

Article

Exploring the Spatiotemporal Patterns of Passenger Flows in Expanding Urban Metros: A Case Study of Shenzhen

Sirui Lv¹, Hu Yang¹, Xin Lu², Fan Zhang³ and Pu Wang^{1,*}

¹ Rail Data Research and Application Key Laboratory of Hunan Province, School of Traffic and Transportation Engineering, Central South University, Changsha 410000, China; lvsirui@csu.edu.cn (S.L.); yanghu@csu.edu.cn (H.Y.)

² College of Systems Engineering, National University of Defense Technology, Changsha 410000, China; lvxin@nudt.edu.cn

³ Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; zhangfan@siat.ac.cn

* Correspondence: wangpu@csu.edu.cn

Abstract: Despite extensive investigations on urban metro passenger flows, their evolving spatiotemporal patterns with the extensions of urban metro networks have not been well understood. Using Shenzhen as a case study city, our study initiates an investigation into this matter by analyzing the evolving network topology of Shenzhen Metro. Subsequently, leveraging long-term smart card data, we analyze the evolving spatiotemporal patterns of passenger flows and develop an analytical approach to pinpoint the major passenger sources of urban metro congestion. While the passenger travel demand and the passenger flow volumes kept increasing with the extension of the urban metro network, the major passenger sources were very stable in space, highlighting the inherent invariance in the evolution of the urban metro system. Finally, we analyze the impact of population and land use factors on passenger flow contributions of passenger sources, obtaining useful clues to foresee future passenger flow conditions.

Keywords: spatiotemporal patterns; expanding urban metro; passenger travel demand; passenger sources; power-law distribution



Citation: Lv, S.; Yang, H.; Lu, X.; Zhang, F.; Wang, P. Exploring the Spatiotemporal Patterns of Passenger Flows in Expanding Urban Metros: A Case Study of Shenzhen. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 267. <https://doi.org/10.3390/ijgi13080267>

Academic Editors: Wolfgang Kainz and Godwin Yeboah

Received: 26 May 2024

Revised: 12 July 2024

Accepted: 22 July 2024

Published: 26 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Characterized by the positive features of high speed, a large capacity, punctuality, and low pollution, the urban metro is widely regarded as a green and efficient transportation mode [1,2] and favored by commuters in big cities [3]. With the rapid growth of passenger travel demand, however, the volumes of passenger flow often exceed the capacities of urban metro sections [4] (e.g., the train load rates of some sections in the Beijing Subway often exceeded 120% [5]). Excessive passenger congestion could cause prolonged train dwell times and an increased likelihood of stranded passengers. This situation, in turn, might result in crowding on platforms [6], accidents such as passengers being caught in screen doors [7], and declines in operational efficiency and service levels [8]. To avoid excessive passenger congestion, many passenger flow control [9–11], route guidance [12–14], and train timetable optimization approaches [15–17] have been proposed and implemented in practices [9,18].

Understanding the spatiotemporal patterns of urban metro passenger flows can facilitate the development of effective congestion mitigation strategies. Some previous studies focused on the analysis of passenger congestion in stations. Yu et al. estimated the passenger flow volume of each station and identified the congested stations of Nanjing Metro [19]. Lu et al. and Peng et al. studied how passenger congestion is formed in stations [20,21], discovering that critical facilities (e.g., security checks, automatic ticket gates) may become overloaded under situations of large passenger flow, leading to passenger congestion in

stations. Moreover, some researchers used video data to monitor the congestion in stations. Zhou et al. analyzed the video data collected at the passages, stairs, and platforms of a station in Ningbo Metro and proposed the Pedestrian Crowd Level (GPC) index to quantify the congestion level [22]. In addition, section congestion, which usually refers to the passenger flow volume of a section exceeding its capacity or the number of passengers on a train exceeding its capacity, was also investigated. Kang et al. located the congested sections and analyzed the impact of land use on section congestion [23]. Haywood and Koning used the survey data and smart card data of the Paris Subway to uncover the relationship between train congestion and travel time [24].

Despite urban metro passenger flows being extensively investigated, previous studies have paid little attention on how passenger congestion is formed. To solve this, the concept of passenger source was proposed [25,26], and previous results indicated that passengers using congested sections mostly originate from a few stations (i.e., major passenger sources) [27]. Leveraging on this discovery, some pioneering works improved passenger flow control approaches based on passenger source information, generating more effective and efficient passenger control schemes [13]. However, the evolving spatiotemporal patterns of passenger sources with the extension of urban metro networks are still not understood well. To fill this research gap, we make a comprehensive analysis on the evolving patterns of the topology, the passenger travel demand, and the passenger sources of an expanding urban metro (i.e., Shenzhen Metro). The used geographic information data and smart card data were collected in November 2014, April 2016, August 2016, December 2016, and January 2018, during which Line 11, Line 7, and Line 9 were successively put into operation. This allows us to explore how the spatiotemporal patterns of passenger flows and passenger sources evolve with the changing network topology.

The remaining sections of this paper are organized as follows. Section 2 introduces the data used in the present study. Section 3 details the methods. Section 4 presents the main results. Section 5 discusses the significance of the contributions of this research to current knowledge and practice. Section 6 concludes the results and findings and discusses the research limitations and future directions.

2. Study Area and Data

2.1. Study Area

Shenzhen is a big city in the south of China. As a core city of the Guangdong–Hong Kong–Macao Greater Bay Area, Shenzhen plays an important role in China’s economic and technological developments and has attracted a large number of new residents in recent years. The studied Shenzhen Metro is the backbone of the city’s public transportation network. The Shenzhen Metro has been expanding since its opening in 2004, connecting more residential, commercial, and industrial areas in the city. In this study, we use Shenzhen as the case study area to explore the spatiotemporal patterns of passenger flows and passenger sources in the expanding urban metro (Figure 1).

2.2. The Smart Card Data of Shenzhen Metro

The obtained smart card data cover five time periods: 10 November to 14 November 2014; 18 April to 22 April 2016; 15 August to 19 August 2016; 5 December to 9 December 2016; and 15 January to 19 January 2018.

In the smart card data, each recorded line of data is generated when a passenger swipes their smart card to enter or exit an urban metro station. Each record documents the card ID, the station ID, the station name, the line ID, the time, the tapping type, and the gate ID (Table 1). The card ID is anonymized to protect the privacy of passengers. The tapping type ‘21’ indicates that the passenger enters a station, whereas the tapping type ‘22’ indicates that the passenger exits a station. The time is recorded in the format of ‘year–month–day–hour–minute–second’. For example, ‘20180115221151’ means that the recorded time is 22:11:51, 15 January 2018 (Table 1).

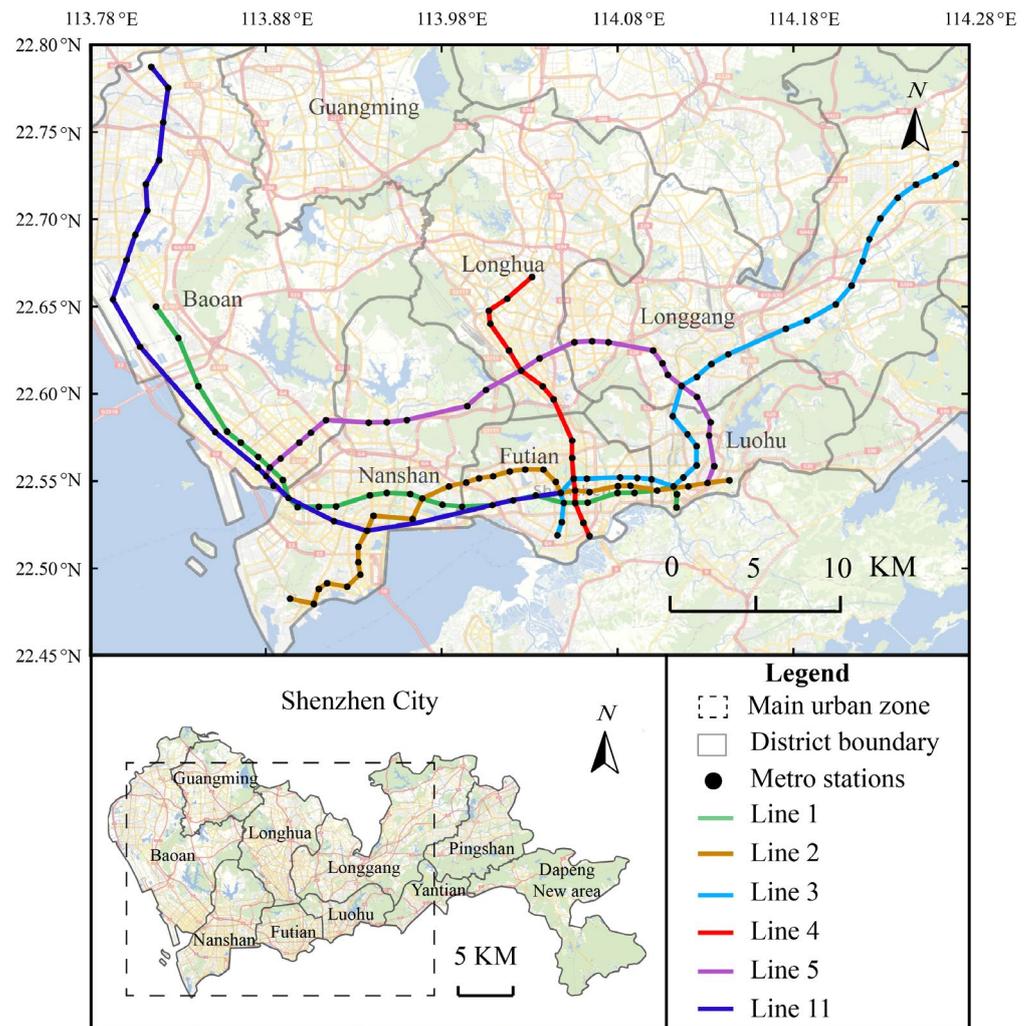


Figure 1. The studied area and the Shenzhen Metro network.

