ZUOXIAN GAN, Ph.D.[1]
(Corresponding author)
E-mail: zxgan@dlmu.edu.cn
JING LIANG, Ph.D.[1]
E-mail: liangjing@dlmu.edu.cn
[1] College of Transportation Engineering
Dalian Maritime University
Dalian 116026, China

Traffic Planning
Original Scientific Paper
Submitted: 7 May 2020
Accepted: 25 Aug. 2020

UNDERSTANDING HUMAN MOBILITY WITHIN METRO NETWORKS – FLOW DISTRIBUTION AND COMMUNITY DETECTION

ABSTRACT

In this paper, smart card data collected from the Nanjing Metro over 2-hour time periods are used to characterize within- and between-day human mobility patterns within the metro network. Results show that the OD (origin to destination) flows can be characterized well by shifted power law distributions with similar exponents around 2, which reflects the fact that a few OD pairs in the system play a dominant role and undertake disproportionately large OD flow distribution. The different exponents signify heterogeneous human movement in within- and between-day ranges. In addition, we analyze the metro community structures over different time periods based on the community detection method using random walks to visualize and understand passenger movement from a spatial perspective. Normalized mutual information is used to compare community partitions over different time-intervals. The results show that the properties of human mobility during different time periods have a similar rhythm, although some nuances exist, and the community structure is usually divided according to the line distribution. This empirical study provides spatiotemporal insights into understanding urban human mobility and some potential applications for transportation management.

KEYWORDS

metro system; OD flows; shifted power law; community partition.

1. INTRODUCTION

With fast-growing urbanization, urban transportation systems become very important in our society. They facilitate daily mobility of people as they enable continuous movement of many passengers. Meanwhile, they ensure the sustainable development of cities as they save energy and reduce carbon emissions [1]. Uncovering the dynamics of human mobility and urban interaction helps many applications, ranging from urban planning to epidemiology [2-4]. Since automatic fare collection (AFC) systems are widely used in urban transportation systems all over the world, it becomes possible to collect real-time transaction information from AFC systems to accurately capture daily movement of people in cities [2, 5-11]. Compared to traditional transport data sources, such as questionnaires and travel diaries, smart card data (SCD) are more extensive and accurate since all transactions are recorded in the system, timestamped and geotagged [10]. These properties ensure SCD a reliable and promising source for the analysis of human mobility and urban interaction at a large scale with low cost.

Benefits from the spatiotemporally explicit SCD records and other types of big-data sources (e.g. GPS devices, taxi trajectories, mobile phone data and social media check-in data) have triggered an explosion in the study of human mobility from a statistical physics perspective during the last decade. Recently, the scaling law of human mobility has drawn much attention from both physics and transportation research [9]. As an important and high-capacity traffic mode in modern cities, public transit systems such as metro and bus networks have been an active research object to explore and analyze urban human mobility. Studies of traffic flow (or trip displacement) distributions of public transit systems underlying the topological context have proved that analysis using statistical physics approach is a useful method for understanding human mobility, in which human travel behavior follows specific mathematical forms such as power law, exponential, gamma and other distributions [2, 9, 12]. A Total of 11.22 million trips from London metro system revealed that intra-urban movement is strongly heterogeneous in terms of volume but not displacement. The results showed that a limited number of central stations involves a large amount of flows and the trips between two stations can be fitted by a power law [2]. Jiang et al. [12] noticed that trip displacement of bus and metro follows the exponential and gamma distribution separately, while the fusion trip displacement follows the power law with an exponential cutoff. Xu et al. [9] studied passenger flows in Beijing Metro System and found that the distributions of departure flow (outflow), total inflow and outflow of station, exogenous flow (flow from outside to the station or flow from the station to outside) of station, and throughflow of station are all heavy-tailed and illustrate significant curvature on log-log scales. These distributions can be characterized by power law functions.

Meanwhile, mobility pattern detection, the so-called clustering, also attracted much attention because it helps urban planners and public transit operators to better understand passenger demand, which is useful to redesign and improve current transportation policies. Based on daily fluctuations in passenger flows, urban transportation stations can be clustered into several categories. Stations in one cluster may have high in-flow morning peak and out-flow afternoon peak volumes, while other stations in another cluster may exhibit the opposite patterns – “high out-flow morning peak and in-flow afternoon peak volumes” [7, 10]. This research has provided valuable insights into human mobility from a spatiotemporal perspective. However, it lacks the identification of the strength of the relationship between stations, which reflects the spatial interaction over time.

In recent years, a common feature of many networks (e.g. social networks, biochemical networks, and internet networks), called “community structure”, has been found [13-14]. Its salient property refers to the division of network nodes into different groups, with dense connections within groups but sparser connection between them [13]. The community detection is able to help us to analyze and visualize the structure of traffic network. Furthermore, the community partitions based on dynamic travel flows will be useful for better understanding human mobility over within- and between-day in a specific city, which may provide suggestions to urban-planners and transport agencies for urban planning and transport management. Unfortunately, few existing studies have applied the community detection approach to investigate urban travel mobility so as to improve transport management. Drawing upon the smart card data from public transit system in September 2010, April 2011, and September 2012, Zhong et al. [4] used spatial network analysis to detect the dynamics of urban structure in Singapore and indicated that the most important communities remain the same although there are some significant changes in flows between communities. Based on the GPS trajectories of more than 6600 taxis in Shanghai, Liu et al. [15] applied a community detection method to reveal sub-regional structures and examine the properties of sub-regions.

The findings from prior studies provide valuable insights in understanding urban human mobility. However, to the best of our knowledge, we are unable to find works that explore people movement within a metro system through combining the technologies of scaling law and commuting detection. Moreover, most previous studies focused on the properties of the node (station) including inflow, outflow, and visitation frequency, while a small body of prior studies paid attention to the properties of linkage between stations such as trip displacement and OD (origin to destination) flows. However, the OD flows are more conductive to understand human mobility characteristics from a space-time perspective [16]. Meanwhile, few existing studies are attentive to whether the transport properties are different during different time periods (e.g. morning, noon, afternoon, etc.) and whether the differences are significant or not. Hence, this study takes this stream of exploring human mobility patterns one step further by examining the properties of OD flows within days and between days.

