Complex Network Analysis of Transportation Systems

Lead Guest Editor: Massimiliano Zanin
Guest Editors: Xiaoqian Sun and Sebastian Wandelt

Complex Network Analysis of Transportation Systems

Complex Network Analysis of Transportation Systems

Lead Guest Editor: Massimiliano Zanin
Guest Editors: Xiaoqian Sun and Sebastian Wandelt

Hongtai Yang , China
Vincent F. Yu , Taiwan
Mustafa Zeybek, Turkey
Jing Zhao, China
Ming Zhong , China
Yajie Zou , China

Contents

WILEY | Hindawi

Research Article

Vulnerability Analysis of Urban Rail Transit Network considering Cascading Failure Evolution

Ranran Sun [iD],[1,2] **Guangyu Zhu** [iD],[1,2] **Bing Liu** [iD],[1,2] **Xiaolu Li** [iD],[1,2] **Yiyuan Yang** [iD],[1,2]
and Jingxuan Zhang [iD][3]

[1]*Beijing Research Center of Urban Traffic Information Sensing and Service Technologies, Beijing Jiaotong University,
Beijing 100044, China*
[2]*Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport,
Beijing Jiaotong University, Beijing 100044, China*
[3]*Planning and Standard Research Institute of the National Railway Administration of the People's Republic of China,
Beijing 100055, China*

Correspondence should be addressed to Guangyu Zhu; gyzhu@bjtu.edu.cn

Vulnerability analysis is the premise of operational risk management and control for the large-scale and complex urban rail transit
network (URTN) under the operation interruption of important stations. The temporary operation interruption of one station in an
emergency may lead to the cascading failure and the paralysis of the whole URTN due to the load of other stations exceeding the
limited capacity. The priority of important stations is proposed by combining its location and function in URTN. In addition,
focusing on the analysis of the travel behaviour of passengers and the synergy of public transport networks, a novel cascading failure
evolution model is established to simulate the cascading failure process of URTN under different attack scenarios. The vulnerability
indicators are constructed to dynamically evaluate the vulnerability of URTN considering cascading failure evolution, which are
different from the traditional vulnerability indicators based on complex network theory. Taking the Beijing urban rail transit network
as an example, the dynamic simulation results show that the cascading failure of URTN is closely related to the temporal-spatial
distribution of passenger flows and malicious attacks are more destructive than random attacks. Compared with the important
stations with the largest betweenness or degree, the interrupted stations with largest intensity have a greater impact on the operational
stability of URTN. Moreover, increasing the capacity coefficient of the station can reduce the vulnerability of URTN.

1. Introduction

With the rapid development of URTN, the safety operation of
URTN is facing new challenges and requirements. Once a
station in the network fails, other stations may change from
normal to failure when the load exceeded limited capacity due to
the strong spatiotemporal correlation between stations, which
may further cause cascading failure and seriously affect the travel
quality of passengers. Therefore, the studies on the evolution
mechanism of cascading failure in the URTN, identification of
important stations, and dynamic assessment of vulnerability
have become significant in optimizing the structure and im-
proving the efficiency of the URTN.

Robustness indicates the ability of the system to operate
under interference [1], and reliability indicates the de-
pendability of the normal operation of the system under
interference [2]. Vulnerability is widely used in measuring
the performance of networks under external disturbances
[3]. The concept of vulnerability is defined as the affection
degree of external disturbance (e.g., unfavourable weather
and equipment failure) on the performance of a system [4],
which has been used by researches to measure the stability of
various networks, such as the Internet [5], computer net-
works [6], social networks [7], and power networks [8]. In
transportation networks, the vulnerability can be defined as
the sensitivity to external disturbances that may decrease the

service performance [9]. Compared with the robustness and the reliability of the transportation network, vulnerability analysis pays more attention to the impact of external interference on the network.

In transportation networks, the studies on vulnerability assessment can be mainly divided into two categories: static and dynamic. Static vulnerability assessment refers to the analysis of the network topology, efficiency, and connectivity under a certain failure state, without considering the impact of external disturbances on the passenger travel behaviour and the propagation process of the passenger flows. On this topic, researchers often use the indices of the number of normal and failed components [10], network efficiency [11], and connectivity [12, 13] to measure the performance of the transportation network during the failure of edges, one or more stations. Many complex systems can be modeled as complex networks for analysis. At present, the research on complex networks has attracted extensive attention and research. When considering network characteristics, typical characteristics are usually used to measure the network. Liu and Song [14] introduced the concepts of degree, clustering coefficient, and average shortest path length in complex network theory into URNT to analyze the network accessibility and traffic efficiency under the failure of transfer stations. Zhang et al. [15] analysed the topological and functional characteristics of the Shanghai subway network and proposed to evaluate the function and connectivity of traffic network based on the function loss parameters and connectivity parameters. Combining the topology of URTN and the characteristics of passenger flow distribution, Lu [16] investigated the cumulative affected traffic flows caused by external disturbances and quantified the vulnerability of the network under different attacks. They also pointed out that the failure of the transfer station is less effective in the resilience of the network. Yang et al. [17] used the relative size of maximum connected subgraph (RSMCS) and Global Network Efficiency (GNE), two common indicators in the field of complex network analysis, to comprehensively evaluate the vulnerability of Beijing urban rail transit network (BRTN) under random and malicious attack strategies. Liu et al. [13] defined the fraction of acceptable trip and total generalized travel cost increase ratio to analyze the impact of link capacity reduction on the performance of urban rail transit network. Cats and Krishnakumari [18] deleted the nodes or edges of URTN according to the order of degree or betweenness. The relative size of the largest connected sunetwork when removing nodes or edges and the largest subnetwork of the original network and the normalized average shortest path are used to describe the change of network performance. Chen et al. [19] used the demand-impedance index to describe the performance curve of URTN in the process of external attack and repair. The results show that BRTN is a scale-free network, and malicious attack is more destructive than random attack. However, when the URTN is disturbed by external disturbance, the analysis of vulnerability cannot be limited to the network topology [20]. Because the external disturbance will also have an impact on travel decision making of passengers. Therefore, it is difficult to reflect the strong

spatiotemporal correlation between nodes and the relationship between external interference and passenger travel behaviour only using evaluation method for static vulnerability.

Dynamic vulnerability assessment refers to the evaluation of network stability, which considers the spatiotemporal dynamic changes of the network structure and the passengers within a certain failure duration [21]. The dynamic change of the passenger distribution and the topology of URTN can be described by cascading failure process during interruption time [22]. Sun et al. [9] evaluated the cascading failure process of URTN based on the coupled map lattice model and found that the failure was difficult to control when the loop was attacked. Liu et al. [23] built a CASCADE model with designing a traffic redistribution method based on the edge weight function, to analyze the survivability of URTN from the perspective of the scale of failure and degree of damage. Xing et al. [24] introduced the ORNL-PSERC-Alaska (OPA) model to describe the cascading failure phenomenon of urban regional transportation network under random attack. The essence of cascading failure is that when one or more nodes fail, the load will be redistributed, which may cause other nodes to fail because the load exceeds their limited capacity. Therefore, the proposed capacity-load model [25] is often used to describe cascading failures. The crux of cascading failure process analysis is load redistribution [26]. In the past, the load of the network was redistributed by constructing a load function of betweenness or degree [27, 28]. With the increasing interest, researchers gradually introduced the distribution characteristics of the passengers into the redistribution during the evolution process of cascading failure. Wu et al. [29] evaluated the survivability of weighted road traffic network considering cascading failure based on the user equilibrium model. Zhang and Wang [30] proposed a method of nonuniform load distribution to the neighborhood and evaluated the survivability of Nanjing Subway Network cascading failure based on the capacity-load model. In addition, Szymula and Bešinović [31] proposed a new network vulnerability model to evaluate the vulnerability of the railway system by finding out the combination of key links that cause the most adverse consequences to passengers and trains. Vulnerability is affected by the total travel cost of passengers, the number of passengers unable to reach the destination, and the cost of train service adjustment.

