r_1 \) is the most frequently visited location. Their equation for \( R_s \) was the following:

\[ R_s^{(k)} = \frac{1}{W_k} \sum_{i=1}^{k} w_i (r_i - r_{cm}^{(k)})^2 \]  \hspace{1cm} (3-3)

Where \( r_i - r_{cm}^{(k)} \) was the Euclidian distance between location \( r_i \) and the center feature \( r_{cm}^{(k)} \), and \( w_i \) was the number of visits of location \( r_i \) while \( W_k \) was the sum of the weights. \( R_{a} \) is also calculated with this equation. When \( k \) equals to the total amount of locations, the outcome of the equation was \( R_{a} \) (Pappalardo et al. 2015).

In order to study the movement patterns of Beijing citizens, in the Microsoft’s GeoLife Trajectory Dataset, all trajectory data of the top 50 users in data size were selected as the sample dataset. Using Rstudio, I automatically extracted O-D (Origin-Destination) pairs of trajectories and the O-D pairs of the same users were dissolved (combined). Before using the Equation 3-3, the number of visits of every location needed to be calculated. Due to the limitation of GPS accuracy, the GPS records of a place may not be in the same exact location (Zhang et al. 2006), I set a tolerance scope in calculation and the tolerance scope is an area around the place within a radius of
30m based on the research of Zhang et al. (2006). With the “Aggregate Points” tool (the aggregation distance is 30m), frequently visited locations were identified through identifying spatial clusters within O-D data of every user and minimum bounding polygons of these clusters were generated in a new layer. I spatial joined the original point features to the “aggregate points” polygon features by person to get the join count amounts of polygons (the visit counts of places). Then I created the centroids of those polygons and replaced the points within every polygon with the centroid of that polygon. After that, I got all the O-D networks with visit frequency fields. The center feature was the median center of the most frequented locations. According to Pappalardo et al. (2015), when k=4, the outcome amounts of returners and explorers met the balance and were close to the reality. With the Equation 3-3 provided by Pappalardo et al. (2015), when k = 4, $R_s$ of all respondents were calculated. $R_a$ of respondents were calculated with the sample equation when k equaled to the total number of places. Finally, a scatter plot of $R_a$ and $R_s$ was drawn to show whether the majority of the sample respondents were explorers or returners. If most of the sample respondents’ $R_s/R_a$ is greater than 0.5 and less than 1, most of the sample respondents’ travel modes were closer to the returner’s. If most of the sample respondents’ $R_s/R_a$ is less than 0.5 and greater than 0, most of the sample respondents’ travel modes were closer to the explorer’s. Otherwise, there was no significant tendency showed in the sample respondents.

Furthermore, I ran a further analysis on whether there was a significant difference between the routine activity radiiuses of explorers and the routine activity radiiuses of returners through statistic tests. I ran the Shapiro-Wilk normality tests to see whether the data were normally distributed. If so, I would use the Student’s t-test, otherwise, I would use the Wilcoxon rank sum test. In addition to comparing the activity radiiuses between different travel modes, I also compared the weekday’s activity radiiuses and weekend’s activity radiiuses of the same travel mode to see whether there were different movement patterns on weekdays and at weekends through a similar statistical process.
3.3.2 Method for Research Question 2 about Low Speed Movement (LSM)

Every trajectory in this research has a single start point and a single destination. About 75% of LSMs last less than one minute. Therefore, LSMs here are more related to the following scenarios, people slow down to avoid collisions with others or people are attracted by surrounding environment so that they stop to take pictures or a rest. This test aims to find out what environmental variables may be related to LSMs. This test assessed correlations between the durations of LSMs and LSM trajectories’ surrounding environmental conditions separately for different environmental variables (the distance to commercial facility, the distance to dining place, the distance to green space, junction density, and population density).

In order to get LSM trajectories’ surrounding environment in terms of those different environmental variables, I firstly created buffers, whose radiuses were 30m, for every LSM trajectory. LSM trajectories with buffers were overlaid on different base maps of different environmental variables and the “Zonal Statistics to Table” was used to get the mean value of every specific environmental variable within the buffer of every LSM trajectory. Then I correlated the durations of LSMs with these data.