Table 1. Information of the used smart card data.

Card ID	Station ID	Station Name	Line ID	Time	Tapping Type	Gate ID
66****8	26****9	Yijing	Line5	20180115221151	21	IGT-119
68****0	26****1	Gushu	Line1	20180115214907	22	AGT-101
25****6	26****9	Yijing	Line5	20180115220820	21	IGT-119
33****0	26****9	Yijing	Line5	20180115220811	22	IGT-119

The preprocessing of the smart card data involves the following steps: an initial removal of duplicate records followed by filtering out entries beyond the operational period. Subsequently, the records of each passenger are sorted chronologically. Each pair of consecutive entry and exit records is identified as a passenger trip, and a total of 55,113,024 passenger trips are identified. For each of the five data collection periods, the average daily passenger flow of each Origin–Destination (OD) pair is calculated.

2.3. The Operational Data of Shenzhen Metro

The basic information of the studied urban metro network is outlined in Table 2. Using the operational data released at the official website (<https://www.szm.net>, accessed on 6 June 2019) of Shenzhen Metro, we estimate the train load rate of each urban metro section. The capacity of each train on Line 3 is 1440 passengers, whereas the capacity of each train

on other lines is 1860 passengers. As shown in Table 3, the train departure interval is progressively reduced to accommodate the growing passenger travel demand.

Table 2. Basic information of Shenzhen Metro.

Time Period	Lines in Operation	The Number of Stations	The number of Sections
November 2014	1, 2, 3, 4, 5	118	252
April 2016	1, 2, 3, 4, 5	118	252
August 2016	1, 2, 3, 4, 5, 11	132	286
December 2016	1, 2, 3, 4, 5, 7, 9, 11	166	380
January 2018	1, 2, 3, 4, 5, 7, 9, 11	166	380

Table 3. Peak hour train departure intervals (in minutes) during the five studied time periods.

Line	November 2014	April 2016	August 2016	December 2016	January 2018
Line 1	4	3	3	3	3
Line 2	8	6	6	5	5
Line 3	6	6	6	6	6
Line 4	6	6	6	6	6
Line 5	9	6	6	5	5
Line 7	\	\	\	10	5
Line 9	\	\	\	10	5
Line 11	\	\	5	5	5

3. Methods

This study develops a framework to explore the evolving spatiotemporal patterns of passenger flows and passenger sources in an expanding urban metro. First, the passenger flow volume of each urban metro section is estimated. Second, the passenger sources of congested sections are identified. Third, the Gini coefficients are employed to evaluate the heterogeneity of section passenger flows and passenger flow contributions of passenger sources. Finally, the evolving spatiotemporal patterns of passenger flows and passenger sources are analyzed. The details of the proposed analytical framework are shown in Figure 2.

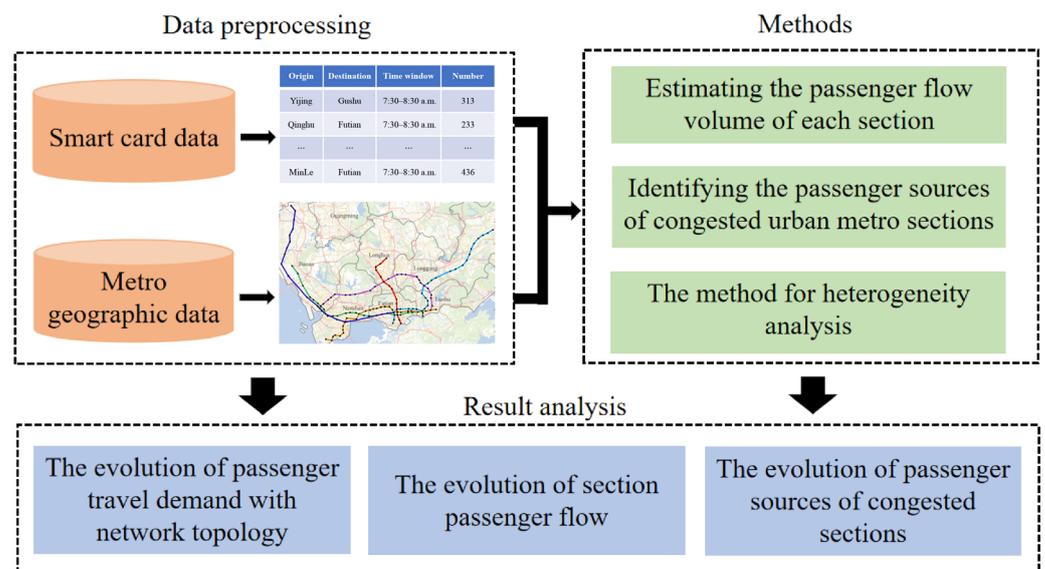


Figure 2. Analytical framework of this study.

3.1. Estimating the Section Passenger Flow Volume

We assume that all passengers use the shortest path measured via travel time, which is calculated using the Dijkstra algorithm [28]. For passengers using a single line, the travel time is the sum of the travel time of each section. For passengers using more than one line, the travel time is composed of the section travel time and the transfer time. The passengers' perceived transfer time is longer than the actual transfer time [29]. Based on the method proposed by Wang et al. [30], the perceived transfer time \bar{T}_r is estimated by multiplying the transfer time T_r with an amplification factor β [31].

$$\bar{T}_r = \beta \cdot T_r \quad (1)$$

Here, we infer a passenger's path using the perceived travel time:

$$\bar{t}(i, j) = \sum_s^n t_s + \sum_r^m \bar{T}_r \quad (2)$$

where $\bar{t}(i, j)$ is the perceived travel time from station i to station j , and t_s is the travel time of section s , n is the number of sections, and m is the number of transfers. After inferring the path of each passenger trip, the passenger travel demand is assigned to the urban metro network, and the passenger flow volume of each section is estimated.

A traversal method is used to determine the value of β . We calculate the root mean square error (RMSE) between the actual average travel time $t(i, j)$ and the estimated travel time $\hat{t}(i, j)$. The parameter β is set to the value generating the minimum RMSE.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_i \sum_j (t(i, j) - \hat{t}(i, j))^2} \quad (3)$$

In Equation (3), N denotes the number of OD pairs, $t(i, j)$ is the average actual travel time from station i to station j , and $\hat{t}(i, j)$ is the estimated travel time from station i to station j :

$$\hat{t}(i, j) = \sum_s^n t_s + \sum_r^m T_r \quad (4)$$

3.2. Identifying the Passenger Sources of Congested Sections

We use the load rate to quantify the level of congestion of an urban metro section. The load rate of an urban metro section k is defined as the ratio of the section passenger flow volume V_k and the section capacity:

$$l_k = V_k / (f_k \cdot C) \quad (5)$$

where f_k is the number of trains passing section k in a given time window, and C is the capacity of a train. The sections with a load rate exceeding 1.0 are identified as the congested sections. Consequently, passenger sources are defined as the boarding stations of the passengers using congested sections.

3.3. Analyzing the Heterogeneity of Passenger Flows

Gini coefficient is widely used to quantify the inequality among the values of a frequency distribution [32]. As shown in Equation (6), we propose a Gini coefficient GN_1 to analyze the heterogeneity of section passenger flows. The value of GN_1 ranges from 0 to 1. A small GN_1 indicates the small differences among the section passenger flow volumes, whereas a large GN_1 implies that the section passenger flow volumes exhibit great variances. Hence, GN_1 evaluates the degree of unevenness of section passenger flows. Similarly, as shown in Equation (7), we propose another Gini coefficient GN_2 to evaluate the heterogeneity of passenger flow contributions of passenger sources. A small GN_2 indicates that the passenger flow contributions of different passenger sources are

similar, whereas a large GN_2 suggests that some passenger sources contribute much more passenger flows to congested sections.

$$GN_1 = \left(\sum_i^m \sum_j^m |s_i - s_j| \right) / \left(2m \sum_{j=1}^m s_j \right) \quad (6)$$

$$GN_2 = \left(\sum_i^y \sum_j^y |p_i - p_j| \right) / \left(2y \sum_{j=1}^y p_j \right) \quad (7)$$

In Equation (6), s_i and s_j represent the passenger flow volumes of sections i and j ; m is the total number of sections. In Equation (7), p_i and p_j represent the passenger flow contributions of passenger sources i and j to congested sections; y is the total number of passenger sources.