The researches of the above scholars provide extremely valuable results for the modeling of cascading failure and vulnerability analysis of URTN and provide a theoretical framework for subsequent research. Considering these findings, we found that the following points need attention:

(i) The existing vulnerability research methods of URTN are usually limited to URTN. In the real world, URTN is coupled with other networks such as bus network, which does not exist in isolation. If the coordination between urban public transport networks is not considered, the impact of emergencies on the operation of URTN is difficult to be accurately evaluated.

(ii) The key to judge whether the cascading failure of URNT occurs is to estimate whether the load of the station exceeds its limited capacity. In the past, most studies used departure frequency and topology metrics for complex networks to define the load of nodes and edges. Although the existing capacity-load model considers the passenger flow of URTN as the load of nodes or edges, the complex travel behavior of passengers under emergencies is not considered in the cascade failure process. Therefore, it is still difficult to assess the dynamic vulnerability of the URTN under emergencies.

(iii) The problem of passenger redistribution is a research difficulty when a station is interrupted due to the disturbance. URTN serves the heterogeneous passengers, and their travel behaviour is complex and changeable when emergencies occur. Simply loading the passengers of adjacent stations according to the proportion of topology metrics for complex networks is difficult to meet the needs of the actual network.

The sequence data of passenger flow distribution of urban rail transit and other public transport are correlated in time and space dimensions, and the affected passengers due to the failure station may change their travel modes and travel paths, which causes the passengers in URTN transferred to other public transportation networks. Therefore, different from previous studies, the main contribution of this study is considered the coordination between the URTN and other public transport networks and the complex travel behaviour of affected passengers. During the analysis of cascade failure evolution, the edge weight function considering the capacity coefficient of passenger flow is used to define the states of stations. In the process of cascading failure, the repairability of stations is considered, that is, with the real-time change of network load, the failure state of each station of urban rail transit network changes, which is different from the previous situation that once the station is interrupted; it will be in failure state for a long time. In addition, the vulnerability indicators are constructed to dynamically evaluate the vulnerability of URTN considering cascading failure evolution when the station is attacked randomly or maliciously, which are different from the traditional vulnerability indicators (such as degree, clustering coefficient and average shortest path length) based on complex network theory.

This article is mainly divided into the following parts: Section 2 defines various attacks and analyses passenger travel behaviour under emergencies. Section 3 constructs the vulnerability evaluation model of URTN, including the cascading failure model and the establishment of the indicators for dynamic vulnerability assessment. Section 4 takes the BRTN as an example to analyze the dynamic vulnerability under the different attack scenarios. Finally, Section 5 summarizes our conclusions.

2. The Definition of Various Attacks and Travel Behaviour Analysis of Affected Passengers

Once a station of urban rail transit system is attacked, it is obviously characterized by high uncertainty, easy to spread, and trigger a chain reaction. At the same time, when the station is interrupted due to the attack, the travel choice behavior of passengers affected by the operation interruption will change. The behavior analysis of passengers when the station is attacked is the need of vulnerability analysis of the URTN.

2.1. The Definition of Various Attacks. Urban rail transit system plays a crucial role in urban public transport system, providing fast and convenient transportation services for the public. It has the characteristics of high construction requirements, high technical complexity, closed passenger transport environment, high operation intensity, and network operation. URTN usually encounters two kinds of attack events, namely, random attack and malicious attack [32], as shown in Figure 1.

In real life, it is difficult to quantitatively analyze the breaking force of the attack on URTN. Therefore, in this article, we define random attack as a random failure of a station in URTN, and the probability of random attack at all stations is equal. Malicious attacks are defined as targeted and destructive attacks, which often occur in stations with large traffic and vital stations in the network.

Under the action of attack events, the service of the station is often interrupted. The interrupted station will lead to the failure of other stations through the connection relationship between stations, resulting in cascading failure. It should be noted that the attack may take place at the stations or within the train operating line of the URTN. In addition, the failure of one position may also lead to the shutdown of the entire line [18]. Therefore, when studying the vulnerability of URTN under different attack scenarios, it is necessary to analyze specific scenarios.

Malicious attacks often occur at pivotal stations of URTN. Timely and efficient management and control of key stations under emergencies will help to improve the vulnerability of URTN. In this article, the important station of the URTN refers to the station that can cause the cascading failure and may have a large disruptive impact on the network. In order to identify the damage strength of the failed station to the network in a specific emergency scenario, it is necessary to traverse each node in the network for cascading failure analysis. By analyzing the damage degree of urban rail transit system, we can identify the pivotal stations in the network. For multiple stations, we also need to study each combination to judge the relative importance of the station [33]. However, this method is inefficient, and it is widely known that the important stations often play a key role in the structure and function of the actual URTN. In graph theory and network analysis, centrality is an index to
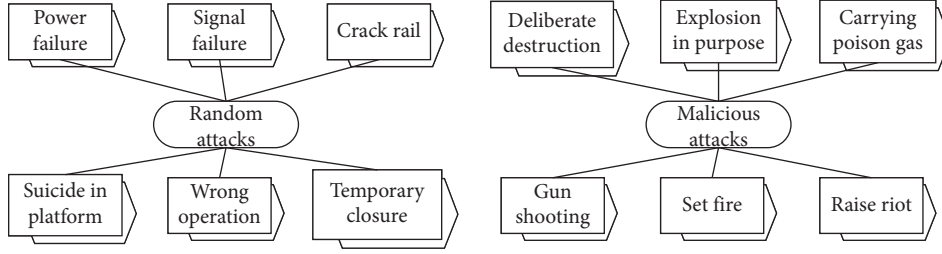
Figure 1: Examples of attack types.

judge the importance/influence of nodes in a network, mainly including degree and betweenness. In addition, the main object of urban rail transit service is passengers, that is, the number of transported passengers also determines whether the station plays a central role. Therefore, this article proposes to evaluate the importance of stations by degree, betweenness, and the intensity of the node.

Before calculating the indicators to evaluate the importance of the station, we first model the URTN based on L-space [34]. L-space method refers to a network modeling method in which stations are regarded as nodes and the connections between adjacent stations are regarded as edges. P-space [35] method regards the station as a node and the line relationship to which the station belongs as an edge. When multiple stations belong to the same operation line, any two stations are connected by one edge. R-space [36] method regards the connecting edges between stations as nodes and stations as edges. Through comparative analysis, it can be concluded that the physical significance of the characteristic parameters of P-space and R-space spatial models is relatively specific, which mainly reflects the transfer characteristics of URTN. However, L-space method can graphically reflect the corresponding relationship between nodes and edges, which is helpful to analyze the structural and functional characteristics of URTN. It is also convenient to integrate passenger travel information into the network and analyze the dynamic evolution characteristics of URTN under emergencies.

According to the following three evaluation indexes, we can rank the importance of the stations in the URTN.

(1) Node Degree. Node degree reflects the association of each node with other nodes in URTN, which is an important attribute of the node and reflects the embodiment of connectivity of the node.

$$k_i = \sum_{j \in N, j \neq i} \sigma_{ij}, \qquad (1)$$

where k_i is the degree of node i. σ_{ij} represents the adjacency relationship between nodes i, j. N is the total number of nodes in the network. If there are directly connected edges between nodes i and j, $\sigma_{ij} = 1$, otherwise, $\sigma_{ij} = 0$.

