

Received February 23, 2020, accepted February 28, 2020, date of publication March 4, 2020, date of current version March 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978279

# Dynamic Robustness Analysis for Subway Network With Spatiotemporal Characteristic of Passenger Flow

YI FAN<sup>1</sup>, FAN ZHANG<sup>1</sup>, SHIHONG JIANG<sup>1</sup>, CHAO GAO<sup>1,2</sup>, ZHANWEI DU<sup>3,4</sup>, ZHEN WANG<sup>2</sup>, AND XIANGHUA LI<sup>1</sup>

<sup>1</sup>College of Computer and Information Science, Southwest University, Chongqing 400715, China

<sup>2</sup>Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an 710072, China

<sup>3</sup>Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR, Shenzhen 518000, China

<sup>4</sup>Department of Integrative Biology, The University of Texas at Austin, Austin, TX 78705, USA

Corresponding authors: Zhen Wang (zhenwang0@gmail.com) and Xianghua Li (li\_xianghua@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant U1803263, Grant 61976181, Grant 11931015, and Grant 81961138010, in part by the Natural Science Foundation of Chongqing under Grant cstc2018jcyjAX0274 and Grant cstc2019jcyj-zdxmX0025, in part by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, under Grant MNR KF-2019-04-034, in part by the Key Area Research and Development Program of Guangdong Province under Grant 2019B010137004, in part by the Key Area Research and Development Program of Shaanxi Province under Grant 2019ZDLGY17-07, and in part by the Fundamental Research Funds for Central Universities under Grant 3102019PJ006.

**ABSTRACT** The robustness is a crucial and essential problem of a subway network (SN), which can help us improve the efficiency of a transportation system. Several existing researches have analyzed the SN robustness based on the rail structure or the static distribution of passenger flow. However, the spatiotemporal characteristic of passenger flow also plays an important role in the SN robustness, since it can trigger some unexpected cascading failures in SN. Therefore, how to characterize the effect of this cascading failure on the SN robustness still remains an important and open problem. In this paper, we address the above problem as follows: (1) we propose a temporal subway network (TSN) to consider the dynamics of passenger flow in SN; (2) we adopt the linear threshold (LT) model to simulate the cascading failure process of TSN and propose a new robustness metric  $R(t)$  to evaluate the effect of this cascading failure on SN robustness. Based on the Shanghai subway smart card data, we carry out extensive experiments to analyze the effects of the cascading failure on the Shanghai SN robustness. Experiments show that the Shanghai TSN robustness varies over time. More significantly, the large volume of passenger flow can increase the impact of failure modes (i.e., random and malicious failure modes) on the Shanghai TSN robustness.

**INDEX TERMS** Subway network, robustness, dynamic passenger flow, cascading failure.

## I. INTRODUCTION

Credited for the advantages of high speed, safety, environmental protection, the urban rail transit has become the first option of transportation [1], [2]. However, some incidents (such as severe weather, sudden disaster, and terrorist attack) may lead to the subway station disruption [3]–[5]. Therefore, it is essential for us to investigate the robustness of subway system, which can largely ensure the reliance and efficiency of service that the subway system provides.

Based on the network science, the robustness analysis of subway network (SN) has made great progress [6], [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Dominik Strzalka.

Currently, there exists two types of methods for SN robustness analysis, namely, the rail structure-based analysis [8]–[11] and passenger flow-based analysis [12]–[15]. The rail structure-based analysis focuses on the topological properties of subway. For example, Derrible *et al.* have studied the subway systems of 33 cities [9], through which they have found the scale-free and small-world characteristics of SN. They have also proposed some suggestions (such as increasing additional transfer stations) to improve the SN robustness. Mouronte *et al.* have analyzed the characteristics of urban bus and subway networks of Madrid [10], through which they have found the structure parameters (such as the shortest distance between stations, betweenness and detection of clusters) can help to improve the robustness of transport

networks. Based on the FMECA method (i.e., failure mode, effects and criticality analysis) [16], Deng *et al.* have proposed a new framework, through which they can study the physical vulnerability of subway system [11]. The results show that train is the most vulnerable functional module in subway system.

Considering the large volume of passenger flow in the SN [17], [18], the network robustness is likely to go beyond the issue of pure topology of rail structure. Therefore, the passenger flow-based analysis pays much attention to the properties of SN with respect to the time and passenger flow. For example, Sun *et al.* have addressed the issue of SN robustness from the perspective of line operation by exemplifying the Shanghai subway system [12]. The result shows that the subway lines with a large passenger flow volume generally have a significant impact on the network vulnerability. Xiao *et al.* have proposed some new dynamic metrics that can reflect the local and global features of node's degree and betweenness, through which they have found the heterogeneity and vulnerability of Beijing subway network vary over time when passenger flow changes dynamically [13].

Furthermore, due to the spatiotemporal characteristic of passenger flow in SN [19], the failure caused by the disrupted stations will diffuse to the global SN or at least to a large part of it [20]–[22], that is, the cascading failures occur in the SN [23], [24]. In order to explore the cascading failure caused by the passenger flow in SN, some cascading failure models (such as the load capacity model [25], coupled map lattice model [26] and linear threshold model [27], [28]) have been improved to simulate this cascading failure process. For example, both Ma *et al.* and Shen *et al.* have proposed an improved coupled map lattices model to analyze the cascading failure process of Xi'an SN [14] and Nanjing SN during a certain time period [15], respectively. Ma *et al.* have found there is a two-stage cascading failure process of a network under the condition of passenger flow overload. Shen *et al.* have found that it is easy for the largest strength station disruption to trigger a global network failure.

However, it appears apparent that the cascading failure will affect more people than that occurs at other time, if stations or lines are disrupted at some time with a large passenger flow. Therefore, it matters to investigate the impact of the cascading failure occurring at different times on the SN robustness. In this paper, we aim to answer two main questions as follows:

- How to reflect the spatiotemporal characteristics of passenger flow in the SN?
- How can the dynamic effects of passenger flow on the SN robustness be characterized?

In order to address the above research problems, we first design a temporal subway network (TSN) to reflect the spatiotemporal characteristic of passenger flow. More specifically, we construct the TSN based on the L-Space method [33]. A subway station is denoted as a node and the pathway between two adjacent stations is formulated as an edge. Each edge weight is calculated by estimating the

whole travel routes, and then assigning passenger flows to each pathway at a specific time. Second, we adopt the linear threshold (LT) model to simulate the cascading failure process of TSNs at different times. Third, to evaluate the dynamic effect of this cascading failure on the network robustness, we propose a new robustness metrics  $R(t)$ . Finally, we construct the Shanghai TSN based on published subway datasets (e.g., the rail structure, running data, traveling data based on the smart card). Moreover, we simulate the cascading failure process of Shanghai TSN by the LT model and analyze the dynamic robustness of such network.

The main contributions of this paper can be listed as follows:

- 1) We propose a temporal subway network (TSN) for characterizing the dynamics of passenger flow.
- 2) We propose a new robustness metric  $R(t)$  to evaluate the dynamic robustness of TSN.
- 3) We carry out extensive experiments to reveal the dynamic robustness of Shanghai TSNs at different times by the LT model and  $R(t)$ .

The rest of this paper is organized as follows. Sec. II first illustrates the formulation of TSN, and then introduces the LT model and robustness metrics. Sec. III implements some analyses to reveal the dynamic robustness of Shanghai TSN. Finally, Sec. IV concludes this paper.

## II. PRELIMINARIES

This section first introduces the formulation of temporal subway network (TSN), followed by some node property metrics of TSN in Sec. II-A. Then, Sec. II-B presents the linear threshold (LT) model. Finally, Sec. II-C describes a new robustness metric.

### A. TEMPORAL SUBWAY NETWORK

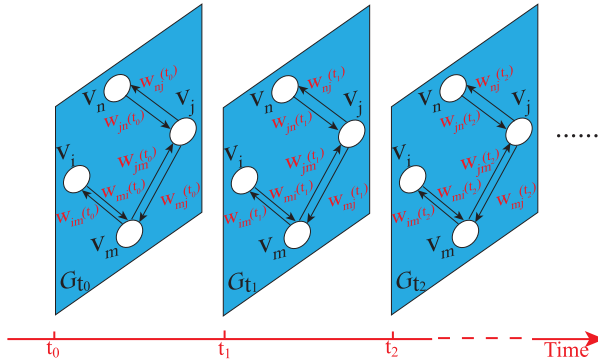
#### 1) FORMULATION OF TEMPORAL SUBWAY NETWORK

In this paper, a temporal subway network (TSN) is constructed by L-Space method [33], through which we can display the connectivity feature and assign the passenger flows to each pathway. More specifically, a station is denoted as a node and the pathway between two adjacent stations is formulated as an edge. Let  $N$  denote the total number of stations.  $A$  denotes the adjacency matrix of TSN whose size is  $N \times N$ , with its elements defined as Eq. (1).

$$A_{ij} = \begin{cases} 1, & \text{there is an edge pointing from } v_j \text{ to } v_i \\ 0, & \text{there is no edge pointing from } v_j \text{ to } v_i \end{cases} \quad (1)$$

Generally speaking, subways have upstream and downstream directions [34]. Based on this fact, Def. 1 defines the TSN. A simple example of the TSNs at different times is shown in Fig. 1.