4. Results

4.1. The Evolution of Passenger Travel Demand with Network Topology

4.1.1. The Network Topology Changes in the Expanding Urban Metro

The Shenzhen Metro expanded in 2016 with Line 11 opening on 28 June, and Lines 7 and 9 opening on 28 October. The obtained geographic information data cover the time periods before and after the opening dates of the new lines, allowing us to analyze the topology changes in the urban metro network. Using the geographic information data, we generate the networks of the Shenzhen Metro during different time periods (Figure 3). In the generated urban metro networks, nodes represent urban metro stations, and links represent urban metro sections. The degree of a node k_i equals to the number of links connecting to it, and the average degree of the network is $\langle k \rangle = \sum_i k_i$. The path length l_{ij} from node i to node j is defined as the least number of links connecting i and j . Consequently, the average path length of the network is $\langle l \rangle = \sum_{ij} l_{ij}$.

Figure 4 shows the network topology changes from 2014 to 2018. First, the number of nodes (stations) increases from 118 to 166, which greatly extends the service area of the Shenzhen Metro (Figure 4a). Accordingly, the number of links (sections) increases from 252 to 380, and the total length of the sections increases from 178 km to 286 km (Figure 4b,c). Second, different lines are more connected with each other with the expanding of the urban metro; the number of transfer stations increases from 13 to 28 (Figure 4d), leading to an increase in average degree $\langle k \rangle$ (Figure 4e). Third, the average path length of the urban metro network decreases, which implies that the transport efficiency is improved from the topological point of view (Figure 4f). In addition, taking the decreasing train departure interval into account (see Table 3), we can arrive at the conclusion that the Shenzhen Metro has evolved with more service areas and higher transport efficiency.

4.1.2. The Evolution of Passenger Travel Demand

As expected, the passenger travel demand keeps growing throughout the four observed years with the expanding of the urban metro network (Figure 5a). However, the average trip distance of passengers first increases and then decreases, reaching the highest value in August 2016 when Line 11 was just put into operation (Figure 5b). The prominent increase in average trip distance from April 2016 to August 2016 could be attributed to the opening of Line 11, which attracted the passengers who lived in the northwestern area of Shenzhen and generated many long-distance trips. Interestingly, the average trip distance has an obvious decrease after Line 7 and Line 9 were put into operation. An explanation is that a large part of Line 7 and Line 9 is located at the central urban area, where short trips are more frequent.

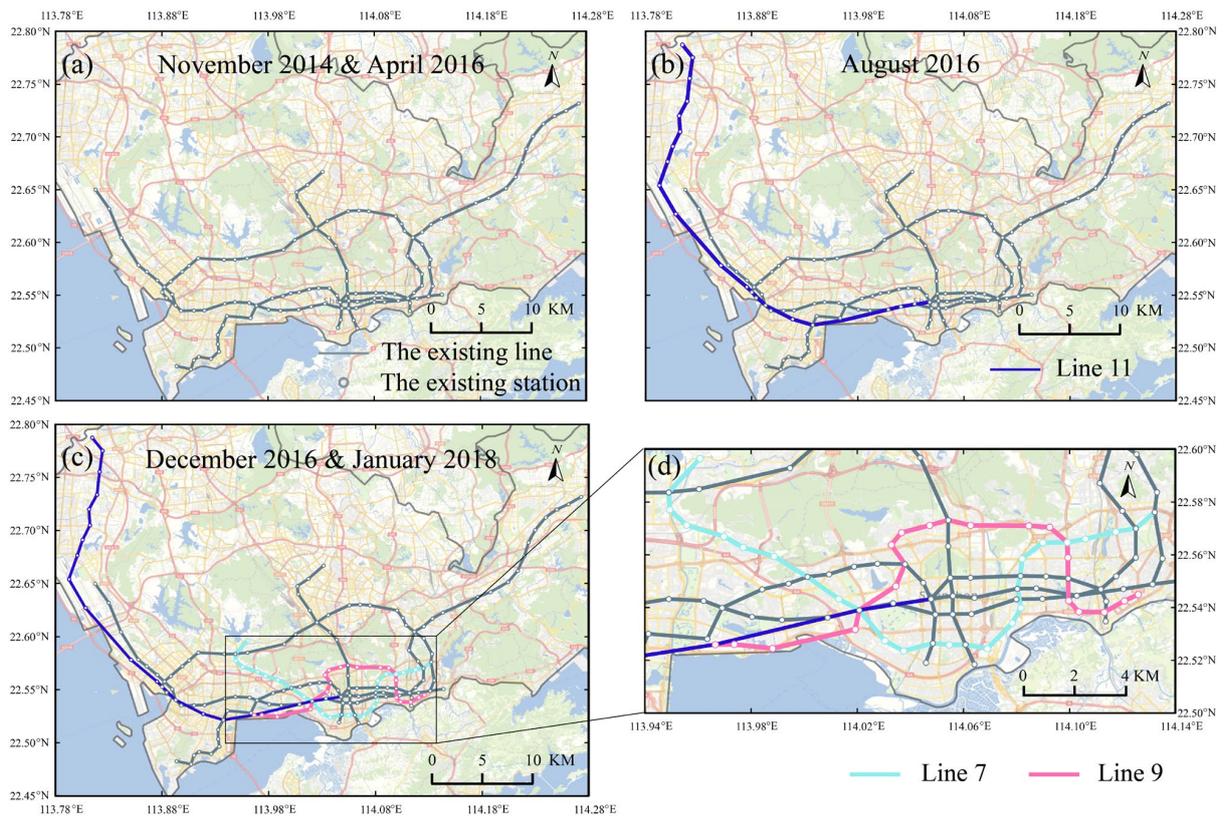


Figure 3. The evolving networks of Shenzhen Metro. (a) The Shenzhen Metro network in November 2014 and April 2016. (b) The Shenzhen Metro network in August 2016. (c) The Shenzhen Metro network in December 2016 and January 2018. (d) Enlarged view of Line 7 and Line 9.

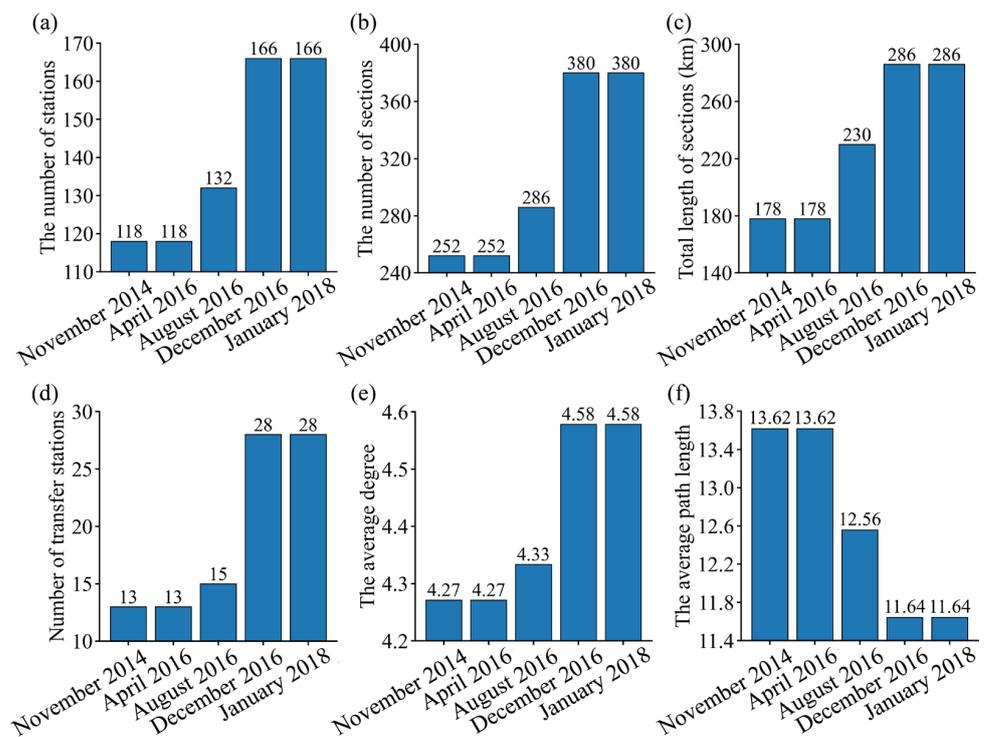


Figure 4. The network topology changes in the expanding urban metro. (a) The number of stations. (b) The number of sections. (c) The total length of the sections. (d) The number of transfer stations. (e) The average degree of the network. (f) The average path length of the network.