(2) Node Betweenness. Node betweenness reflects the number of shortest paths through a node in a network and reflects the importance of node as a "bridge," which is a measure of graph centrality based on shortest path.

$$B_i = \sum_{o,d \in N} \frac{g_i(L_{\text{short}}^{o,d})}{g(L_{\text{short}}^{o,d})}, \qquad (2)$$

where B_i is the betweenness of node i. $g_i(L_{\text{short}}^{o,d})$ is the number of shortest paths through node i between o and d. $g(L_{\text{short}}^{o,d})$ is the number of shortest paths between o and d.

(3) Intensity of Node. The service subject of URTN is the passengers, so the intensity of node is introduced to reflect the passenger transport capacity of each station. The intensity of node is the sum of weights of edge directly connected to the node.

$$\varpi_i = \sum_{j=1, j \neq i}^{N} \sigma_{ij} \lambda_{ij}, \qquad (3)$$

where ϖ_i is the intensity of node i. λ_{ij} is the weight that is defined as the passenger flow on e_{ij} of the direct connective edge of node i and j.

2.2. Travel Behaviour Analysis of Affected Passengers. In the normal operation of the URTN, passengers may choose a relatively familiar route to travel, but when the network is interrupted due to external interference, the travel choice behaviour of passengers affected by the operation interruption will change. Part of passengers will be transferred to other modes of transportation, such as bus, which requires flexible service modes to adapt to the changing passenger demand [37]. If the interrupted station is the origin or the destination, passengers may cancel the trip or change the origin or the destination of the trip. When the fault occurs at the intermediate station of a trip, passengers can change the travel mode or path.

In this article, passengers whose travel path contains invalid stations are called affected passengers and the passenger travel path includes urban rail transit path, bus path, and the combined path. By analyzing the travel behaviour of passengers under emergencies, this article divides the travel schemes of passengers under emergencies into four categories: waiting for the station to recover (P_1), reselecting the shortest path of rail transit (P_2), choosing the shortest bus path (P_3), and choosing the combined mode of bus and rail transit (P_4), as shown in Figure 2.

In addition, before modeling passenger travel behaviour, we first put forward the following assumptions:

Origin — Wait for the station to recover

Destination --→ Reselect the shortest path of rail transit

Failed station — Choose the shortest bus path

Bus station --→ Choose the combined mode of road bus and rail transit
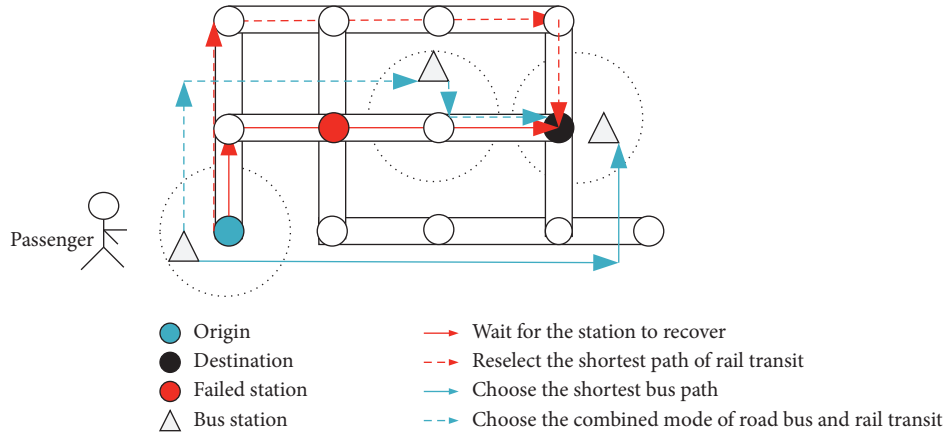
FIGURE 2: Travel schemes of passengers.

(i) The travel demands of passengers during the study period are all rigid, i.e., passengers will not cancel their trips even if the operation is interrupted. In addition, the manager will not change the vehicle operation plan of the stations not affected by the interruption.

(ii) During the entire period from interruption to recovery, passengers of all stations can complete their trips through the corresponding routes except the failed station.

(iii) Passengers rely on public transport, namely, only considering strong dependence and correlation between the URTN and the bus network.

(iv) It is assumed that passengers will select the shortest path in each scheme and then determine the probability of finally selecting a travel scheme according to the utility function.

When studying the vulnerability of URTN, the traditional method only redistributes the affected passengers by invalid stations to other rail transit paths, ignoring the diversity of passenger travel choice behaviour, which is inconsistent with the actual situation [38]. After the station operation is interrupted, passengers may choose other travel paths and modes, resulting in the loss of some passengers. However, at the same time, the lost passenger flow will not be redistributed to other rail transit feasible paths, thus reducing the failure risk of other stations. In order to simplify the model and reduce the complexity of calculation, this article only considers two travel modes: rail transit and bus travel.

The research on passenger travel choice behaviour is mainly based on utility function. The discrete choice model based on utility theory can make a scientific and reasonable explanation for specific behaviour decisions, such as the choice between multipath, waiting, or changing travel mode under emergencies. The multinomial logit model (MNL) [39] is actually a model describing probability selection, which can obtain the travel probability of different modes of transportation through the utility function.

MNL model is the basic form of the logit model. Based on the theory of probability, the MNL model with j options can be expressed as follows: $\partial_j = \exp(bV_j)/\sum_j \exp(bV_j)$, ∂_j represents the probability of choosing the jth option and b is the parameter.

The judgment of travel utility is mainly based on experience. Considering the impact of travel time, travel cost, transfer times, and station failure time on the travel scheme, the utility function of each travel scheme for each passenger is constructed as follows:

$$V\left(P_j^{o,d}\right) = \alpha_1 T_j^{o,d} + \alpha_2 C_j^{o,d} + \alpha_3 H_j^{o,d} + \alpha_4 T_e, \quad j = 1, 2, 3, 4, \tag{4}$$

where $V(P_j^{o,d})$ is the utility function of the passengers choose the j th travel scheme between o and d. α_i is the model coefficient, $i = 1, 2, \ldots, 4$, which can be calibrated by the data of resident travel survey. $T_j^{o,d}$ and $C_j^{o,d}$ represent the travel time and cost of selecting the j th travel scheme, respectively. $H_j^{o,d}$ and T_e represent the transfer times of the jth travel scheme and the estimated duration of operational disruption of a station in the route, respectively. It should be noted that when passengers do not choose to wait for the station to recover, $\alpha_4 = 0$.

$$\partial\left(V\left(P_j^{o,d}\right)\right) = \frac{\exp\left(V\left(P_j^{o,d}\right)\right)}{\sum_j \exp\left(V\left(P_j^{o,d}\right)\right)}, \tag{5}$$

where $\partial(V(P_j^{o,d}))$ represents the probability of choosing the jth travel scheme to travel for each passenger between o and d.

When the station fails, we need to obtain the range of affected passengers and further analyze the changes of their travel behaviour. Due to the different location of the failed station in the rail transit path, the scope of searching the feasible passenger travel path is different. Combined with the actual situation, this article sets the scope of searching the feasible passenger travel path that different affected passengers can choose to reduce the complexity of searching travel path in the process of allocating passenger flow:

(1) When the failure station is the origin of the travel path, affected passengers who start from this station cannot directly enter the URTN system. These passengers can travel through three schemes: choosing the initial shortest rail transit path $M_{\text{shortest}}^{o,d}$ and waiting for the station to recover, taking the shortest bus path directly and choosing the combined mode of bus and rail transit. These travel schemes of affected passengers can be expressed as follows:

$$P_j^{o,d} = \begin{cases} M_{\text{shortest}}^{o,d}, \\ B_{\text{shortest}}^{o',d'}, \\ \min\left\{ B^{o',i'} + M^{i,d} \right\}, \end{cases} \quad (6)$$

where o and d represent the origin and destination of one trip, respectively. o' and d' represent the nearest bus stop to o and d, respectively. i represents a new origin rail station for passengers to reenter the rail transit system. Similarly, i' stands for the nearest bus stop to i. $P_j^{o,d}$ represents the travel scheme for the affected passengers. $B^{o'i'}$ is the bus path from o' to i'. $M^{i,d}$ is the rail transit path from the new origin station to the initial destination.