3.3.3 Method for Research Question 3 about Trajectory Typology

Due to the limitation of GPS accuracy, trajectory data usually have error records (outliners) so that it is hard to detect different trajectory typologies automatically with some parameters. In order to identify the typologies of trajectories, I mapped those trajectories into polylines on the satellite image and manually categorized them. Through converting the trajectories from sequences of points to lines with “XY to Line” tool, the patterns of trajectories were mapped and categorized into two types: the circular trajectory and easiest route trajectory. Through simulating shortest paths
with O-D pairs of trajectories as a control group to compare with actual trajectories, trajectories of the shortest path and the least cost route trajectory, which followed both the shortest path and the easiest path, were detected. Trajectories of Redundant routes are either longer than the shortest network distances or more tortuous than the easiest path. Redundant route trajectories were identified by both distance comparisons with the shortest paths and turn amount comparisons with the easiest paths. Then the type of backtracking path, which was a round trip following the same path, was separated from the circular trajectories as well.

Boxplot is an efficient way to describe the pattern of data through showing the distribution of values while highlighting the median and the majority value by the box as well as the minimum and maximum by the nodes on both sides of the line (Figure 3-18). Furthermore, by drawing boxplots of different trajectory typologies, the commonality and differences of different types can be well described and visualized (Dodge et al. 2014). After categorizing, descriptive statistics including the breakdown of different trajectory types, boxplots of the durations of different trajectory types was drawn. In addition to the durations of different trajectory types, boxplots of trajectories’ surrounding environmental conditions was drawn separately for different environmental variables (the distance to commercial facility, the distance to dining place, the distance to green space, mixed land use degree, junction density, and population density) and different trajectory types.
Since sample amounts of different types of trajectories are not balanced, it will need more samples in the future to balance the dataset so that the analysis can deal with all different types of trajectories together with methods like one-way ANOVA. This part of research focuses more on the comparisons among trajectory types with sufficient data (See Section 4.3) to figure out the environmental influences on pedestrian route choice. Before assessing whether there are significant differences among different kinds of trajectories in terms of a series of environmental variables (the distance to commercial facility, the distance to dining space, the distance to green space, mixed land use degree, and population density and junction density) with statistics approaches, Shapiro-Wilk normality tests for datasets of different variables were ran first to decide what kinds of statistical tests to use. If the data are normally distributed, then I would use the Student's t-test. Otherwise, I used the Wilcoxon signed-rank test (which is a non-parametric test) instead. Moreover,
for this part of the analysis, I present some follow-up approach illustrations in Chapter 4 with the result of categorizing and descriptive statistics.

3.4 Limitation

This research has three limitations: 1). Most of the GPS data are self-reported and collected by different devices within different conditions. Thus, not all the GPS data are perfectly accurate and much of trajectory data lack contextual information, for example, demographic information. 2). Due to the limited availability of data in China, it was necessary to make assumptions in the analysis and interpolations in the base map generations, which may need to be refined in the future. 3). This thesis studies only one city, Beijing. More cross comparisons between different cities and works on other cities will need to be done in the future.

3.5 Validation

This research result was validated by three approaches. Firstly, ground truth and daily life experiences were used to validate the research results. If the results are not strongly against ground truth and daily life experiences, the results will regarded as reliable. Besides, the outcomes were also compared with results of related research in other theoretical perspectives in a way of triangulation. Thirdly, this thesis has several member checks with my advisor and committee members from different professional fields, including landscape architecture, geography, and information science, to ensure the credibility. Furthermore, external audits was also used in validation and a professor not involved in the research process evaluated both the process and product of the research study in order to ensure the findings, interpretations and conclusions were supported by the data and methodology.
Chapter 4

