A TALE OF TWO STATIONS: ANALYZING METRO RIDERSHIP WITH BIG DATA

Liang Ma
Center for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
Skempton Building, South Kensington Campus
SW7 2AZ London, United Kingdom
Tel: +8613803310711; Email: marina.ma0711@gmail.com

Qingning Chen
Innovation Center of Technology
Beijing Tsinghua Tongheng Urban Planning & Design Institute
Floor 22, Building Jia1, Qinghe Jiayuan
Haidian District, Beijing, China
Tel: +8613996281201; Email: chenqingning@thupdi.com

Dr. Ke Han, Corresponding Author
Center for Transport Studies
Department of Civil and Environmental Engineering
Imperial College London
605 Skempton Building, South Kensington Campus
SW7 2AZ London, United Kingdom
Tel: +44 (0)20 7594 5682; Email: k.han@imperial.ac.uk

Yong Gao
Beijing Branch of Shenzhen Urban Transport Planning Center
768 Creative Industrial Park, Haidian District, Beijing, China
Tel: +8613661399034; Email: gaoyong@sutpc.com

Dr. Dong Li
Innovation Center of Technology
Beijing Tsinghua Tongheng Urban Planning & Design Institute
Floor 22, Building Jia1, Qinghe Jiayuan
Haidian District, Beijing, China
Tel: +8613683544077; Email: lidong@thupdi.com

Word count: 5,247 words text + 9 tables/figures x 250 words (each) = 7,497 words

Submission date: 1 August 2017

ABSTRACT

This paper presents a multi-dimensional case study of the Beijing metro system. In particular, we examine two non-transfer stations, Zaoying and Jiangtai, which are on the same metro line in central Beijing. Multi-source and heterogeneous data are integrated to analyze and diagnose the drastically different metro ridership at the two stations. These include transit smart card data, taxi GPS data, network data, Point of Interest data, demographic data, online second-hand property price data, cell phone signalling data, and bike sharing data. The different utilization of metro system at these two locations is attributed to a number of factors pertaining to transportation infrastructure, built environment, demographic composition, commuting patterns, and connectivity of multi-modal transit networks. The findings suggest the importance of local accessibility of the metro stations as well as its connectivity with the rest of the transit system, in order to maximize the transport capability of the metro system. Our analysis also highlights the benefit of collecting and analyzing fine-granularity data in order to identify key bottlenecks and inefficiencies in the transportation system, as conventional macroscopic transportation planning data do not sufficiently capture the local accessibility and mobility in an urban environment.

Keywords: Urban rail transit, Big data, Urban planning, Mobility and accessibility, Shared bike

1. INTRODUCTION

The rapid development of Information and Communication Technologies (ICT) in recent years has enabled the acquisition, storage, and processing of large amount of data in extremely high granularities. ICT-based data analytics, such as data mining and visualization, has been widely applied in energy, medicine, and social science. Data-driven research and development bring paradigm-shifting innovation to the conventional way of scientific exploration, and underpin the technological aspect of urban studies.

The city-scale deployment of ICT infrastructure and widespread utilization of internet-based services have led to the accumulation of data on urban dynamics and individual behavior, with large quantity, high frequency, and significant diversity. Data frequently seen in recent urban studies include: map data, point of interest (POI) data, floating car data, passenger flow data, cell phone (signalling) data, location based service data, camera/video image data, environment monitoring data, meteorological data, and social activity data (*1*). These urban data are widely sourced and contain rich and diverse information that supports multi-scale, multi-dimensional, and fine-granularity analyses.

According to the technical approach employed, the application of urban data can be categorized as: (1) analysis of fine-granularity characteristics from a multi-disciplinary perspective; (2) development of domain-specific models/methodologies based on cross-disciplinary studies; and (3) pattern recognition, characterization, and prediction based on machine learning methods.

The first category is mostly seen in the fields such as geography, economics, sociology, and transportation. This type of application tends to extract relevant information from vast data sets, refine the spatio-temporal granularities of existing studies, extend the humanistic scale in conventional urban analysis, and enrich the context in which traditional urban studies are conducted. Examples of this type include city-wide network analysis (*2-3*), urban infrastructure dynamics (*4-5*), regional characteristics analysis (*6*), and individual behaviour analysis (*7-9*). Additionally, urban data typically contain rich spatial and temporal information to conduct general spatial analysis such as buffer analysis, spatial overlay analysis, network analysis, spatial statistics analysis, and spatial econometric analysis (*10*). Zhen et al. (*11*) use social network data from Sina Micro-blog to interpret the hierarchical structure of the Chinese city network. Gao (*12*) considers cell phone signalling data and employs spatio-temporal visualization, space-time kernel density estimation, and spatio-temporal autocorrelation analysis to explore and visualize intra-urban and inter-urban mobility patterns.

The second category mainly focuses on improving the calibration and validation of existing models and refining their temporal and spatial granularities. Liu et al. (*13*) mine cell phone signalling data to infer activity-based trip chain of individuals based on their residence and employment locations. The extracted trips are then compared with those from activity-based transportation models, thereby validating the real-world relevance and accuracy of the latter. Hamilton and Sankaranarayanan (*14*) rely on RFID technology to dynamically simulate the location and occupancy of buses as well as passenger demands at bus stops, which leads to a technological framework for dynamic bus operation system. Hood et al. (*15*) define utilities of using different routes in a network based on GPS data and static route information, then establish logit route choice model to understand the decision-making of travellers. Hou et al. (*16*) perform detailed calibration of weather effects with loop detector data and automated surface observing system data. The calibrated models are then fed into the weather-integrated dynamic traffic assignment simulation system to estimate and predict traffic state under inclement weather conditions.