Against this background, this study aims to enrich the existing literature on human mobility within a city by considering within- and between-day OD flows based on AFC data pertaining to the metro system of Nanjing, China. Since flows distribution and flows spatial-organization are two important aspects in structure features of passenger movement and together provide a basic understanding of human mobility characteristics, the research is organized around two objectives. First, we obtain the spatiotemporal OD flow size distribution pattern on weekdays (Monday to Thursday, Friday) and weekends (Saturday and Sunday). This part should provide evidence about the global statistical characteristics of OD flow distributions across time.

Second, community detection is applied to discover the community structure of the transportation network. This part aims to explore the spatial interaction during different time-intervals and provide the analysis of heterogeneity among stations of a specific transit network.

The remainder of this paper is organized as follows. Section 2 introduces the empirical dataset of Nanjing Metro. Section 3 presents and analyzes the statistical properties of OD flows patterns on weekdays (Monday to Thursday, Friday) and weekends (Saturday and Sunday). Section 4 explores and visualizes community structures of the metro networks across time. Finally, Section 5 provides conclusions of the work.

2. DATASETS

Our study uses data from more than 20 million SCD records from the metro system of Nanjing, China. Nanjing is the capital of Jiangsu Province and one of the central cities in eastern China. As of April 2015, the Nanjing Metro System has 6 lines and 112 stations (*Figure 1*). It carries 717 million passengers annually and its share in 2015 was about 34.8% of the passenger volume of all types of public transit, being a major transportation mode in this modern city.

The SCD records, provided by the Nanjing Metro Corporation, were collected by the AFC system of Nanjing Metro from 13 April to 26 April 2015. Each raw journey record includes smart-card ID, tap-in time, tap-out time, line number of boarding, line number of alighting, station ID of boarding, station ID of alighting, and trip duration. For the purpose of analysis in the study, we divided the dataset into 2-hour time periods to analyze human mobility in different time periods of the day in terms of the departure time. Then, a passenger OD flow matrix is constructed to calculate the aggregated movement in the metro network. Every element of this matrix represents the number of passengers travelling from an origin metro station to a destination station over the given time period.

3. OD FLOW DISTRIBUTION ANALYSIS

3.1 Shifted power law function

One goal of this study is to figure out the global characteristics of the urban OD flow distributions over different time-intervals. It helps us to analyze whether human mobility differs within days (e.g. peak hours, mid-day, and evening) and between days (Monday to Thursday, Friday, Saturday and Sunday). For example, the complementary cumulative distribution function (CCDF, $F(x)=P(X>x)$) of

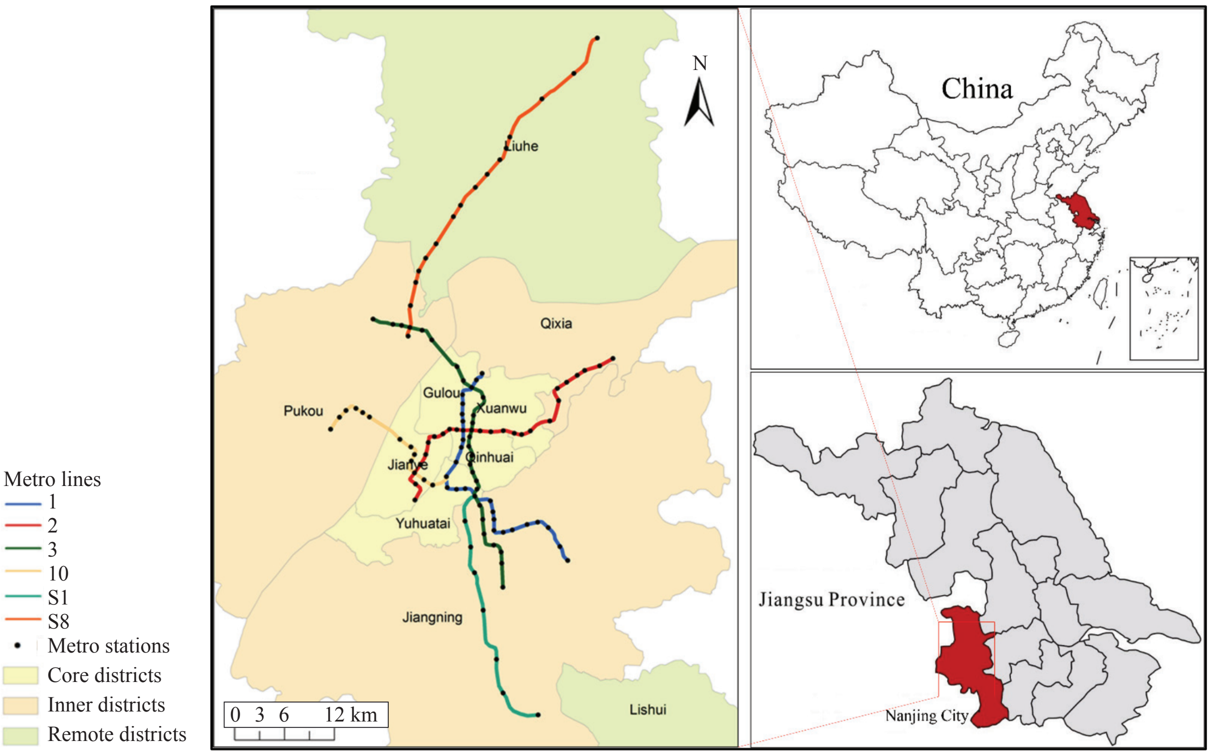


Figure 1 – Location of the case study area

the number of trips at different time spans of Monday to Thursday, $P(N \geq n)$, is plotted in *Figure 2*. The corresponding CCDF of the number of trips at other time intervals is not shown in the paper due to length limitations. The plot on the lin-lin scale in *Figure 2* shows that the CCDF of the number of trips sharply declines at first and then smooths out over a wide range. Meanwhile, the plot of the CCDF on log-log scales approximates a line, which meets the power law distribution. Normally, A quantity x obeying a power law can be expressed as:

$$P(x) = Ax^{-\alpha} \tag{1}$$

where A is a constant parameter, α is the scaling parameter, also known as the exponent.

