



Understanding coupling dynamics of public transportation networks

Zhihao Zheng¹, Zhiren Huang¹, Fan Zhang² and Pu Wang^{1*}

*Correspondence:

wangpu@csu.edu.cn

¹School of Traffic and Transportation Engineering, Central South University, Changsha, P.R. China

Full list of author information is available at the end of the article

Abstract

Subway and bus networks work as an integrated multiplex transportation system and play an indispensable role in modern big cities. Even though a variety of works have investigated the coupling dynamics of multiplex transportation networks, empirical data that validates the determinant coupling factors are still lacking. In this paper, we employ smartcard data of 2.4 million subway and bus passengers in Shenzhen, China to study the coupling dynamics of subway and bus networks. Surprisingly, the coupling of subway and bus networks is not notably influenced by the time-varying speed ratio of the two network layers but is jointly determined by the distribution of travel demands and transportation facilities. Our findings highlight the important role of real travel demand data in analyzing the coupling dynamics of multiplex transportation networks. They also suggest that the speed ratio of different network layers, which was regarded as a key factor in determining coupling strength, has a negligible effect on travelers' route selections, and thus the coupling dynamics of multiplex transportation networks.

Keywords: Multiplex networks; Coupling; Spatial networks; Data analysis

1 Introduction

Multiplex networks are ubiquitous in modern human societies. Representative types of multiplex networks range from social networks [1] and biological networks [2] to power grids [3] and transportation networks [4, 5]. In multiplex networks, network layers are not isolated entities but interact with each other as an integrated system [6]. In recent years, a variety of new and interesting properties of percolation processes [7–11], cascade failures [8, 12], epidemic spreading [13–20], and coupling dynamics [21–24] on multiplex networks have been discovered. Coupling is a very important and interesting research topic because it occurs on real networks in our daily life. In big cities, several layers of transportation infrastructure work together to supply the continuously growing urban mobility demand [25]. The coupling of different layers of transportation infrastructure is generated by travelers using more than one mode of transportation to complete a trip. The most common examples include the “Park and Ride” mode [26] that generates coupling between road and subway networks and passenger transfer that generates coupling between bus and subway networks.

Previous works discovered a variety of interacting features of multiplex transportation networks. Morris and Barthélemy found that the shortest path length decreases as the

coupling increases on a spatial network [21]. Gallotti and Barthélemy studied how the riding, walking and waiting time of trips vary with distance [22]. Zhou *et al.* proposed a layered network model for transportation systems [23]. The authors found the transition point of cooperation between layered networks [23]. Strano *et al.* analyzed the coupling property between street and subway networks in London and New York [24]. An optimal operation speed is determined for London subway [24]. How operation speed of subway influences passenger travel time and road traffic congestion was studied using CDR (Call Details Records) data [27]. The results indicated that travel demand plays crucial quantitative and qualitative roles [27]. How interconnections between network layers influence traffic congestion [28] and how speed ratio influences network capacity and travel time [29] were also investigated.

Despite the rapid developments in coupling dynamics of multiplex transportation networks, most existing works are based on simulated networks or simulated travel demands. This may lead to findings inconsistent with empirical evidence. First, the actual topology of a transportation network is usually much more complicated than the topology of a simulated network. Second, the obtained coupling results usually deviate from real situations when using uniformly (randomly) generated travel demands [27]. Third, mobile phone data [30, 31] cannot be used to capture the transfers between different network layers either. Finally, the inconvenience of transfers and individual preferences cannot be evaluated and incorporated in building the models. In summary, previous works had to make strong assumptions on travelers' routes and mode-selection behavior due to the lack of data.

The recent emerging General Transit Feed Specification (GTFS) service provides both static and real-time transit data, which include transit route data, schedule data, real-time bus position data, etc. Geographic transit service data formatted in GTFS have been widely used in transportation analysis and applications [32]. Wong used GTFS data to analyze transit headways for single transport agency and multiple transport agencies [33]. Hadas employed GTFS data and geographic information data of transportation networks to analyze the connectivity indicators (network coverage, operation speed, intersection coverage, stop transfer potential and route overlap) of public transportation systems [34]. Tao studied the spatial-temporal travel behavior of bus passengers using GTFS data [35]. Lee *et al.* used GTFS data to investigate the symmetry of boarding and alighting in time and space [36]. Huang *et al.* combined GTFS data and taxi GPS data to identify and anticipate large crowd gatherings [37].

Using a set of data similar to GTFS data, we studied the coupling dynamics of the multiplex transportation network in Shenzhen, China. The geographic information systems (GIS) data of subway and bus networks, and the transit route data were used to obtain accurate topology of the multiplex transportation network. Smartcard data of bus and subway passengers were analyzed to estimate dynamical passenger travel demands. The GPS data of buses were used to estimate the road traffic condition. Real transfers between subway and bus networks can be captured using the real large-scale data. In the present study, we focus on one important question: What are the major determining factors for the coupling strength of subway and bus networks?

2 Data

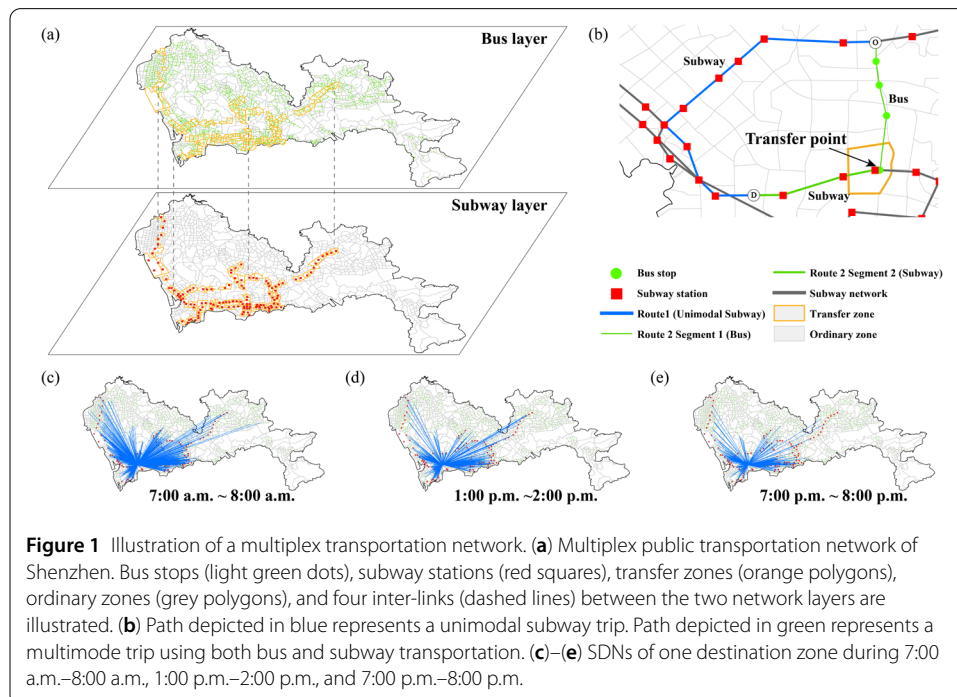
In this section, we briefly introduce the datasets used. More details about the used data are provided in the Additional file 1.

The GIS data of subway and bus networks provide geographic locations of each bus stop and each subway station in Shenzhen, China (Fig. S6). In the studied multiplex public transportation network, the number of bus stops (9114 in total) is much larger than the number of subway stations (132 in total).

Bus transit route data and subway transit route data provide the stop information of all buses and the station information of all subway lines. The data also provide the basic attributes (length, running time) of the subway sections.