The third category generally concerns with mining data for pattern analysis, identification of cluster characteristics, and prediction based on supervised or unsupervised machine learning. A typical application is the estimation and prediction of road congestion level using GPS data. Thianniwet et al. (*17-18*) employ sliding window technique and J48 decision tree method to characterize and identify road congestion levels based on taxi GPS data and opinion survey. Diker and Nasibov (19) estimate the road congestion level with GPS data based on Fuzzy Neighborhood Density-Based Spatial Clustering of Applications with Noise (FN-DBSCAN). Ma et al. (*20*) predict large-scale traffic congestion with a Restricted Boltzmann Machine and Recurrent Neural Network framework. Other applications of this category could include dynamic detection of traffic accidents, characterization of transport infrastructure utilization, traffic flow prediction, and identification of urban mobility patterns, with a wide range of data sources including social media (*21*), transit card data (*22*), traffic sensor data (*23*), and cell phone (signalling) data (*24*).

Most of the aforementioned studies rely on a single source of data or fixed spatio-temporal granularity for the analysis. There is a lack of multi-dimensional analysis of urban mobility, particularly in microscopic scale, based on multi-source and heterogeneous data, which is accomplished in this paper. This paper illustrates the relationship between public transit, urban planning, demographic characteristics, and emerging transportation modes using a multi-dimensional analysis with big data sources. In particular, we select two metro stations on Line 14 in Beijing for our case study, and utilize the following sources of data:

- Transit smart card data including metro card data (6 million records) and bus card data (4.3 million records);
- Taxi GPS data (1.5 million records), which contain the origin and destination of loaded trips of over 60,000 taxis in Beijing;
- Network data (geographic information, transportation infrastructure);
- Point of Interest (POI) data, including 422,000 records in 15 categories;
- Population and demographic data;
- Online second-hand property price data of 1,2316 communities in Beijing;
- Cell phone signaling data, mainly covering 17 million cell phone users with approximately 70G data per day; and
- Bike sharing data, which contain origin-destination information of 1.3 million bike trips.

Using such a ‘big-data’ approach, we analyze the drastically different metro ridership at the two stations along the same metro line, from the perspectives of transportation planning, built environment, and demographic composition. The results not only reveals the key reasons for the aforementioned different metro ridership, but also highlights some key elements in improving urban mobility. Such a comprehensive and multi-dimensional analysis of urban transportation is rarely seen in the literature, and this paper has demonstrated the necessity of integrating multi-source and heterogeneous data when analyzing complex urban problems. This paper also highlights the importance of collecting and analyzing fine-granularity data. Conventional macroscopic transportation planning data with aggregate information and relatively low update frequency do not sufficiently capture the characteristics and inherent interdependencies of sub-systems in an urban environment. Instead, data on individual activities and behaviours provide much detailed spatio-temporal dynamics of individual commuters, and help to identify key bottlenecks and inefficiencies in the transportation system.

The rest of the paper is organized as follows. Section 2 presents some basic information about the two metro stations of interest, including their passenger flow, geographic information, and built environment. Multiple data sources are utilized in Section 3 for an in-depth and comprehensive analysis of the two locations in terms of transportation planning, accessibility, and mobility. Section 4 summarizes the findings of this paper and extracts further insights for city planning and policy appraisal.

2. COMPARISON OF TWO METRO STATIONS IN BEIJING

This paper focuses on two metro stations in Beijing, namely *Zaoying* and *Jiangtai*, which are located in the Chaoyang District, and are both intermediate, non-transfer stations on Metro Line 14. Both stations were opened for public access at the end of 2014. In this section, we perform a preliminary comparison of the two stations in terms of their time-varying passenger flow and nearby public transit ridership based on transit (metro, bus) smart card data and taxi GPS data.

2.1. Time-varying metro ridership

To quantify the difference between the metro ridership at the two stations on typical working days, we consider the metro card data collected on 24 Sep (Thursday) and 29 Sep (Tuesday) 2015. These days are chosen as they well represent the commute patterns on working days, and are free of interference from inter-city commuting and recreational activities concentrated in the afternoon peak of Friday and morning peak of Monday. The dataset records the inflow and outflow of every metro station in Beijing at a time resolution of 30 min. The time-varying passenger flows at Zaoying and Jiangtai are shown in Figure 1.

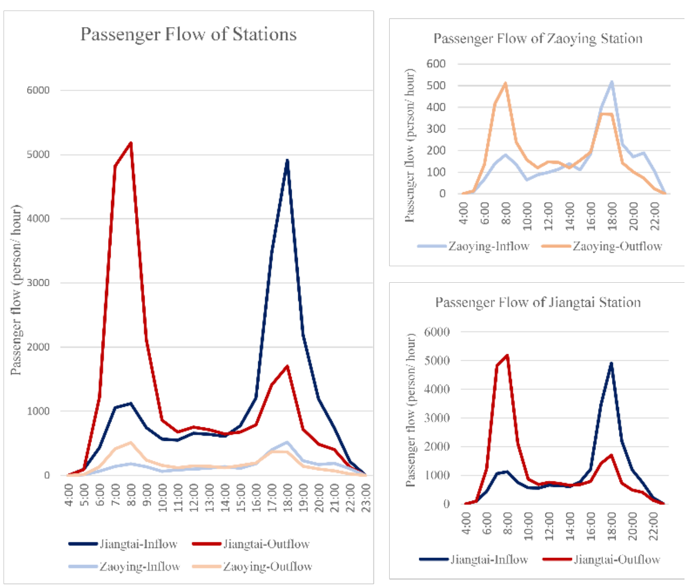


FIGURE 1 Passenger inflows and outflows at the two metro stations.