Analysis and Results

In this chapter I present the findings of three research questions introduced in Chapter 2. The findings are ordered by research questions and organized in following sections: 1) The 4.1 section presents findings about the travel mode of Beijing citizens, including travel mode identification, comparisons between different travel modes and the factors related to people’s travel modes; 2) The section 4.2 introduces the findings about environmental factors that affect pedestrians’ movement suspensions; 3) The last section, section 4.3, presents findings about different types pedestrian trajectories and the environmental factors that affect pedestrian route choice and help us to identify trajectory types from GPS data in practice.
4.1 The Travel Mode of Beijing Citizens

As a result of travel mode analysis, 26 of the total sample of 50 respondents from GeoLife dataset are found to be returners while 24 sample respondents are explorers, so there is no significant tendency in sample respondents’ movement pattern in terms of being returner versus explorer.

![Diagram of Travel Mode Analysis](image.png)

Figure 4-1. Diagram of Travel Mode Analysis.

However, when using Wilcoxon rank sum test to compare the difference between explorers’ routine activity radiuses and returners’ routine activity radiuses, there is a significant ($p = 3.497e-07 < \alpha = 0.05$) difference. Returners’ routine activity radiuses are significantly greater than explorers’ routine activity radiuses.
Table 4-1. Routine Activity Radius Comparison.

<table>
<thead>
<tr>
<th></th>
<th>Shapiro-Wilk normality test</th>
<th>Wilcoxon rank sum test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w</td>
<td>p</td>
</tr>
<tr>
<td>Explorers’ routine activity radiuses</td>
<td>0.94847</td>
<td>0.2135</td>
</tr>
<tr>
<td>Returners’ routine activity radiuses</td>
<td>0.88341</td>
<td>0.00974</td>
</tr>
</tbody>
</table>

If $P < 0.05$, the data does not follow the normal distribution.

Figure 4-2. Boxplots of Total Activity Radiiuses on Weekdays and at Weekends.
Table 4-2. Weekday Activity Radiuses and Weekend Activity Radiuses Comparison.

<table>
<thead>
<tr>
<th></th>
<th>Weekday Activity Radius</th>
<th>Weekend Activity Radius</th>
<th>p (Weekday Activity Radius &lt; Weekend’s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explorer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shapiro-Wilk normality test</td>
<td>Wilcoxon signed rank test</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td>p</td>
<td>Normal-Distributed?</td>
<td></td>
</tr>
<tr>
<td>0.79147</td>
<td>0.0001265</td>
<td>F</td>
<td>0.02467</td>
</tr>
<tr>
<td>0.85806</td>
<td>0.002047</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>Returner</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shapiro-Wilk normality test</td>
<td>Student’s t- test</td>
<td></td>
</tr>
<tr>
<td>Weekday Activity Radius</td>
<td>0.94115</td>
<td>0.1731</td>
<td>T</td>
</tr>
<tr>
<td>Weekend Activity Radius</td>
<td>0.93484</td>
<td>0.1251</td>
<td>T</td>
</tr>
</tbody>
</table>

In Shapiro-Wilk normality test, if P < 0.05, the data does not follow the normal distribution.

According to the boxplots (Figure 4-2), the medians of weekend activity radiuses are greater than those of weekday activity radiuses. Furthermore, the difference between explorers’ radius medians is greater than the difference between returners’ radius medians. In order to test whether these are significant differences, I compared weekday activity radius and weekend activity radius of both types of people with test statistics. The Wilcoxon signed rank test was used for comparison of the data of explorers, because the Shapiro-Wilk normality tests showed the data did not fit the normal distribution. The Student’s t- test was used for comparison of the data of returners, because the Shapiro-Wilk normality tests showed the data fitted the normal distribution. The results
indicate that both explorers and returners have greater activity radiuses during weekends relative to weekdays.