The GIS data of traffic zones in Shenzhen were provided by Shenzhen Transportation Authority. In this dataset, the geographic area of Shenzhen is partitioned into 996 traffic zones, within which 844 zones contain at least one bus stop or one subway station (Fig. 1(a)).

Smartcard data recorded the trips of more than 2.4 million smartcard users on 915 bus lines and 6 subway lines in Shenzhen (Figs. S1–S4). Each smartcard has a unique ID number. During the 1-week data collection period (August 8–15, 2016), 48 million transaction records were generated. The time and station that a passenger enters or exits the subway were recorded, whereas only the information of the bus boarded and the boarding time were recorded for bus passengers. The time that a bus arrived at each bus stop was inferred by matching the GPS coordinates of the bus to the stops along the bus route. This generates the timetable of the bus. Consequently, the boarding time of each bus passenger was compared with the timetable to identify the bus stop where each passenger got on board. Given that bus passengers do not tap cards when getting off a bus, the destination information is not available. In the following section, we use a four-step method to infer the destination of each bus trip.



3 Methods

3.1 Estimating the destination of a bus trip

We propose a four-step method to estimate the destinations of the bus trips. The method is elaborated as follows:

Step 1: Two assumptions proposed by Barry *et al.* [38] are used: (1) a passenger will start a new trip near the destination of the current trip. Therefore, the bus stop which is closest to (and within 2 km of) the boarding stop or station of the next trip is inferred as the destination stop of the current bus trip. Here, 2 km is selected as the maximum radius to search candidate stops, because previous works [38–41] suggest that the maximum radius of individual walking activities is 1–2 km. (2) at the end of a day, a passenger will return to the origin of his/her first trip of the day. Therefore, the bus stop which is closest to (and within 2 km of) the boarding stop or station of the first trip of a day is inferred as the destination stop of the last trip of the day.

Step 2: If a passenger only has one bus trip during a day, the bus stop which is closest to (and within 2 km of) the origin of the first trip in the next day is inferred as the destination [39]. If for a passenger the destination of the last bus trip cannot be found within 2 km of the origin of the first trip of the day, the bus stop which is closest to (and within 2 km of) the origin of the first trip in the next day is inferred as the destination [39].

Step 3: If the destination of a bus trip cannot be identified in Step 1 and 2, the historical smartcard records of the bus passenger [39] are used. Bus trips that share the same route, same boarding stop, and similar boarding time with the unidentified bus trip are analyzed. The similar historical bus trip that has the most similar boarding time to the unidentified trip is identified. And the destination of the historical bus trip is inferred as the destination of the unidentified bus trip.

Step 4: If the destination of a bus trip cannot be identified in Step 1, 2 and 3 (approximately 9% of all bus trips), the most frequent alighting bus stop after the boarding stop is selected as the destination of the bus trip.

Using the proposed method, destinations are identified for 97% of the bus trips. Once the destination of a bus trip is estimated, we infer the alighting time of the bus passenger according to the arrival time of the bus at the bus stop.

3.2 Identifying transfers in the multiplex transportation network

The methods to identify transfers can be classified into two types [40–48]. One method uses the time interval between the starting times of two consecutive trips as the criterion to determine if two consecutive trips belong to a whole trip. The other method uses the time interval between the end time of the first trip and the starting time of the following second trip as the criterion. The advantage of the first method is that the estimation of the time interval is accurate because the boarding time of a bus or subway passenger is always recorded. However, this method may not work well when the travel time of the first trip is too short or too long. The second method avoids the drawback of the first method, however, the second method depends on accurate estimation of the destinations of bus trips. Here, we combine the two methods to identify transfers:

Step 1: If the time interval between the end of the first trip and the start of the second trip is within 30 min, a transfer is identified between the two consecutive trips; otherwise, the two trips are identified as separate trips.

Step 2: If the destination information of the first trip is not available, another criterion is applied. If the time interval between the starting times of two consecutive trips is within

60 min, a transfer is identified between the two consecutive trips; otherwise, the two trips are identified as separate trips.

In the present study, only multimode trips with transfers between subway network and bus network are used to estimate the coupling strength between subway network and bus network.

3.3 Generating time-varying single-destination networks

Each observational day is split into 17 1-hour time windows from 6:00 a.m. to 10:00 p.m. that cover the main service period of public transportation in Shenzhen. Passenger trips are grouped into a time window according to the time (1-hour time window) when the trip started. Single-destination network (SDN) of each time window is generated based on the trips started during the time window (Figs. 1(c)–(e)). In an SDN, nodes represent traffic zones, and an origin zone is connected to the unique destination zone if at least one qualified trip is observed between the two zones during the time window. Qualified trips include:

- (a) Multimode trips that use both bus and subway services.
- (b) Bus trips with alternative multimode routes connecting the origin and destination zones.
- (c) Subway trips with alternative multimode routes connecting the origin and destination zones.

These criteria ensure that all unimodal trips have alternative multimode routes to the destination zone. In the following analysis, only qualified trips are used. For each SDN, there are several origin zones and one destination zone. For each destination zone, there are 17 SDNs generated during a day.

In the present study, traffic zones containing at least one bus stop and one subway station are defined as transfer zones (124 in total), and the other traffic zones are defined as ordinary zones (Fig. 1(b)). Here, we only analyzed the SDNs with destinations located at transfer zones (Fig. S5). This is because both bus-to-subway trips and subway-to-bus trips exist in these SDNs. While for the SDNs with destinations located at ordinary zones, only bus-to-subway trips or subway-to-bus trips can be observed. The transfer zones are popular destinations, roughly 70% of the trips are captured by the 124 SDNs.

4 Results

For each SDN i of a time window t , we measure the coupling strength $\lambda_i(t)$ between the subway network and the bus network as

$$\lambda_i(t) = \frac{n_i(t)^{\text{coupled}}}{n_i(t)}, \quad (1)$$

where $n_i(t)^{\text{coupled}}$ and $n_i(t)$ denote the number of multimode trips and the total number of trips in i . Given that SDNs may have different scales of coupling strengths, we normalize the coupling strength $\lambda_i(t)$ of each SDN when analyzing the temporal patterns of $\lambda_i(t)$:

$$\lambda_i^{\text{norm}}(t) = \frac{\lambda_i(t) - \lambda_i^{\min}}{\lambda_i^{\max} - \lambda_i^{\min}}, \quad (2)$$

where λ_i^{\max} and λ_i^{\min} denote the maximum and minimum values of $\lambda_i(t)$ in the day.