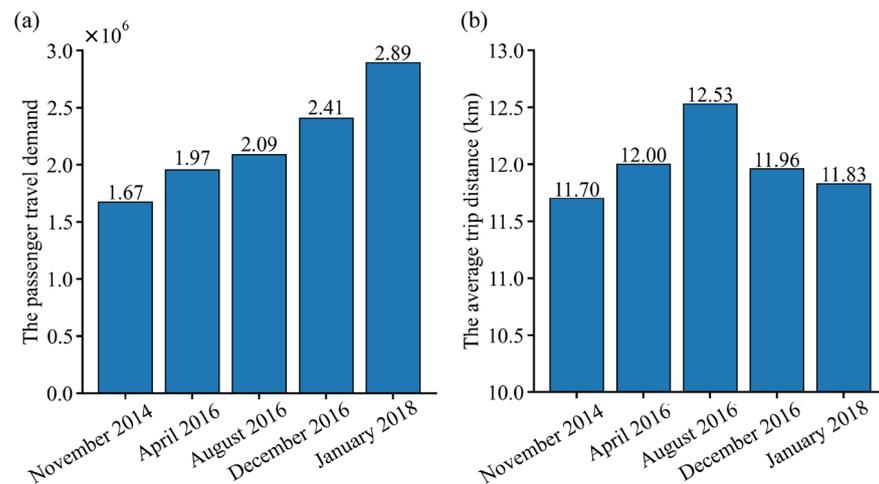


Figure 5. The evolution of passenger travel demand. (a) The passenger travel demand. (b) The average trip distance.

The temporal pattern of passenger travel demand is quite stable across the five studied time periods from 2014 to 2018 (Figure 6). The passenger flow volume exhibits two peaks in the morning and evening. In addition, the hourly passenger flow volume keeps increasing throughout the four years (Figure 6), which is consistent with the growing pattern of passenger travel demand (Figure 5a). In December 2016, when Line 7 and Line 9 were just put into operation, the passenger travel demand had an obvious rise in the evening period. An explanation is that many stations of Line 7 and Line 9 are located at the central urban area of Shenzhen, where more trips are generated in the evening by passengers who work late or engage in recreational activities.

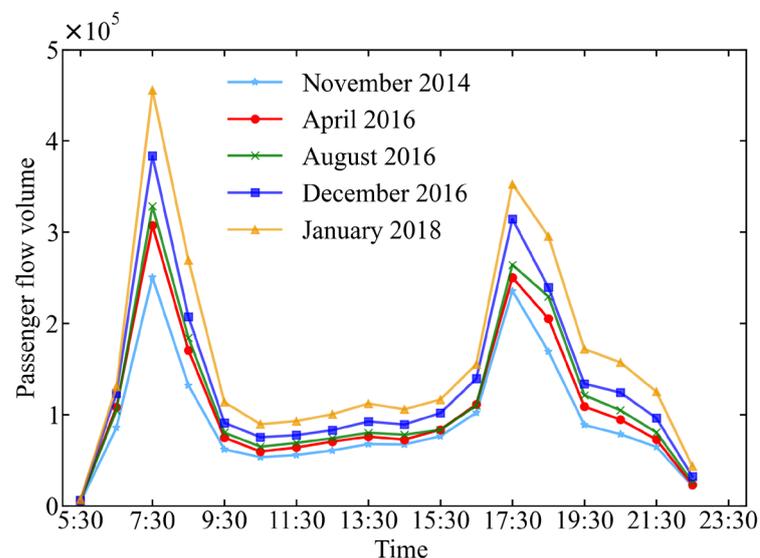


Figure 6. The average hourly passenger flow volumes during the operation period (5:30 a.m.–11:30 p.m.).

The OD passenger travel demand is defined by the daily average number of passengers taking trains from one station (origin) to another station (destination). As shown in Figure 7, the OD passenger travel demand can be well approximated using a power-law distribution across each of the five studied time periods, with minor variations observed in the power-law coefficients. While the majority of OD pairs indicate relatively lower passenger flow volumes, a few specific OD pairs exhibit significantly high passenger travel demand. For instance, in November 2014, the OD passenger travel demand registered 6564 individuals

from Ping'zhou Station to Shenzhen University Station, whereas the average OD passenger travel demand was merely 121. This observation underscores the heterogeneous spatial distribution of passenger travel demand, demonstrating a universal pattern irrespective of the network scale and the operation plan.

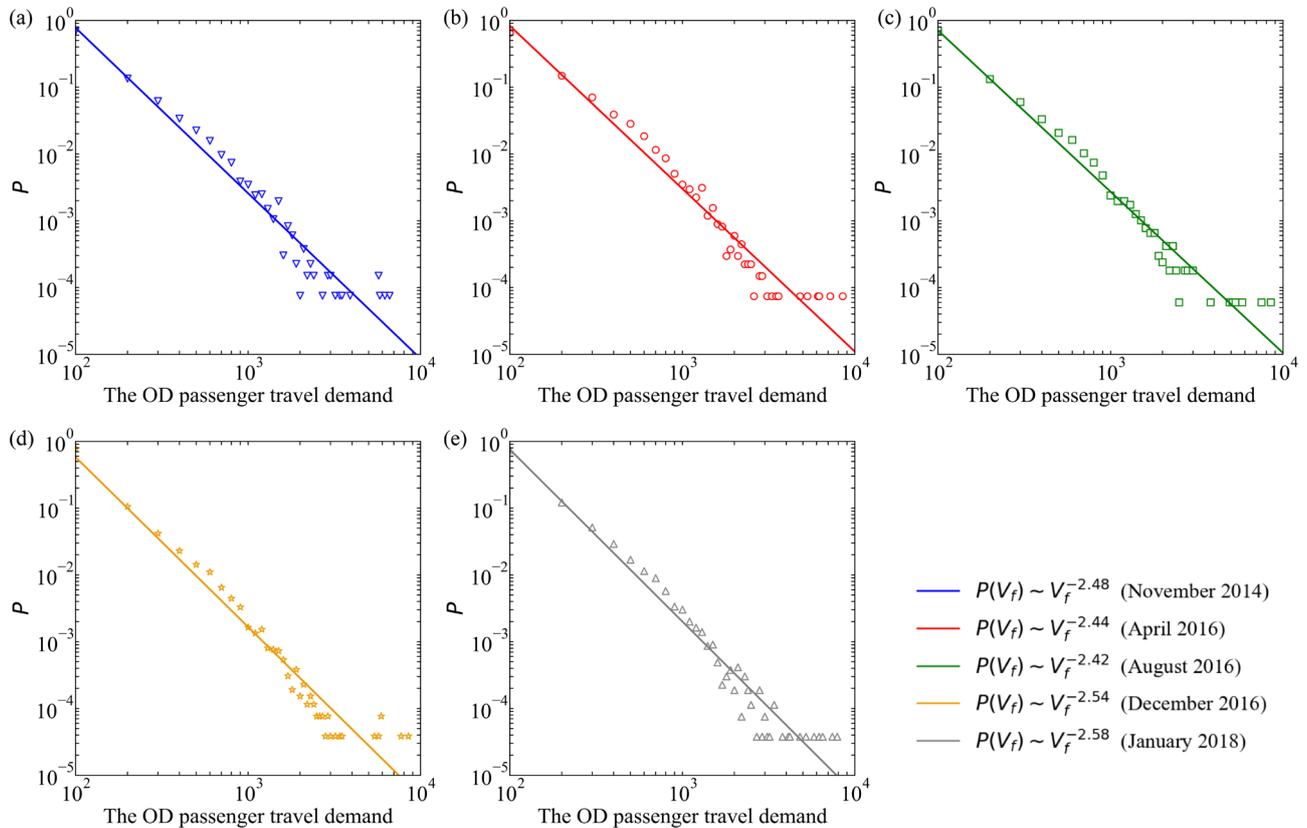


Figure 7. Distribution of OD passenger travel demand. (a) November 2014; (b) April 2016; (c) August 2016; (d) December 2016; (e) January 2018.

4.2. The Evolution of Section Passenger Flow

We use a morning peak hour (7:30–8:30 a.m.) as a case study time window. Using the passenger flow assignment method proposed by Si et al. [29] (see Section 3.2), we estimate the passenger flow volume of each urban metro section. While the passenger flow volumes of the majority of sections are less than 20,000 passengers/h, the passenger flow volumes of a few sections exceed 40,000 passengers/h (Figure 8a). The volumes of section passenger flow can be well approximated using a truncated power-law distribution:

$$P(f) = (f + f^0)^{-\beta} \exp\left(-\frac{f}{f_{cut}}\right) \tag{8}$$

where $\beta = 0.16$, $f^0 = 1801143.35$, and the exponential cutoff $f_{cut} = 13110.15$. The coefficient of determination [33] (R^2) between the data and the truncated power-law function is 0.89, showing a good fitting performance. As the section passenger flow exceeds the exponential cutoff, $P(f)$ exhibits a rapid decline. The heterogeneity of the section passenger flow is fundamentally driven by the heterogeneously distributed passenger travel demand, which naturally generates more passenger flows in some sections but less passenger flows in the others.