(2) When the failure station is in the middle of the rail transit path, affected passengers can complete their trips by the following ways. If there are other rail transit paths between origin and destination, passengers may choose another subshortest rail transit path or wait for the station to recover. If there is already no rail transit path between the origin and the destination, passengers may choose bus alone or the combined mode of bus and rail transit. In this scenario, the travel schemes for these affected passengers to complete this trip can be expressed as follows:

$$P_j^{o,d} = \begin{cases} M_{\text{shortest}}^{o,d}, \\ M_{\text{sub-shortest}}^{o,d}, \\ B_{\text{shortest}}^{o',d'}, \\ \min\left\{ B^{o',i'} + M^{i,d}, M^{o,i} + B^{i',d'} \right\}, \end{cases} \quad (7)$$

where $M_{\text{sub-shortest}}^{o,d}$ represents the shortest path of rail transit reselected bypassing the failure station. Station i is a reselected transfer station in the path between the origin station and the destination. $M^{o,i}$ and $B^{i',d'}$ are the rail transit path from the origin o to the station i and bus path from the bus stop i' nearest i to bus stop d' nearest the destination d.

(3) When the failure station is the destination of the trip, affected passengers cannot directly arrive at the destination by urban rail transit. In this scenario, passengers may wait for the station to recover. In addition, passengers may change travel mode and choose the shortest bus path or choose the shortest combined path of urban rail transit and bus. The travel schemes for these passengers to complete their trip can be expressed as follows:

$$P_j^{o,d} = \begin{cases} M_{\text{shortest}}^{o,d}, \\ B_{\text{shortest}}^{o',d'}, \\ \min\left\{ M^{o,i} + B^{i',d'} \right\}. \end{cases} \quad (8)$$

K-shortest path algorithm (KSP) [40] is mainly used to search the shortest path of the above passenger travel schemes. In the process of travel, when an emergency occurs at a station, in addition to waiting for operation recovery, passengers also hope to get decision-making references, such as suboptimal and suboptimal routes. Therefore, it is necessary to extend the shortest path problem, which is called K shortest paths. Assuming that $k = 3$, take Figure 3 as an example to explain the behaviour of passenger selecting the travel path under emergencies. The "K" here is used to determine the number of paths searched by the algorithm. Under the normal operation of urban rail transit, 1-2-5 is the shortest path of rail transit, 1-3-5 is the second short path, and 1-2-4-5 is the third shortest path, according to the data of urban rail transit network. If the station 2 is suddenly interrupted, passengers can still choose 1-2-5 after the station is restored and choose the subshortest path bypassing the interrupted station. In addition, passengers can consider bus travel. Passengers can choose a bus station nearest to the rail transit station to assist in completing the trip. According to the data of bus network and URTN, the same method is adopted to obtain the shortest bus path of passengers and the shortest path combining bus and rail transit. Therefore, the optional path schemes for passengers can be determined under emergencies.

According to the shortest path of each scheme searched, the travel time, travel fare, and the transfer times of the shortest path can be calculated. In this article, the calculation of travel time ignores the walking time of passengers and only considers the operation time of vehicles. There is a fixed timetable for urban rail transit operation, which can easily obtain the operation time between stations. The travel time of bus is determined by the average link travel time by estimating approach with the bus operation data. In addition, the travel cost is determined by the ticket price.

3. Vulnerability Assessment of the URTN Based on the Cascading Failure Evolution Model

In this article, the vulnerability assessment of URTN is mainly divided into two parts: The cascading failure evolution rules are defined based on the constructed capacity-load model of URTN, and the redistribution algorithm of passenger flow in the cascade failure process is designed. Then, the vulnerability assessment index is constructed, and the vulnerability level is calculated by weighted analysis.
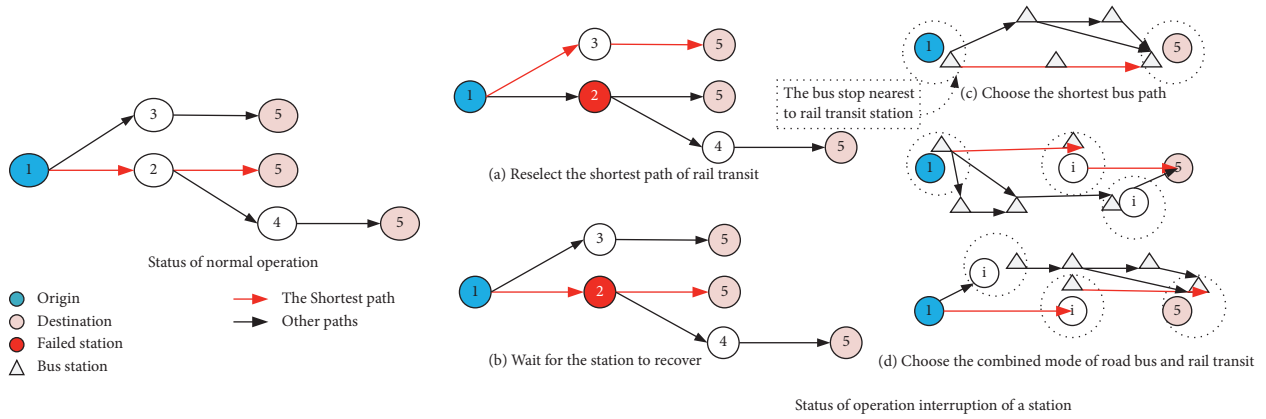
FIGURE 3: Passenger travel paths.

3.1. Cascading Failure Evolution Model

3.1.1. The Capacity-Load Model.
Whether a station in the URTN is in the state of normal operation, by judging whether its load exceeds its capacity, if the load of the station is less than its capacity, the station is in the normal state; otherwise, it is in the failure state. In this article, the station capacity is defined as the maximum number of passengers per unit time. The capacity-load model suitable for URTN is proposed based on the research of Corman and D'ariano [41].

$$C_i = (1 + \varepsilon) \sum_{j=1, j\neq i}^{N} \sigma_{ij} \lambda_{ij}, \qquad (9)$$

where C_i is the capacity of station i, ε is the capacity co-efficient, which is used to adjust the station capacity. Because in the actual operation of urban rail transit, a certain amount of passenger flow overload is allowed. δ_{ij} represents the relationship of connection between station i and station j. If there is a direct connection edge between station i and j, $\sigma_{ij} = 1$; otherwise $\sigma_{ij} = 1$. λ_{ij} represents the initial flow of edge e_{ij}.

For example, the intensity of station i in Figure 4 is $\omega_i = \lambda_{ai} + \lambda_{ci} + \lambda_{di} + \lambda_{ei}$. The capacity of station i is: $C_i = (1 + \varepsilon)\omega_i$. If the intensity is less than the capacity of the station, the station is in the normal operation state. On the contrary, it is defined that the station changes from the normal state to the failure state, and the affected passenger flow generated by the failure station is redistributed according to the travel decision-making behaviour. It should be noted in this article that although the station whose capacity exceeds its load is regarded as a failure state, it still has the ability to transport passengers.

3.1.2. Algorithm of Passenger Flow Redistribution.
In case of an emergency at a station, if the operation is interrupted, the station is in the failure state. The affected passenger flow selects different travel schemes according to the travel utility, that is, passenger flow redistribution. In the process of passenger redistribution, when the intensity of other stations is greater than its maximum carrying capacity, the
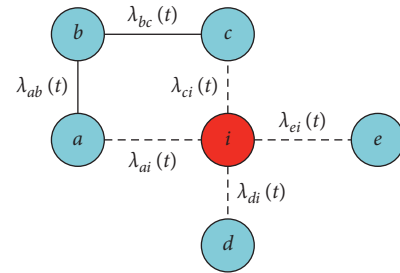
FIGURE 4: The analysis of traffic redistribution algorithm.

transportation capacity of the station is also regarded as invalid. The passengers affected by the failure station shall participate in the next redistribution.