Considering that few empirical phenomena obey power laws for all values of x, more often the power law applies only for values greater than the threshold x_{min} [17]. Thus, only the tails of the distributions are fitted by power laws and a large number of observed values are thrown away [17, 18]. Here, we use the "Mandelbrot law", also called "shifted power law" (SPL), to fit the within- or between-day OD pairs distributions instead of the basic power law [19]. This method can well fit the distributions of samples without eliminating many small observed values. It is expressed as

$$P(x) = A(x+\beta)^{-\alpha} \tag{2}$$

where β is a constant, the so-called shifting coefficient. For $\beta << x$, *Equation 2* can approximate a normal power law (*Equation 1*). While if $\beta >> x$, it approximates:

$$P(x) = A \cdot \exp(x+\beta)^{-\alpha} \tag{3}$$

indicating an exponential distribution. When β changes from 0 to ∞, the distribution varies from a power law to an exponential distribution. The typical SPL functions can be shown with β lower than 100 because the SPL function shows a rather good linear line on a semi-log scales indicating an

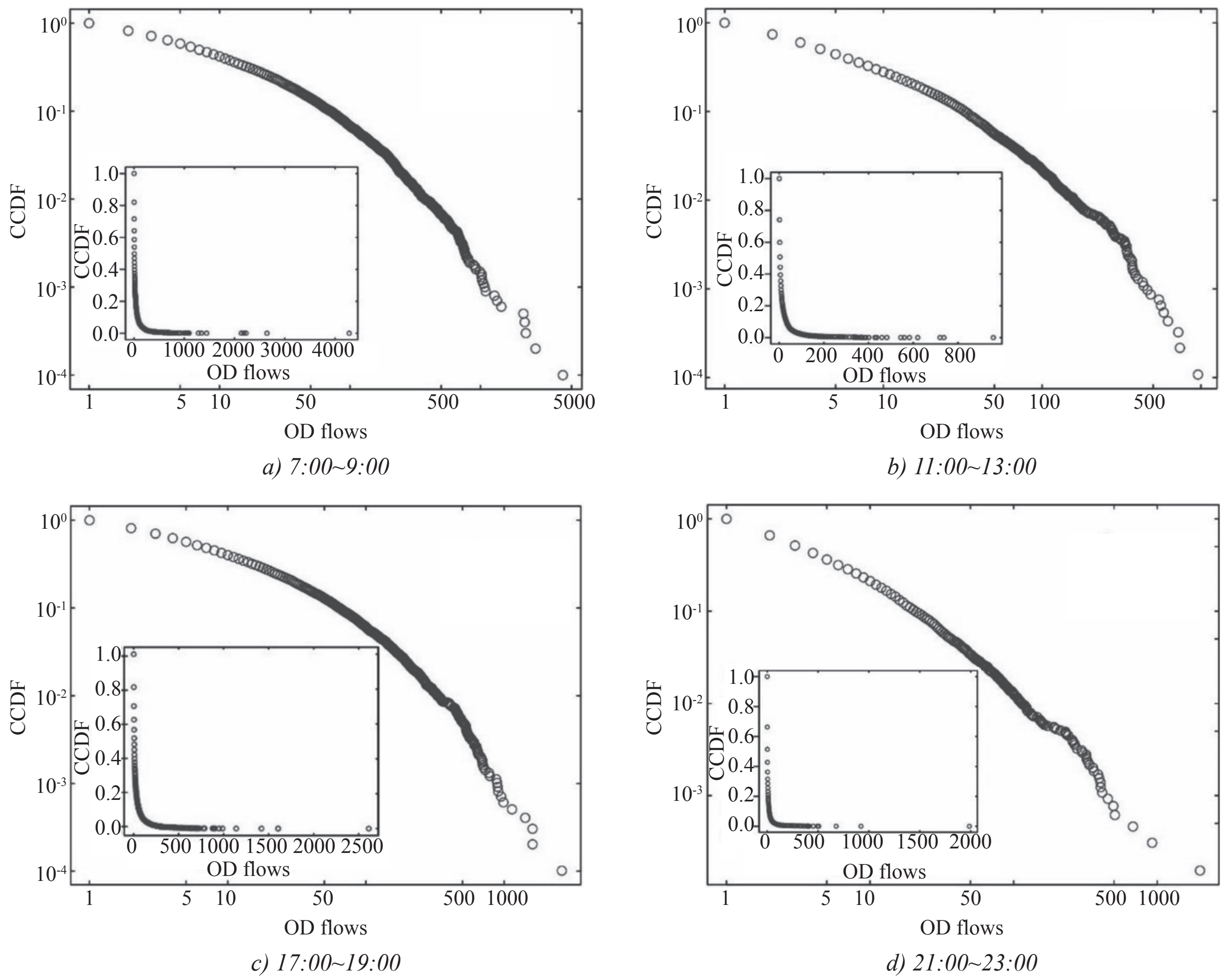


Figure 2 – Distribution of OD flows from Monday to Thursday on a lin-lin scale
Insets: the CCDF is plotted on log-log scale

approximately exponential distribution when β is larger than 100. The exponent α and shifting coefficient β can be fitted by nonlinear iteration [19].

After taking the logarithm, *Equation 2* can be expressed as:

$$\ln P(x) = \ln A - \alpha \ln(x+\beta) \tag{4}$$

3.2 Temporal difference of the OD flow distributions

The SPL function can be plotted as a linear line with slope α on the log-log scale. Temporal differences of OD flow distributions are examined across eight 2-hour time periods. *Figure 3* represents four CCDF of OD flow distributions from Monday to Thursday. The values of the key coefficients α and β of all time-intervals are presented in *Table 1*.