Due to the small sample size, this analysis does not provide a definitive answer for the assessment of the major travel mode of Beijing citizens. However, because the sample data is randomly selected, the results are still worth noting. Yanqi Wang pointed out that Beijing citizens, especially the young generations (which are the major respondents of GeoLife Trajectory Dataset) have heavy pressure on their lives and spend a lot of time at work (Wang 2002) -- probably a cause for the difference between weekday and weekend travel radius. Moreover, people spend a lot of time in commuting because of the traffic (Wang, Lei, and Shi 2002). This explains why both explorers and returners often have greater activity radiuses at weekends. According to the routine activity radius comparison between returners’ and explorers’, returners’ routine activity radiuses are usually greater than explorers’ routine activity radiuses. It can be inferred that a user’s travel mode may be impacted by his/her job – housing distance. If he/she needs to travel long distances to get to his/her routine locations on weekdays, he/she may not prefer long distance travel and chooses not to travel much at his/her spare time, so he/she may act as a returner.

4.2 The Spatial Structure of LSMs in Trajectories

Every trajectory in this research has a single start point and a single destination. About 75% of LSMs are last less than one minute. Therefore, LSMs here are more related to the scenarios like the instance when people slow down to avoid collisions with others or people are attracted by the surrounding environment, prompting them to stop to take pictures or spend a few minutes standing and observing. Since all the data don’t fit the normal distribution (Table 4-3), Spearman's rank correlations were used to find out the relationship between the length of LSM durations and different environmental factors. According to the tests, the lengths of LSM durations are negatively
correlated with mean distance to commercial facility and mean distance to dining facilities. It means that when pedestrians walk near commercial facilities or dining places, they usually walk slower, probably because the spaces are crowded so that pedestrians need to slow down to avoid collisions, or pedestrians are attracted by shops and dining choices – prompting them to stop to check out the details.

Table 4-3. Shapiro-Wilk normality test

<table>
<thead>
<tr>
<th></th>
<th>LSM Duration</th>
<th>Mean Distance to Commercial Facility</th>
<th>Mean Distance to Dining Place</th>
<th>Mean Distance to Green Space</th>
<th>Mean Junction Density</th>
<th>Mean Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>w</td>
<td>0.24005</td>
<td>0.66447</td>
<td>0.49621</td>
<td>0.6017</td>
<td>0.90958</td>
<td>0.2089</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
</tr>
</tbody>
</table>

If P < 0.05, the data does not follow the normal distribution

Table 4-4. Spearman's Correlation between LSM Duration and Environment Factors

<table>
<thead>
<tr>
<th></th>
<th>Mean Check-in Density</th>
<th>Mean Distance to Commercial Facility</th>
<th>Mean Distance to Dining Facility</th>
<th>Mean Distance to Green Space</th>
<th>Mean Junction Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman's rank correlation coefficient</td>
<td>-0.00760455</td>
<td>-0.09077601</td>
<td>-0.09079136</td>
<td>0.03891323</td>
<td>0.02443266</td>
</tr>
<tr>
<td>p</td>
<td>0.842</td>
<td><strong>0.01707</strong></td>
<td><strong>0.01706</strong></td>
<td>0.3074</td>
<td>0.5217</td>
</tr>
</tbody>
</table>
4.3 Typologies of Pedestrian Trajectory

After mapping all the trajectories into polylines on the satellite image, I categorized them into five types:

Type A (least cost): Trajectories that follow the shortest (shortest distance) and easiest (fewest turns) paths.

Type B (easiest): Trajectories that follow the easiest path only.

Type C (shortest): Trajectories that follow the shortest path only.

Type D (redundant): Trajectories of the unclosed redundant routes whose lengths are longer than the Manhattan distances between the origins and destinations and turns are more than necessary.

Type E (circular): Trajectories of circular paths.

Type F (backtracking): Trajectories of backtracking paths whose inbound trips and outbound trips follow the same paths.

In total 172 trajectories, there are 54 least cost path trajectories, 8 shortest path trajectories, 10 easiest path trajectories, 20 circular route trajectories, 4 backtracking path trajectories, and 76 redundant routes trajectories (Figure 4-3). Through mapping the duration patterns of different types of trajectories (Figure 4-10), I discover that the redundant route trajectories and the circular trajectories usually have long durations with 48 trajectories of redundant and circular movement lasting more than an hour. Furthermore, the redundant route trajectories and the circular trajectories usually have more complicated spatial patterns than other types of trajectories. The short duration movement, especially those lasting between 10 to 20 minutes, mostly followed the shortest paths (shortest distance) or easiest paths (fewest turns).
Figure 4-3. Breakdown of Trajectory Types.