*Definition 1:* The TSN is denoted as a directed and weighted graph  $G_t = \langle V, E, W \rangle$ , where  $V = \{v_i | i \in [1, N]\}$  and  $E = \{e_{ij} = (v_i, v_j) | i \neq j\}$  represent a set of nodes and a set of edges, respectively.  $W = \{w_{ij}(t) | i \neq j\}$  denotes a set



**FIGURE 1.** An example of the TSNs at different times. At different times, the structure topologies of TSNs are the same, but the edge weights vary over time.

of edge weights.  $w_{ij}(t)$  stands for the number of passengers passing through  $e_{ij}$  at time  $t$ .

In order to assign passenger flows to each pathway at time  $t$ , we first estimate the whole travel routes based on the original-destination (OD) data of passengers taking the subway. Such data consists of many OD pairs  $(v_i, v_j, w_{ij})$  [35], [36]. Each OD pair denotes there are  $w_{ij}$  passengers from  $v_i$  and  $v_j$ . Generally, some factors (e.g., the length of a route, the transfer time and the weather condition) may affect the route choice of passengers [37], [38]. In this paper, if there are more than one routes from  $v_i$  to  $v_j$ , the route will be selected according to the behaviors and purposes of passenger travel [39], [40], such as the shortest route [41], the minimum number of transfer and the shortest transfer time. If there are still more than one routes that meet the above screening criteria, the route will be the one randomly selected from them.

Then, we estimate the time when passenger departs from the first station of the whole travel routes. Generally, when passengers enter the automatic fare gate, it takes a while for them to walk to the platform and wait until the train arrives [42], [43]. Passengers can only alight or board during the dwell time of the train at the platform. Therefore, we calculate the departure time of passengers ( $T^d$ ) by Eq. (2).

$$T^d = \begin{cases} t^{in} + t^{walk}, & \text{if a train has arrived} \\ t^{ac} + t^{dwell}, & \text{if no train arrives} \end{cases} \quad (2)$$

where  $t^{in}$  denotes the check-in time of passenger.  $t^{walk}$  denotes the walking time span from an automatic fare gate to a platform and  $t^{ac}$  denotes the closest arrival time of the subway train which is greater than  $(t^{in} + t^{walk})$ . All the arrival time in each station can be calculated based on the timetable of metro system.  $t^{dwell}$  denotes the dwell time span of the train at the platform, which can also be collected from the timetable of metro system. In this paper, when the passenger arrival coincides with a train stop at the platform, we set  $t^{dwell} = 0$ .

Finally, we estimate the time span when passengers passing through the pathway between any two adjacent stations. Based on this, we can calculate the moments when passengers arrive at each corresponding station of the whole travel route,

which starts from  $T_d$ , so that the passenger flows can be assigned to the corresponding pathway at time  $t$ . If passengers need not transfer, this time information can be collected by the timetable of metro system. If passengers need to transfer, we should add the transfer time, which is calculated by Eq. (3).

$$T_i^f(m, n) = \alpha \frac{DIS_i(m, n)}{NWS} + t^{dwell} \quad (3)$$

where  $DIS_i(m, n)$  denotes the distance from  $v_i$  in Line  $m$  to  $v_i$  in Line  $n$ .  $NWS$  denotes the normal walk speed of passengers [44].  $\alpha$  is the congestion coefficient. If passengers arrive at the platform when a train has just stopped at the platform, we set  $t^{dwell} = 0$ .

## 2) NODE PROPERTY METRICS OF NETWORK

Lots of metrics have been proposed to estimate the centrality of nodes in a network [45], [46]. In this paper, we utilize four node property metrics (i.e., degree, betweenness, strength and flow betweenness) to evaluate the property of node in the TSN. The details of such metrics are as follows:

### a: DEGREE

In  $G_t$ ,  $v_i$  has the incoming and outgoing degrees. Incoming degree  $D_i^{in}$  denotes the number of edges that point to  $v_i$  and outgoing degree  $D_i^{out}$  denotes the number of edges from  $v_i$  to other nodes, which are given by Eq. (4) and Eq. (5), respectively. In addition, the degree of  $v_i$  is given by Eq. (6).

$$D_i^{in} = \sum_{j=1, j \neq i}^N A_{ij} \quad (4)$$

$$D_i^{out} = \sum_{j=1, j \neq i}^N A_{ji} \quad (5)$$

$$D_i = D_i^{in} + D_i^{out} \quad (6)$$

### b: BETWEENNESS

In  $G_t$ , the betweenness of  $v_i$  ( $B_i$ ) is defined as Eq. (7).

$$B_i = \sum_{o, d \in V, o \neq d} \frac{\sigma_i^{od}}{\sigma^{od}} \quad (7)$$

where  $\sigma_i^{od}$  is the number of the shortest paths from  $v_o$  to  $v_d$  that pass through  $v_i$  in  $G_t$ .  $\sigma^{od}$  denotes the number of the shortest paths from  $v_o$  to  $v_d$ . The larger  $B_i$  is, the more important the connection role of  $v_i$  in  $G_t$  becomes.

### c: STRENGTH

The strength of a node  $v_i$  ( $S_i(t)$ ) is the sum of the weight of edges that  $v_i$  shares with other nodes at time  $t$ . Therefore, in  $G^t$ ,  $v_i$  has the incoming and outgoing strengths. The incoming strength  $S_i^{in}(t)$  denotes the sum of the weight of edges that point to  $v_i$  at time  $t$  and the outgoing strength  $S_i^{out}(t)$  denotes the sum of the weight of edges that from  $v_i$  to other nodes at time  $t$ , which are given by Eq. (8) and Eq. (9), respectively. In addition, the total strength of  $v_i$  ( $S_i(t)$ ) is given by Eq. (10).

A large  $S_i(t)$  means that there are lots of passengers passing through  $v_i$  at time  $t$ .

$$S_i^{in}(t) = \sum_{j=1, j \neq i}^N w_{ij}(t) \quad (8)$$

$$S_i^{out}(t) = \sum_{j=1, j \neq i}^N w_{ji}(t) \quad (9)$$

$$S_i(t) = S_i^{in}(t) + S_i^{out}(t) \quad (10)$$

**d: FLOW BETWEENNESS**

The flow betweenness of  $v_i$  ( $F_{B_i}(t)$ ) in  $G_t$  equals the passenger flow ratio of the shortest paths from all nodes to all other nodes that pass through  $v_i$  in  $G_t$  at time  $t$  and is defined as Eq. (11).

$$F_{B_i}(t) = \sum_{o,d \in V, o \neq d} \frac{f\sigma_i^{od}}{f\sigma^{od}} \quad (11)$$

where  $f\sigma_i^{od}$  is the sum of the passenger flow volume on the shortest paths from  $v_o$  to  $v_d$  that pass through  $v_i$  at time  $t$ .  $f\sigma^{od}$  denotes the sum of the passenger flow on the shortest paths from  $v_o$  to  $v_d$  at time  $t$ . When  $F_{B_i}(t)$  is large, there are lots of passengers on the shortest path that pass through  $v_i$  at time  $t$ .

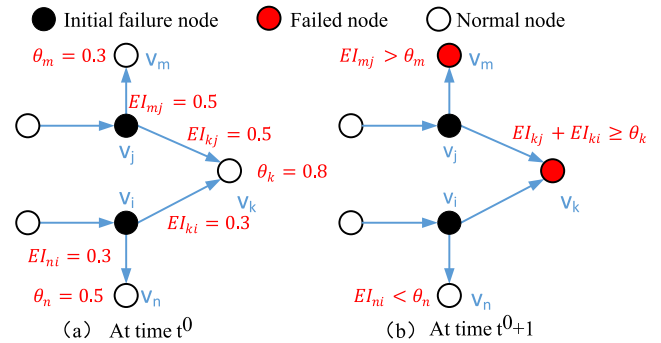
**B. LINEAR THRESHOLD MODEL**

In this paper, the cascading failure process of TSN is modeled based on the linear threshold (LT) model [27], [28]. In detail, each node  $v_i$  has two states (i.e., the failed state and normal state). Each node  $v_i$  has a threshold  $\theta_i$  which is selected randomly in the interval [0,1] [29], [30].  $v_i$  is influenced by its incoming neighbors  $v_j$  based on the edge influence  $El_{ij}$ . The calculation of  $El_{ij}$  is shown in Eq. (12).

$$El_{ij} = \frac{w_{ij}(t)}{\max_{T \in [t_1, t_2, t_3, \dots]} \{S_i^{in}(T)\}} \quad (12)$$

where  $\max_{T \in [t_1, t_2, t_3, \dots]} \{S_i^{in}(T)\}$  represents the maximum incoming strength of SN at different times. In addition,  $\sum_{v_j \in \Gamma(V)} El_{ij} \leq 1$ . Here,  $\Gamma(V)$  denotes a set of the incoming neighbors. At time step  $t^0$ , each currently normal node  $v_i$  will become failed at time step  $t^0 + 1$  if and only if the total edge influence of its failed and incoming neighbors is at least  $\theta_i$ , that is,  $\sum_{v_j \in \Gamma_i(V)} El_{ij} \geq \theta_i$ . Here,  $\Gamma_i(V)$  denotes a set of the incoming neighbors of  $v_i$  that are failed at time step  $t^0$ . Fig. 2 illustrates an example of the above process.