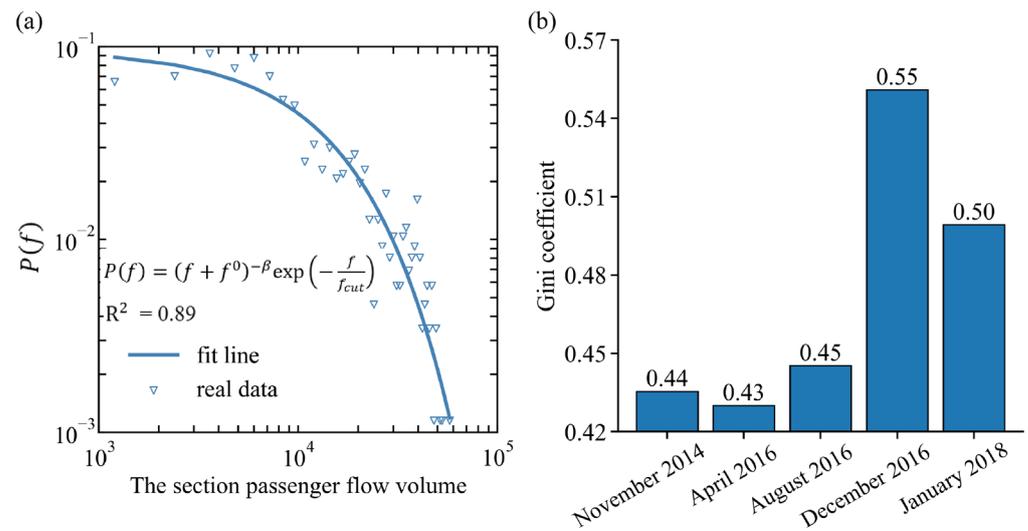


Figure 8. The evolution of section passenger flow. (a) Distribution of section passenger flow. (b) The Gini coefficient of section passenger flow.

The Gini coefficients of the section passenger flows remain at a notably high level throughout the four years of observation (Figure 8b). Additionally, intriguing variations in the Gini coefficient emerge, predominantly influenced by the evolving network topology and passenger travel demand. Notably, a slight decrease in the Gini coefficient occurs from November 2014 to April 2016, indicating a relatively more even spatial distribution of section passenger flows during this interval. However, with the opening of Line 11, there is an observed increase in the Gini coefficient from April 2016 to August 2016. This rise can be attributed to the typically lower initial passenger flow volumes of the new line [34], accentuating the heterogeneity of the section passenger flows. A similar increase in the Gini coefficient is witnessed from August 2016 to December 2016, coinciding with the opening of Line 7 and Line 9. Finally, a subsequent decrease in the Gini coefficient from December 2016 to January 2018 aligns with our expectations. This decrease corresponds to the gradual increase in section passenger flows in the three new lines, contributing to a more balanced distribution of section passenger flows within the urban metro network.

Despite the expanded network size and the enhanced section capacity, the Shenzhen Metro becomes more congested. This is manifested in the increasing number of congested sections from 2014 to 2018 (Figure 9a). The more sufficient use of urban metro infrastructure can increase the profit of the operation company [35,36] and alleviate traffic congestion on roads [37–39]. However, we must note that the increasing passenger congestion downgrades the level of service of the urban metro. As shown in Figure 9b, the number of passengers using congested sections keeps increasing from 2014 to 2018.

To analyze the evolving spatial distribution of section passenger flow, we analyze the change in section passenger flow between two consecutive time periods.

$$fc_i(t) = f_i(t) - f_i(t-1) \quad (9)$$

where $f_i(t)$ is the passenger flow volume of section i in time period t (e.g., April 2016) and $f_i(t-1)$ is the passenger flow volume of section i in time period $t-1$ (e.g., November 2014). For each consecutive time periods $t-1$ and t , we identify the congested sections in time period $t-1$, and then analyze their passenger flow condition in time period t .

First, with the growth of passenger travel demand, the congested sections become more congested from November 2014 to April 2016 (Figure 10a). Subsequently, with the opening of Line 11, the passenger flow volumes of some sections of Line 1 decrease in August 2016, implying that Line 11 can share some of the transportation loads posed on Line 1 (Figure 10b). Yet, the other congested sections become more congested, which is probably caused by the new passengers induced by Line 11 (Figure 10b). With the opening

of Line 7 and Line 9, three congested sections on Line 4 become less congested, which is also a sign of passenger flow diversion (Figure 10c). Finally, as expected, the congested sections become more congested with the growing passenger travel demand from December 2016 to January 2018 (Figure 10d). In summary, the opening of new lines may temporarily alleviate the congestion of some sections, but it will eventually induce more regular passengers to the urban metro.

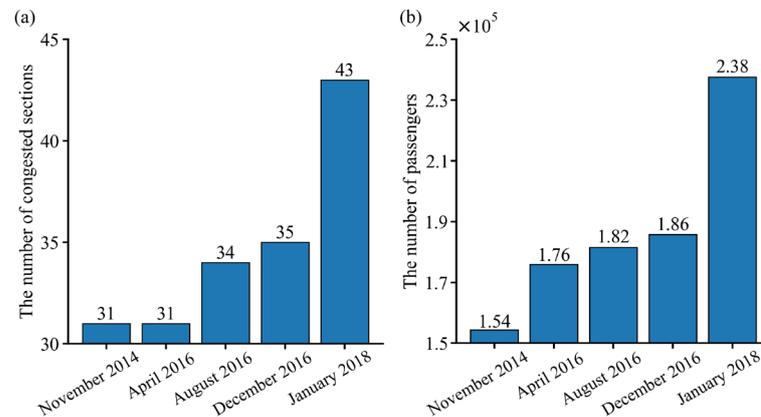


Figure 9. The number of congested sections (a) and the number of passengers using congested sections (b) throughout the five studied time periods.

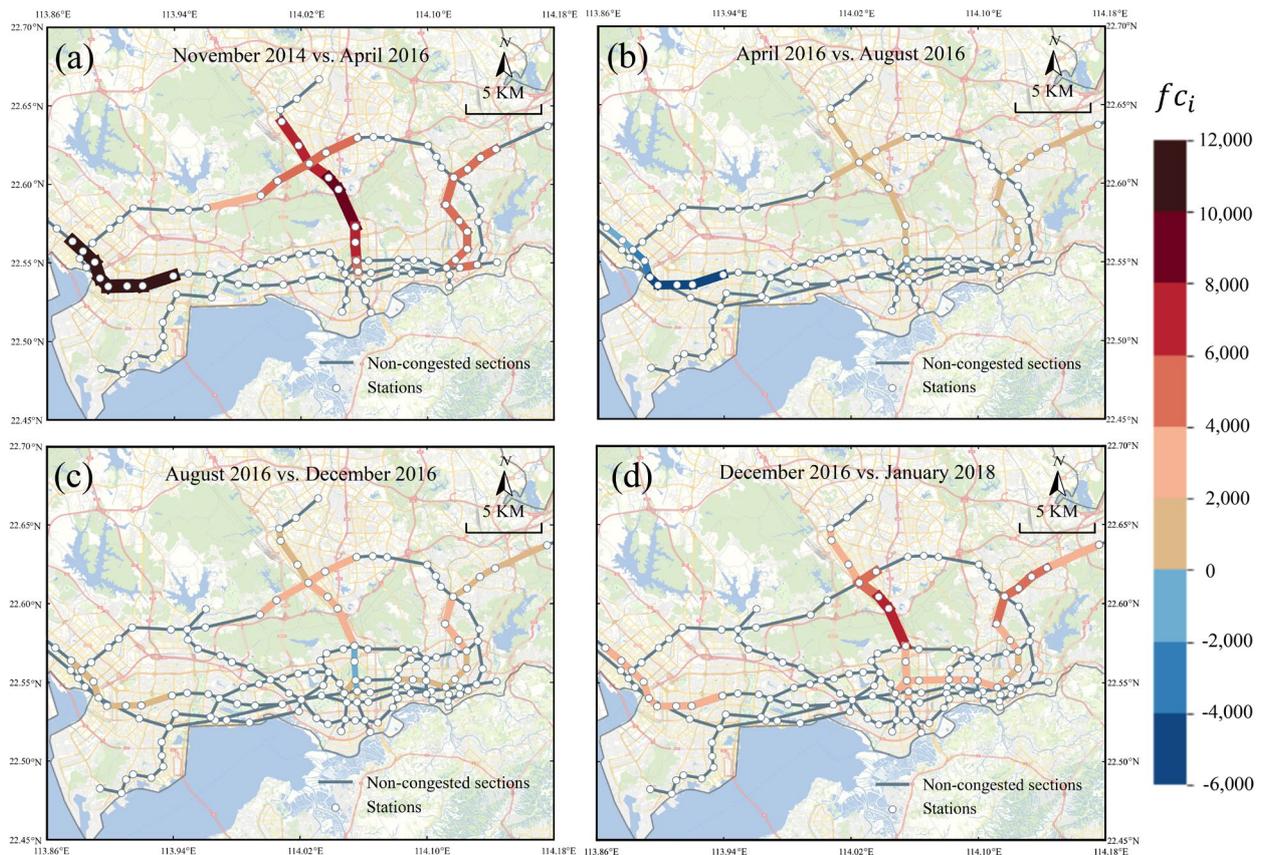


Figure 10. The changes in passenger flow volumes of the congested sections during a morning peak hour (7:30–8:30 a.m.). (a) Comparison between November 2014 and April 2016. (b) Comparison between April 2016 and August 2016. (c) Comparison between August 2016 and December 2016. (d) Comparison between December 2016 and January 2018.