- A (least cost)
- B (easiest)
- C (shortest)
- D (redundant)
- E (circular)
- F (backtracking)

Figure 4-4. A Sample of Type A Trajectory.
Figure 4-5. A Sample of Type B Trajectory.

Figure 4-6. A Sample of Type C Trajectory.
Figure 4-7. A Sample of Type D Trajectory.

Figure 4-8. A Sample of Type E Trajectory.
Figure 4-9. A Sample of Type F Trajectory.

Figure 4-10. Boxplots of Durations of Trajectories (separated by types).
Figure 4-11. Boxplots of Junction Density (separated by types).

Figure 4-12. Boxplots of Mixed Land Use (separated by types).
Figure 4-13. Boxplots of Population Density (separated by types).

Figure 4-14. Boxplots of Distances to Different Types of POIs (separated by types).
The numbers of different types of trajectories in the sample are not balanced with most trajectories belonging to the redundant route trajectories (76 in total) and the least cost path trajectories (54 in total). Furthermore, according to the boxplots (Figure 4-11 – 4-14), the shortest path trajectories seem to be greatly different from other types of trajectories in various aspects of their surrounding environment. Thus, here I present how the redundant route trajectories differ from the shortest path trajectories and the least cost trajectories.

As for the statistical analysis between the redundant route trajectories and the least cost path trajectories, I ran the Shapiro-Wilk normality tests first and I found that most of the data do not fit the normal distribution (Table 4-5). So I used the Wilcoxon signed rank tests for the comparisons. According to these Wilcoxon signed rank tests, there is no significant difference between the redundant route trajectories and the least cost path trajectories with reference to environmental factors (Table 4-6). This may be because there is significant overlap between the trajectories of redundant and least cost path in their main route choice criteria – walking on major streets, so they share similar environmental features. In terms of the rest of the redundant route trajectories, their redundant parts are mostly located on campuses and green spaces whose environment is also less complicated as the environment of the major streets.
Figure 4-15. A Mapping of the Least Cost Paths and Redundant Routes.
Table 4-5. Shapiro-Wilk normality test

<table>
<thead>
<tr>
<th></th>
<th>Mean Distance to Commercial Facility</th>
<th>Mean Distance to Dining Place</th>
<th>Mean Distance to Green Space</th>
<th>Mean Mixed Land Use Degree</th>
<th>Mean Junction Density</th>
<th>Mean Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redundant paths</td>
<td>w</td>
<td>0.63928</td>
<td>0.55826</td>
<td>0.72021</td>
<td>0.93418</td>
<td>0.86125</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>1.85e-12</td>
<td>7.278e-14</td>
<td>8.152e-11</td>
<td>0.0006217</td>
<td>5.892e-07</td>
</tr>
<tr>
<td>Least cost paths</td>
<td>w</td>
<td>0.72147</td>
<td>0.79414</td>
<td>0.93186</td>
<td>0.83529</td>
<td>0.96556</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>0.001605</td>
<td>0.0123</td>
<td>0.4665</td>
<td>0.03874</td>
<td>0.8469</td>
</tr>
</tbody>
</table>

Table 4-6. Comparison between Redundant Routes and the Least Cost Paths.