For an initial set of failed nodes, the cascading failure spreads out deterministically in discrete steps. Initial failure nodes are used to simulate network under different failure modes in real life [31], [32]. When there is no failed node in  $G_t$ , such process will be terminated. Let  $F$  denote the total number of failed nodes at this time. In this paper, we call  $F$  as the cascading failure size of  $G_t$ .



**FIGURE 2.** An example of the cascading failure process based on the linear threshold (LT) model. (a)  $v_i$  and  $v_j$  are initial failure nodes at time step  $t^0$ .  $v_i$  and  $v_j$  are the incoming neighbor node of  $v_n$  and  $v_m$ , respectively. Both  $v_i$  and  $v_j$  are the incoming neighbor nodes of  $v_k$ . Therefore, the states of  $v_n$  and  $v_m$  are affected by  $v_i$  and  $v_j$ , respectively. The state of  $v_k$  is affected by  $v_i$  and  $v_j$ . (b) shows the results of the states change of  $v_n$ ,  $v_m$  and  $v_k$ . Since, the edge influence  $El_{mj}$  is greater than the threshold  $\theta_m$ ,  $v_m$  is failed at time step  $t^0 + 1$ . Similarly, we can see  $v_k$  is failed at time step  $t^0 + 1$ . However,  $El_{ni} \leq \theta_n$ , so  $v_n$  is still in normal state at time  $t^0 + 1$ .

**C. EVALUATION OF ROBUSTNESS**

In order to evaluate the dynamic effect of the cascading failure caused by the passenger flow on the SN robustness, we propose a robustness metrics  $R(t)$ .  $R(t)$  couples the relative size of largest component [47], [53] and operational efficiency [13], [48] by using a coupling coefficient  $\varepsilon$ , which is given by Eq. (13).

$$R(t) = \varepsilon \cdot LC + (1 - \varepsilon) \cdot \frac{OE(t) - F_{OE}(t)}{\max_{T \in [t_1, t_2, t_3, \dots]} \{OE(T)\}} \quad (13)$$

where  $R(t)$  denotes the robustness of  $G_t$ .  $LC$  denotes the relative size of largest component, which is defined in Eq. (14).  $OE(t)$  denotes the operational efficiency of  $G_t$  as defined in Eq. (15).  $\max_{T \in [t_1, t_2, t_3, \dots]} \{OE(T)\}$  denotes the maximum value of the operational efficiency at different times  $T$ . In particular, the networks at the time set  $[t_1, t_2, t_3, \dots]$  should reflect the difference of passenger flow volume.  $F_{OE}(t)$  denotes the operational efficiency of  $G_t$  when there are failed nodes in  $G_t$ .  $\varepsilon$  is a coupling coefficient to evaluate the importance of the effect of network topology on the robustness of  $G_t$ .  $(1 - \varepsilon)$  is used to evaluate the importance of the effect of passenger flow on the robustness of  $G_t$ . In this paper,  $\varepsilon$  is quantified as Eq. (16).

$$LC = \frac{N'}{N} \quad (14)$$

where  $N'$  is the number of nodes in the largest component after some nodes are failed.  $N$  is the total number of nodes.  $LC$  is used to characterize the robustness of  $G_t$  from the perspective of network scale. The higher the value of  $LC$  is, the more robust a network becomes when suffering attack.

$$OE(t) = \frac{1}{N(N-1)} \sum_{i,j=1(i \neq j)}^N \frac{w_{ij}(t)}{d_{ij}} \quad (15)$$

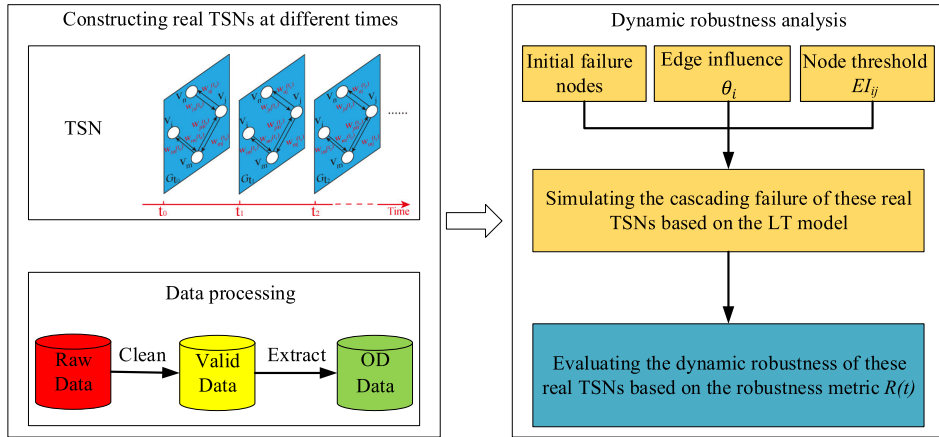


FIGURE 3. The flowchart of evaluating the dynamic robustness of some real TSNs.

where  $OE(t)$  denotes the operational efficiency of  $G_t$ .  $d_{ij}$  is the number of edges along the shortest path from  $v_i$  to  $v_j$  in  $G_t$ .  $w_{ij}(t)$  denotes the weight of edges along the shortest path from  $v_i$  to  $v_j$  at time  $t$ .  $OE(t)$  is used to estimate the averaged passenger flow volume of edges on the shortest path. The higher the value of  $OE(t)$  is, the more passengers are transported in the pathway between two stations at time  $t$ .

$$\varepsilon = \frac{\sum_{v_i \in V} S_i(t)}{\max_{T \in \{t_1, t_2, t_3, \dots\}} \{\sum_{v_i \in V} S_i(T)\}} \quad (16)$$

where  $\sum_{v_i \in V} S_i(t)$  denotes the total sum of node strengths in  $G_t$ .  $\max_{T \in \{t_1, t_2, t_3, \dots\}} \{\sum_{v_i \in V} S_i(T)\}$  denotes the maximum value of the total sum of node strengths in the TSNs at different times.

Based on the above-mentioned works, Fig. 3 presents the flowchart evaluating the dynamic robustness of a real TSN. Specifically, we extract the OD data from our raw data firstly by cleaning and processing it. Then, we use the OD data to construct real TSNs at different times based on the formulation of TSN in Sec. II-A1. In addition, we use the LT model to simulate the cascading failure of these real TSNs. Finally, we evaluate the effect of the cascading failure on these real TSNs robustness based on the proposed metric  $R(t)$ .

### III. SIMULATION AND ANALYSIS

In this section, we first introduce the datasets and statistical analysis in Sec. III-A. Then, following the construction of a Shanghai TSN in Sec. III-B, we carry out some experiments to explore the dynamic robustness of Shanghai TSN in Sec. III-C.

#### A. DATASETS AND STATISTICAL ANALYSIS

##### 1) DATASET

The contents datasets use in this work are (1) Shanghai subway lines and stations data, (2) Shanghai subway running data and (3) Shanghai subway smart card data. The details of the three datasets are as follows.

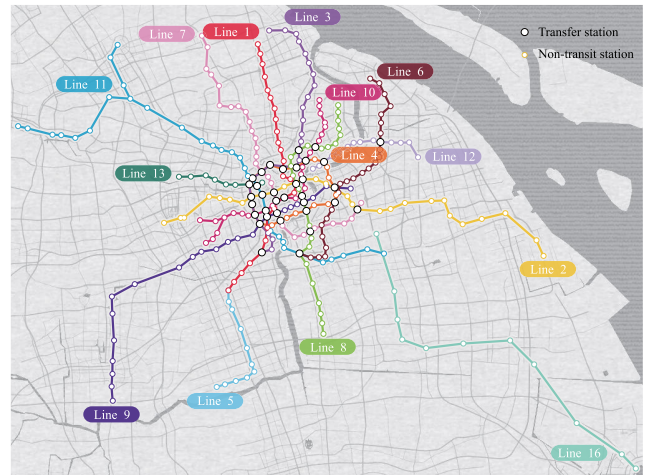


FIGURE 4. The positions of Shanghai subway lines and stations, in which each line is highlighted with different colors. There are 14 lines, 35 transfer stations and 254 non-transfer stations.