Therefore, passenger flow redistribution is the key to the cascading failure of URTN, and the main algorithms are user equilibrium assignment, system optimal assignment model, and all-or-nothing allocation algorithm. User equilibrium assignment is based on passengers knowing exactly the traffic state of the network and choosing the shortest path. The optimal allocation of the system requires the passengers on the road network to be allocated according to the minimum average cost or total cost of the system. The all-or-nothing allocation algorithm does not consider the congestion of the road network. It directly allocates the passengers in each OD pair to the shortest path, which is the most basic allocation algorithm. The calculation is quite simple and only needs to be completed at one time. In the emergency scenario, the affected passengers need to make travel decisions again according to their travel utility and tend to choose an optimal travel path of each travel scheme. Due to the heavy calculation in the process of passenger assignment in large URTN, this article proposes a passenger redistribution algorithm based on all-or-nothing allocation algorithm. The specific steps are as follows:

(1) *Basic Preparation.* Firstly, determine the station set $N = \{1, 2, \ldots, n\}$ and edge set $E = \{1, 2, \ldots, e\}$. Then, according to the AFC data (data collected by automatic fare collection system), the passengers of each OD pair in the initial network is obtained. By searching the shortest rail transit path of each OD

pair and distributing the initial passengers of each OD pair and distributing the initial passengers of each OD pair to each edge of the URTN, the passengers of each edge λ_{ij} can be obtained. According to the passengers of each edge, the initial intensity ϖ_k and maximum capacity of each station C_k can be obtained, $k \in N$. According to the location of the emergency, determine the initial operation interruption station i and estimate the operation interruption duration of the station T_e.

(2) Extract the passengers whose shortest rail transit path contains the station in abnormal operation state and determine the affected passengers $q_i^{o,d}(t)$. According to Section 2.2, search the travel options $P_j^{o,d}(t)$ for passengers. Then, according to the utility function $V(P_j^{o,d}, t)$, calculate the probability $\sigma_{P_j}^{o,d}(t)$

and affected passengers of each OD pair $q_i^{o,d}(t)\sigma_{P_j}^{o,d}(t)$, choosing the travel scheme $P_j^{o,d}(t)$. In addition, update the travel path of passengers.

(3) Determine the time step Δt. Extract the affected passengers who still choose urban rail transit (including the combined travel mode of bus and rail transit) and redistribute the affected passengers according to the time $T_{P_j}^{o,b}(\cdot)$ of passing through the station of URTN.

Using Figure 4 as an example to illustrate the redistribution algorithm of affected passengers. When station i is the failure station, the weight of the edge connecting the station a and the station b at time $t + \Delta t$ is given as

$$\lambda_{ab}(t+\Delta t) = \begin{cases} \lambda_{ab}(t), & t+\Delta t < T_{P_j}^{o,b}(a), \\ \lambda_{ab}(t) + \sum_o \sum_d \sum_j \mu_{P_j}^{o,d}(t)\delta_{P_j}^{o,d}(e_{ab},t)q_i^{o,d}(t)\sigma_{P_j}^{o,d}(t), & t+\Delta t \geq T_{P_j}^{o,b}(a), \end{cases} \quad (10)$$

where $\lambda_{ab}(t)$ and $\lambda_{ab}(t+\Delta t)$ are the weight of directly connected edges between the station a and the station b at time t and after passenger flow redistribution loading at time $t + \Delta t$, respectively. $\mu_{P_j}^{o,d}(t)$ indicates whether the travel path P_j includes the rail transit route. If P_j contains rail transit route, then $\mu_{P_j}^{o,d}(t) = 1$; otherwise, $\mu_{P_j}^{o,d} = 0$. $\delta_{P_j}^{o,d}(e_{ab},t)$ indicates whether edge e_{ab} will be passed. If the P_j will pass through edge e_{ab}, $\delta_{P_j}^{o,d}(e_{ab},t) = 1$; otherwise $\delta_{P_j}^{o,d}(e_{ab},t) = 0$. $q_i^{o,d}(t)$ represents the affected passenger flow between station o and d when the station i failed. $\sigma_{P_j}^{o,d}(t)$ is the probability that passengers between station o and station d choose the path P_j. $T_{P_j}^{o,b}(a)$ is the time of passengers choosing the path P_j leave the station a.

(4) Compare the sizes of $\varpi_k(t+\Delta t)$ and $C_k(t+\Delta t)$. If $\exists \varpi_k(t+\Delta t) > C_k(t+\Delta t)$, then consider k as a failure station and record the set of failure stations $V(t+\Delta t)$. The reaffected passengers will be calculated and redistributed at the next time step. Let $t = t + \Delta t$, repeat steps (2–4). Otherwise, the process of passenger redistribution ends. In addition, it should be noted that when $t + \Delta t < T_{P_j}^{o,b}(a)$, the affected but not reaffected passengers passing the station a is not redistributed at time $t + \Delta t$ and will be judged whether to complete the redistribution at the next time step according to the time $T_{P_j}^{o,b}(\cdot)$.

According to the above analysis, during the cascade failure evolution of URTN, the passenger flow redistribution algorithm is as shown in Figure 5:

3.2. Construction of Vulnerability Assessment Indicators.
Previous studies mainly used the metrics of complex network topology to evaluate the vulnerability of URTN under

different attack strategies, such as the size of the largest connected subnetwork when removing the failure stations. However, the dynamic vulnerability is not only related to the size of the affected area of URTN but also related to its functional characteristics. After the station is interrupted, the travel choice behaviour of passengers is affected, which is easy to cause local or large-scale traffic congestion and seriously affect the operation efficiency of the network. Therefore, this article establishes the vulnerability evaluation index of URTN by analyzing the travel behaviour and congestion propagation process of passengers under emergencies.

3.2.1. Loss Flow Ratio.
The affected passengers due to the failure station may change their travel modes and travel paths, which causes the passenger flow in URTN transferred to other public transportation networks. This article defines the passenger flow transferred from the URTN to bus networks, which is the loss flow, and the loss flow ratio is defined as the loss flow to the total passenger flow of the URTN.

$$\eta_1(t) = \frac{R_0 - R_1(t)}{R_0}, \quad (11)$$

where $\eta_1(t)$ is the loss flow ratio of URTN. $R_1(t)$ and R_0 are the passengers flow to be loaded on the URTN at time t and the total passengers of the URTN, respectively.