It can be seen from Figure 1 that the passenger flow at the Jiangtai Station far exceeds that of Zaoying, especially during morning and afternoon peaks. The average daily passenger volume at Zaoying is 6,373 (ranking 72/76 in Chaoyang District and 255/268 in Beijing) while that of Jiangtai is 44,590 (ranking 30/76 in Chaoyang District and 84/268 in Beijing). Overall, the metro ridership in the two stations differs by a factor of 7 (44,590/6,373). Moreover, the passenger flow of Jiangtai during morning peak (7:00-9:00) is 9.76 times the flow of Zaoying;

even during night operation (20:00-22:00) the Jiangtai flow is 5.3 times the Zaoying flow. Such a stark contrast of metro ridership is disproportional to the population in the vicinity of these two stations, which differs only by a factor of 2 (see Section 2.3). This entails further investigation with additional transport, infrastructure, and land use data.

2.2. Utilization of public transportation

Additional to the contrast of metro ridership, a quantitative depiction of the public transit capacity and utilization is considered for an in-depth understanding of the travel pattern in these two areas. Additional bus smart card data and taxi GPS data are used to calculate the modal share between metro, bus and taxi in the 1-km buffer of the two stations. We calculate the average modal share for taxi, bus, and metro as follows.

For every metro station in Beijing, we focus on trips that start or end within the station buffer.[1] The outbound modal share O_i^{mode} and inbound modal share I_i^{mode} of a given station i are defined as follows.

$$O_i^{\text{mode}} = \frac{\sum_{j\neq i} T_{ij}^{\text{mode}}}{\sum_{j\neq i} T_{ij}}\ ,\ I_i^{\text{mode}} = \frac{\sum_{j\neq i} T_{ji}^{\text{mode}}}{\sum_{j\neq i} T_{ji}} \qquad \text{mode} = \text{taxi, bus, metro} \qquad (1)$$

where T_{ij}^{mode} denotes the number of person trips from station i's buffer to station j's buffer by a certain mode (taxi, bus, or metro), $T_{ij} = T_{ij}^{\text{taxi}} + T_{ij}^{\text{bus}} + T_{ij}^{\text{metro}}$. Regarding T_{ij}^{taxi}, the number of taxi trips is converted to the number of person trips with a conversion factory of 1.3 person/taxi (*25*). The outbound and inbound modal shares of Zaoying and Jiangtai are shown in Table 1.

TABLE 1 Outbound and Inbound Modal Shares of the Two Stations

	O_i^{taxi}	O_i^{bus}	O_i^{metro}	I_i^{taxi}	I_i^{bus}	I_i^{metro}
Zaoying	18%	38%	44%	17%	34%	49%
Jiangtai	5%	54%	41%	4%	54%	42%

Table 1 shows much lower share of taxi trips in Jiangtai, indicating more vigorous public transportation (metro, bus) activities than Zaoying. In addition, the city-wide modal share of taxi is 3%, which, compared to the overall 17% taxi share of Zaoying, further highlights the low efficacy of public transportation near Zaoying.

2.3. Geographic, built environment, and population characteristics

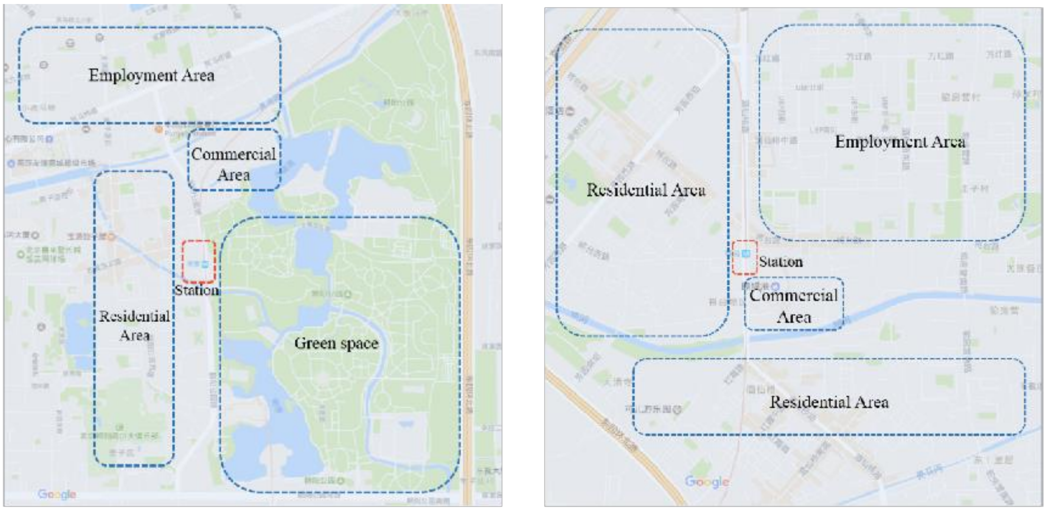
Before investigating the significantly different metro ridership and public transportation usage of Zaoying and Jiangtai areas, we first study their location, built environment, and population.

We begin with the geographic location. Over the past several decades, Beijing has developed in a monocentric spatial pattern, which means that land price, population density, and development intensity are negatively correlated to the distance from the city center (*26*). Both stations of interest are located to the north of the Chang'an Avenue where civil infrastructure, economy, and financial income are superior to the southern part of Beijing (*27*). Moreover,

[1] Here, we ignore those trips whose origin or destination is not within the buffer of any metro station, as we recon these trips to be irrelevant to the metro.

Zaoying station is closer to the city center than Jiangtai station.

In terms of land use, both stations are primarily surrounded by industrial, residential, and business areas (Figure 2). Notably, Zaoying is immediately adjacent to a green space, while Jiangtai is 1 km away from the nearest park. The presence of a major green space near Zaoying might have some negative impact on the mobility around the station due to limited access to transportation infrastructure.