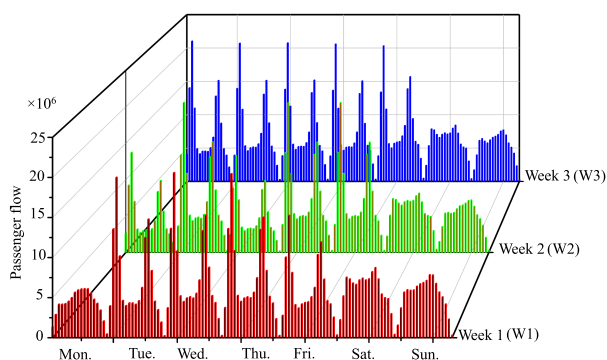
**Dataset 1 (Shanghai Subway Lines and Stations Data):** This dataset is collected by the Shanghai Metro<sup>1</sup> in April 2015 and records the information of 14 subway lines and stations. In addition, the subway stations are numbered. More specifically, stations with the same name but in different lines are represented by the same ID. Conversely, stations with different names are represented by different ID. Due to the two Pudian Road are located on different areas, they are seen as two different nodes and numbered separately. Based on the latitude and longitude information of all stations, Fig. 4 visualizes the rail structure of Shanghai SN.

**Dataset 2 (Shanghai Subway Running Data):** This dataset is collected by the Shanghai Metro<sup>1</sup>, which contains some information of the train operations in April 2015. More specifically, it records the number of trains in each line, the departure station, the departure time of the departure station in each line, the departure intervals, the dwell time

<sup>1</sup><http://service.shmetro.com/>

**TABLE 1.** Examples of three datasets. In Dataset 1, “Corresponding Lines” denotes the line a station belongs to. In Dataset 2, we list the departure intervals, the dwell time and the arrival time of trains in two different departure stations. In Dataset3, Card\_ID denotes the smart card number of passengers; “In\_” and “Out\_” represent check in and check out, respectively.

Dataset 1							
Example	Station ID	Station Name		Corresponding Lines			
1	1	Xingzhuang		Line 1, Line 8			
2	289	Dishui Lake		Line 16			
Dataset 2							
Example	Departure Station ID (Name)	Departure Interval		Dwell Time	Arrival Time		
1	1 (Xinzhuang)	3min		30s	[5:30, 5:33, 5:36, .....]		
2	289 (People’s Square)	6min		30s	[6:00, 6:06, 6:12, .....]		
Dataset 3							
Example	Card_ID	In_Line	Out_Line	In_Station ID (Name)	Out_Station ID (Name)	In_Time	Out_Time
1	602141128	Line 1	Line 1	1 (Xinzhuang)	235 (East Changji Road)	07:51:08	09:07:57
2	2001530605	Line 2	Line 10	48 (Guanglan Road)	224 (Guoquan Road)	08:44:19	09:33:05



**FIGURE 5.** The dynamics distribution of three-week passenger flow. Week 1 is from April 6 to April 12, Week 2 is from April 13 to April 19, and Week 3 is from April 20 to April 26. The statistics range of passenger flow is from 6:00 to 24:00 every day. The correlation analyses show that passenger travel mode is regular and similar every week.

of each train in each station, and the arrival time of each train in each station.

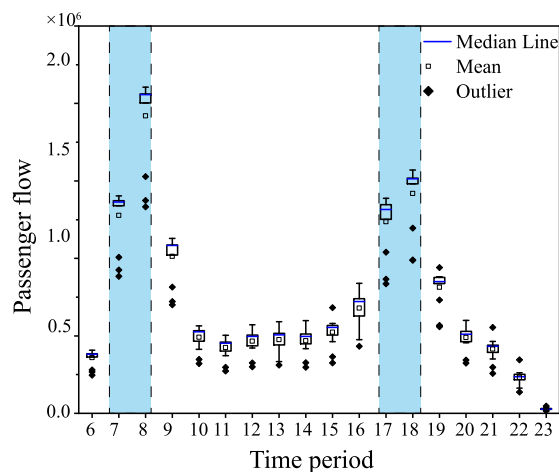
*Dataset 3 (Shanghai Subway Smart Card Data):* This dataset is collected by the Shanghai Public Transportation Card Co. Ltd and released by organizing committee of Shanghai Open Data Apps.<sup>2</sup> Furthermore, this dataset includes 11 million passengers and 123 million trips’ records during April 2015. Each trip record includes the check-in and check-out date, time, station and so on.

As some existing examples from 13th April 2015 can demonstrate the details of the above three datasets, we list some ones in Table 1.

2) STATISTICAL ANALYSIS

Existing studies have shown the dynamic features of human behaviors, such as the human mobility [49] and online behaviors [50]. The spatiotemporal characteristic of passenger flow greatly affects its own dynamic distribution [51], [52], which can help us construct the Shanghai TSN. We select three weeks’ data (i.e., April 6, 2015 to April 26, 2015) from *Dataset 3* for passenger flow statistics. As shown in Fig. 5,

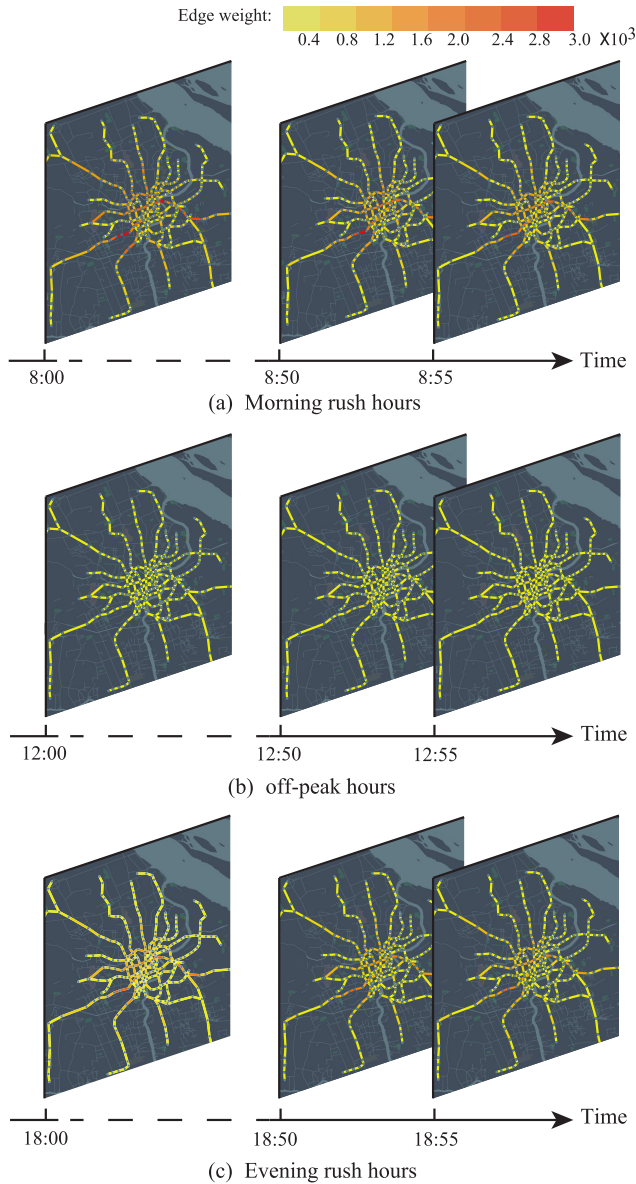
<sup>2</sup><http://data.sh.gov.cn/>



**FIGURE 6.** The distribution of passenger flow during different time periods. The x-axis denotes 18 hours in a day and the range is from 6:00 to 24:00. The number on the x-axis is time period. For example, 6 denotes the time period 6:00-7:00. The y-axis denotes the passenger flow volume. The result shows that the passenger flow during each time period has obvious differences and there are two rush hours each day (i.e., morning and evening rush hours).

April 6 is an exception, which is known as Tomb Sweeping Festival, when the residents switch back and forth between working mode (from Monday to Friday) and holiday mode (from Saturday to Sunday) [53]. Moreover, we use the *Pearson correlation coefficient* [54] to analyze the correlation of three weeks’ data. The results show that the correlation coefficient of two weeks is 0.89, 0.86 and 0.95 between W1 and W2, W1 and W3, W2 and W3, respectively. Based on this, we can conclude that the distribution of weekly passenger flow is regular and similar.

In particular, as shown in Fig 6, we use the box-whisker plot to analyze the difference in the distribution of passenger flow during different time periods on weekday. According to the result, we observe the passenger flow data is close to its mean or median, with few outliers. Therefore, the passenger flow data among the three-week is similar. Obviously, the distribution of passenger flow characterizes the morning and evening rush hours on weekdays. Based on the above



**FIGURE 7.** 36 Shanghai TSNs at different times on weekdays. (a), (b) and (c) are constructing Shanghai TSN during the morning rush hour (8:00-9:00), off-peak hour (12:00-13:00) and evening rush hour (18:00-19:00), respectively. The result shows that the edge weights change obviously during the rush hours.

analyses, this paper divides each weekday into several travel periods [55], which is shown in Table 2.