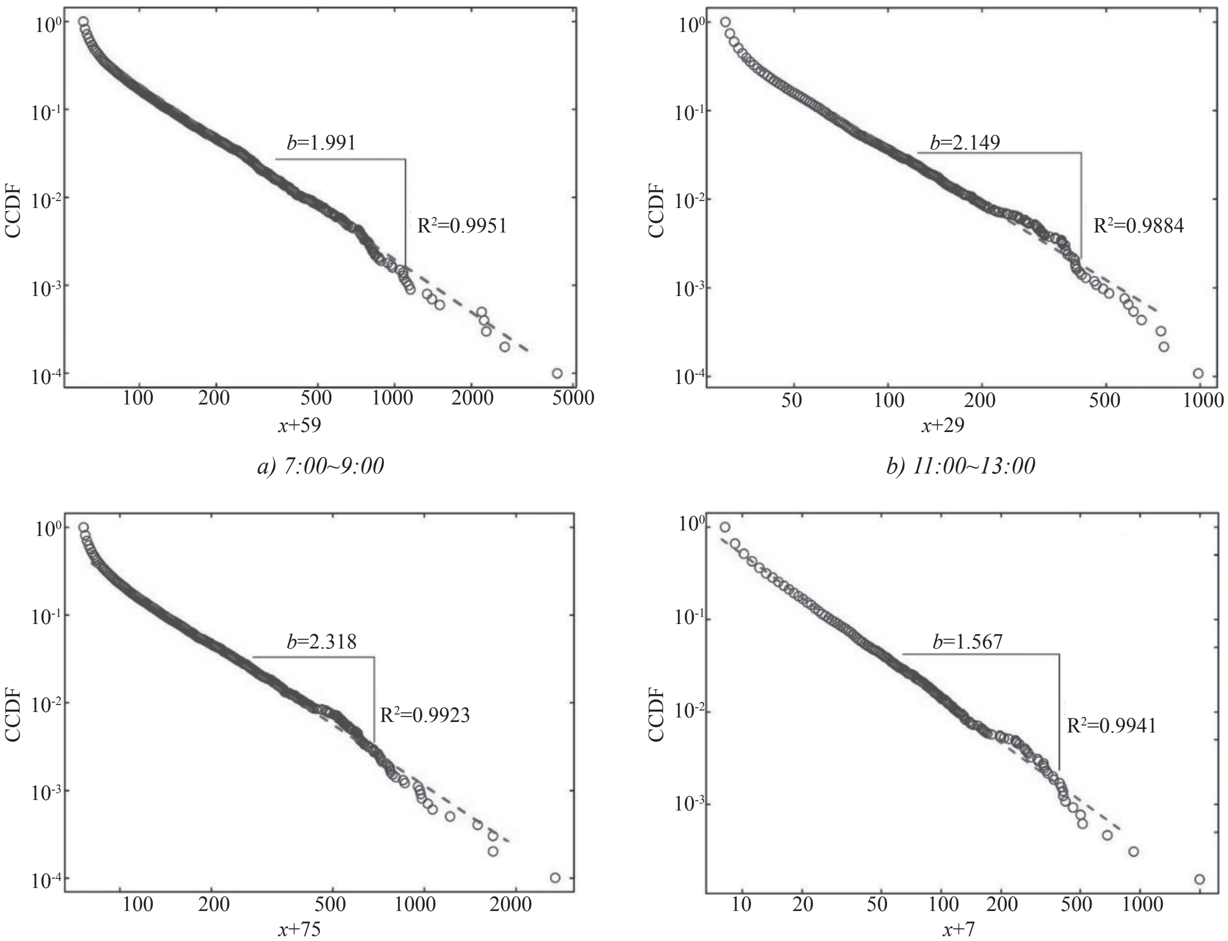
In *Figure 3*, the OD flow distributions can be approximately fitted by SPL using the exponents 1.991, 2.149, 2.318, and 1.567, respectively. On the one hand, the shifted power law distributions indicate significant heterogeneity in human movement in the metro network over different time-intervals, revealing that human movements in Nanjing metro represent heterogeneous flows. Since it has been confirmed that the heterogeneous flow organization exists in metro systems over a day, it also provides further evidence that the heterogeneity also exists over different time periods of a day [2, 9]. On the other hand, the exponential decay rate α explains the decay difference. The larger the α, the greater the rate of decay, which means the OD flow distributions in the network are more balanced and the differences of OD flows are smaller. *Figure 3* and *Table 1* show that the α for the morning peak (7:00~9:00) and the evening (after 19:00) is lower than that of other time-intervals on weekdays. However, this is not applicable for weekends, and α for the morning peak is not lower than that of mid-day.

Refocusing on the OD flow distributions during the morning peak in *Figure 3*, the exponents on 21:00~23:00 are slightly lower than those on other time spans, which demonstrates that the OD flow distributions in the evening may be more unbalanced and a few ODs may have more trips than those in other time spans. The results from *Table 1* further indicate that the OD flow distribution in the evening (19:00~23:00) is the most unbalanced, regardless of the day of the week, while the most balanced time periods for different days of the week are not entirely the same.