4.3. The Evolution of Passenger Sources of Congested Sections

Passenger congestion may cause various problems such as crowding risks, low service level and low operation efficiency [6,40]. This calls for an in-depth exploration on how passenger congestion is formed. To achieve this, we analyze the passenger flow volume g_i contributed by each passenger source i to the congested sections. As shown in Figure 11a, the passenger flow contributions of passenger sources (G) follow truncated power-law distributions across the five studied time periods:

$$G = (g + g^0)^{-\beta} \exp\left(-\frac{g}{g_{cut}}\right) \tag{10}$$

where the fitted parameters are shown in Table 4. When the ranks of passenger sources exceed the exponential cutoff g_{cut} (ranges from 11 to 14), the passenger flow contributions of the passenger sources exhibit a sharp decline. This indicates that most passenger sources contribute limited passenger flows to congested sections, while a few passenger sources are the key passenger sources causing urban metro congestion.

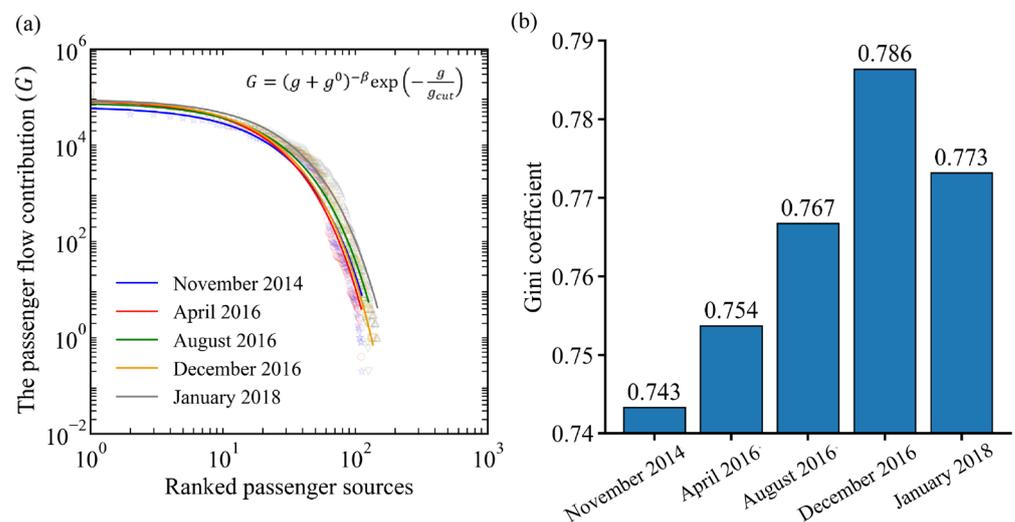


Figure 11. The evolution of passenger flow contributions of passenger sources. (a) Passenger flow contributions of passenger sources across the five studied time periods. The scattered points of different colors represent the actual data, and corresponding colored lines represent the fitted functions. (b) The Gini coefficient of passenger flow contributions.

Table 4. The fitted parameter values across the five studied time periods in Equation (10).

Time Period	β	g^0	g_{cut}	R^2
November 2014	−1.19	10,364.16	12.39	0.92
April 2016	−1.24	10,081.03	11.06	0.92
August 2016	−1.16	16,132.38	13.18	0.98
December 2016	−1.22	11,908.00	11.54	0.95
January 2018	−1.17	17,820.46	14.70	0.99

The passenger flow contributions of passenger sources are heterogeneously distributed, manifesting in the high Gini coefficients throughout the four observational years (Figure 11b). The Gini coefficient increases from November 2014 to December 2016, during which three new lines were put into operation. An explanation is that the passengers living in the vicinities of the major passenger sources may be attracted to visit the regions covered by the new stations. Consequently, more passenger flows are generated at the major passenger sources and the heterogeneity of passenger flow contributions to congested sections is enhanced. This, however, inevitably increases the management pressure at the major

passenger sources. On the other hand, the Gini coefficient decreases from December 2016 to January 2018, during which no new lines were put into operation. An explanation is that the passenger travel demand of the new lines gradually increases during this period, the section passenger flows are more evenly distributed, and the heterogeneity of the passenger flow contributions of passenger sources is reduced.

The passenger sources are sorted in a descending order according to their passenger flow contributions to the congested sections. The top ranked stations that cumulatively contribute 50% of the passenger flows of congested sections are defined as the major passenger sources of the urban metro congestion. Figure 12 shows the spatial distribution of the identified major passenger sources of Shenzhen Metro. During 7:30–9:30 a.m., the major passenger sources are mainly distributed in the northern part of Line 1, Line 3, and Line 4 (Figure 12a,b). An explanation is that many residential districts are located at these regions, where a large number of passengers travel to the central urban area during morning rush hours. These passengers contribute to the majority of the passenger flows of the congested sections. During 5:30–7:30 p.m., the major passenger sources are mainly distributed in the central urban area of Shenzhen (Figure 12c,d), where a lot of passengers return to home during evening rush hours.

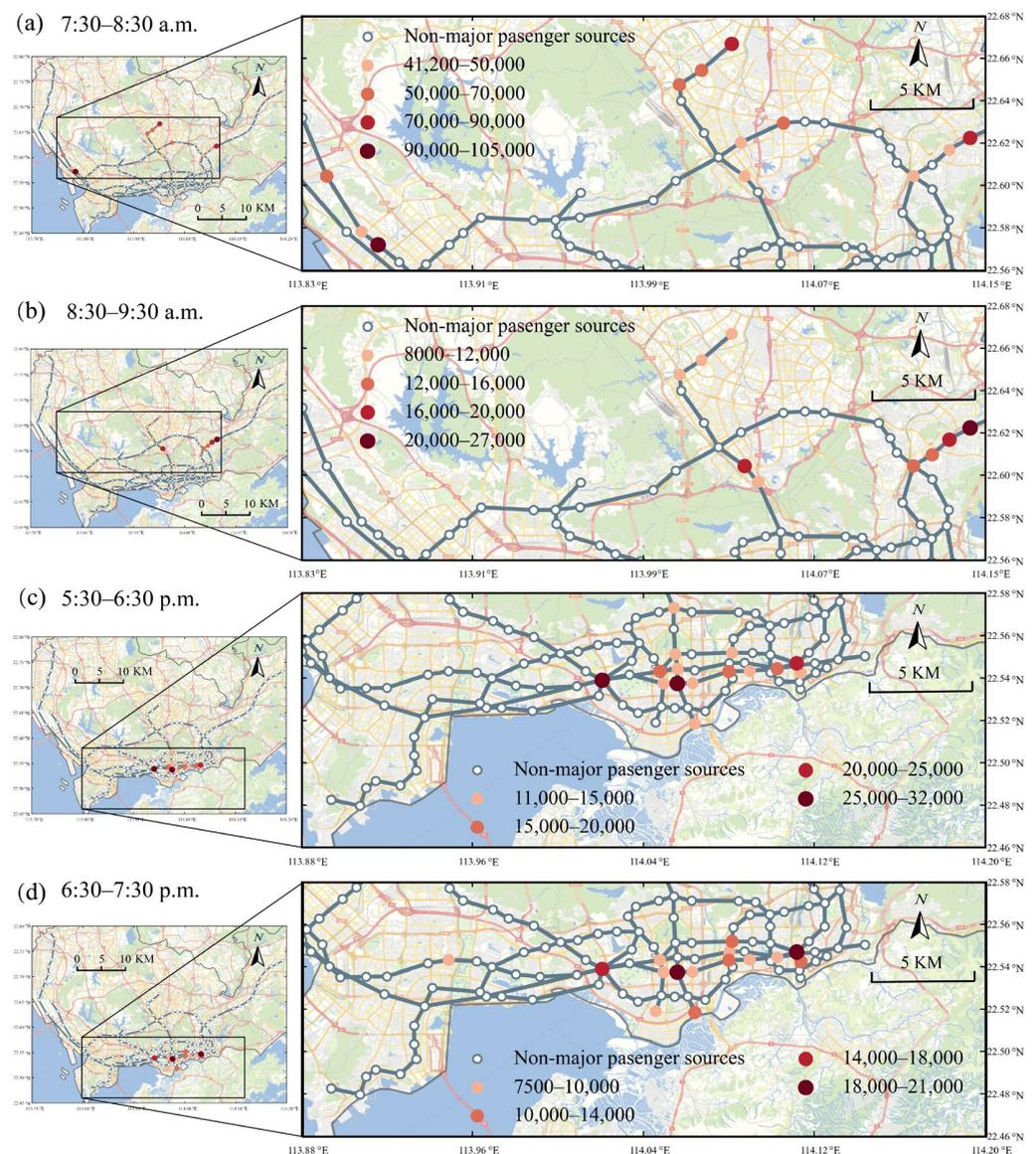


Figure 12. The major passenger sources of Shenzhen Metro in January 2018.

Next, we analyze the impact of population and land use factors on passenger flow contributions of passenger sources. Here, we define the service area of an urban metro station as the 1 km circular region around the station. Subsequently, using the fine-grained WorldPop data provided by the WorldPop Project (<https://www.worldpop.org>, accessed on 12 May 2024), we estimate the population within the service area of each station, which is a factor that potentially influences the volumes of passenger flow. In addition, leveraging the Point of Interest (POI) data provided by the Gaode Map (<https://lbs.amap.com>, accessed on 15 January 2018), we estimate the numbers of different types of POIs within the service area of each station to evaluate the land use properties (Table 5).