**B. SHANGHAI TEMPORAL SUBWAY NETWORK**

Before constructing the Shanghai TSN, some relevant assumptions are listed as follows to simplify the network construction process:

- 1) The departure time of each departure station in each line is 5:30 am.
- 2) The departure interval between two adjacent subway trains during the rush hours is 3 minutes and during the off-peak hours is 6 minutes.

**TABLE 2.** Travel periods partition on weekdays.

Time period	Start time	End time
Morning off-peak hours	06:00:00	07:00:00
Morning rush hours	07:00:00	09:00:00
Noon off-peak hours	09:00:00	17:00:00
Evening rush hours	17:00:00	19:00:00
Evening off-peak hours	19:00:00	00:00:00

- 3) The dwell time span of each subway train at the platform is 30 seconds.
- 4) The normal walking speed of each passenger is 60m/s [44].
- 5) When passengers enter the automatic fare gate, they will immediately go to the platform to take the subway.
- 6) The congestion coefficient during the rush hours is 0.8 and during the off-peak hours is 1.

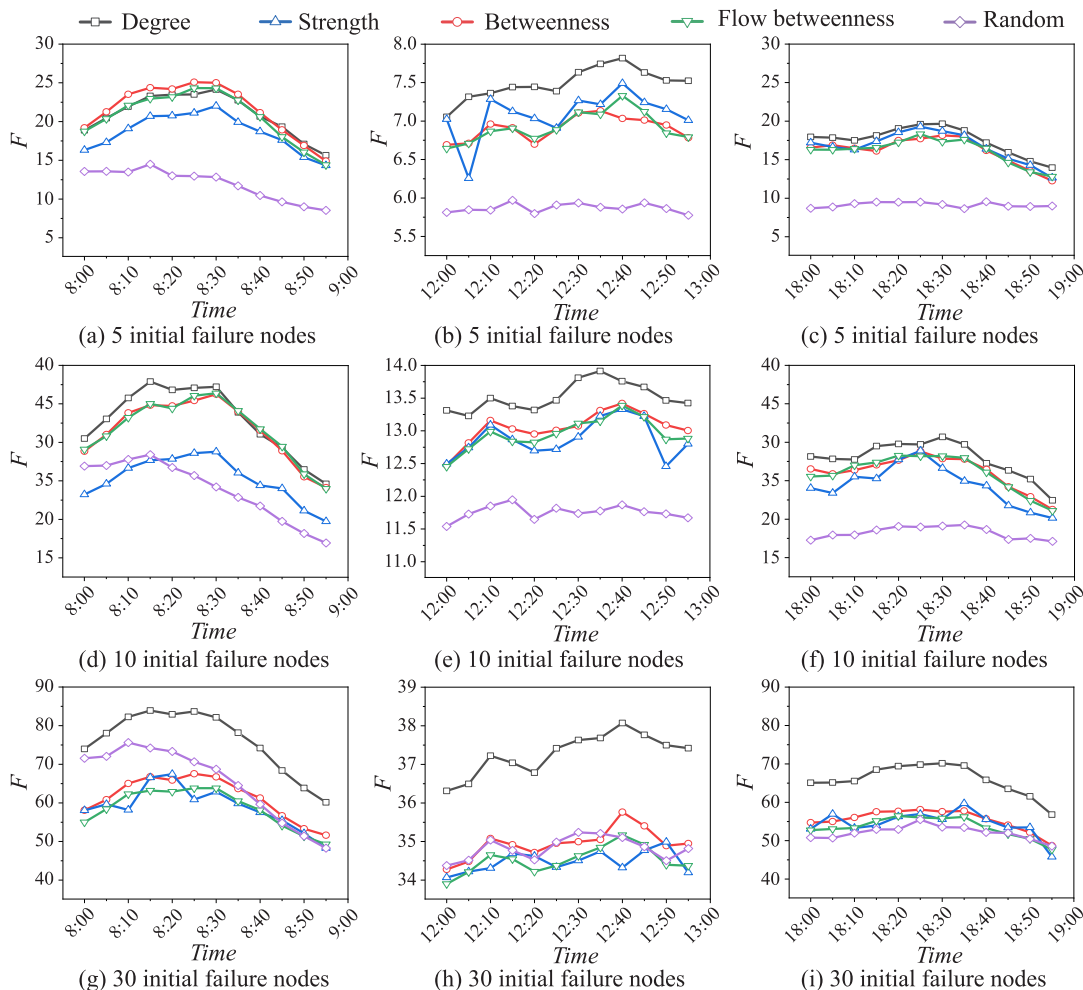
where 1) and 3) are the value with highest frequency in the departure time of each departure station in each line and the dwell time span of each subway train at the platform, respectively. In 2), 3 minutes and 6 minutes are calculated by rounding the average value of all the departure interval between two adjacent stations during the rush hours and off-peak hours, respectively.

As the distribution of passenger flow varies over time, the TSN differs at different moment. Despite this, it is unrealistic and unnecessary for us to construct TSNs at every moment. In Sec. III-A2, we have found that the changes of passenger flows on weekdays can reflect the characteristics of it. Therefore, with the partition of travel time periods in Table 2, we construct a corresponding Shanghai TSN every 5 minutes based on the data of 8:00-9:00 (morning rush hour), 12:00-13:00 (off-peak hour) and 18:00-19:00 (evening rush hour) on April 13th (Monday) in the Dataset 3. Fig. 7 shows 36 Shanghai TSNs at different times on weekdays.

**C. DYNAMIC ROBUSTNESS ANALYSIS**

In order to trigger the linear threshold (LT) model, we first define five kinds of failure modes based on the node property metrics in Sec. II-A2, which are shown as follows.

- (1) **Degree failure mode** is defined as selecting the failed nodes from a network based on the descending order of node degree.
- (2) **Betweenness failure mode** is defined as selecting the failed nodes from a network based on the descending order of node betweenness.
- (3) **Strength failure mode** is defined as selecting the failed nodes from a network based on the descending order of node strength.
- (4) **Flow betweenness failure mode** is defined as selecting the failed nodes from a network based on the descending order of node flow betweenness.
- (5) **Random failure mode** is defined as selecting the failed nodes from a network randomly.



**FIGURE 8.** (a)-(c), (d)-(f) and (g)-(i) show the results of the cascading failure size ( $F$ ) caused by 5, 10 and 30 initial failure nodes for different failure modes at different failure times ( $Time$ ), respectively. The result shows that different failure modes have different effect on the cascading failure of Shanghai TSN.

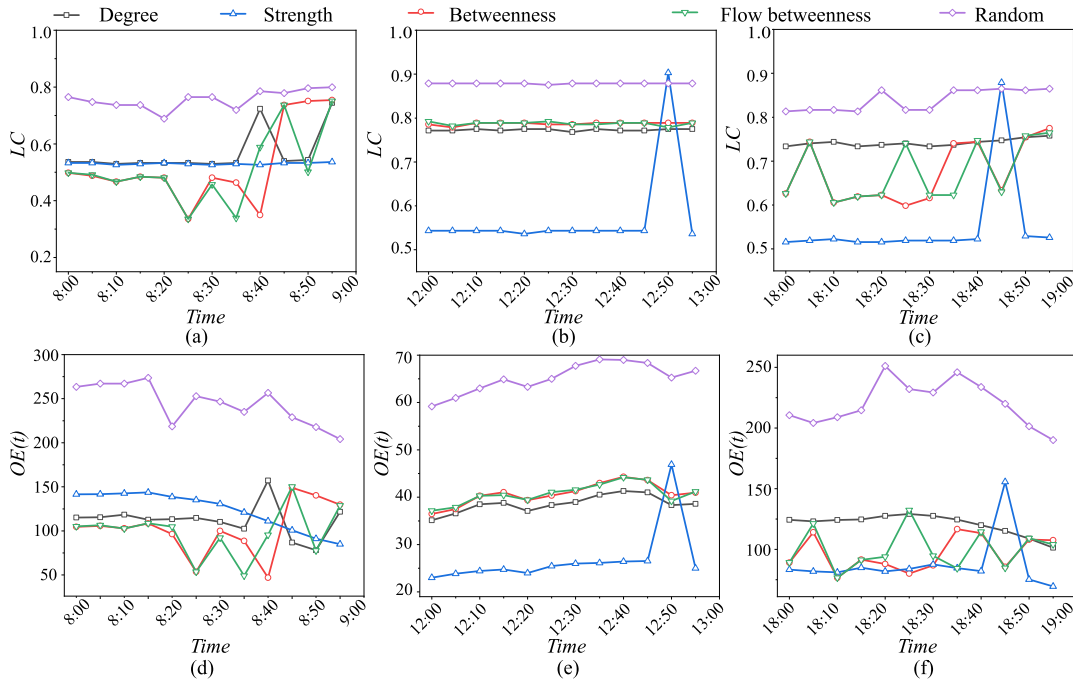
Then, we select 5, 10 and 30 initial failure nodes under each failure mode to trigger the cascading failure in each Shanghai TSN, respectively. In order to eliminate the influence of the node threshold  $\theta_i$  in the LT model, we repeat each cascading failure experiment caused by one failure mode for 500 times. The final result is the average of these 500 experiments.