Using the smartcard data and the bus GPS data, we estimate the time-varying operation speeds of subway network and bus network for each SDN. We find that the operation speed of subway is stable during the whole day, whereas the operation speeds of buses are influenced by road traffic conditions (Fig. S7). Here, the speed ratio of an SDN i during time window t is defined as the average trip speed on the subway network over the average trip speed on the bus network. The average speed of a trip is calculated using the trip distance divided by the travel time:

$$\bar{v}_i^s(t) = \frac{1}{N_s} \sum v_{i,j}^s, \tag{3}$$

$$\bar{v}_i^b(t) = \frac{1}{N_b} \sum v_{i,k}^b, \tag{4}$$

$$\beta_i(t) = \frac{\bar{v}_i^s(t)}{\bar{v}_i^b(t)}, \tag{5}$$

where $v_{i,j}^s$ is the speed of subway trip j and $v_{i,k}^b$ the speed of bus trip k , $\bar{v}_i^s(t)$ and $\bar{v}_i^b(t)$ are the average speed on the subway layer and bus layer of SDN i during time window t . When comparing $\beta_i(t)$ of different SDNs, we normalize the speed ratio using the maximum–minimum normalization method:

$$\beta_i^{\text{norm}}(t) = \frac{\beta_i(t) - \beta_i^{\text{min}}}{\beta_i^{\text{max}} - \beta_i^{\text{min}}}, \tag{6}$$

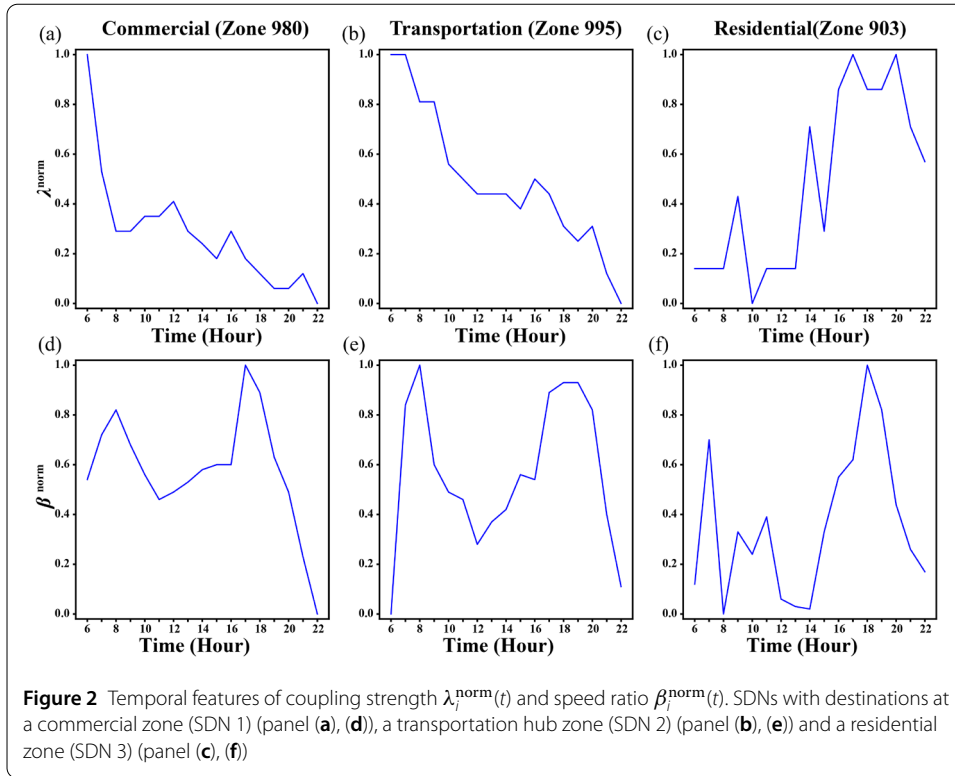
where β_i^{max} and β_i^{min} denote the maximum and minimum values of $\beta_i(t)$ in the day.

4.1 Preliminary observation on three (case studies) SDNs

Three SDNs with destination zones located at a commercial zone (SDN 1), a transportation hub zone (SDN 2), and a residential zone (SDN 3) are analyzed (Fig. 2). The coupling strengths $\lambda_i^{\text{norm}}(t)$ of the SDN 1 and SDN 2 reached the maximum in the morning rush hours (6:00 a.m.–7:00 a.m.), whereas the coupling strength of SDN 3 reached the maximum in the evening rush hours (5:00 p.m.–7:00 p.m.) (Figs. 2(a)–(c)). The speed ratio $\beta_i^{\text{norm}}(t)$ is larger in the morning and evening rush hours for all three SDNs, because the operation speeds of buses are lower in peak hours. The speed ratio $\beta_i^{\text{norm}}(t)$ was considered as the key factor influencing the mode-selection behavior of travelers and, thus, the coupling strength of layered networks [21, 24, 29]. However, our empirical results indicate that passengers did not react to speed ratio changes (Figs. 2(d)–(f)). No strong correlation is found between coupling strength $\lambda_i^{\text{norm}}(t)$ and speed ratio $\beta_i^{\text{norm}}(t)$ for all three SDNs (Figs. 3(a), (d), (g)).

One intuition is that as trip distance increases, a passenger is more likely to use both bus network and subway network. Therefore, the correlation between average trip distance and coupling strength is analyzed. Here, average trip distance is defined as the average Euclidean distance from each origin to the single destination of an SDN, weighted by the number of trips from each origin:

$$d_i(t) = \sum_{o \in O_i(t)} \frac{n_{o_i}(t)}{n_i(t)} d_{o_i}, \tag{7}$$



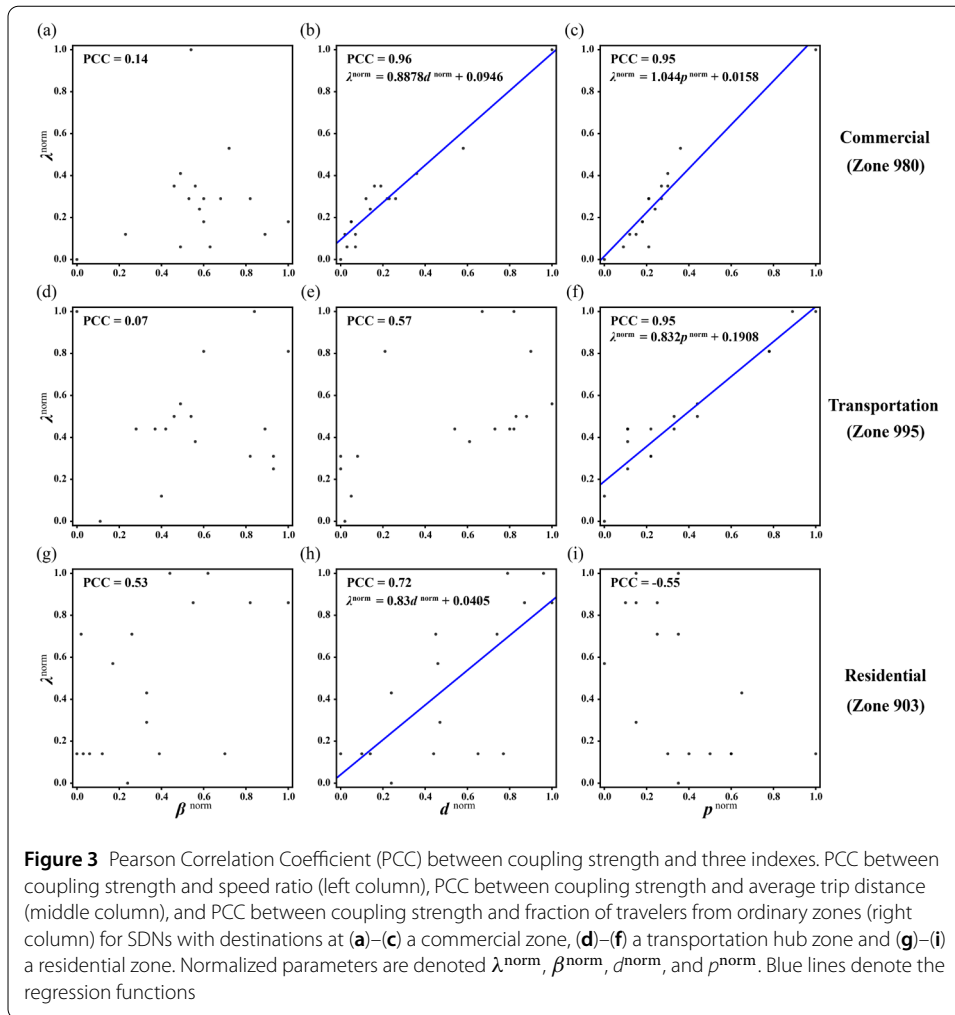
where $O_i(t)$ represents the set of origins of SDN i during time window t , $n_{o_i}(t)$ represents the number of trips from origin o , and d_{oi} represents the Euclidean distance between o and the destination zone. We normalize the average trip distance using the maximum–minimum normalization method:

$$d_i^{\text{norm}}(t) = \frac{d_i(t) - d_i^{\text{min}}}{d_i^{\text{max}} - d_i^{\text{min}}}, \tag{8}$$

where d_i^{max} and d_i^{min} denote the maximum and minimum values of $d_i(t)$ in the day. We observe good positive correlations between coupling strength $\lambda_i^{\text{norm}}(t)$ and average trip distance $d_i^{\text{norm}}(t)$ for SDN 1 and SDN 3 (Figs. 3(b), (h)). However, the correlation between $\lambda_i^{\text{norm}}(t)$ and $d_i^{\text{norm}}(t)$ is weak for the SDN 2 (Fig. 3(e)). Possible explanations are: there are abundant direct public transportation service to the transportation hub, and the transportation hub serves more as the destination rather than the transfer place for passengers (44,931 trips vs. 5262 trips on Monday).

Another experience in our daily life is that trips originating from suburban areas may have a higher possibility of using both subway and bus transportation. This is because no subway service is available in suburban areas, and travelers may first use bus transportation to reach a subway station. Therefore, we define the origin composition index p , the fraction of trips from ordinary zones:

$$p_i(t) = \frac{n_i(t)^{\text{ordinary}}}{n_i(t)}, \tag{9}$$



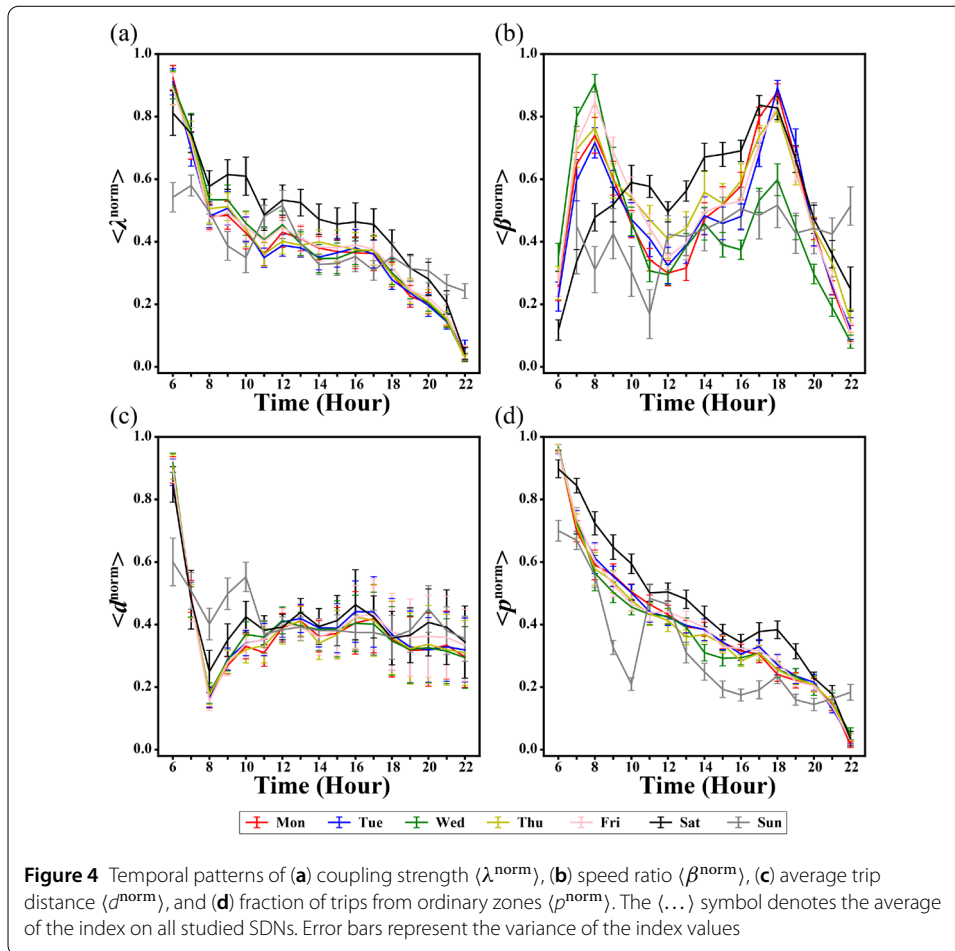
where $n_i(t)^{\text{ordinary}}$ and $n_i(t)$ denote the number of trips originating from ordinary zones and the total number of trips of SDN i in time window t . We normalize $p_i(t)$ using the maximum–minimum normalization method:

$$p_i^{\text{norm}}(t) = \frac{p_i(t) - p_i^{\text{min}}}{p_i^{\text{max}} - p_i^{\text{min}}}, \quad (10)$$

where p_i^{max} and p_i^{min} denote the maximum and minimum values of $p_i(t)$ in the day. Strong positive correlations are observed between $\lambda_i^{\text{norm}}(t)$ and $p_i^{\text{norm}}(t)$ for SDN 1 and SDN 2 (Figs. 3(c), (f)). However, a negative correlation is observed for SDN 3 (Fig. 3(i)). Possible reason could be that the trip distance is relatively short (the average trip distance is 8 km during a day). Even when a large fraction of passengers come from ordinary zones, they can use direct public transportation service or the travel time cannot be obviously reduced when transferring to the faster subway network.

4.2 Statistical analysis on all 124 SDNs

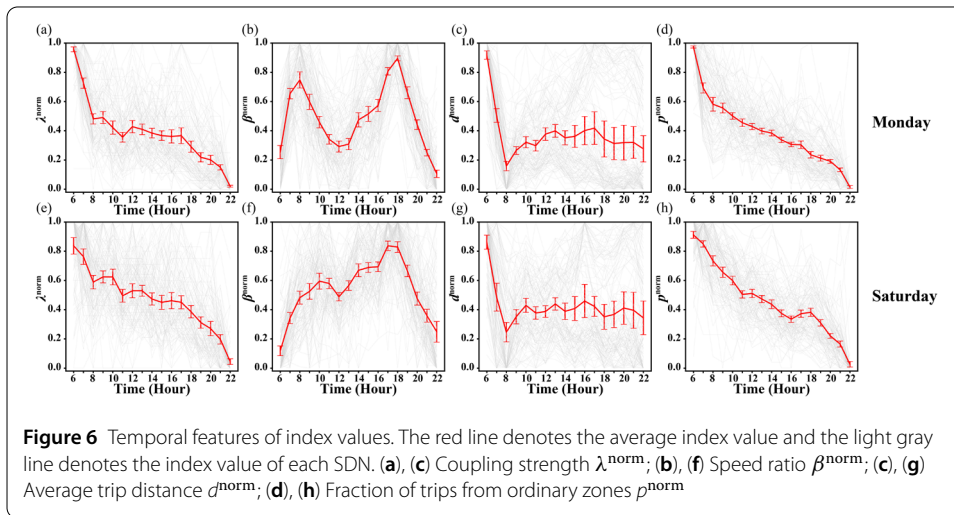
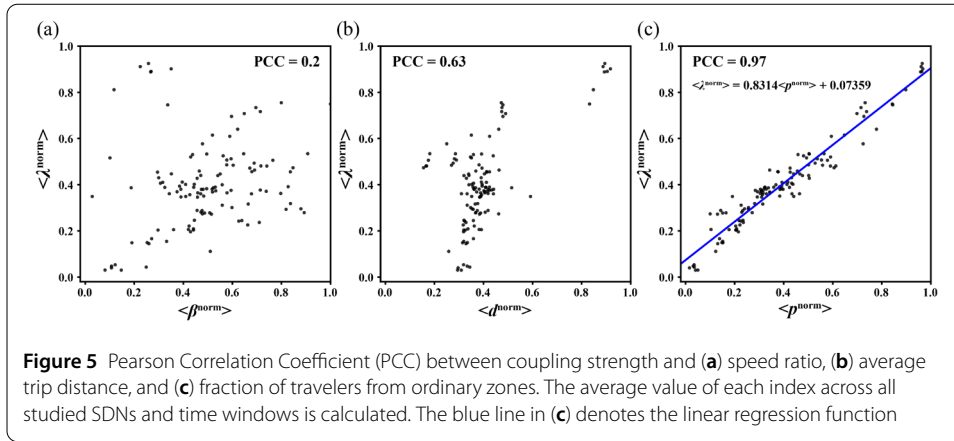
The results of the three case studies indicate that no single index can determine the coupling strength of all SDNs. In addition, different land use features may influence the route