Table 5. The coefficient of multiple linear regression model.

Time	Type	Coefficient	<i>p</i>	Type	Coefficient	<i>p</i>
7:30–9:30 a.m.	Catering services	22.158	0.061	Domestic services	−5.270	0.651
	Companies	−12.025	0.014	Residences	38.717	0.032
	Public facilities	−161.492	0	Population	0.026	0.425
5:30–7:30 p.m.	Catering services	5.802	0.157	Domestic services	−13.330	0.001
	Companies	5.132	0.003	Residence	−3.729	0.549
	Public facilities	72.129	0	Population	0.025	0.026

The multiple linear regression (MLR) approach [41] is used to analyze the correlations between the passenger flow contributions of passenger sources and the land use and population factors. Based on the F-statistic and the associated *p*-values (all smaller than 0.05), the generated MLR models are significant. As indicated in Table 5, the population and land use factors have different impacts on the passenger flow contributions of passenger sources during the morning (7:30–9:30 a.m.) and evening (5:30–7:30 p.m.) peak hours. In the morning, the number of residences has a positive impact, while the number of companies and the number of public facilities have negative impacts on the volumes of passenger flow. Conversely, in the evening, the number of companies and the number of public facilities have positive impacts, while the number of domestic services has a negative impact on the volumes of passenger flow. This result highlights that the abundance of a certain type of POIs around a station may enhance the station’s contribution to urban metro congestion at a specific time window. Moreover, the discovery provides a guideline to foresee the passenger flow conditions of the stations to be opened in the future, which could potentially facilitate more reasonable planning and the more efficient management of expanding urban metros.

5. Discussion

The first prominent feature of the present study relies on the analysis of long-term smart card data. There were a number of research works focusing on the analysis of passenger flows in urban metros; however, most of them were based on short-term smart card data. During the observation period, the structure of the studied urban metro network does not change, and the passenger travel demand is quite stable. Hence, most of those studies only analyzed an isolated slice of passenger flow status during a certain time period. The dynamic process of the passenger flow evolving with the extension of the urban metro network was not sufficiently investigated. In this study, we try to fill this research gap by employing long-term smart card data that spans four years, uncovering the evolving spatiotemporal patterns of passenger flows.

Indeed, we obtained some new and exciting discoveries by analyzing the long-term smart card data. We discovered a counter intuitive phenomenon, where building new lines does not alleviate urban metro congestion. This phenomenon could be explained by the growing attractiveness of the expanding urban metro to passengers. In addition, we discovered the inherent invariance in the evolution of the urban metro system. Although the topology, the structure of the urban metro network, the passenger travel demand, and

the number of congested sections kept changing across the four observational years, the spatial distribution of the major passenger sources, which is mainly determined by the land use pattern, was quite stable.

The second prominent feature of this study relies on the proposed passenger source analytical approach, which is applied to uncover the inherent laws driving the formation of urban metro congestion. This reflects an improvement to previous research that mostly focused on passenger flow analysis. Given that the identified major passenger sources are the key contributors of urban metro congestion, the passenger source information could facilitate the improvement of urban metro operation strategies (e.g., short-turning plan, passenger flow control). To better show the important role of major passenger sources in alleviating urban metro congestion, we designed a simple proof-of-concept simulation experiment.

In the designed simulation experiment, the passenger travel demand is reduced to alleviate urban metro congestion, and two strategies are tested. In strategy S_1 , a certain number of passengers from the major passenger sources are randomly selected and controlled to not use the urban metro. In strategy S_2 , a certain number of passengers from all urban metro stations are randomly selected and controlled to not use the urban metro. Taking the congested section “Bao’an Center Station-Xin’an Station” as an example, we estimate the required number of controlled passengers to decrease its passenger flow by a fraction f (ranging from 4% to 20%). As shown in Table 6, much fewer passengers are controlled when using strategy S_1 compared with using strategy S_2 . Obviously, controlling passengers from major passenger sources can alleviate the congestion more effectively. Given that the operator cannot forbid passengers from using the urban metro, we could not directly apply the strategies of this simulation experiment in practice. However, the simulation result highlights the importance of pinpointing major passenger sources. In practice, more sophisticated operation strategies can be developed based on the passenger source information to improve operation efficiency and reduce management cost.

Table 6. Comparison of strategy S_1 and strategy S_2 .

Reduction Rate of Section Passenger Flow in the Case Study Congested Section	The Number of Controlled Passengers Using Strategy S_1	The Number of Controlled Passengers Using Strategy S_2
4%	2074	15,494
8%	3567	18,774
12%	5055	24,837
16%	6543	32,277
20%	8241	34,944

6. Conclusions

Exploring the spatiotemporal patterns of passenger flows and passenger sources in expanding urban metros is essential for the sustainable development of urban transportation. With the opening of new lines, the Shenzhen Metro has evolved to have more service areas and a higher transportation efficiency. However, the passenger travel demand grows at a faster speed, and the urban metro becomes more congested throughout the four observational years. Despite that the opening of new lines may temporarily alleviate the congestion of some urban metro sections, it will eventually induce more passenger travel demand. To this end, we propose a passenger source analytical approach to explore the inherent laws driving the formation of urban metro congestion. Interestingly, only a few key passenger sources contribute to the major urban metro congestion, and the simulation experiment suggests that developing operation strategies targeting these key passenger sources may achieve a better congestion mitigation effect. Finally, we explore the impact of land use and population factors on the passenger flow contributions of passenger sources. The results can help in anticipating the passenger flow condition of the stations to be opened in the future and facilitate the more reasonable planning and more efficient management of the expanding urban metros.

Due to the limitation of the obtained data, we only investigate the evolving network topology and congestion patterns of the urban metro. The interplay between the urban metro and other transportation systems has not been uncovered. When more transportation data are available, a similar approach can be used to explore the evolution patterns of a multi-layer urban transportation system. Moreover, the proposed approach can be applied in joint operations of buses and urban metro trains. For example, transportation authorities could appropriately increase the number of buses passing through major passenger sources to alleviate urban metro congestion. Finally, it would be interesting to develop powerful machine learning models to predict the passenger flow contributions of passenger sources in the future, based on the land use and population distribution patterns.

Author Contributions: Conceptualization, Pu Wang; data curation, Hu Yang and Fan Zhang; formal analysis, Sirui Lv; funding acquisition, Pu Wang; investigation, Sirui Lv; methodology, Sirui Lv; project administration, Pu Wang; supervision, Xin Lu and Fan Zhang; visualization, Sirui Lv; writing—original draft, Sirui Lv and Hu Yang; writing—review and editing, Xin Lu and Pu Wang. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Hunan Provincial Natural Science Fund for Distinguished Young Scholars, grant number 2022JJ10077; the National Natural Science Foundation of China, grant number 71871224; and the 2021 Science and Technology Progress and Innovation Plan of Department of Transportation of Hunan Province, grant number 202102.