Figure 8 compares the cascading failure size ( $F$ ) of Shanghai TSNs under various failure modes at different failure times. Figs. 8(a)-(f) show that the cascading failure size of Shanghai TSN is smaller under the random failure mode than other failure modes. This finding suggests that when the number of initial failure nodes is relatively fewer (i.e., 5 or 10 initial failure nodes), the cascading failure is unlikely to occur under the random failure mode in Shanghai TSN, but it is likely to occur under other failure modes instead. However, as shown in Figs. 8(g)-(i), when the number of initial failure nodes is large (e.g., 30 initial failure nodes), the Shanghai TSN under random failure mode can also cause larger cascading failure. Moreover, the gap of the cascading failure size of the Shanghai TSN under the degree failure

mode and other failure modes becomes increasingly large. According to this, we can conclude that with the initial failure nodes increasing, the larger cascading failure in Shanghai TSN is prone to occur under the degree failure mode. In addition, when the initial failure nodes and failure modes are identical, the cascading failure size of Shanghai TSN during the rush hours is far greater than that during other hours. Meanwhile, there also exists fluctuation among each cascading failure size of Shanghai TSN during the rush hours. In particular, during the morning rush hour, this fluctuation range is relatively large. For example, the cascading failure size of Shanghai TSN at 8:15 is larger than that at 8:40. These results suggest that the effect of passenger flow on the cascading failure size of Shanghai TSN varies over time. The larger volume of passenger flow is, the more likely the cascading failure will occur when the network is under the same failure modes.

Finally, we use the robustness metrics  $LC$ ,  $OE(t)$  and  $R(t)$  to evaluate the effect of the above cascading failures on the Shanghai TSN robustness. More specifically, since the node





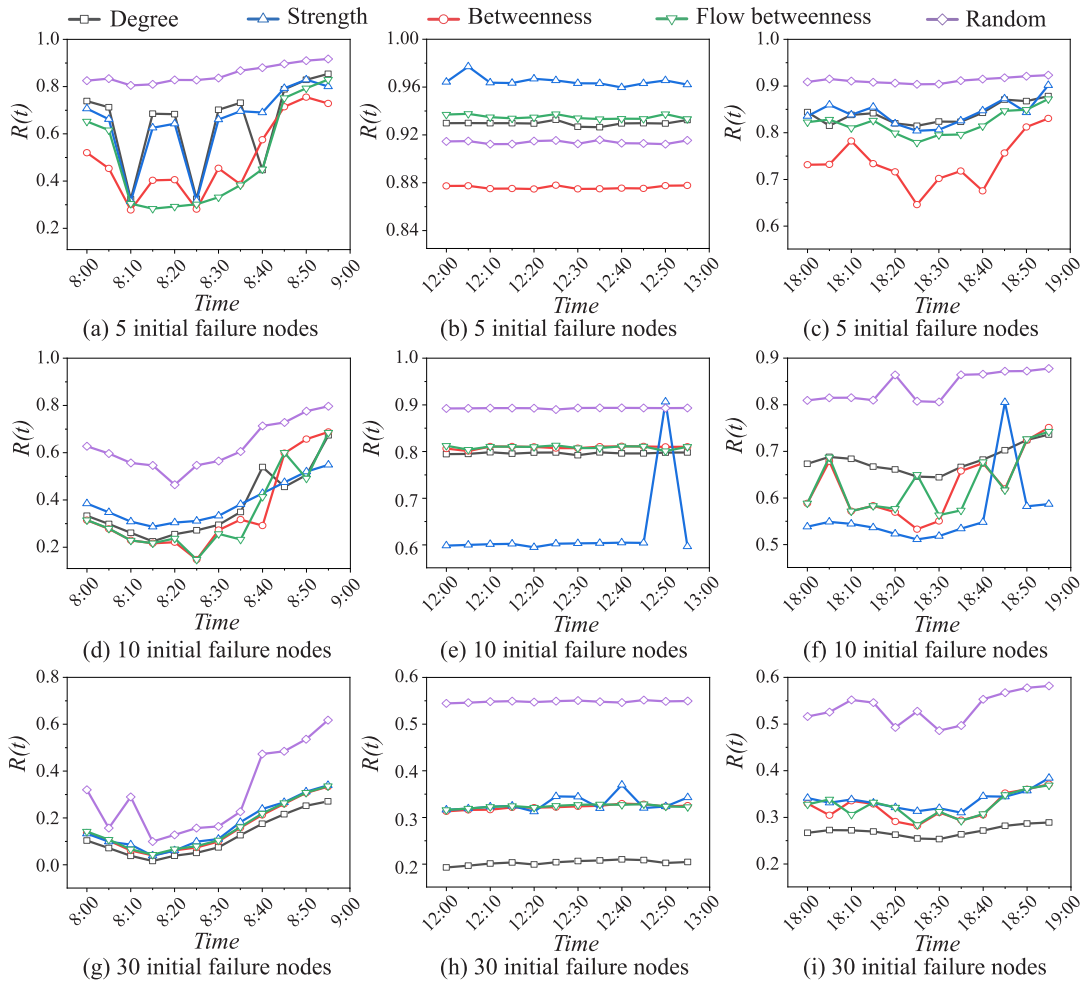
**FIGURE 9.** (a)-(c) are the relationship between the different failure times (*Time*) and robustness metric *LC*, and (d)-(f) are that between the different failure times (*Time*) and the metric *OE(t)*. After removing the final failure nodes in the cascading failure experiments where the initial failure nodes are 10, we find that both *LC* and *OE(t)* have their own limitations when evaluating the Shanghai TSN robustness. *LC* cannot reflect the effects of passenger flow on Shanghai TSN robustness. *OE(t)* cannot reflect the dynamic effects of passenger flow on the Shanghai TSN robustness at different failure times.

threshold  $\theta_i$  is randomly selected in each experiment, the final failure nodes may be different even if the network is under the same failure mode. Therefore, we record the final failure nodes in each cascading failure experiment and sort the nodes based on the frequency of each failure node in experiments. Then, nodes are removed from the networks based on the above sequence. Based on this, we can evaluate the effect of the above cascading failures on the Shanghai TSN robustness by analyzing the dynamic changes in robustness metrics.

Figure 9 uses *LC* and *OE(t)* to compare the Shanghai TSN robustness under various failure modes at different failure times. The specific method we use is to remove the final failure nodes in the cascading failure experiments where the initial failure nodes are 10. According to the changes in *LC* and *OE(t)*, we find that different failure modes have different effects on Shanghai TSN robustness. For example, Figs. 9(a)-(f) show that the Shanghai TSN is more robust under the random failure mode than other failure modes. Therefore, this finding shows that both *LC* and *OE(t)* can reflect the effects of failure modes on the network robustness to a certain extent. However, even during the rush hours, the changes in *LC* is unapparent when the network is under the degree failure mode, strength failure mode or random failure mode. This is because *LC* is used to evaluate the network robustness from the perspective of network scale, without considering the impact of passenger flows on the network robustness. Actually, although the number of final

failure nodes is the same, the passenger flow volume is changing with time. In particular, for some stations in special functional areas (e.g., the working area), the passenger flow may be concentrated at a certain moment. Therefore, if these stations disrupt at certain times with a large passenger flow, the network robustness will be more affected than that at other times. *OE(t)* evaluates the network robustness from the perspective of passenger flow. The higher the value of *OE(t)* is, the more robust the network is at time *t*. From the changes in *OE(t)*, we can observe that *OE(t)* is small during the off-peak hours. However, since the final failure nodes during the off-peak hours are far fewer than that during rush hours, the conclusion that the Shanghai TSN is the most vulnerable during the off-peak hours is against our normal cognition. Therefore, we propose a new robustness metric *R(t)* to overcome the problems caused by *LC* and *OE(t)* in evaluating the dynamic robustness of Shanghai TSN.