and mode selection behavior of travelers. In the present study, without confident and accurate land use information, we focus on the general trend rather than the case feature of the determining factor of the coupling strength.

We first find that average coupling strength $\langle \lambda_i^{\text{norm}}(t) \rangle$ is high in morning rush hours and decreases afterwards for both weekdays and weekends (Fig. 4(a)). This can be explained by that most studied SDNs attract early morning commuters who must use multimode transport to ensure punctuality. Compared with weekday results, $\langle \lambda_i^{\text{norm}}(t) \rangle$ is lower in morning, but larger for other time windows on weekends. This can be explained by that there are less commuters but more afternoon activities on weekends.

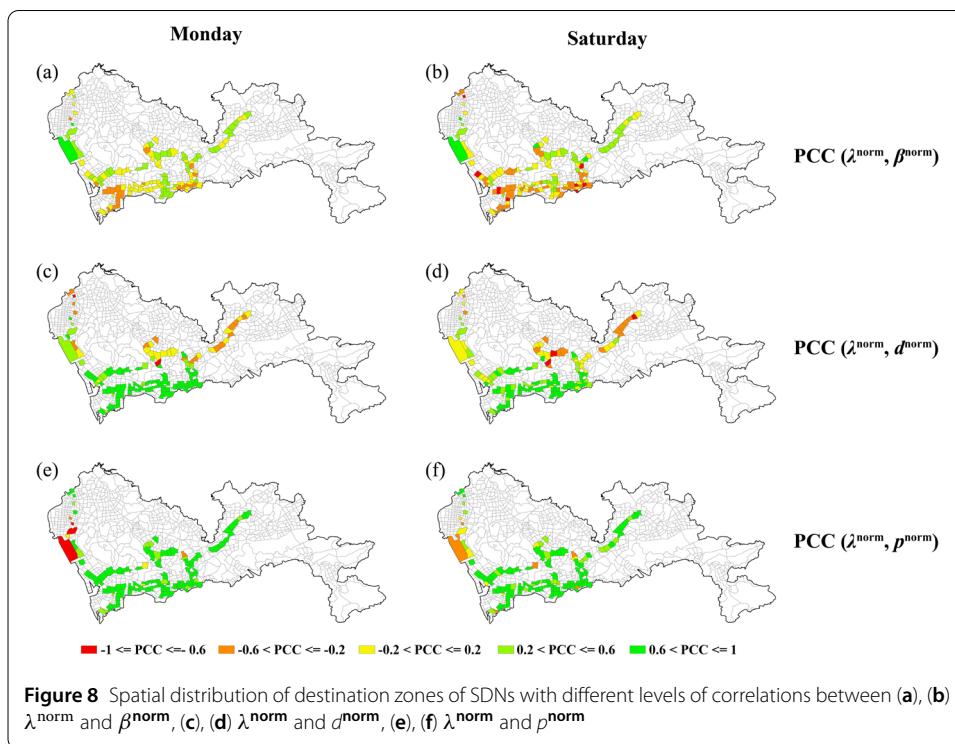
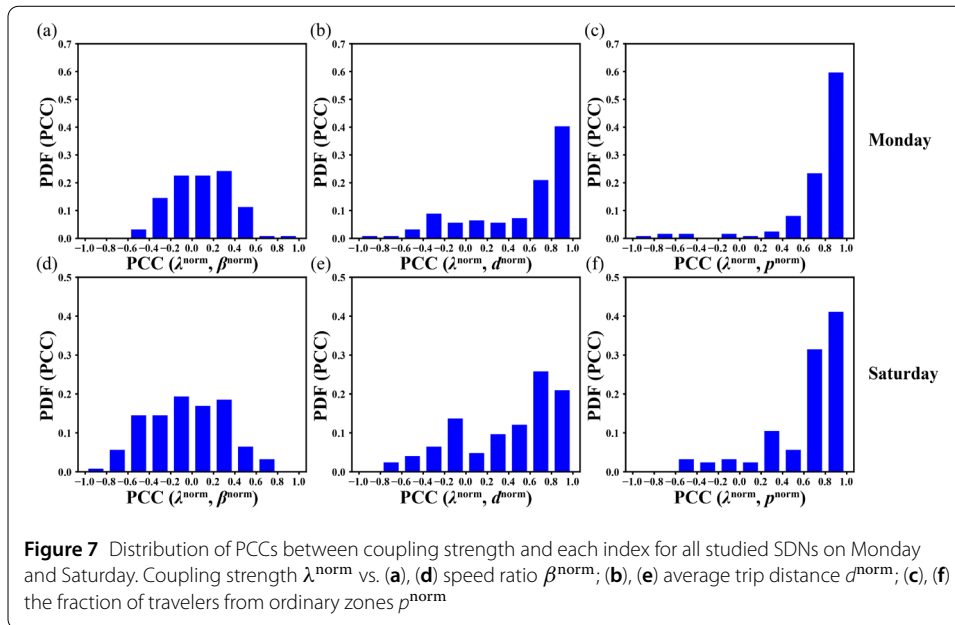
The average speed ratios on weekdays and Saturday have similar temporal patterns. The speed ratio did not have a peak on Sunday evening (Fig. 4(b)). The coupling strength did not increase with the speed ratio during evening rush hours (Figs. 4(a), (b)). No prominent correlation is found between the coupling strength and the speed ratio on the studied 124 SDNs (Fig. 5(a)). The average trip distances on the studied SDNs show similar patterns for all weekdays and weekends (Fig. 4(c)). We observed a moderate positive correlation between coupling strength $\langle \lambda_i^{\text{norm}}(t) \rangle$ and average trip distance $\langle d_i^{\text{norm}}(t) \rangle$ (Fig. 5(b)). The origin composition index $\langle p_i^{\text{norm}}(t) \rangle$ decreases slightly faster than $\langle \lambda_i^{\text{norm}}(t) \rangle$ after the morning rush hours (Fig. 4(d)). A strong positive correlation is observed between $\langle \lambda_i^{\text{norm}}(t) \rangle$ and $\langle p_i^{\text{norm}}(t) \rangle$ (Fig. 5(c)). This finding suggests that $\langle p_i^{\text{norm}}(t) \rangle$ could be a dominant factor deter-