In addition, to confirm the relationships between the OD flow distributions and exponent α, we numerically calculated the proportions of the OD flows at 2-hour time periods. *Table 2* shows some of the results (only the results from Monday to Thursday are described here because of space limitation, and the results of other time spans are available upon request). The average number of trips (μ) and the corresponding standard deviations (σ) are calculated first, and then the proportions of the number of ODs (PoN) and the proportions of the OD trips (PoT). The value of μ is all trips (OD flows) divided by the number of OD pairs involving passengers with respect to a specific time interval (OD pairs without passenger flows are not included). We only focus on those trips between metro stations that exceed the value of the average number of trips (μ) plus their standard deviations (σ). Then, the OD trips that exceed threshold values, $\mu+\sigma$ and $\mu+2\sigma$, are calculated. According to the results in

Table 1 – The exponents α and shifting coefficients β of OD flow distributions

Time of day	Monday to Thursday		Friday		Saturday		Sunday	
	α	β	α	β	α	β	α	β
7:00~9:00	1.991	59.210	2.018	59.709	2.186	39.176	2.095	24.111
9:00~11:00	2.306	36.472	2.264	35.947	2.194	40.004	2.131	30.729
11:00~13:00	2.149	29.108	2.255	37.866	2.065	37.952	2.107	37.397
13:00~15:00	2.128	31.532	2.141	36.557	1.944	34.651	1.979	36.483
15:00~17:00	2.304	41.319	2.399	58.005	1.954	37.193	2.001	39.856
17:00~19:00	2.218	74.99	2.361	91.047	2.007	43.476	2.062	41.278
19:00~21:00	1.991	23.732	2.003	34.308	1.859	24.643	1.906	24.571
21:00~23:00	1.567	7.182	1.628	13.767	1.680	15.213	1.632	9.296



a) 7:00~9:00 *b) 11:00~13:00*

c) 17:00~19:00 *d) 21:00~23:00*

Figure 3 – Comparison of distribution of OD flows from Monday to Thursday. The red lines denote the fitting of the data with exponents of 1.991, 2.149, 2.318, and 1.567, respectively. All R^2 values are higher than 0.98

Tables 1 and 2, it is hard to find a simple positive or negative relation between the exponent α and the mean value μ and standard deviations σ. Nevertheless, we find that with a slightly lower exponent α, there are fewer ODs (PoN) whose trips are more than $\mu+\sigma$ (or $\mu+2\sigma$) but would account for relatively more OD trips (PoT). Taking the time periods of 7:00~9:00 and 9:00~11:00 on Monday to Thursday as an example for within-day OD flow distributions analysis, we observed that the top 5% (2.24%) of ODs account for 50.63% (35.1%) of all trips at 7:00~9:00 (α=1.991), while 5.6% (2.74%) of ODs account for only 48.12% (34.61%) of all trips at 9:00~11:00 (α=2.306).

Table 2 – Proportions of the OD flows during different time-intervals on Monday to Thursday

Time of day	Mean (μ)	S.D. (σ)	$\geq\mu+\sigma$		$\geq\mu+2\sigma$	
			PoN	PoT	PoN	PoT
7:00~9:00	31.102	99.771	5%	50.63%	2.24%	35.1%
9:00~11:00	15.155	40.455	5.6%	48.12%	2.74%	34.61%
11:00~13:00	13.855	36.976	5.45%	48.28%	2.72%	35.39%
13:00~15:00	15.434	42.047	5.48%	48.54%	2.46%	33.89%
15:00~17:00	17.07	44.488	5.96%	49.35%	2.75%	34.55%
17:00~19:00	27.247	75.904	5.69%	51.38%	2.81%	36.93%
19:00~21:00	13.489	40.474	4.89%	47.95%	2.29%	34.35%
21:00~23:00	10.200	39.322	3.41%	44.02%	1.53%	31.9%

As mentioned above, the higher the shifting coefficient β, the greater the tendency of being an exponential distribution; and the value of β within a typical SPL function is normally between 1 and 100. *Table 1* shows that all values of β are lower than 100 and higher than 1. The values of a at 17:00~19:00 are the highest among different time periods, which means the OD flow distribution of this time period is closer to an exponential distribution than those of other time periods. In contrast, the OD distributions at 21:00~23:00 from Monday to Thursday and Sunday are closest to a power law distribution due to their lowest values for β.

Overall, we can see that a small number of OD pairs in the metro system undertake the majority of OD flows, and the OD flow distributions can be fitted by shifted power law with different exponents. It means that mobility behavior within the metro network is centralized although this network itself is decentralized. It can be explained by the fact that some metro stations (actually the surrounding areas of the stations) are frequently linked due to huge travel demand (e.g., commuting, shopping, leisure, and entertainment activities) between these areas, while other metro stations are rarely connected. It also indicates that the metro network is a scale-free network under the preferential attachment mechanism. Besides, most exponents are very close and much around 2 (excluding 21:00~23:00). In other words, although there are nuances of the human mobility across time, the distribution of OD flows in the metro system in most time periods has a similar rhythm.

4. COMMUNITY DETECTION OF THE METRO NETWORK

In addition to analyzing the statistics and temporal patterns of the OD flows from a space perspective, we also investigate community detection. It identifies groups based on the strength of spatial flow distributions between each node, which is useful to reflect the OD travel demand and spatial interaction.

4.1 Community detection algorithm

In order to detect the community structure of the metro network based on OD flows, a similarity-based cluster analysis algorithm, *WalkTrap*, is applied [14]. *WalkTrap* uses a distance measure based on random walks and can be computed efficiently. The idea behind this algorithm is that random walks tend to get trapped in a community, strongly weighted parts of a network. A random walker, also called an agent, moves from one vertex to another, and at each time step, the next vertex is selected at random picking a neighbor of the current vertex. Defining P_{ij}^t as the probability of the agent going from vertex i to vertex j in t steps within a set of vertices V, three basic rules of *WalkTrap* are described as follows:

1) If two vertices i and j are in the same community, the probability P_{ij}^t should be high. But a higher value of P_{ij}^t does not necessarily state that i and j are in the same community.
2) The agent has a higher probability to go to high degree vertices since P_{ij}^t is affected by the degree $d(j)$.
3) Two vertices in a same community tend to "see" all other members in the same way. When there is another vertex k in this community, we will probably conclude that $\forall k, P_{ik}^t \simeq P_{jk}^t$ if vertices i and j are in the same community.