Zaoying station   Jiangtai station

FIGURE 2 Built environment surrounding Zaoying and Jiangtai stations.

Moreover, based on the POI data, both stations are close to regional business centers, with 1,101 POIs and 1,281 POIs within the corresponding buffers. Overall, the numbers of POIs in these two locations are comparable.

Finally, in terms of population size, we consider Traffic Analysis Zones (TAZs) to be basic space units. TAZs are a typical geographic units used in conventional transportation planning models. The TAZ data from Beijing Transport Institute (*25*) are used to infer the residence and employment population within 1-km buffer of stations. The resident and employment population for each TAZ is assumed to be uniformly distributed in the TAZ. Based on this, we use the area of the intersection of the 1-km buffer and each TAZ to calculate the contribution of that TAZ to the population of the buffer. This process is straightforward and further details are omitted in this paper. The estimated population within the 1-km buffer is 28,093 (17,411 residence, 10682 employment) for Zaoying and 58,301 (38,328 residence, 19,973 employment) for Jiangtai. The total population of Jiangtai is 2.08 times that of Zaoying.

As a summary of this section, we observe that

1. The metro ridership at Jiangtai is far greater than that of Zaoying (by a factor of 7);
2. The modal share of public transportation (bus & metro) indicates a stronger public transit activity in the vicinity of Jiangtai station than Zaoying;
3. Both stations are located in the more developed, northern part of the city, and Zaoying is closer to the city center than Jiangtai;
4. The numbers of POIs within 1-km radius from the stations are similar (1,101 for Zaoying and 1,281 for Jiangtai), which means that the two locations have comparable intensity of commercial, recreational, and industrial activities;

5. The population (residence & employment) of the Jiangtai buffer is approximately twice that of the Zaoying buffer.

Several interesting and important questions arise from these observations. The geographical locations of the two stations, composition of population in their neighbouring areas, and the number of POIs nearby, are not at all reflected in the stations' ridership. They also fail to explain the different take up of public transportation in these areas. These issues will be analyzed in greater detail in the next section, using a rich set of transportation, demographic, and land use data.

3. IN-DEPTH ANALYSIS USING BIG-DATA SOURCES

In this section, in order to address the issues raised from Section 2, we investigate in detail and comprehensively the characteristics of the Zaoying and Jiangtai areas pertaining to demographic composition (Section 3.1), connectivity and accessibility of public transit system (Section 3.2), spatial commute patterns (Section 3.3), and emerging transport modes (Section 3.4).

3.1. Demographic information

The demographic characteristics of an area, including age, social status, and employment rate, tend to have a strong effect on the commute pattern of that area.

The Sixth China Population Census data reveal the proportion of senior citizen (defined to be 65 or above) at various regions in the city (Figure 3). It shows that the Maizidian sub-district, where Zaoying is located, has a relatively high senior citizen proportion (32.5%), ranking second among all sub-districts between the 2nd and 5th Ring Roads. In contrast, the neighboring sub-districts of Jiangtai Station, namely Jiuxianqiao and Jiangtai, have senior citizen proportions of only 9.8% and 4.7%, respectively. In addition, the senior population size, indicated by the size of the bubbles in Figure 3, also suggests a much larger senior population in Zaoying area. Such a significantly different senior population may have a direct influence on the travel patterns as senior citizens make less frequent commute trips, and have lower propensity towards the metro compared to the younger, working class. This fact partly contributes to the observation made earlier regarding the usage of the metro stations (see Section 2.1).