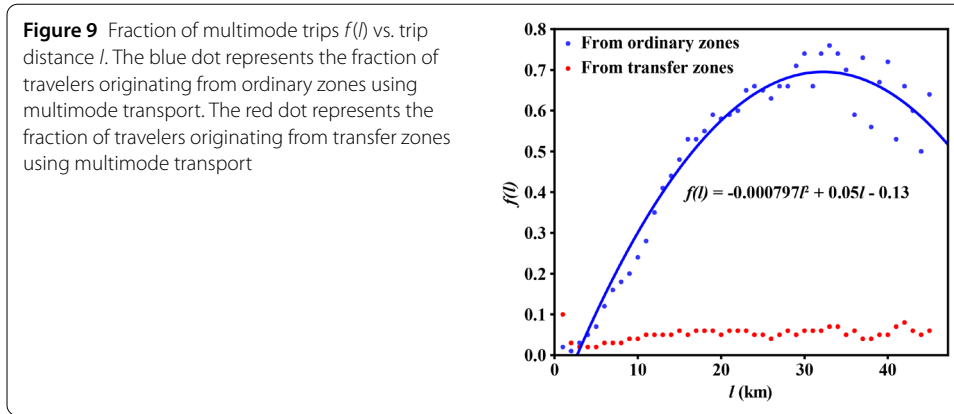
mining the coupling strength, though negative correlations between $\lambda_i^{\text{norm}}(t)$ and $p_i^{\text{norm}}(t)$ may exist in a few SDNs (Fig. 3(i), Figs. 7(c), (f)).

Using Monday and Saturday as examples, temporal distributions of all indexes for each SDN are analyzed (Fig. 6). Pearson Correlation Coefficient (PCC) between $\lambda_i^{\text{norm}}(t)$ and $\beta_i^{\text{norm}}(t)$, PCC between $\lambda_i^{\text{norm}}(t)$ and $d_i^{\text{norm}}(t)$, and PCC between $\lambda_i^{\text{norm}}(t)$ and $p_i^{\text{norm}}(t)$ are measured (Fig. 7). The speed ratio index and the origin composition index of all SDNs follow similar trends. While several SDNs show different trends for the coupling index and the distance index (Fig. 6). As shown in Fig. 7, origin composition index $p_i^{\text{norm}}(t)$ shows strong positive correlation with coupling strength $\lambda_i^{\text{norm}}(t)$ for most SDNs, while fewer SDNs are found to have such strong correlation between distance index $d_i^{\text{norm}}(t)$ and coupling index $\lambda_i^{\text{norm}}(t)$. The speed index $\beta_i^{\text{norm}}(t)$, which was considered a key factor in coupling dynamic studies, has no prominent correlation with coupling index $\lambda_i^{\text{norm}}(t)$ for the majority of SDNs.

The PCC values were also illustrated in space using the destination zone of each SDN (Fig. 8). No prominent spatial pattern is found for the correlation between speed ratio index $\beta_i^{\text{norm}}(t)$ and coupling index $\lambda_i^{\text{norm}}(t)$ (Figs. 8(a), (b)). Passengers seem to have a random reaction with the time-varying travel speed on different SDNs. Interestingly, the destination zones of SDNs with strong positive correlations between distance index $d_i^{\text{norm}}(t)$



and coupling index $\lambda_i^{\text{norm}}(t)$ are located in the central urban area. The destination zones of SDNs with weak or negative correlations between $d_i^{\text{norm}}(t)$ and $\lambda_i^{\text{norm}}(t)$ are located in the periphery area (Figs. 8(c), (d)). The peripheral destination zones attract long-distance travelers with less use of multimode transport, which is probably caused by that these zones attract many long-distance travelers from the city center with wide accessibility to public transportation. The destination zones of SDNs with strong positive correlations between $p_i^{\text{norm}}(t)$ and $\lambda_i^{\text{norm}}(t)$ are widely distributed across the investigated area. A few



exceptions with weak correlations between $p_i^{\text{norm}}(t)$ and $\lambda_i^{\text{norm}}(t)$ are also found. This implies that though origin composition index $p_i^{\text{norm}}(t)$ is the dominant factor determining the coupling strength, the coupling strength is also determined by other factors.

4.3 A new index showing higher correlations with coupling strength

For SDNs that have moderate or weak correlations between $p_i^{\text{norm}}(t)$ and $\lambda_i^{\text{norm}}(t)$, we suppose that the distance index can be a secondary factor to influence the coupling. We respectively analyze the fraction of multimode trips $f(l)$ for travelers originating from transfer zones and ordinary zones (Fig. 9). On one hand, multimode trips originating from transfer zones account for less than 2% of the overall multimode trips, independent of trip distance l . Hence, two modes of transportation are barely used in trips connecting two zones with subway stations. This is consistent with our intuition. Subway usually has a faster speed than buses. Even when a mixing use of subway and bus is faster than a unimodal subway trip, travelers are not likely to do so due to the inconvenience of transfers [49, 50]. On the other hand, for trips originating from ordinary zones, multimode trips are increasingly favored as the trip distance increases. The finding can be explained by the decreasing availability of direct buses to destinations and the increasing attraction of reduced travel times provided by the subway. A transition point, however, is observed when the trip distance exceeds 33 km. It is probably because multimode trips are replaced by customized bus lines connecting major remote districts.

Taken together, the use of multimode transport majorly exists within trips from ordinary zones to transfer zones. Neither the trip distance nor the speed ratio has a large impact on passengers' routing behavior for trips originating from transfer zones. While for travelers traveling from ordinary zones to transfer zones, the trip distance plays an important role in determining whether to use a multimode transport. These findings motivate us to incorporate the distance of trips originating from ordinary zones into the origin composition index $p_i(t)$. Hence, we propose a new index $\xi_i(t)$, the trip distance weighted fraction of passengers from ordinary zones. We use the regression function $f(l)$ shown in Fig. 9 to calculate the weight for the origin composition index $p_i(t)$:

$$w_i(t) = \frac{-0.000797(d_i^{\text{ord}}(t))^2 + 0.05d_i^{\text{ord}}(t) - 0.13}{f_{\text{max}}}, \tag{11}$$

$$\xi_i(t) = w_i(t) \times p_i(t), \tag{12}$$

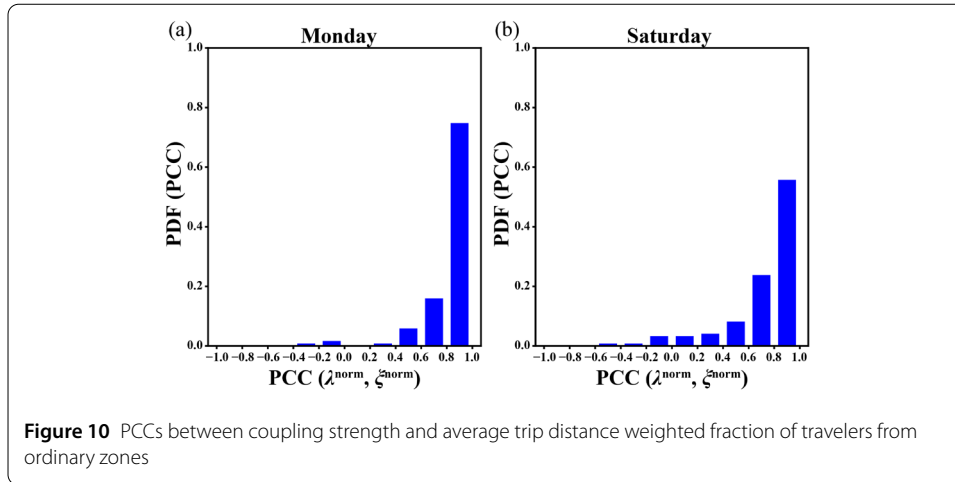


Table 1 Average PCC values and R^2 values between measured indexes across studied SDNs