<table>
<thead>
<tr>
<th></th>
<th>Wilcoxon signed rank test (p value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean Distance to Commercial Facility</td>
</tr>
<tr>
<td>Two tailed</td>
<td>0.8784</td>
</tr>
<tr>
<td>Redundant path &gt;</td>
<td>Least cost paths</td>
</tr>
<tr>
<td>Redundant path &lt;</td>
<td>Least cost paths</td>
</tr>
</tbody>
</table>
In comparing the redundant route trajectories and the shortest path trajectories, for the same reason as the former comparisons (Table 4-7), I used the Wilcoxon signed rank tests for the comparisons. Two types of trajectories show significant differences in mean distance to commercial facility, mean distance to dining place, mean mixed land use degree, and mean junction density (Table 4-8). It seems that the shortest path trajectories go through more complicated areas as many of them go across blocks. The shortest path trajectories are more close to crowded or attractive area (smaller mean distance to commercial facility and mean distance to dining place) and they cross more diverse surrounding environments (greater mixed land use degree and mean junction density).

There are two kinds of redundant route trajectories. Some of the redundant route trajectories mainly follow the major street network. The reason why these people choose longer routes is probably because they consider the complexity of the walking environment as a more important criteria than travel distance in terms of route choice. Pedestrians chose a less complicated and less crowded environment to walk so that they can save more time even though they walked longer. Furthermore, they may want to walk more safely and have a better sense of direction by choosing a less diverse environment, so they are more willing to walk along major streets instead of walking across blocks.

In terms of the rest of the redundant route trajectories, their redundant parts are mostly located on campuses and green spaces whose environment is less complicated. These people might walk through longer distances in campuses and green spaces for recreation or exercise purposes. Thus, for recreation and exercise purposes, people may prefer to walk in green spaces or other similar environment.

As I mentioned before, people are willing to walk longer than necessary mainly under two circumstances. One is people choose longer paths to avoid negative environmental factors (for example, crowding), the other is people do not choose the shortest paths for exercise and recreation purposes. From the perspective of spatial pattern analysis, trajectories under these two different circumstances are different in their tortuosity values and surrounding environment, which may help
us to identify different activities through identifying different trajectory types. Tortuosity is a measurement that describes how twisted a curve is, compared to the straight line between its start point and end point. The arc-chord ratio of the length of a trajectory to the length of the shortest path between two ends of it is used to estimate the trajectory tortuosity. If the arc-chord ratio of a trajectory is close to 1, then this trajectory is more likely to be chosen in the pedestrian’s movement to avoid negative environmental factors. Otherwise, when the tortuosity is closer to infinity, the activity of this trajectory is more for exercise or recreational purposes. In order to discover the route choice preferences in redundant routes, with the correlation method, I also test the hypothesis that in terms of redundant routes, people prefer to walk in green spaces or less complicated environments for exercises or recreation activities. Therefore, I correlate the tortuosity of the trajectory (Independent variable) with the ratio of the surrounding environmental factors of the redundant route trajectory (RT) to those of the shortest path trajectory (ST) between the two ends of it. Variables and results of this correlation are shown in Table 4-9. According to the results, the tortuosity is positively correlated with the ratio of mean distance to commercial facilities of RT to that of ST and the ratio of mean distance to dining spaces of RT to that of ST. Furthermore, the tortuosity is negatively correlated with the ratio of mean distance to green spaces of RT to that of ST and the ratio of mixed land use degree of RT to that of ST. In other words, it shows that in terms of redundant routes, people prefer to walk in green spaces or less complicated environment when they are walking for recreation or exercise (when the tortuosity of the trajectory is high). Moreover, with these findings, I argue that by studying the tortuosity of the trajectory and the percentage of the trajectory in green space, researchers can identify exercise trajectories from GPS logs of walking.
Figure 4-16. A Mapping of the Shortest Paths and Redundant Routes
Table 4-7. Shapiro-Wilk normality test.

<table>
<thead>
<tr>
<th></th>
<th>Mean Distance to Commercial Facility</th>
<th>Mean Distance to Dining Place</th>
<th>Mean Distance to Green Space</th>
<th>Mean Mixed Land Use Degree</th>
<th>Mean Junction Density</th>
<th>Mean Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortest Route</td>
<td>w 0.62346</td>
<td>0.66809</td>
<td>0.67817</td>
<td>0.89399</td>
<td>0.66489</td>
<td>0.30861</td>
</tr>
<tr>
<td></td>
<td>p 9.478e-13</td>
<td>6.612e-12</td>
<td>1.051e-11</td>
<td>9.745e-06</td>
<td>5.719e-12</td>
<td>&lt; 2.2e-16</td>
</tr>
</tbody>
</table>

Table 4-8. Comparison between Redundant Routes and the Shortest Paths.