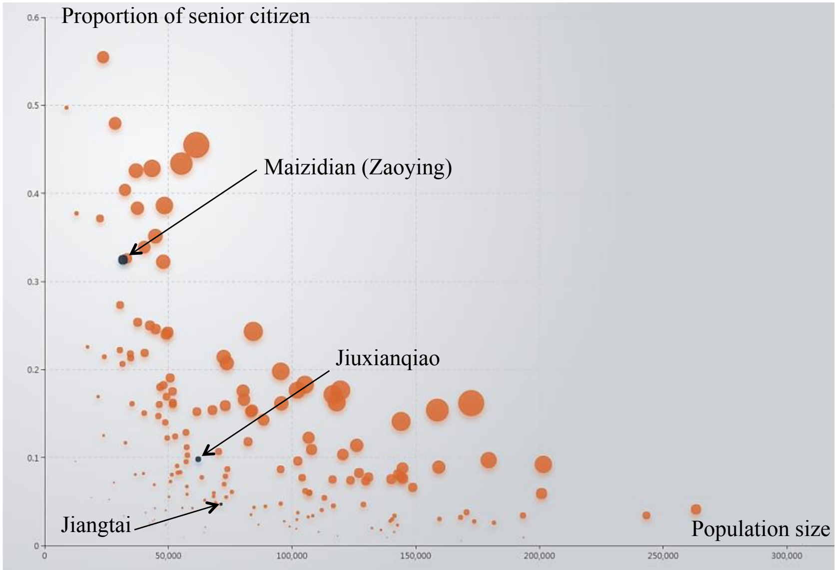


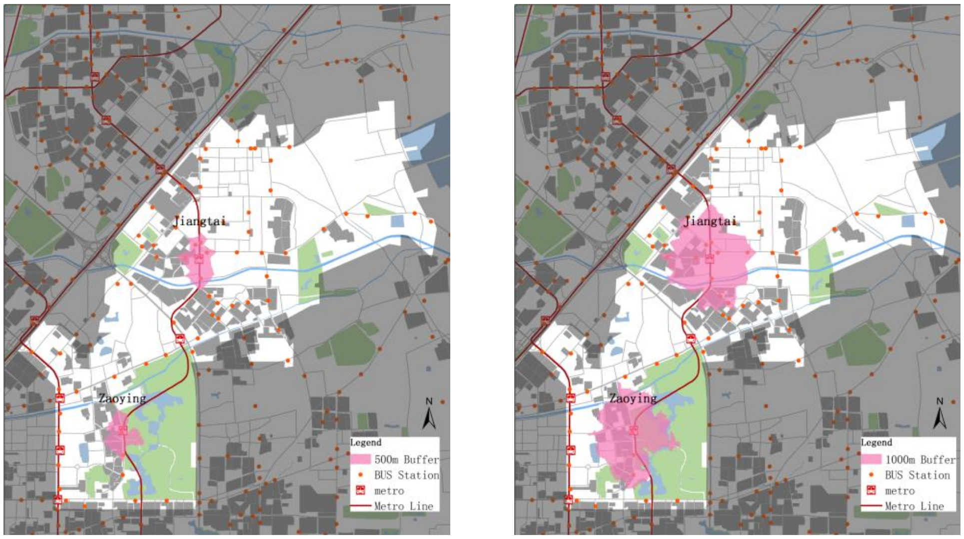
FIGURE 3 Population size and proportion of senior citizens in Beijing sub-districts (bubble size indicates the size of senior population).

The income structure and the car ownership also influence the preference of transport modes. The residents in communities with higher price tend to have higher average income (*28*) and higher car ownership (*29*). Therefore, the online second-hand property pricing data in September 2015 are selected to reflect these two factors. The average price of communities in Maizidian sub-district, where Zaoying is located, is shown to be 17.8% higher than that of Jiuxianqiao and Jiangtai sub-districts. As a result, the higher proportion of high-income residents may be another reason for Zaoying to have lower metro ridership.

3.2. Connectivity with bus transit system

The preliminary analysis on the modal share in the two metro station buffers (see Section 2.2) suggests insufficient public transport activity near Zaoying. In this section, we investigate their connectivity with the bus transit system. This provides further insights on the utilization of the metro, which is an integral part of the metro-bus multi-modal transportation system.

We begin by examining the number of bus stops in the neighborhood of the two metro stations. Figure 4 shows the bus stops within 500 m and 1 km from the metro stations. There are respectively 2 and 4 bus stops within 500 m and 1 km from the Zaoying station, which are far below the average in the Chaoyang District (500 m: 3 bus stops; 1 km: 9 bus stops). The numbers of bus stops around Jiangtai station (500 m: 5 bus stops; 1 km: 11 bus stops) are considerably higher than those of Zaoying and above the average of Chaoyang District.



500 m station buffers 1000 m station buffers

FIGURE 4 Bus stops adjacent to the two metro stations. The pink areas indicate the buffers, which are obtained based on street network distance instead of Euclidean distance.

It is clear that the bus transit network and the metro network are relatively isolated at the Zaoying station. The low accessibility to nearby bus transit network makes it difficult for Zaoying station to fully realize its potential as a metro-bus transfer hub in a multi-modal transportation system.

3.3. Spatial commute patterns

In this section, we analyze the commute patterns around the two metro stations based on signalling data. Signalling data refers to location information registered by mobile phone devices during calls, text messages, and location updates. Such a data enjoys large spatio-temporal coverage and high penetration rate as well as long-term and continuous tracking of locations. Signalling data can be used to extract useful information of individuals such as residential and work locations by mining their daily commute activities (*30-31*). We group the signalling data by International Mobile Subscriber Identity (IMSI). The location with the highest frequency of occurrence at night period (0:00-6:00) is identified as the residential location, and the location with the highest frequency in the working hours (10:00-16:00) as the work location for a single subscriber.

Using TAZ as the basic spatial unit, we compute the Commuting Connectivity (CC) between a TAZ and the metro station buffer. The CC is defined to be the sum of the following quantities:

- The number of commuters who work in the TAZ and reside in the metro station buffer; and
- The number of commuters who reside in the TAZ and work in the metro station buffer.

The signalling data suggest that the population working or residing within the Jiangtai station buffer is approximately twice that of Zaoying station buffer. This is consistent with the (static) population result presented in Section 2.3. Figure 5 indicates the CC between each TAZ and the station buffers. For the Zaoying area, TAZs with high CC are located in nearby regions such as Sanyuanqiao, Liangmaqiao and Tuanjiehu (indicated by darker colors). For the Jiangtai area, TAZs with high CC are distributed more outwards from the station.