An agglomerative approach can be used to detect communities using random walks as described in [14]. First, the metro network is divided into n communities and every community includes only a single vertex. This is the first community partition $\Omega_1=\{\{v\},v\in V\}$. Second, computing the distance between all adjacent vertices. Third, repeating the following processes to create the next community partition, at each step k:

1) Choose two communities C_1 and C_2 in Ω_1 according to a criterion based on the distance between them.
2) Merge C_1 and C_2 into a new community $C_3=C_1\cup C_2$ and produce the new partition $\Omega_{k+1}=(\Omega_k\setminus\{C_1,C_2\})\cup\{C_3\}$.
3) Update the distances between communities.

The algorithm will be finished after n-1 steps and each step defines a community partition Ω_k of the metro network. Then a network modularity score Q is used to quantify and explore which community partition is the best as measured by

$$Q=\frac{1}{W}\sum_{i}^{n}\sum_{j}^{n}\left(w_{ij}-\frac{s_i^{out}s_j^{in}}{W}\right)\delta(C_i,C_j) \quad (5)$$

where w_{ij} is the OD flows from i to j, $W=\sum_{i}^{n}\sum_{j}^{n}w_{ij}$, s_i^{out} is the sum of the outflows of node i, s_j^{in} is the sum of the inflows of node j; $\delta(C_i,C_j)=1$ when $C_i=C_j$, and otherwise, $\delta(C_i,C_j)=0$. The best community partition is considered to be the one that maximizes Q [13].

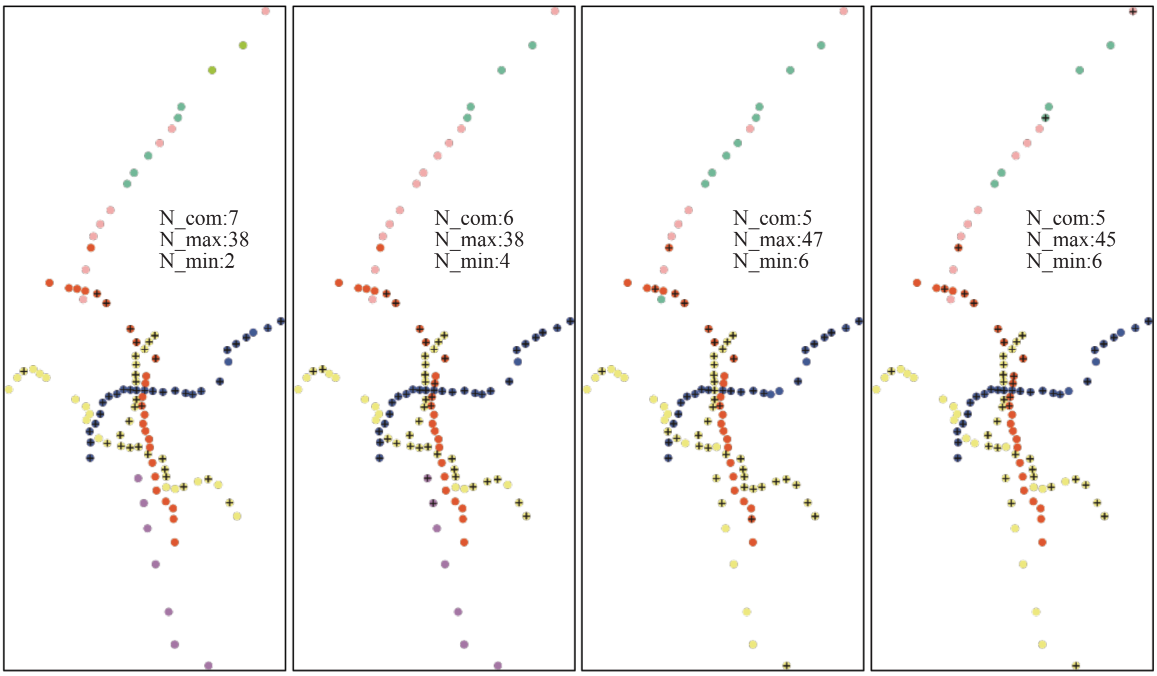
4.2 Community detection results

Combined with the above community detection algorithm, the Igraph Library for network analysis is used to divide communities and find the best partition with high modularity [20]. The traffic community divisions of Nanjing metro network in the morning peak are illustrated in *Figure 4* as an example and traffic community divisions at other time periods are available upon request due to space limitations. Meanwhile, the stations related to the top 5% of OD pairs accounting for about 50% of the trips are highlighted with the symbol "+" in the plots.

Figure 4 shows the following results: first, the community structures for the morning peak on weekdays (Monday to Thursday and Friday) and on weekends (Saturday and Sunday) are similar, with respect to not only the overall layout of community division but also the statistical properties. For the morning peaks, there are more traffic communities on weekdays (7 and 6 on Monday to Thursday and Friday, respectively) than on weekends (5 on both Saturday and Sunday). The largest traffic community for the morning peak is found on Saturday, which contains 47 metro stations on lines 1, 10, and S1, while the smallest traffic community for the morning peak contains only 2 metro stations from Monday to Thursday.

Second, it is obvious that the traffic communities are divided according to the spatial distribution of metro lines. The finding in the present study is in line with the previous studies. Even in a more sophisticated metro system such as Beijing Metro, the community structure is also divided according to the line distribution. Moreover, this rhythm is feasible because different ratios of the total number of trips are considered (e.g. 20%, 50%, 80%). The present study further confirms the similarity across different time periods. Meanwhile, the comparison between the work of Sun et al. [8] and this study shows that the community structure of a relatively sophisticated metro system would display a property with more traffic line division in a metro line than that of a metro system with smaller magnitude. Nevertheless, both works indicate that the spatial partition of the community structure of the metro network is unlike that of the taxi mobility network which is usually geographically adjacent [15]. A major reason may be that relatively more travel by metro concerns long-distance trips and two geographically adjacent metro stations of different lines actually involve a big detour.