Data Availability Statement: The data that support the findings of this study are available on request from the first author (lvsirui@csu.edu.cn) or corresponding author (wangpu@csu.edu.cn).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wu, T.; Li, M.; Zhou, Y. Measuring Metro Accessibility: An Exploratory Study of Wuhan Based on Multi-Source Urban Data. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 18. [\[CrossRef\]](#)
2. Dou, M.; Wang, Y.; Dong, S. Integrating Network Centrality and Node-Place Model to Evaluate and Classify Station Areas in Shanghai. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 414. [\[CrossRef\]](#)
3. Zeng, J.; Tang, J. Combining Knowledge Graph into Metro Passenger Flow Prediction: A Split-Attention Relational Graph Convolutional Network. *Expert Syst. Appl.* **2023**, *213*, 118790. [\[CrossRef\]](#)
4. Ding, X.; Guan, S.; Sun, D.J.; Jia, L. Short Turning Pattern for Relieving Metro Congestion during Peak Hours: The Substance Coherence of Shanghai, China. *Eur. Transp. Res. Rev.* **2018**, *10*, 28. [\[CrossRef\]](#)
5. Jiang, M.; Li, H.; Xu, X.; Xu, S.; Miao, J. Metro Passenger Flow Control with Station-to-Station Cooperation Based on Stop-Skipping and Boarding Limiting. *J. Cent. South Univ.* **2017**, *24*, 236–244. [\[CrossRef\]](#)
6. Shi, J.; Yang, L.; Yang, J.; Zhou, F.; Gao, Z. Cooperative Passenger Flow Control in an Oversaturated Metro Network with Operational Risk Thresholds. *Transp. Res. Part C-Emerg. Technol.* **2019**, *107*, 301–336. [\[CrossRef\]](#)
7. Yuan, F.; Sun, H.; Kang, L.; Wu, J. Passenger Flow Control Strategies for Urban Rail Transit Networks. *Appl. Math. Model.* **2020**, *82*, 168–188. [\[CrossRef\]](#)
8. Azadpeyma, A.; Kashi, E. Level of Service Analysis for Metro Station with Transit Cooperative Research Program (TCRP) Manual: A Case Study—Shohada Station in Iran. *Urban Rail Transit* **2019**, *5*, 39–47. [\[CrossRef\]](#)
9. Huan, N.; Yao, E.; Zhang, J. Demand-Responsive Passenger Flow Control Strategies for Metro Networks Considering Service Fairness and Passengers' Behavioural Responses. *Transp. Res. Part C-Emerg. Technol.* **2021**, *131*, 103335. [\[CrossRef\]](#)
10. Zhang, P.; Sun, H.; Qu, Y.; Yin, H.; Jin, J.G.; Wu, J. Model and Algorithm of Coordinated Flow Controlling with Station-Based Constraints in a Metro System. *Transp. Res. Part e-Logist. Transp. Rev.* **2021**, *148*, 102274. [\[CrossRef\]](#)
11. Wang, X.; Wu, J.; Yang, X.; Guo, X.; Yin, H.; Sun, H. Multistation Coordinated and Dynamic Passenger Inflow Control for a Metro Line. *IET Intell. Transp. Syst.* **2020**, *14*, 1068–1078. [\[CrossRef\]](#)
12. Ding, H.; Di, Y.; Zheng, X.; Liu, K.; Zhang, W.; Zheng, L. Passenger Arrival Distribution Model and Riding Guidance on an Urban Rail Transit Platform. *Phys. A* **2021**, *571*, 125847. [\[CrossRef\]](#)
13. He, K.; Xu, Z.; Wang, P.; Deng, L.; Tu, L. Congestion Avoidance Routing Based on Large-Scale Social Signals. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2613–2626. [\[CrossRef\]](#)
14. Lei, Y.; Lu, G.; Zhang, H.; He, B.; Fang, J. Optimizing Total Passenger Waiting Time in an Urban Rail Network: A Passenger Flow Guidance Strategy Based on a Multi-Agent Simulation Approach. *Simul. Model. Pract. Theory* **2022**, *117*, 102510. [\[CrossRef\]](#)
15. Xue, H.; Jia, L.; Li, J.; Guo, J. Jointly Optimized Demand-Oriented Train Timetable and Passenger Flow Control Strategy for a Congested Subway Line under a Short-Turning Operation Pattern. *Phys. A* **2022**, *593*, 106957. [\[CrossRef\]](#)

16. Wang, Y.; Tang, T.; Ning, B.; van den Boom, T.J.J.; De Schutter, B. Passenger-Demands-Oriented Train Scheduling for an Urban Rail Transit Network. *Transp. Res. Part C-Emerg. Technol.* **2015**, *60*, 1–23. [[CrossRef](#)]
17. Liu, R.; Li, S.; Yang, L. Collaborative Optimization for Metro Train Scheduling and Train Connections Combined with Passenger Flow Control Strategy. *Omega-Int. J. Manag. Sci.* **2020**, *90*, 101990. [[CrossRef](#)]
18. Xu, X.; Li, H.; Liu, J.; Ran, B.; Qin, L. Passenger Flow Control with Multi-Station Coordination in Subway Networks: Algorithm Development and Real-World Case Study. *Transp. B-Transp. Dyn.* **2019**, *7*, 446–472. [[CrossRef](#)]
19. Yu, W.; Bai, H.; Chen, J.; Yan, X. Analysis of Space-Time Variation of Passenger Flow and Commuting Characteristics of Residents Using Smart Card Data of Nanjing Metro. *Sustainability* **2019**, *11*, 4989. [[CrossRef](#)]
20. Lu, J.; Ren, G.; Xu, L. Analysis of Subway Station Distribution Capacity Based on Automatic Fare Collection Data of Nanjing Metro. *J. Transp. Eng. Part A-Syst.* **2020**, *146*, 04019067. [[CrossRef](#)]
21. Peng, J.; Wei, Z.; Yang, Y.; Wang, W.; Qiu, S.; Wang, S. What Size of Aisle Is Necessary? A System Dynamics Model for Mitigating Bottleneck Congestion in Entrance Halls of Metro Stations. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22923–22936. [[CrossRef](#)]
22. Zhou, J.; Wu, Y.; Mao, X.; Guo, S.; Zhang, M. Congestion Evaluation of Pedestrians in Metro Stations Based on Normal-Cloud Theory. *Appl. Sci.* **2019**, *9*, 3624. [[CrossRef](#)]
23. Kang Graduate, Y.; Yi, C.; Lee, S. Correlation between Characteristic of Subway Network and Congestion in Subway Train. In Proceedings of the International Symposium on City Planning (ISCP 2013), Sendai City, Miyagi Prefecture, Japan, 22–24 August 2013.
24. Haywood, L.; Koning, M. The Distribution of Crowding Costs in Public Transport: New Evidence from Paris. *Transp. Res. Part A-Policy Pract.* **2015**, *77*, 182–201. [[CrossRef](#)]
25. Lv, S.; Yang, H.; Zhang, F.; Wang, P. Identifying the Bottlenecks of Urban Metros and Analyzing the Passenger Source. In Proceedings of the 23rd Joint COTA International Conference of Transportation Professionals (CICTP 2022), Changsha, China, 14–17 July 2022; pp. 625–2635.
26. Wang, C.; Wang, P. Data, Methods, and Applications of Traffic Source Prediction. In *Transportation Analytics in the Era of Big Data*; Ukkusuri, S.V., Yang, C., Eds.; Springer: New York, NY, USA, 2019; pp. 105–120.
27. Lv, S.; Wang, K.; Yang, H.; Wang, P. An Origin–Destination Passenger Flow Prediction System Based on Convolutional Neural Network and Passenger Source-Based Attention Mechanism. *Expert Syst. Appl.* **2024**, *238*, 121989. [[CrossRef](#)]
28. Dijkstra, E.W. A Note on Two Problems in Connexion with Graphs. *Numer. Math.* **1959**, *1*, 269–271. [[CrossRef](#)]
29. Si, B.; Zhong, M.; Liu, J.; Gao, Z.; Wu, J. Development of a Transfer-Cost-Based Logit Assignment Model for the Beijing Rail Transit Network Using Automated Fare Collection Data. *J. Adv. Transp.* **2013**, *47*, 297–318. [[CrossRef](#)]
30. Wang, K.; Guo, B.; Yang, H.; Li, M.; Zhang, F.; Wang, P. A Semi-Supervised Co-Training Model for Predicting Passenger Flow Change in Expanding Subways. *Expert Syst. Appl.* **2022**, *209*, 118310. [[CrossRef](#)]
31. Wang, K.; Wang, P.; Huang, Z.; Ling, X.; Zhang, F.; Chen, A. A Two-Step Model for Predicting Travel Demand in Expanding Subways. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19534–19543. [[CrossRef](#)]
32. Bandyopadhyay, S. The Absolute Gini Is a More Reliable Measure of Inequality for Time Dependent Analyses (Compared with the Relative Gini). *Econ. Lett.* **2018**, *162*, 135–139. [[CrossRef](#)]
33. Nakagawa, S.; Schielzeth, H. A General and Simple Method for Obtaining R² from Generalized Linear Mixed-Effects Models. *Methods Ecol. Evol.* **2013**, *4*, 133–142.
34. Liu, S.; Yao, E.; Cheng, X.; Zhang, Y. Evaluating the Impact of New Lines on Entrance/Exit Passenger Flow of Adjacent Existing Stations in Urban Rail Transit System. *Transp. Res. Procedia* **2017**, *25*, 2625–2638. [[CrossRef](#)]
35. Reddy, A.; Lu, A.; Wang, T. Subway Productivity, Profitability, and Performance. *Transp. Res. Record* **2010**, *25*, 48–58. [[CrossRef](#)]
36. Zhou, Q.; Mao, B.; Peng, S.; Huang, J.; Tian, P. Analysis of the Impacts of Passenger Demand on the Profitability of Different Types of Urban Rail Transit. *Promet* **2023**, *35*, 71–86. [[CrossRef](#)]
37. Yang, J.; Chen, S.; Qin, P.; Lu, F.; Liu, A.A. The Effect of Subway Expansions on Vehicle Congestion: Evidence from Beijing. *J. Environ. Econ. Manag.* **2018**, *88*, 114–133. [[CrossRef](#)]
38. Guo, G.; Zhang, T. A Residual Spatio-Temporal Architecture for Travel Demand Forecasting. *Transp. Res. Part C-Emerg. Technol.* **2020**, *115*, 102639. [[CrossRef](#)]
39. Gu, Y.; Jiang, C.; Zhang, J.; Zou, B. Subways and Road Congestion. *Am. Econ. J.-Appl. Econ.* **2021**, *13*, 83–115. [[CrossRef](#)]
40. Xu, X.; Liu, J.; Li, H.; Jiang, M. Capacity-Oriented Passenger Flow Control under Uncertain Demand: Algorithm Development and Real-World Case Study. *Transp. Res. Part e-Logist. Transp. Rev.* **2016**, *87*, 130–148. [[CrossRef](#)]
41. Huang, Y.; Lin, T.; Zhang, G.; Zhu, W.; Hamm, N.A.S.; Liu, Y.; Zhang, J.; Yao, X. Exploring the Relationship between the Spatial Distribution of Different Age Populations and Points of Interest (POI) in China. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 215. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.