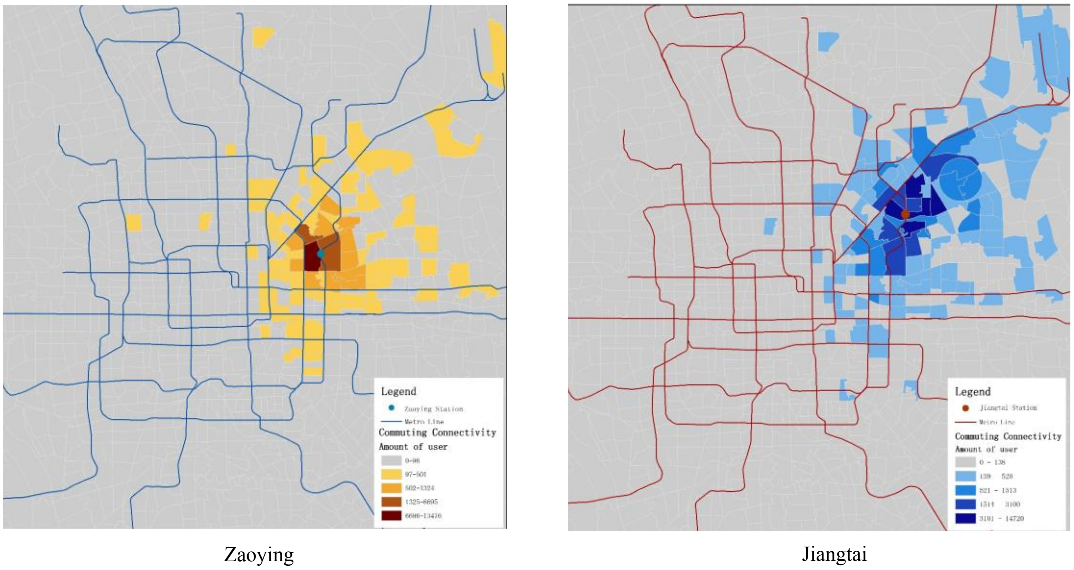


FIGURE 5 TAZ-based Commuting Connectivity with the metro station areas.

Besides the TAZ-based CC, we further investigate the direction-based CC. To do this, we first categorize all the relevant TAZs into North, South, East, and West according to the line segment connecting the center of the TAZ and the metro station (Figure 6). Then, the direction-based CC is defined to be the sum of CCs of relevant TAZs in the given direction (Table 2).

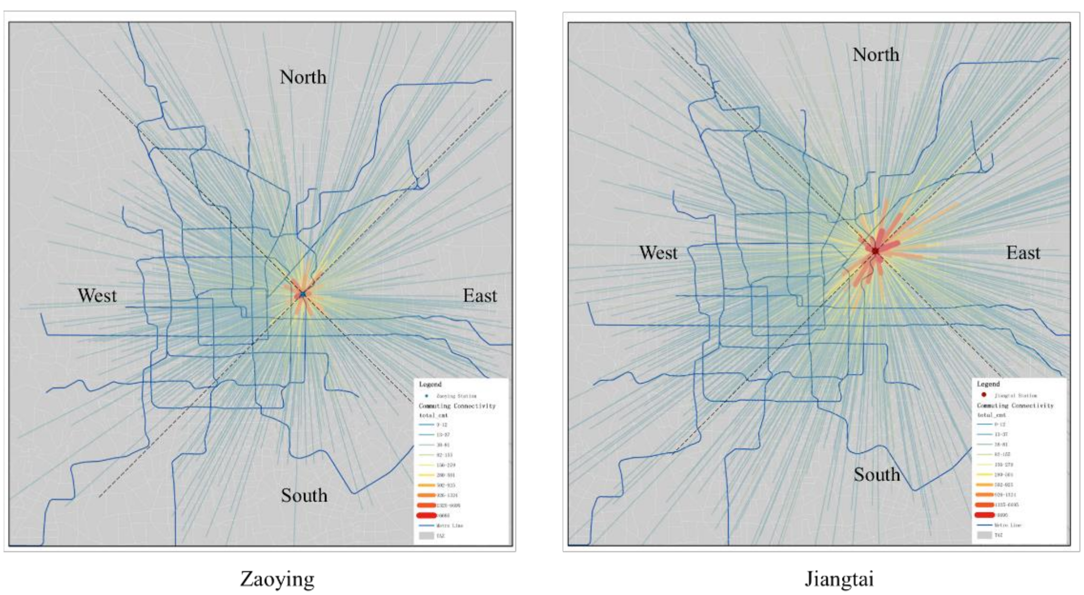


FIGURE 6 TAZ-based CC indicated by the width and color of the line segment connecting the station and the TAZ.

TABLE 2 Direction-based CC for Zaoying and Jiangtai.

	Zaoying		Jiangtai	
	Number of TAZs	Sum of TAZ-based CC	Number of TAZs	Sum of TAZ-based CC
South	485	12,381	574	39,916
West	700	27,883	735	36,824
North	505	16,355	413	32,021
East	316	16,742	284	34,859
North + South	**990**	**28,736**	**987**	**71,937**
East + West	**1,016**	**44,625**	**1,019**	**71,683**

Figure 5, Figure 6 and Table 2 reveal distinct spatial commute patterns around Zaoying and Jiangtai areas. High-CC TAZs ($\geq 1{,}000$) associated with Zaoying are within close proximity (1 km) to the station, while those associated with Jiangtai are distributed much further away (3 km). Additionally, Table 2 suggests that the primary commuting direction associated with Zaoying is East-West, while for Jiangtai the commuting directions of East-West and North-South are comparable. The Metro Line 14, where Zaoying and Jiangtai stations are both located, is North-South bound; this suggests a misalignment between the metro line and the main commuting direction of Zaoying. This highlights the low accessibility of Zaoying to/from its neighbouring areas. In particular, areas like Sanyuanqiao, Liangmaqiao, and Tuanjiehu are close to Zaoying with approximately 10 min by car or taxi, but they take 30 min to reach by metro. Such a low metro accessibility to places with high commuting demand makes metro a generally expensive choice for travelers to/from Zaoying. This also explains the significant dependence on taxi around Zaoying area (see Table 1).

3.4. Bike sharing to increase local mobility

The commuting distance between the Zaoying station and nearby TAZs with high CC is typically below 1 km (see Figure 5). Such short-distance commute may be supplemented by shared bike scheme, which is booming in most Chinese cities (*32*). Shared bikes, due to their low economic and environmental costs as well as easy access for short-distance commute, is widely used as the first-mile or last-mile travel mode in a multi-modal urban transportation system. The shared bike companies have started to enter the market since Sep 2016 (notice that the metro ridership data were collected in Sep 2015).