<table>
<thead>
<tr>
<th></th>
<th>Wilcoxon signed rank test (p value)</th>
<th>Mean Distance to Commercial Facility</th>
<th>Mean Distance to Dining Place</th>
<th>Mean Distance to Green Space</th>
<th>Mean Mixed Land Use Degree</th>
<th>Mean Junction Density</th>
<th>Mean Population Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two tailed</td>
<td>0.01399</td>
<td>0.001095</td>
<td>0.7223</td>
<td>0.0003198</td>
<td>2.512e-14</td>
<td>0.1768</td>
</tr>
<tr>
<td>Redundant path &gt; Shortest Route</td>
<td>0.00696</td>
<td>0.0005477</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Redundant path &lt; Shortest Route</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0001599</td>
<td>1.256e-14</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4-9. Correlations between the Tortuosity of the Trajectory (Independent Variable) and the Ratio of the Surrounding Environmental Factors of RT to those of ST.

<table>
<thead>
<tr>
<th>Tortuosity</th>
<th>Ratio of mean distance to commercial facilities of RT to that of ST</th>
<th>Ratio of mean distance to dining places of RT to that of ST</th>
<th>Ratio of mean distance to green spaces of RT to that of ST</th>
<th>Ratio of junction density of RT to that of ST</th>
<th>Ratio of mix land use degree of RT to that of ST</th>
<th>Ratio of population density of RT to that of ST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shapiro-Wilk normality test</td>
<td>w</td>
<td>0.7625</td>
<td>0.78447</td>
<td>0.77316</td>
<td>0.76772</td>
<td>0.85726</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>7.978e-10</td>
<td>2.905e-09</td>
<td>1.478e-09</td>
<td>1.077e-09</td>
<td>4.301e-07</td>
</tr>
<tr>
<td>Spearman's rank correlation</td>
<td>Ratio of Length (Length of the redundant route/length of the shortest route)</td>
<td>coefficient</td>
<td>0.3809086</td>
<td>0.2760923</td>
<td>-0.2309795</td>
<td>-0.1531626</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>0.0006318</td>
<td>0.01508</td>
<td>0.04327</td>
<td>0.1836</td>
<td>0.02287</td>
</tr>
</tbody>
</table>
Chapter 5

Conclusions and Discussions

5.1 Conclusions

There are five major outcomes from this analytical investigation: (1) A person’s travel mode, whether the person is an explorer or a returner, is impacted by his/her job and his/her job-housing distance; (2) Pedestrian route choice is shaped by external environmental factors, and trip distance and the amount of turns are not always the dominant criteria in route choice; (3) Among all six environmental factors, commercial facilities and dining places are the most effective environmental elements that can influence people’s route choice; (4) People prefer to walk in green spaces or less complicated environment (away from crowded areas around commercial facilities and dining spaces) when they are walking for recreation or exercising. They may choose curvilinear paths that are longer than the shortest paths to walk through more green spaces; (5) Exercise trajectories can be identified from GPS logs of walking by studying the tortuosity of the trajectory and the percentage of the trajectory in green space.