Third, the majority of stations related to the top 5% of OD pairs is located in Metro Lines 1 and 2. This result is in accordance with the reality that



a) Monday to Thursday *b) Friday* *c) Saturday* *d) Sunday*

N_com – number of communities
N_max – number of stations in the largest community
N_min – number of stations in the smallest community

Figure 4 – Results of the metro network community division for the morning peak (7:00~9:00)
Note: Different color represents different community

Metro Lines 1 and 2 are the two earliest metro lines in operation. On the one hand, they run through the areas with dense population and intensive commercial and work places, and thus become the important inbound and outbound routes. On the other hand, the operation of these two lines (2005 and 2010, respectively) is far earlier than that of other lines (after July 2014); thus, compared to other metro lines, people are more familiar with Metro Lines 1 and 2 and would rather complete their trips by taking lines 1 and 2 (considering the data in the paper was collected in April 2015). Besides, it is not surprising to find that most stations related to the top 5% of OD pairs are distributed in core districts. *Figure 4* also illustrates that, compared to the morning peak hours on weekdays, more stations are involved in the top 5% of OD pairs for the morning peak hours in weekends. And these extra stations are mostly located in remote suburbs. It indicates that OD flows are more concentrated in several key stations during the AM rush hours of weekdays than that during the AM rush hours of weekends, and the volumes of boarding/alighting passengers of some stations in the city center reduce but the ridership of tap-in and/or tap-out of stations in peripheral areas increases when the time changed from weekdays to weekends.

Moreover, based on metro network community division for all time, we can further identify more details about the difference of community structures within-day (from morning to evening) and between-day (from Monday to Sunday). For example, for a day, the community structure of the evening (21:00~23:00) usually has most communities and the smallest community. It is common for all four types of days. In the evening (after 19:00), the community structures of Monday to Thursday and Sunday are similar, while that of Friday and Saturday are also similar. However, it is different for the morning peak (*Figure 4*).

4.3 Community partition comparison

In order to compare the difference between different within-day and between-day community structures, an information theoretic criterion, normalized mutual information, is used [13]. It can be expressed as

$$I(A,B)=\frac{-2\sum_{i=1}^{C_A}\sum_{i=1}^{C_B}N_{ij}\log\left(\frac{N_{ij}N}{N_{i.}N_{.j}}\right)}{\sum_{i=1}^{C_A}N_{i.}\log\left(\frac{N_{i.}}{N}\right)+\sum_{j=1}^{C_B}N_{.j}\log\left(\frac{N_{.j}}{N}\right)},\tag{6}$$

where C_A, C_B are the number of communities of A and B. N is a confusion matrix where the rows correspond to the communities of A and the columns correspond to the communities of B. N_{ij}, the element of N, is the number of nodes in the community i that also appear in community j. $N_{i.}$ and $N_{.j}$ are the sum of row i and column j, respectively. The value of $I(A,B)$ varies from 0 to 1. The larger the value of I, the greater the structural similarity between the two community partitions.

Based on the community partitions of the morning peak (7:00~9:00) and the community partitions of Monday to Thursday, we calculate the values of $I(A,B)$ as examples for within- and between-day analysis. *Table 3* shows that the biggest similarity of community partitions for the morning peak is between Monday to Thursday and Friday. The community partitions between Saturday and weekdays are quite different, because the corresponding values of $I(A,B)$ are the lowest. The results in *Table 3*

Table 3 – Difference between community partitions for the morning peak (7:00~9:00)

Day	Monday to Thursday	Friday	Saturday	Sunday
Monday to Thursday	1			
Friday	0.964	1		
Saturday	0.857	0.853	1	
Sunday	0.914	0.914	0.932	1

Table 4 – Difference between community partitions of different time periods on Monday to Thursday

Time of day	7:00~9:00	9:00~11:00	11:00~13:00	13:00~15:00	15:00~17:00	17:00~19:00	19:00~21:00	21:00~23:00
7:00~9:00	1							
9:00~11:00	0.833	1						
11:00~13:00	0.815	0.867	1					
13:00~15:00	0.811	0.831	0.936	1				
15:00~17:00	0.848	0.831	0.907	0.904	1			
17:00~19:00	0.845	0.919	0.869	0.846	0.854	1		
19:00~21:00	0.753	0.783	0.836	0.821	0.864	0.859	1	
21:00~23:00	0.904	0.841	0.799	0.769	0.806	0.849	0.768	1

partially support the above analysis about the community partitions for the morning peak based on *Figure 4*.

5. DISCUSSION

In this section, we highlight the key findings from the distributions of OD flow distributions and the community structures of a metro network, using trip datasets collected in Nanjing, China. Since edge weight, instead of the outflow/inflow of node (station) or trip displacement, plays an important role in network properties, we use OD flows during different time periods to strengthen the understanding of urban human mobility and travel regulation.

The main findings of this study can be summarized as follows: (a) the distributions of metro OD flows tend to follow the shifted power law, which reflects the fact that a few OD pairs undertake disproportionately large traffic; (b) the OD flow distributions over 2-hour time periods are checked, and then the temporally dynamic parameters are able to capture the dynamic changes of OD trips in a week; (c) the different but similar exponents indicate that human movement heterogeneities in terms of flow size do exist in within- and between-day ranges, but the disparities between different time periods are not large; (d) the paper proves that the community partitions of metro network over different time periods are similar but with slight difference, and commonly divided according to the spatial distribution of metro lines; (e) the difference of between-days community partitions is not as obvious as that of within-days.

In summary, the results of both scaling law and commuting detection show that human mobility within a metro system during different time periods generally has a similar pattern but with intricate nuances. The first part on the scaling law of OD flows reveals the fact that passenger mobility patterns within the metro network are strongly heterogeneous in terms of volume and a small body of OD pairs undertake disproportionately large traffic. It indicates that the OD flows are very unevenly distributed and a few OD pairs play a dominant role in the linkage of metro stations. For transportation administrators, more attention should be paid on these OD pairs. Flexible and appropriate management policies, such as an integration of full-length and short-turn services, or a combination of express services stopping only at the dominated stations and slower services stopping at all stations, should be applied. Moreover, the second part on community detection provides clear evidences to illustrate that community structures are obviously divided according to the line distribution, which further supports the strategy that the integration of full-length and short-turn services or a combination of express and slower services is feasible and meaningful since the community structure is strongly related to travel demand. The result of commuting detection also indicates that the differences of community partitions within a day are more obvious than those between days. Therefore, relatively more efforts should be given on the daily metro system management (e.g. daily train timetable, an integration of full-length and short-turn services) to meet the complex situations in a day.