We use the origin-destination (OD) information of shared-bike trips to investigate the effect of bike sharing on the local mobility around Zaoying station, and the subsequent impact on the metro ridership at the Zaoying station. In particular, we focus on bike trips that start (or end) within 200 m-buffer of Zaoying, and plot the heat map of their corresponding destinations (or origins) (Figure 7). In particular, trips to/from Sanyuanqiao and Liangmaqian areas, which carry high local travel demands (see Figure 6), comprise 22% of the total bike trips. As a result of the implementation of the bike sharing scheme since Sep 2016, the ridership in Zaoying station has increased by 80 % based on metro data collected in Nov 2016.

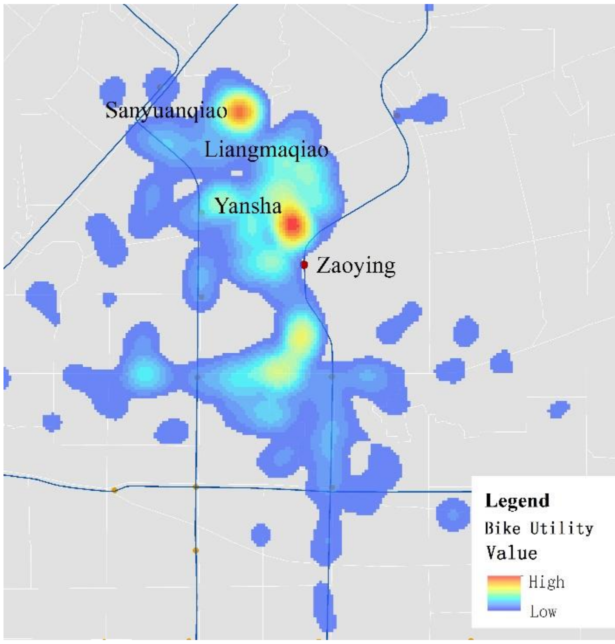


FIGURE 7 The bike use intensity around Zaoying station. The color indicate the number of bike trips that start or end within 200 m from the station.

4. CONCLUSION AND DISCUSSION

This paper presents a comprehensive and multi-dimensional analysis of the public transit system at two locations in Beijing. We investigate the public transit capabilities near Zaoying and Jiangtai metro stations, both located on the Metro Line 14. The former has much lower metro ridership than the latter (by a multiplicative factor of 7). In addition, a much higher dependency on taxi as a private transportation mode is also seen in Zaoying area (overall 17% for Zaoying vs. 5% for Jiangtai). These statistics suggest that the public transportation in Jiangtai area is much more vigorous than Zaoying area; this is disproportional to their local population and in spite of their comparable geographic location and commercial activities. Through a data-intensive and multi-dimensional analysis, we analyze factors that influence the local mobility of these areas and reasons behind the much lower metro ridership in Zaoying. In particular, the following findings are made.

- The Zaoying area is predominantly resided by senior population, who make less frequent commute trips, and have lower propensity towards the metro compared to the younger, working class.
- The higher property price around Zaoying station suggest a potential higher average income and car ownership of nearby residents, which contribute to lower metro ridership.
- Zaoying station has much lower accessibility from the bus transit network than the Jiangtai station. It is therefore difficult for Zaoying to be fully utilized as part of the metro-bus multi-modal transit system.
- Commute patterns to/from the two areas, which are extracted from signalling data, shows that most commuters to (from) Zaoying travel from (to) areas within close proximity (1 km), while those associated with Jiangtai are distributed much further away (3 km). Moreover, the primary commuting direction associated with Zaoying is East-West, which is misaligned with Metro Line 14 (North-South).

- The low accessibility of the Zaoying station from nearby areas such as Sanyuanqiao and Liangmadian is mitigated by the introduction of shared bike services.OD information of shared-bike trip data show that the most frequent bike-sharing trips are to/from nearby locations with high commuting demands. Furthermore, the emergence of shared bikes has mitigated the low connectivity between the Zaoying station and other transit systems. As a result, the metro ridership at Zaoying has significantly increased (by 80%) after introduction of the shared bike scheme.

The aforementioned big-data analyses reveal some key insights regarding urban planning. Conventional transportation planning approaches tend to focus on aggregated characteristics like travel time, fare, and comfort. However, the drastic difference in the metro ridership at Zaoying and Jiangtai stations suggests vital roles played by the local accessibility of the metro station as well as its connectivity with the rest of the transit system.

This paper also highlights the importance of collecting and analyzing fine-granularity data, beyond conventional aggregate, static transportation planning data. Data on individual activities and behaviors, such as cell phone signaling data and bike sharing data, provide much detailed spatio-temporal dynamics of individual commuters, which help transportation planners to identify key bottlenecks and propose effective solutions.