		λ vs. β	λ vs. d	λ vs. p	λ vs. ξ
Monday	PCC	0.09	0.52	0.72	0.81
	R^2	0.09	0.49	0.64	0.71
Tuesday	PCC	0.06	0.47	0.7	0.77
	R^2	0.11	0.45	0.61	0.68
Wednesday	PCC	0.34	0.48	0.7	0.76
	R^2	0.2	0.47	0.62	0.64
Thursday	PCC	0.13	0.49	0.71	0.77
	R^2	0.11	0.47	0.6	0.67
Friday	PCC	0.14	0.47	0.7	0.77
	R^2	0.11	0.47	0.6	0.68
Saturday	PCC	-0.04	0.4	0.63	0.72
	R^2	0.14	0.37	0.53	0.61
Sunday	PCC	-0.1	0.28	0.6	0.65
	R^2	0.17	0.29	0.48	0.52

where $d_i^{ord}(t)$ is the average distance of trips from ordinary zones on SDN i and f_{max} the local maximum value of $f(l)$ show in Fig. 9. We normalize the weighted origin composition index for each SDN using the maximum–minimum normalization method:

$$\xi_i^{norm}(t) = \frac{\xi_i(t) - \xi_i^{min}}{\xi_i^{max} - \xi_i^{min}}, \tag{13}$$

where ξ_i^{max} and ξ_i^{min} denote the maximum and minimum values of $\xi_i(t)$ in the day. We obtained higher correlations between the calibrated index $\xi_i(t)$ and coupling strength $\lambda_i(t)$ for most SDNs (Fig. 10). Table 1 summarizes the average PCC values between coupling strength and the studied indexes and the R^2 values of linear fits between coupling strength and the indexes. In general, the proposed index $\xi_i(t)$ improves the correlation between coupling index $\lambda_i(t)$ and origin composition index $\xi_i(t)$.

5 Conclusions and discussion

In summary, we found that the coupling between the studied bus network and subway network is mainly generated by the long-distance trips originating from ordinary zones. Multimode trips are favored from an ordinary zone to a transfer zone when trip distances are long. Counterintuitively, the coupling strength of the studied subway network and bus

network is weakly correlated with the speed ratio of the two networks. This can be caused by that passengers do not have a full view of global traffic states. This can be also caused by that passengers select the routes according to personal favor or local traffic conditions.

The present study is limited by the lack of data in diverse cities of different sizes and different transportation structures. However, we believe the results will have a lasting impact in the research areas of multiplex networks and urban transportation. We expect our work will attract the interest of researchers from related research communities. Similar analyses can be done for other cities to generalize the findings and open new research directions. For example, understanding the microcosmic route selection behavior of passengers can help understand why passengers use buses or multimode transport between two transfer zones with subway connection; investigating the land use types of destination zones can help further understand different temporal features of travel demand to specific destination areas.

Nowadays, many operational practices, such as the synchronization of public transport schedules, are implemented to improve the connections between transportation networks and attract more people to public transportation. However, our results imply that the use of multiplex transportation networks is not determined by the time-varying service conditions but by the distribution of mobility demand and transportation facilities. In other words, most passengers select two networks to complete their trips, because they have no direct transportation services to their destinations. Therefore, the coupling strength of two network layers will increase when the mobility demands of passengers from long-distant ordinary zones increase. These mobility demands are actually caused by the lack of necessary transportation facilities. Our finding highlights the need for a better understanding of passenger traffic demand, and more attention needs to be paid to the planning of multiplex transportation networks rather than to the operational aspects of current practices.

Additional material

Additional file 1: Supplementary information for: Understanding coupling dynamics of public transportation networks. (PDF 821 kB)

Acknowledgements

Not applicable.

Funding

National Natural Science Foundation of China (No. 61473320). Project of Innovation-driven Plan in Central South University (No. 2016CSX014). National Key Research and Development Program of China (No. 2016YFB1200400). Hunan Science and Technology Plan Project (No. 2015RS4011).

List of abbreviations

GIS, Geographic Information System; CDR, Call Detail Records; GPS, Global Positioning System; SDN, Single-destination Network; PCC, Pearson Correlation Coefficient.

Availability of data and materials

The data used in this article will not be shared because of the confidentiality agreement with the data provider.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ZZ, ZH, FZ analyzed the data. ZZ, PW wrote the manuscript. PW designed the research. All authors read and approved the final manuscript.

Author details

¹School of Traffic and Transportation Engineering, Central South University, Changsha, P.R. China. ²Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, P.R. China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 December 2017 Accepted: 26 June 2018 Published online: 11 July 2018