3.2.2. Ratio of Node Failure.
The ratio of node failure refers to the ratio of the number of failure stations to the total number of network stations in the process of cascading failures in URTN at time t.
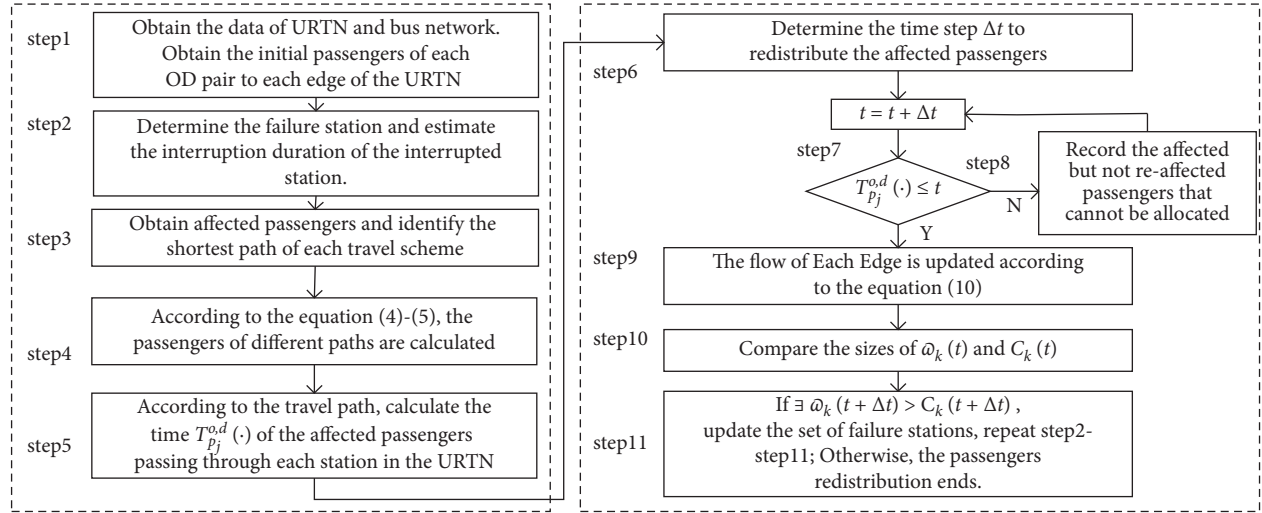
FIGURE 5: Passenger flow redistribution algorithm.

$$\eta_2(t) = \frac{N_1(t)}{N_0}, \tag{12}$$

where $\eta_2(t)$ is the ratio of node failure. N_0 and $N_1(t)$ are the total number of stations and failure stations at time t, respectively.

The ratio of node failure reflects the change of the affected area of URTN, and the ratio of passengers lost reflects the change of the transportation function of the URTN. Therefore, in order to comprehensively assess the vulnerability and identify the vulnerable stations in the URTN, this article constructs a weighted comprehensive vulnerability assessment index.

$$\eta_3 = \omega_1\eta_1 + \omega_2\eta_2, \tag{13}$$

where ω_1 and ω_2 are the weight of η_1 and η_2, respectively. The larger the value of η_3, it indicates that the attacked station is easy to lead to the failure of other stations, that is, the vulnerability of the station is high, and the vulnerability of urban rail transit network is also high.

Combined with the characteristics of the structure and function of the URTN, the assessment index of vulnerability is constructed. The algorithm for the vulnerability assessment is shown in Figure 6. The load of each station is calculated based on the redistribution of passengers. The presence of a failed station is determined by whether the load of any station exceeds its capacity.

4. Case Study

This method of vulnerability assessment is applied to the Beijing Rail Transit Network (BRTN) in January 2016. The BRTN model is composed of 16 operation lines and 262 stations, which is built based on L-space, as shown in Figure 7(a).

4.1. Preparatory Work.
At first, the OD distribution flows of the entire network during the period of 7 : 00–9 : 00 is calculated based on the operation data of the BRTN on January 21, 2016. According to the KSP search algorithm, we get the k-shortest path for passengers of each OD pair, $K = 3$. Selecting the shortest path and assigning the initial passengers of each OD pair to each edge of the BRTN according to the all-or-nothing allocation algorithm as show in Figure 7(b), the initial intensity of each station can be obtained as shown in Figure 7(c).

4.2. Set the Scene of Various Attacks on the Station.
In the actual scene, emergencies often occur in a specific station, resulting in the operation interruption of the station. According to the calculation method of the importance of the station designed in Section 2.1, the degree, betweenness, and intensity of the stations are calculated by formulation (1)–(3) firstly and the ranking results corresponding to the top 5 are given, as shown in Table 1. It can be seen that Xizhimen has the largest degree and betweenness. The betweenness of Dongdan is the largest except Xizhimen Station. Therefore, Xizhimen and Dongdan are the structural important station in the BRTN. The intensity of Guomao accounts for about 3% of the intensity of all stations in the network, which indicates that Guomao is the station with the largest intensity.

This article selects Xizhimen with the largest degree, Dongdan with the largest betweenness, and Guomao with the largest intensity as three different types of attack objects to study cascading failure mechanism and network vulnerability dynamic evaluation. Combined with the survey of Beijing residents, the MNL model is calibrated using the least square method. The parameter calibration results are as follows: $\alpha_1 = -0.115$, $\alpha_2 = -0.145$, $\alpha_3 = -0.313$, $\alpha_4 = -0.086$, respectively. Assuming that the interruption time of the attacked station is $T_d = 30\,\text{min}$, the vulnerability of BRTN under malicious attack and random attack of three types of important stations is simulated and analysed.

After determining the attacked station, extract the range of affected passengers, that is, extracting the passengers whose shortest rail transit path contains the attacked station.
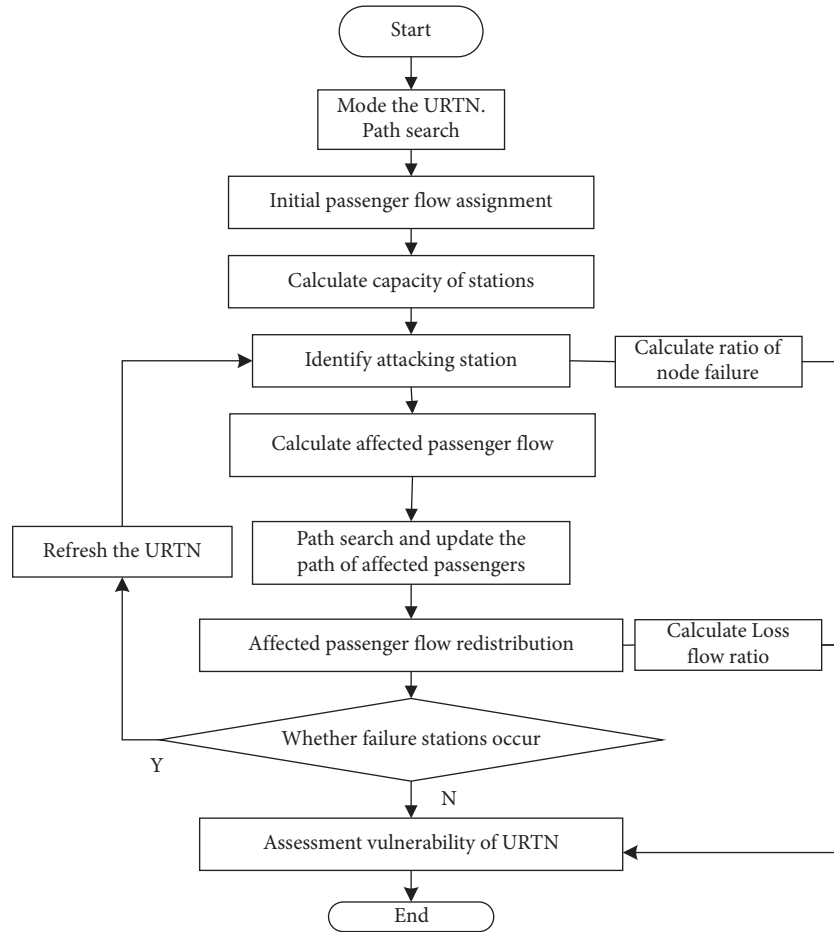
Figure 6: Vulnerability assessment algorithm.
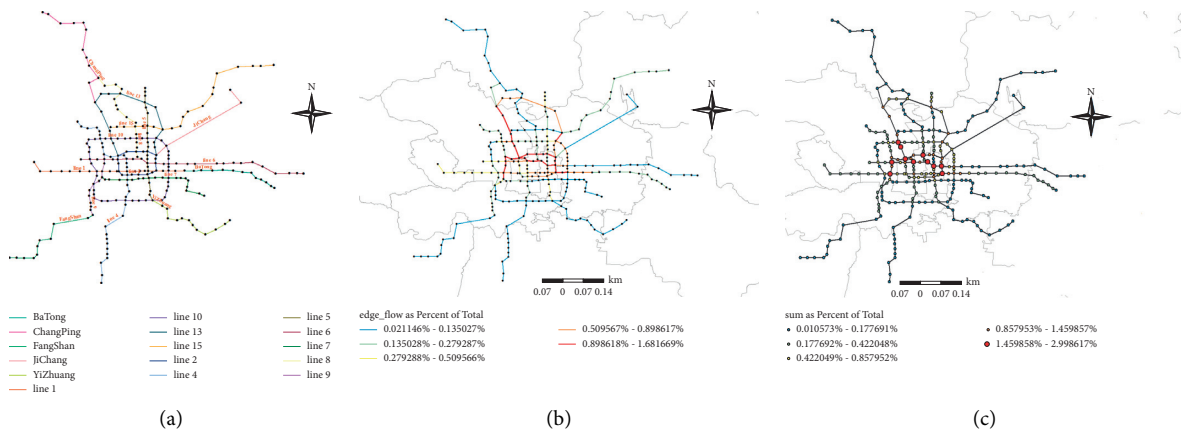
Figure 7: Topology and initial passengers of Beijing subway network (a–c).