REFERENCE

1. Wang, J., C. Li, , Z. Xiong, and Z. Shan. Survey of data-centric smart city. *Journal of Computer Research and Development*, Vol. 51, No. 2, 2014, pp. 237-259.
2. Lee, K., W. S. Jung, , J. S. Park, and M. Y. Choi. Statistical analysis of the Metropolitan Seoul Subway System: Network structure and passenger flows. *Physica A: Statistical Mechanics and its Applications*, Vol. 387, No. 24, 2008, pp. 6231-6234.
3. Soh, H., S.Lim, T.Zhang, X. Fu , G. K. K. Lee, T. G. G. Hung, P.Di, S. Prakasam, and L.Wong. Weighted complex network analysis of travel routes on the Singapore public transportation system. *Physica A: Statistical Mechanics and its Applications*, Vol. 389, No. 24, 2010, pp. 5852-5863.
4. Zhong, C., S. M. Arisona, X. Huang, M. Batty, and G. Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, Vol. 28, No. 11, 2014, pp. 2178-2199.
5. Wang, D., C.Wang, D.Xie, W. Zhong, M. Wu, W. Zhu, J. Zhou, and Y. Li. Comparison of retail trade areas of retail centers with different hierarchical levels: A case study of east nanjing road, wujioachang, anshan road in Shanghai. In *Urban Planning Forum*, Vol. 3, 2015, pp. 50-60.
6. Hollenstein, L., and R. Purves. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, No. 1, 2010, pp. 21-48.
7. Jang, W. Travel time and transfer analysis using transit smart card data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2144, 2010, pp. 142-149.
8. Nishiuchi, H., J. King, and T. Todoroki. Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data. *International Journal of Intelligent Transportation Systems Research*, Vol. 11, No. 1, 2013, pp. 1-10.
9. Shen, Y., and Y. Chai. Study on commuting flexibility of residents based on GPS data: A case study of suburban mega-communities in Beijing. *Acta Geographica Sinica*, Vol. 67, No. 6, 2012, pp. 733-744.
10. Páez, A., and D. M. Scott. Spatial statistics for urban analysis: a review of techniques with examples. *GeoJournal*, Vol. 61, No. 1, 2005, pp. 53-67.
11. Zhen, F., B. Wang, and Y. Chen. China's city network characteristics based on social network space: An empirical analysis of Sina Micro-blog. *Acta Geographica Sinica*, Vol. 67, No. 8, 2012, pp. 1031-1043.
12. Gao, S. Spatio-temporal analytics for exploring human mobility patterns and urban dynamics in the mobile age. Spatial Cognition & Computation, Vol. 15, No. 2, pp. 86-114.
13. Liu, F., D. Janssens, J. Cui, Y. Wang, G. Wets, and M. Cools. Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Systems with Applications*, Vol. 41, No. 14, 2014, pp. 6174-6189.
14. Hamilton, P., and S. Sankaranarayanan. Intelligent agent based RFID system for on demand bus scheduling and ticketing. *International Journal of Future Computer and Communication*, Vol.2, No. 5, 2013, pp. 399.
15. Hood, J., E. Sall, and B. Charlton. A GPS-based bicycle route choice model for San Francisco, California. *Transportation letters*, Vol. 3, No. 1, 2011, pp. 63-75.

16. Hou, T., H. Mahmassani, R. Alfelor, J. Kim, and M. Saberi. Calibration of traffic flow models under adverse weather and application in mesoscopic network simulation. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2391, 2013, pp. 92-104.
17. Thianniwet, T., S. Phosaard, and W. Pattara-Atikom. Classification of road traffic congestion levels from GPS data using a decision tree algorithm and sliding windows. In *Proceedings of the world congress on engineering*, Vol. 1, 2009, pp. 1-3.
18. Thianniwet, T., S. Phosaard, and W. Pattara-Atikom.The optimization of an intelligent traffic congestion level classification from motorists' judgments on vehicle's moving patterns. *WorldAcademy of Science, Engineering and Technology, International Science Index*, Vo. 53, No. 5, 2001, pp. 5.
19. Diker, A. C., and E. Nasibov. Estimation of traffic congestion level via fn-dbscan algorithm by using gps data. In *Problems of Cybernetics and Informatics (PCI)*, 2012 IV International Conference. IEEE. 2012, pp. 1-4.
20. Ma, X., H. Yu, Y. Wang, Y. Wang. Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS one*, Vol. 10, No. 3, 2015, e0119044.
21. Gu, Y., Z. Qian, and F. Chen. From Twitter to detector: Real-time traffic incident detection using social media data. *Transportation research part C: emerging technologies*, Vol. 67, 2016, pp. 321-342.
22. Ma, X., Y. Wu, Y. Wang, F. Chen, and J. Liu. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies*, Vol. 36, 2013, pp. 1-12.
23. Wibisono, A., W. Jatmiko, H. A. Wisesa, B. Hardjono, and P. Mursanto. Traffic big data prediction and visualization using fast incremental model trees-drift detection (fimt-dd). *Knowledge-Based Systems*, Vol. 93, 2016, pp. 33-46.
24. Widhalm, P., Y. Yang, M. Ulm, S. Athavale, and M. C. González. Discovering urban activity patterns in cell phone data. *Transportation*, Vol. 42, No. 4, 2015, pp. 597-623.
25. Beijing Transport Institute. *Report on the 5th Beijing Comprehensive Transportation Survey*, Beijing, 2015.
26. Lu, M., Z. Chen, , X. Zhu, and X. Xu. *China's regional development: review and prospect*, Routledge, 2013.
27. Yang, X. *Action of South Beijing: Rebuild A New South Beijing*. Beijing Daily, 2012.
28. Zhang, W. An Analysis of the Factors that Influence the Urban Residential Location Selection. Progress in Geography, Vol. 20, No.2, 2001, pp. 268-275.
29. Zhang, Z., W. Jin, H. Jiang, Q. Xie, W. Shen, and W. Han. Modeling heterogeneous vehicle ownership in china: a case study based on the chinese national survey.*Transport Policy*,Vol. 54, 2017, pp. 11-20.
30. Ran, B. Use of cellphone data in travel survey and transportation planning. *Urban Transport of China*, Vol. 1, 2013, pp. 72-81.
31. Zhou, X., X. Chen, and T. Zhang. Impact of megacity jobs-housing spatial mismatch on commuting behaviors: A case study on central districts of Shanghai, China. *Sustainability*, Vol. 8, No. 2, 2016, pp. 122.

32. Beijing Mobike Technology Corporation, and Beijing Tsinghua Tongheng Urban Planning & Design Institute. *The white paper of urban development with bike sharing*. Beijing, 2017.