However, the current situations are as follows: (1) High population density appeared in most areas of downtown Beijing and people keep migrating from rural and suburban areas to the urban section of Beijing. (2) The green space network in Beijing is imbalanced in that its structure is heavily based on urban level parks and traditional imperial gardens which also serve as tourist attractions. There is less attention given to community level open spaces. Therefore, most of the nodes in current Beijing’s green space network are overcrowded and the network is not dense and dispersed enough for citizens as they cannot find good quality green spaces near their frequently visited locations (homes or offices). (3) Due to the imbalanced development between core urban
area and suburbs, the core urban area has a much higher house price and more job opportunities than the suburbs. Many of Beijing’s citizens, especially the young generation, work at downtown but live far away from the downtown and they spend a lot of time in commuting. Thus, many of Beijing’s citizens can only enjoy the green spaces of parks on weekends. On weekdays, they are tired after spending a lot of time at work and the green spaces are not accessible enough for them because of the low density, so the citizens are less likely to go outdoors and enjoy outdoor activities on weekdays. According to the findings of this thesis, in order to improve the walkability of Beijing and provide a better environment for its citizens, here are some suggestions for future planning projects:

1. Since explorers are used to having even long-distance travels so that their outdoor activities and destination choices are less influenced by the distance factor. The existing green spaces and public amenities are more accessible to explorers than returners. In order to improve the accessibility for returners, future urban developments and planning should pay more attention to the green space and public amenity planning at the community level to develop a denser and greener infrastructure network across the city and encourage returners to enjoy outdoor activities.

2. Future planning should balance the development of urban, suburban and rural areas. This can be achieved through building new attractions and creating more work opportunities in rural and suburban areas and improving the connectivity between suburban, rural and urban areas. Furthermore, steps can be taken to gradually decentralize urban areas and release the population pressure in urban areas of center cities and to increase the spatial quality of urban areas and encourage people to enjoy outdoor green spaces.

3. In future planning, the density of green spaces should be linked to the distribution of population density in the area. The importance given to traditional imperial gardens
and tourist attractions in the existing green space network should be decreased. The
green space network should incorporate community level parks, as those tourist
attractions (imperial gardens for example) are already overloaded and residents do not
prefer to have leisure or recreational walks in areas with high population densities.

4. The differences of the underlying structures of redundant routes and shortest paths also
indicate that in Beijing, most of the supportive facilities of social activities are still
located inside the blocks and not along the streets. In future planning projects, planners
and designers should find creative ways to maintain this unique urban character of
Beijing.

5.2 Future Studies

The research presented above offers alternative perspective for future research:

Comparative Research

The route selection process and way finding behavior are the results of people’s internal
cognitive process and the influences of external environmental factors. The internal process
depends on people’s physical / mental abilities, spatial knowledge, spatial memory and personal
background including cultural background and educational background. This research is less
focused on the influence of cultural and social factors related to internal processes. Also this study
is limited to Beijing. Based on the analysis algorithms provided by this research, future research
could analyze the situations in other cities. By comparing the results of different cities of different
cultures, the influence of cultural and social factors will be revealed. In addition to carrying on this
kind of study in cities of different cultures, researchers should also focus on conducting studies in
cities with similar conditions with respect to culture, society, and nature environment to generate
larger sample sizes for more credible results.
Simulation Research

This research is an analytical investigation trying to provide clues for future research based on simulation and prediction model building. With factors identified in this research, researchers of related fields, including GIScience and computer science may be able to develop models or prediction tools to help planners, governments and related service providers create activity friendly urban environments.

Trajectory Study on Spatio-Temporal Pattern with Social Media Data

This research talks more about less dynamic issues, including the character of underlying structure and people’s travel models than examining the trajectory under a real-time scenario. Since human movement is sensitive to spatio-temporal changes and in order to represent trajectories better with more contextual information, methodologies that can analyze and visualize spatio-temporal patterns of travel trajectories need to be developed. Even though in recent years, trajectory research has benefited from the development of new tools, for example, GPS devices, GIS, social media and sensors, there is the need to develop new tools for trajectory studies that can take into account the spatio-temporal pattern of human movement.
References


Dodge, Somayeh, Gil Bohrer, Keith Bildstein, Sarah C. Davidson, Rolf Weinzierl, Marc J. Bechard, David Barber, Roland Kays, David Brandes, Jiawei Han, and Martin Wikelski. 2014. "Environmental drivers of variability in the movement ecology of turkey vultures (Cathartes aura) in North and South America." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 369 (1643):20130195. doi: 10.1098/rstb.2013.0195.