Table 1: Results of top 5 importance nodes.

No.	Station	Degree	No.	Station	Betweenness	No.	Station	Intensity (%)
1	Xizhimen	5	1	Xizhimen	0.3169	1	Guomao	2.999
2	Dongsi	4	2	Dongdan	0.3112	2	Xizhimen	1.945
3	Xuanwumen	4	3	Baishiqiaonan	0.2104	3	Caoyangmen	1.918
4	Guloudajie	4	4	Pinganli	0.2049	4	Junshibowuguan	1.917
5	Yonghegong	4	5	Dongsi	0.1973	5	Zhicunlu	1.911

According to Section 2.2, search the travel options for passengers. According to the utility function, the probability and affected passengers of each OD pair choosing the travel scheme are calculated. In addition, update the travel path of passengers. According to the passenger redistribution algorithm provided in Section 3.1, the affected passengers are redistributed to the BRTN.

4.3. Assessment Vulnerability of BRTN

4.3.1. Vulnerability Assessment of Different Types of Attacking Stations.
Taking the capacity coefficient $\varepsilon = 0.25$ as an example, this article compares and analyses the cascading failures of three different types of important stations under malicious attack, as shown in Figure 8. It can be seen that attacking the stations with the largest betweenness, the largest degree, and the largest intensity, the duration of cascading failure is 18, 23, and 26. The peak failure ratios of stations are 0.42, 0.39, and 0.32, and the time steps to reach the peak are 3, 3, and 11, respectively. It can be seen from this set of data that the stations with the largest intensity have the longest duration of cascading failure process and the lowest ratio of node failure. In contrast, the station with the largest betweenness have the shortest duration, but its ratio of node failure is the highest and the time response to the highest ratio of node failure is the fastest. In the cascading failure process, when the ratio of node failure reaches the highest, the time of attacking the stations with the largest degree is also extremely short. In addition, when the capacity coefficient is constant, in the early stage of cascading failure (when the interruption duration of the station is 1–7 in Figure 8), attacking the station with the largest betweenness may cause a lager failure ratio of stations in the network. Furthermore, attacking the station with the largest betweenness can quickly recover, and attacking the station with the largest intensity requires more recovery time from Figure 8.

This is mainly because the station with the largest betweenness and the station with the largest degree play an important role in the topology of URTN. When these two types of stations are maliciously attacked, the connectivity of the network decreases, resulting in affected passengers being over reassigned to paths with limited capacity. It involves a wide range of passengers, so it is easy to cause cascade failure in a large area and the speed is fast. With the decline of network connectivity, a large number of passengers tend to choose bus in the later stage. With the loss of passenger flow, the status of urban rail transit stations quickly returns to normal. When the station with the largest intensity is attacked maliciously, the range of affected passengers is that the travel path includes the station. When the station with the largest degree or the largest betweenness is maliciously attacked, although the number of passengers affected is not as large as when the attack intensity is the largest, the range of passengers involved is relatively wide due to the station with the largest degree or the largest betweenness. Therefore, in the early stage, the proportion of failure stations is lower than the other two scenarios. Thus, with the redistribution of

passengers, the load of each station is balanced and the network gradually recovers.

Considering the loss of passenger flow, in the early stage of cascading failure, attacking the station with the largest intensity may cause a lager loss of passenger flow. It is mainly because the station serves the most passengers and the alternative routes of urban rail network are limited, resulting in some passengers changing their travel mode. In the middle stage of cascading failure, the loss flow ratio of the attacked station with the largest betweenness and the attacked station with the degree increases faster. This is mainly due to the increase in the affected area of the network, which leads to the change of travel mode and the loss of a large number of passengers. However, in the whole cascade failure process, the station with the largest intensity loses many passengers. To sum up, we can know that the cascading failure caused by different attacking will have different destructive effects on the network in different periods. Attacking the station with the largest betweenness and the largest degree may cause serious damage to the network rapidly. Attacking the station with the largest intensity will cause greater damage to the transportation function of the network because the loss flow ratio is high and the cascade failure process is long.

Therefore, in the actual operation process, when the network is under malicious attacks, the managers should increase the control strength of passenger flow in stations with the largest intensity. For example, managers should combine timetable and passenger flow control measures to relieve the pressure of massive stations and reduce the risk of failure of the largest intensity stations by increasing temporary buses. In addition, the risk management and control of the stations with the large betweenness and large degree should be strengthened to reduce the accidents as much as possible, such as increasing the inspection frequency of stations or tracks. Last but not least, when the travel operation of important stations is interrupted, it should be controlled in time to prevent large-scale paralysis of urban rail transit and massive loss of passenger flow.

Besides, Figure 9 shows the results of vulnerability assessment under both malicious and random attacks. We find that malicious attacks are more likely to cause a larger failure ratio and a loss flow ratio than random attacks. Further speaking, malicious attacks will accelerate the deterioration of the network and aggravate the degree of deterioration. Therefore, during the operation of urban rail transit system, we should strengthen the safety protection of the station to avoid malicious attacks, such as arson. In addition, although the damage degree of random attack is less than that of malicious attack, when some stations are attacked randomly, it will have an impact on the connectivity and function of the network that cannot be ignored.

4.3.2. Vulnerability Assessment under Different Capacity Coefficients.
Figure 10 shows the process of cascading failure after the station with the largest betweenness interrupted in the BRTN. As can be seen from Figure 10(a), when the capacity coefficient $\varepsilon < 0.35$, the failure rate of the station

(a)