6. CONCLUSION

This study demonstrates that big spatiotemporal travel data collected from AFC systems offer great opportunities for in-depth analysis of complex human mobility and travel characteristics in public transit systems. It contributes not only in extending the research field by applying both scaling law and community detection methods on the mobility analysis within the metro network, but also in providing useful insight for urban and transportation planning (e.g. a majority of OD flows are concentrated to a few OD pairs and traffic communities are obviously divided according to the metro line distribution, etc.). This study is limited in the sense that only a single metro network is studied due to the difficulty of data collection. Moreover, metro trips only represent a part of urban travel and show human mobility characteristics from a specific perspective. Additional modes, such as bus and public bicycles, need to be studied in future to make further understanding of urban transport properties.

ACKNOWLEDGEMENT

This research is supported by the China Postdoctoral Science Foundation (No. 2020M670732) and the Fundamental Research Funds for the Central Universities (No. 3132020159).

甘佐贤[1]
邮箱：zxgan@dlmu.edu.cn
梁晶[1]
邮箱：liangjing@dlmu.edu.cn
[1] 大连海事大学交通运输工程学院，大连，116026

理解地铁网络内的出行模式：客流分布和社群识别

摘要

本文以南京地铁采集的智能卡数据为基础，对地铁网络中人的一天中不同时段和一周中不同日期的客流模式进行了分析。结果表明，*OD*客流（起点至终点）可以很好地采用指数在*2*左右的漂移幂律分布来表征，这反映了地铁系统中少数的*OD*对起主导地位并承担了大量的*OD*流量分布。不同的指数表明一天中不同时段和一周中不同日期客流的移动存在异质性。另外，我们采用基于随机游走的社群识别方法分析了不同时段的地铁站点社群结构，进一步从空间角度分析了客流的移动规律。基于信息论准则的标准互信息被用来比较不同时段的社群结构差异，结果表明地铁内部不同时段客流移动特性虽然存在细微差别，但总体具有相似规律，而且社群结构往往按照地铁线路布设进行划分。本文研究为理解城市居民出行模式提供了一种基于时空视角的见解，继而有利于交通管理实践。

关键词

地铁；*OD*客流；漂移幂律分布；社群结构

REFERENCES

[1] Foell S, Phithakkitnukoon S, Veloso M, Kortuem G, Bento C. Regularity of Public Transport Usage: A Case Study of Bus Rides in Lisbon. Portugal. *Journal of Public Transportation*. 2016;19(4): 161-71.

[2] Roth C, Kang SM, Batty M and Barthélemy M. Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS ONE*. 2011;6(1): e15923.

[3] Mari L, Bertuzzo E, Righetto L, Casagrandi R, Gatto M, Rodriguez-Iturbe I, Rinaldo A. Modelling cholera epidemics: The role of waterways, human mobility and sanitation. *Journal of the Royal Society Interface.* 2012;67(9): 376-88.

[4] Zhong C, Arisona SM, Huang X, Batty M, Schmitt G. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*. 2014;28(11): 2178-2199.

[5] Gan Z, Yang M, Feng T, Timmermans H. Understanding urban mobility patterns from a spatiotemporal perspective: Daily ridership profiles of metro stations. *Transportation*. 2020;47(1): 315-336.

[6] Kim K, Oh K, Lee YK, Kim S, Jung JY. An analysis on movement patterns between zones using smart card data in subway networks. *International Journal of Geographical Information Science*. 2014;28(9): 1781-1801.

[7] Reades J, Zhong C, Manley ED, Milton R, Batty M. Finding Pearls in London's Oysters. *Built Environment.* 2016;42(3): 365-381.

[8] Sun L, Ling X, He K, Tan Q. Community structure in traffic zones based on travel demand. *Physica A*. 2016;457: 356-363.

[9] Xu Q, Mao BH, Bai Y. Network structure of subway passenger flows. *Journal of Statistical Mechanics: Theory and Experiment.* 2016;3: 033404.

[10] El Mahrsi MK, Côme E, Oukhellou L, Verleysen M. Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems*. 2017;18(3): 712-728.

[11] Wang Y, de Almeida Correia GH, de Romph E, Timmermans H. Using metro smart card data to model location choice of after-work activities: An application to Shanghai. *Journal of Transport Geography.* 2017;63: 40-47.

[12] Jiang S, Guan W, Zhang W, Chen X, Yang L. Human mobility in space from three modes of public transportation. *Physica A*. 2017;483: 227-238.

[13] Newman ME, Girvan M. Finding and evaluating community structure in networks. *Physical Review E*. 2004;69: 026113.

[14] Pons P, Latapy M. Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*. 2005;10(2): 191-218.

[15] Liu X, Gong L, Gong Y, Liu Y. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography.* 2015;43: 78-90.

[16] Gan Z, Yang M, Feng T, Timmermans HJP. Examining the relationship between built environment and metro ridership at station-to-station level. *Transportation Research Part D: Transport and Environment.* 2020;82: 102332.

[17] Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM Review*. 2009;51: 661-703.

[18] Liang X, Zheng X, Lv W, Zhu T, Xu K. The scaling of human mobility by taxis is exponential. *Physica A.* 2012;391: 2135-2144.

[19] Chang H, Su BB, Zhou YP, He DR. Assortativity and act degree distribution of some collaboration networks. *Physica A*. 2007;383: 687-702.

[20] Csardi G, Nepusz T. The igraph software package for complex network research. *Complex System*. 2006;1695: 1-9.