Figure 10 uses *R(t)* to compare the Shanghai TSN robustness under various failure modes at different failure times. Compared with other failure modes, Shanghai TSN under the random failure mode is the more robust. This finding is understandable because the initial failure nodes in other failure modes are either the hub nodes or the nodes with large volume of passenger flow. In addition, with the number of initial failure nodes increasing, the Shanghai TSN robustness downtrend under each failure mode becomes more and more obvious. In particular, when the



**FIGURE 10.** (a)–(i) are the relationship between the robustness metric ( $R(t)$ ) and different failure times (Time) after removing the failed nodes caused by different failure modes. The results show that during the rush hours, the Shanghai TSN robustness is not only related to the failure modes but also to the dynamic distribution of passenger flow. However, during the off-peak hours, the passenger flow has little impact on the Shanghai SN robustness.

number of initial nodes is 30, the Shanghai TSN robustness under the degree failure mode appears the weakest. Meanwhile, as shown in Figs. 10(a), (d), (g), (c), (f) and (i), we can find that the changing trend of the Shanghai TSN robustness varies over time during the rush hours. However, as shown in Figs. 10(b), (e) and (h), during the off-peak hours, the changing trend of the Shanghai TSN robustness is basically a straight line. Therefore, Shanghai TSN is the most robust under the random failure modes during the off-peak hours. These findings suggest that when the networks have a large volume of passenger flow, the Shanghai TSN robustness can vary over time apparently. However, when the passenger flow volume in the networks is comparatively small, the Shanghai TSN robustness basically depends on the change of network topology [56]. Furthermore, through comparing the Shanghai TSN robustness under the same failure mode at different times, we find that the passenger flow volume can increase the impact of failure modes on the Shanghai TSN robustness. When the passenger flow volume is large, it is tougher for Shanghai TSN to tolerate the cascading failure caused by each failure mode.

#### IV. CONCLUSION

To reflect the dynamic effect of passenger flow on the subway network robustness, we first propose a temporal subway network (TSN). Then, we take advantage of the linear threshold (LT) model to characterize the cascading failure process of TSN. In addition, we propose a new robustness metric  $R(t)$  to evaluate the effect of this cascading failure on the TSN robustness. By simulating the above methods to 36 Shanghai TSNs, the main and practical findings of this paper are as follows:

- Different failure modes will cause different cascading failure. The cascading failure rarely occurs in the Shanghai TSN under the random failure mode, but easily occurs under other malicious failure mode (i.e., degree, betweenness, strength, flow betweenness oriented failure modes).
- Under the same failure mode, Shanghai TSN during the rush hours will cause a larger cascading failure than that during the off-peak hours.
- When the passenger flow volume is comparatively small, the Shanghai TSN robustness basically depends

on the change of network topology. Moreover, the large volume of passenger flow can increase the influence of failure modes on the Shanghai TSN robustness.

From the above findings, we state that at different times, it is necessary for us to adopt different measures to manage stations. We need to focus not only on the hub stations of the subway system, but also the stations with a large volume of passenger flow at a certain moment. In particular, during the rush hours, due to the large passenger flow volume, the SN robustness varies over time. Therefore, to improve the SN robustness, we suggest that (1) updating the changes of passenger flow in time to grasp the dynamics of passenger travel patterns [57]; (2) increasing the travel routes between two stations to decrease the edge influence  $El_{ij}$  [58], so as to reduce the cascading failure size; (3) identifying the priority restoration stations based on the passenger flow volume and rail structure [59].

## REFERENCES

- [1] Y. Xing, J. Lu, S. Chen, and S. Dissanayake, "Vulnerability analysis of urban rail transit based on complex network theory: A case study of Shanghai Metro," *Public Transp.*, vol. 9, no. 3, pp. 501–525, Sep. 2017.
- [2] K. He, X. Cheng, S. Zhang, H. Yang, Y. Yao, M. Peng, and W. Cong, "Critical roof opening longitudinal length for complete smoke exhaustion in subway tunnel fires," *Int. J. Thermal Sci.*, vol. 133, pp. 55–61, Nov. 2018.
- [3] Q.-C. Lu and S. Lin, "Vulnerability analysis of urban rail transit network within multi-modal public transport networks," *Sustainability*, vol. 11, no. 7, p. 2109, Apr. 2019.
- [4] M. Li, H. Wang, and H. Wang, "Resilience assessment and optimization for urban rail transit networks: A case study of Beijing subway network," *IEEE Access*, vol. 7, pp. 71221–71234, 2019.
- [5] Y. Li, X. Wang, S. Sun, X. Ma, and G. Lu, "Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 306–328, Apr. 2017.
- [6] M. E. J. Newman, "Network structure from rich but noisy data," *Nature Phys.*, vol. 14, no. 6, pp. 542–545, Mar. 2018.
- [7] N. Galiana, M. Lurgi, B. Claramunt-López, M.-J. Fortin, S. Leroux, K. Cazelles, D. Gravel, and J. M. Montoya, "The spatial scaling of species interaction networks," *Nature Ecol. Evol.*, vol. 2, no. 5, pp. 782–790, Apr. 2018.
- [8] M. Liang, F. Liu, C. Gao, and Z. Zhang, "Robustness analysis of the complex network," in *Proc. 6th Data Driven Control Learn. Syst. (DDCLS)*, May 2017, pp. 638–643.
- [9] S. Derrible and C. Kennedy, "The complexity and robustness of metro networks," *Phys. A, Stat. Mech. Appl.*, vol. 389, no. 17, pp. 3678–3691, Sep. 2010.
- [10] M. L. Mouronte and R. M. Benito, "Structural properties of urban bus and subway networks of Madrid," *Netw. Heterogeneous Media*, vol. 7, no. 3, pp. 415–428, Oct. 2012.
- [11] Y. Deng, Q. Li, and Y. Lu, "A research on subway physical vulnerability based on network theory and FMECA," *Saf. Sci.*, vol. 80, pp. 127–134, Dec. 2015.
- [12] D. (Sun and S. Guan, "Measuring vulnerability of urban metro network from line operation perspective," *Transp. Res. Part A: Policy Pract.*, vol. 94, pp. 348–359, Dec. 2016.
- [13] X.-M. Xiao, L.-M. Jia, and Y.-H. Wang, "Correlation between heterogeneity and vulnerability of subway networks based on passenger flow," *J. Rail Transp. Planning Manage.*, vol. 8, no. 2, pp. 145–157, Sep. 2018.
- [14] F. Ma, F. Liu, K. Yuen, P. Lai, Q. Sun, and X. Li, "Cascading failures and vulnerability evolution in bus-metro complex bilayer networks under rainstorm weather conditions," *Int. J. Environ. Res. Public Health*, vol. 16, no. 3, p. 329, Jan. 2019.
- [15] Y. Shen, G. Ren, and B. Ran, "Cascading failure analysis and robustness optimization of metro networks based on coupled map lattices: A case study of Nanjing, China," *Transportation*, Nov. 2019, doi: 10.1007/s11116-019-10066-y.
- [16] M. Bertolini, M. Bevilacqua, and R. Massini, "FMECA approach to product traceability in the food industry," *Food Control*, vol. 17, no. 2, pp. 137–145, Feb. 2006.
- [17] J. Lu, G. Ren, and L. Xu, "Analysis of subway station distribution capacity based on automatic fare collection data of Nanjing Metro," *J. Transp. Eng. A, Syst.*, vol. 146, no. 2, Feb. 2020, Art. no. 04019067.
- [18] Yu, Bai, Chen, and Yan, "Analysis of space-time variation of passenger flow and commuting characteristics of residents using smart card data of Nanjing Metro," *Sustainability*, vol. 11, no. 18, p. 4989, Sep. 2019.
- [19] C. M. Liu, C. Gao, and Y. C. Xin, "Measuring the diversity and dynamics of mobility patterns using smart card data," in *Knowledge Science, Engineering and Management (Lecture Notes in Computer Science)*. Cham, Switzerland: Springer, 2018, pp. 438–451.
- [20] F. Ma, F. Liu, K. Yuen, P. Lai, Q. Sun, and X. Li, "Cascading failures and vulnerability evolution in bus-metro complex bilayer networks under rainstorm weather conditions," *Int. J. Environ. Res. Public Health*, vol. 16, no. 3, p. 329, Jan. 2019.
- [21] A. E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 66, no. 6, Dec. 2002.
- [22] C. Gao, Z. Su, J. Liu, and J. Kurths, "Even central users do not always drive information diffusion," *Commun. ACM*, vol. 62, no. 2, pp. 61–67, Jan. 2019.
- [23] S. Hong, J. Zhu, L. A. Braunstein, T. Zhao, and Q. You, "Cascading failure and recovery of spatially interdependent networks," *J. Stat. Mech., Theory Exp.*, vol. 2017, no. 10, Oct. 2017, Art. no. 103208.
- [24] S. Hong, B. Wang, X. Ma, J. Wang, and T. Zhao, "Failure cascade in interdependent network with traffic loads," *J. Phys. A, Math. Theor.*, vol. 48, no. 48, Nov. 2015, Art. no. 485101.
- [25] Z.-H. Zhang, Y. Song, L. Xia, Y.-W. Li, L. Zhang, and G.-P. Jiang, "A novel load capacity model with a tunable proportion of load redistribution against cascading failures," *Secur. Commun. Netw.*, vol. 2018, pp. 1–7, Jun. 2018.
- [26] K. Kaneko, "Overview of coupled map lattices," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 2, no. 3, pp. 279–282, Jul. 1992.
- [27] T. Zhu, B. Wang, B. Wu, and C. Zhu, "Maximizing the spread of influence ranking in social networks," *Inf. Sci.*, vol. 278, pp. 535–544, Sep. 2014.
- [28] X. Li, P. Wang, X.-J. Xu, and G. Xiao, "Universal behavior of the linear threshold model on weighted networks," *J. Parallel Distrib. Comput.*, vol. 123, pp. 223–229, Jan. 2019.
- [29] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2003, pp. 137–146.
- [30] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 284–287.
- [31] S. Hong, H. Yang, T. Zhao, and X. Ma, "Epidemic spreading model of complex dynamical network with the heterogeneity of nodes," *Int. J. Syst. Sci.*, vol. 47, no. 11, pp. 2745–2752, Mar. 2015.
- [32] Z. Chen, W.-B. Du, X.-B. Cao, and X.-L. Zhou, "Cascading failure of interdependent networks with different coupling preference under targeted attack," *Chaos, Solitons Fractals*, vol. 80, pp. 7–12, Nov. 2015.
- [33] J. Sienkiewicz and J. A. Holyst, "Statistical analysis of 22 public transport networks in Poland," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 72, no. 4, Oct. 2005, Art. no. 046127.
- [34] Y. Ding and S. I. Chien, "Improving transit service quality and headway regularity with real-time control," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1760, no. 1, pp. 161–170, Jan. 2001.
- [35] X. Zhou, H. Zhang, G. Ji, and G. Tang, "A method to mine movement patterns between zones: A case study of subway commuters in Shanghai," *IEEE Access*, vol. 7, pp. 67795–67806, 2019.
- [36] X. Li, J. Kurths, C. Gao, J. Zhang, Z. Wang, and Z. Zhang, "A hybrid algorithm for estimating origin-destination flows," *IEEE Access*, vol. 6, pp. 677–687, Feb. 2018.
- [37] Y. Zhang, E. Yao, J. Zhang, and K. Zheng, "Estimating metro passengers' path choices by combining self-reported revealed preference and smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 76–89, Jul. 2018.
- [38] G. Jiang, M. Fosgerau, and H. K. Lo, "Route choice, travel time variability, and rational inattention," *Transp. Res. Procedia*, vol. 38, pp. 482–502, 2019.
- [39] B. Bartosiewicz and I. Pieliesiak, "Spatial patterns of travel behaviour in Poland," *Travel Behav. Soc.*, vol. 15, pp. 113–122, Apr. 2019.
- [40] C. Gao and J. Liu, "Network-based modeling for characterizing human collective behaviors during extreme events," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 1, pp. 171–183, Jan. 2017.