References

- Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. *Proc Natl Acad Sci USA* 107:13636–13641. <https://doi.org/10.1073/pnas.1004008107>
- Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:312. <https://doi.org/10.1038/nrn2618>
- Rosato V, Issacharoff L, Tiraticco F et al (2008) Modelling interdependent infrastructures using interacting dynamical models. *Int J Crit Infrastructures* 4:63. <https://doi.org/10.1504/IJICIS.2008.016092>
- Kurant M, Thiran P (2006) Layered complex networks. *Phys Rev Lett* 96:138701. <https://doi.org/10.1103/PhysRevLett.96.138701>
- Zou SR, Zhou T, Liu AF et al (2010) Topological relation of layered complex networks. *Phys Lett A* 374:4406–4410. <https://doi.org/10.1016/j.physleta.2010.08.073>
- Radicchi F, Arenas A (2013) Abrupt transition in the structural formation of interconnected networks. *Nat Phys* 9:717–720. <https://doi.org/10.1038/nphys2761>
- Buldyrev SV, Parshani R, Paul G et al (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464:1025–1028. <https://doi.org/10.1038/nature08932>
- Parshani R, Buldyrev SV, Havlin S (2011) Critical effect of dependency groups on the function of networks. *Proc Natl Acad Sci USA* 108:1007–1010. <https://doi.org/10.1073/pnas.1008404108>
- Gao J, Buldyrev SV, Stanley HE, Havlin S (2011) Networks formed from interdependent networks. *Nat Phys* 8:40–48. <https://doi.org/10.1038/nphys2180>
- Lee KM, Kim JY, Cho WK et al (2012) Correlated multiplexity and connectivity of multiplex random networks. *New J Phys* 14:033027. <https://doi.org/10.1088/1367-2630/14/3/033027>
- Brummitt CD, D'Souza RM, Leicht EA (2012) Suppressing cascades of load in interdependent networks. *Proc Natl Acad Sci USA* 109:E680–E689. <https://doi.org/10.1073/pnas.1110586109>
- Brummitt CD, Lee KM, Goh KI (2012) Multiplexity-facilitated cascades in networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 85:045102. <https://doi.org/10.1103/PhysRevE.85.045102>
- Ruan Z, Tang M, Liu Z (2012) Epidemic spreading with information-driven vaccination. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 86:036117. <https://doi.org/10.1103/PhysRevE.86.036117>
- Ruan Z, Hui P, Lin H, Liu Z (2013) Risks of an epidemic in a two-layered railway-local area traveling network. *Eur Phys J B* 86:13. <https://doi.org/10.1140/epjb/e2012-30292-x>
- Granel C, Gómez S, Arenas A (2013) Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys Rev Lett* 111:128701. <https://doi.org/10.1103/PhysRevLett.111.128701>
- Funk S, Jansen VAA (2010) Interacting epidemics on overlay networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 81:036118. <https://doi.org/10.1103/PhysRevE.81.036118>
- Saumell-Mendiola A, Serrano MA, Boguñá M (2012) Epidemic spreading on interconnected networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 86:026106. <https://doi.org/10.1103/PhysRevE.86.026106>
- Funk S, Gilad E, Watkins C, Jansen VAA (2009) The spread of awareness and its impact on epidemic outbreaks. *Proc Natl Acad Sci USA* 106:6872–6877. <https://doi.org/10.1073/pnas.0810762106>
- Dickison M, Havlin S, Stanley HE (2012) Epidemics on interconnected networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 85:066109. <https://doi.org/10.1103/PhysRevE.85.066109>
- Wang W, Tang M, Yang H et al (2014) Asymmetrically interacting spreading dynamics on complex layered networks. *Sci Rep* 4:5097. <https://doi.org/10.1038/srep05097>
- Morris RG, Barthelemy M (2012) Transport on coupled spatial networks. *Phys Rev Lett* 109:128703. <https://doi.org/10.1103/PhysRevLett.109.128703>
- Gallotti R, Barthelemy M (2015) Anatomy and efficiency of urban multimodal mobility. *Sci Rep* 4:6911. <https://doi.org/10.1038/srep06911>
- Gu CG, Zou SR, Xu XL et al (2011) Onset of cooperation between layered networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 84:026101. <https://doi.org/10.1103/PhysRevE.84.026101>
- Strano E, Shai S, Dobson S, Barthelemy M (2015) Multiplex networks in metropolitan areas: generic features and local effects. *J R Soc Interface* 12:20150651. <https://doi.org/10.1098/rsif.2015.0651>
- Sun L, Liu L, Xu Z et al (2015) Locating inefficient links in a large-scale transportation network. *Physica A* 419:537–545. <https://doi.org/10.1016/j.physa.2014.10.066>
- Noel EC (1988) Park-and-ride: alive, well, and expanding in the United States. *J Urban Plan Dev*. <http://ascelibrary.org.ezproxy.uctn.cl/doi/pdf/10.1061/%28ASCE%290733-9488%281988%29114%3A1%282%29>
- Chodrow PS, Al-Awwad Z, Jiang S, González MC (2016) Demand and congestion in multiplex transportation networks. *PLoS ONE* 11:1–10. <https://doi.org/10.1371/journal.pone.0161738>
- Tan F, Wu J, Xia Y, Tse CK (2014) Traffic congestion in interconnected complex networks. *Phys Rev E, Stat Nonlinear Soft Matter Phys* 89:062813. <https://doi.org/10.1103/PhysRevE.89.062813>
- Du WB, Zhou XL, Chen Z et al (2014) Traffic dynamics on coupled spatial networks. *Chaos Solitons Fractals* 68:72–77. <https://doi.org/10.1016/j.chaos.2014.07.009>
- He K, Xu Z, Wang P et al (2016) Congestion avoidance routing based on large-scale social signals. *IEEE Trans Intell Transp Syst* 17:2613–2626. <https://doi.org/10.1109/TITS.2015.2498186>

31. Wang P, Hunter T, Bayen AM et al (2012) Understanding road usage patterns in urban areas. *Sci Rep* 2:1001. <https://doi.org/10.1038/srep01001>
32. Ma T, Motta G, Liu K (2017) Delivering real-time information services on public transit: a framework. *IEEE Trans Intell Transp Syst* 18:2642–2656. <https://doi.org/10.1109/ITITS.2017.2656387>
33. Wong J (2013) Leveraging the general transit feed specification for efficient transit analysis. *Transp Res Rec* 2338:11–19. <https://doi.org/10.3141/2338-02>
34. Hadas Y (2013) Assessing public transport systems connectivity based on Google transit data. *J Transp Geogr* 33:105–116. <https://doi.org/10.1016/j.jtrangeo.2013.09.015>
35. Tao S, Rohde D, Corcoran J (2014) Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. *J Transp Geogr* 41:21–36. <https://doi.org/10.1016/j.jtrangeo.2014.08.006>
36. Lee SG, Hickman M (2013) Are transit trips symmetrical in time and space? *Transp Res Rec* 2382:173–180. <https://doi.org/10.3141/2382-19>
37. Huang Z, Wang P, Zhang F, Gao J, Schich M (2018) A mobility network approach to identify and anticipate large crowd gatherings. *Transp Res, Part B, Methodol* 114:147–170. <https://doi.org/10.1016/j.trb.2018.05.016>
38. Barry JJ, Newhouser R, Rahbee A, Sayeda S (2007) Origin and destination estimation in New York city with automated fare system data. *Transp Res Rec* 1817:183–187. <https://doi.org/10.3141/1817-24>
39. Trépanier M, Tranchant N, Chapleau R (2007) Individual trip destination estimation in a transit smart card automated fare collection system. *J Intell Transp Syst* 11:1–14. <https://doi.org/10.1080/15472450601122256>
40. Munizaga MA, Palma C (2012) Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transp Res, Part C, Emerg Technol* 24:9–18. <https://doi.org/10.1016/j.trc.2012.01.007>
41. Munizaga M, Devillaine F, Navarrete C, Silva D (2014) Validating travel behavior estimated from smartcard data. *Transp Res, Part C, Emerg Technol* 44:70–79. <https://doi.org/10.1016/j.trc.2014.03.008>
42. Bagchi M, White PR (2005) The potential of public transport smart card data. *Transp Policy* 12:464–474. <https://doi.org/10.1016/j.tranpol.2005.06.008>
43. Bagchi M, White P (2004) What role for smart-card data from bus systems? *Proc ICE Munic Eng* 157:39–46. <https://doi.org/10.1680/muen.157.1.39.36598>
44. Hofmann M, O'Mahony M (2005) Transfer journey identification and analyses from electronic fare collection data. In: *IEEE conf intell transp syst proceedings, ITSC*, pp 825–830. <https://doi.org/10.1109/ITSC.2005.1520156>
45. Seaborn C, Attanucci J, Wilson NHM (2009) Using smart card fare payment data to analyze multi-modal public transport journeys in London. *Transp Res Rec* 2121(1):55–62. <https://doi.org/10.3141/2121-06>
46. Jang W (2010) Travel time and transfer analysis using transit smart card data. *Transp Res Rec* 2144:142–149. <https://doi.org/10.3141/2144-16>
47. Devillaine F, Munizaga M, Trépanier M (2012) Detection of activities of public transport users by analyzing smart card data. *Transp Res Rec* 2276:48–55. <https://doi.org/10.3141/2276-06>
48. Ma X, Wu YJ, Wang Y et al (2013) Mining smart card data for transit riders' travel patterns. *Transp Res, Part C, Emerg Technol* 36:1–12. <https://doi.org/10.1016/j.trc.2013.07.010>
49. Jansen LN, Pedersen MB, Nielsen OA (2002) Minimizing passenger transfer times in public transport timetables. In: *Proc seventh conf Hong Kong soc transp stud*, pp 229–239
50. Guo Z, Wilson NHM (2011) Assessing the cost of transfer inconvenience in public transport systems: a case study of the London underground. *Transp Res, Part A, Policy Pract* 45:91–104. <https://doi.org/10.1016/j.tra.2010.11.002>

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