[41] V. G. Cerf, D. D. Cowan, R. C. Mullin, and R. G. Stanton, "A lower bound on the average shortest path length in regular graphs," *Networks*, vol. 4, no. 4, pp. 335–342, 1974.

[42] H. Cai, J. Zhu, C. Yang, W. Fan, and T. Xu, "Vulnerability analysis of metro network incorporating flow impact and capacity constraint after a disaster," *J. Urban Planning Develop.*, vol. 143, no. 2, Jun. 2017, Art. no. 04016031.

[43] X.-Y. Xu, J. Liu, H.-Y. Li, and Y.-F. Zhou, "Probabilistic model for remain passenger queues at subway station platform," *J. Central South Univ.*, vol. 20, no. 3, pp. 837–844, Mar. 2013.

[44] R. W. Bohannon and A. Williams Andrews, "Normal walking speed: A descriptive meta-analysis," *Physiotherapy*, vol. 97, no. 3, pp. 182–189, Sep. 2011.

[45] C. Gao, J. Liu, and N. Zhong, "Network immunization and virus propagation in email networks: Experimental evaluation and analysis," *Knowl. Inf. Syst.*, vol. 27, no. 2, pp. 253–279, Jul. 2010.

[46] C. Gao, Z. Wang, X. Li, Z. Zhang, and W. Zeng, "PR-index: Using the h-index and PageRank for determining true impact," *PLoS ONE*, vol. 11, no. 9, Sep. 2016, Art. no. e0161755.

[47] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000.

[48] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Phys. Rev. Lett.*, vol. 87, no. 19, Oct. 2001, Art. no. 198701.

[49] C. Gao and J. Liu, "Modeling and restraining mobile virus propagation," *IEEE Trans. Mobile Comput.*, vol. 12, no. 3, pp. 529–541, Mar. 2013.

[50] X. Li, Z. Wang, C. Gao, and L. Shi, "Reasoning human emotional responses from large-scale social and public media," *Appl. Math. Comput.*, vol. 310, pp. 182–193, Oct. 2017.

[51] J. Zhao, Q. Qu, F. Zhang, C. Xu, and S. Liu, "Spatio-temporal analysis of passenger travel patterns in massive smart card data," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3135–3146, Nov. 2017.

[52] J. Song, C. Zhao, S. Zhong, T. A. S. Nielsen, and A. V. Prishchepov, "Mapping spatio-temporal patterns and detecting the factors of traffic congestion with multi-source data fusion and mining techniques," *Comput., Environ. Urban Syst.*, vol. 77, Sep. 2019, Art. no. 101364.

[53] X. Li, J. Guo, C. Gao, Z. Su, D. Bao, and Z. Zhang, "Network-based transportation system analysis: A case study in a mountain city," *Chaos, Solitons Fractals*, vol. 107, pp. 256–265, Feb. 2018.

[54] D. G. Bonett and T. A. Wright, "Sample size requirements for estimating pearson, Kendall and spearman correlations," *Psychometrika*, vol. 65, no. 1, pp. 23–28, Mar. 2000.

[55] T. Tang, X. Kong, M. Li, J. Wang, G. Shen, and X. Wang, "VISOS: A visual interactive system for spatial-temporal exploring station importance based on subway data," *IEEE Access*, vol. 6, pp. 42131–42141, 2018.

[56] E. Frutos Bernal and A. Martín del Rey, "Study of the structural and robustness characteristics of madrid metro network," *Sustainability*, vol. 11, no. 12, p. 3486, Jun. 2019.

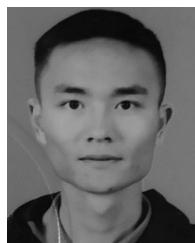
[57] K. Sohn, "Optimizing train-stop positions along a platform to distribute the passenger load more evenly across individual cars," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 994–1002, Jun. 2013.

[58] X.-B. Cao, C. Hong, W.-B. Du, and J. Zhang, "Improving the network robustness against cascading failures by adding links," *Chaos, Solitons Fractals*, vol. 57, pp. 35–40, Dec. 2013.

[59] S. Hong, C. Lv, T. Zhao, B. Wang, J. Wang, and J. Zhu, "Cascading failure analysis and restoration strategy in an interdependent network," *J. Phys. A, Math. Theor.*, vol. 49, no. 19, Apr. 2016, Art. no. 195101.



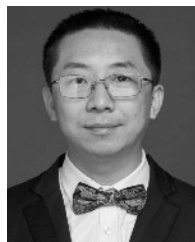
**FAN ZHANG** is currently pursuing the bachelor's degree with the College of Computer and Information Science, Southwest University, Chongqing, China. His research interests include data-driven modeling and statistical analysis.



**SHIHONG JIANG** is currently pursuing the master's degree with the College of Computer and Information Science, Southwest University, Chongqing, China. His research interests include statistical analysis and data-driven modeling.



**CHAO GAO** is currently a Professor with the College of Computer and Information Science, Southwest University, Chongqing, China. He has been a Research Fellow of the Humboldt University of Berlin, Germany, and a Postdoctoral Research Fellow of the Department of Computer Science with Hong Kong Baptist University. His current research interests include complex social networks analysis, nature-inspired computing, and data-driven complex systems modeling.



**ZHANWEI DU** received the Ph.D. degree from the Department of Computer Science and Technology, Jilin University, in 2015. He is currently a Researcher with the Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR. He is also a Postdoctoral Fellow of The University of Texas at Austin. He held a postdoctoral position (Hong Kong scholar) for one year at Hong Kong Baptist University. His current research interests include complex networks, smart city of traffic networks, and epidemic disease propagation.



**ZHEN WANG** received the Ph.D. degree from Hong Kong Baptist University, Hong Kong, in 2014. He is currently a Professor with the Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University. He has authored or coauthored over 100 research articles and four review papers with over 11 000 citations. His current research interests include complex networks, evolutionary game, and data science.



**XIANGHUA LI** received the Ph.D. degree from Jilin University, China, in 2009. She is an Associate Professor with the College of Computer and Information Science, Southwest University, Chongqing, China. Her research interests include nature-inspired computing and statistical analysis.



**YI FAN** is currently pursuing the master's degree with the College of Computer and Information Science, Southwest University, Chongqing, China. His research interests include statistical analysis and complex computational problem solving.