(b)
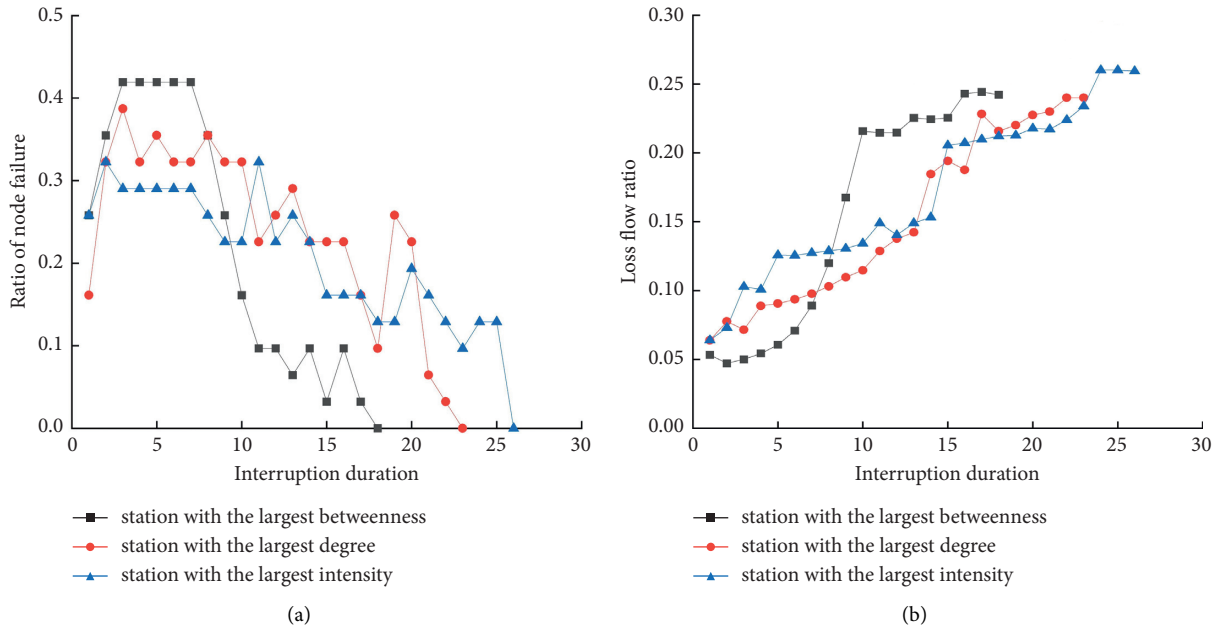
FIGURE 8: The impact of maliciousness on network vulnerability ($\varepsilon = 0.25$). (a) Ratio of node failure. (b) Loss flow ratio.

(a)

(b)
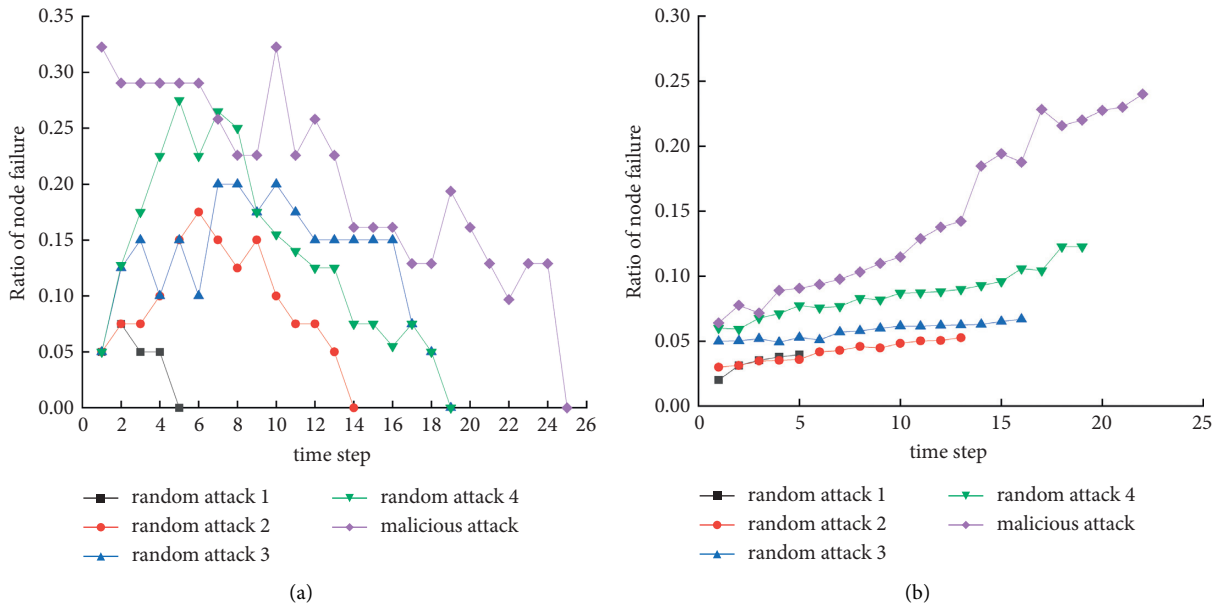
FIGURE 9: The vulnerability of the BRTN under different attacks: (a) ratio of node failure and (b) loss flow ratio.

reaches the peak after approximately 4-5 minutes of operation interruption, and the peak range is $\eta_2 \in [0.412, 0.508]$. When $\varepsilon \geq 0.35$, the failure rate of the network station reaches the peak after approximately 7 minutes of operation interruption, and the range is $\eta_2 \in [0.288, 0.397]$. In addition, it can be seen from Figure 10(b) that when the capacity coefficient $\varepsilon \geq 0.35$, the loss flow of the BRTN does not exceed 15.4%. Therefore, increasing the capacity coefficient of stations in BRTN can improve the robustness of the network and reduce the impact of operation interruption on network traffic efficiency.

Figures 11 and 12 show the BRTN cascading failure process after the station with the largest degree and the largest intensity of station interruption, respectively. It can be seen that with the increase of the capacity coefficient, the vulnerability of the BRTN under cascading failure has decreased. Therefore, reasonably determining the capacity coefficient can increase the robustness of the BRTN. There are many ways to improve the capacity coefficient, such as adjusting the train operation plan and improving the vehicle carrying capacity to increase the supply capacity of urban rail transit system.

(a)

(b)
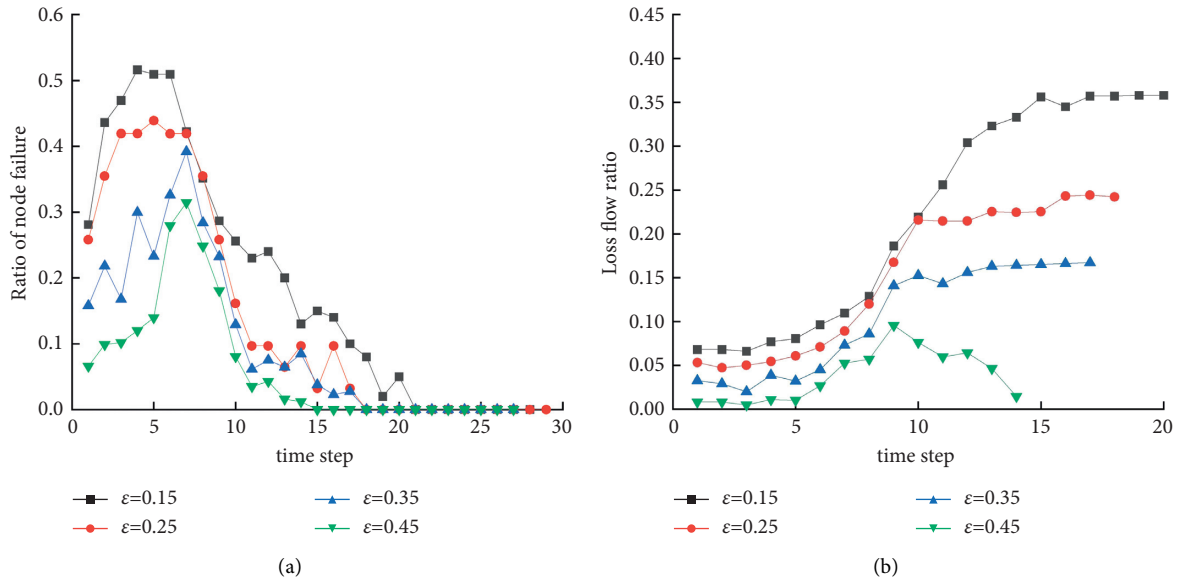
FIGURE 10: The vulnerability under the station with largest betweenness interruption. (a) Ratio of node failure. (b) Loss flow ratio.

(a)

(b)
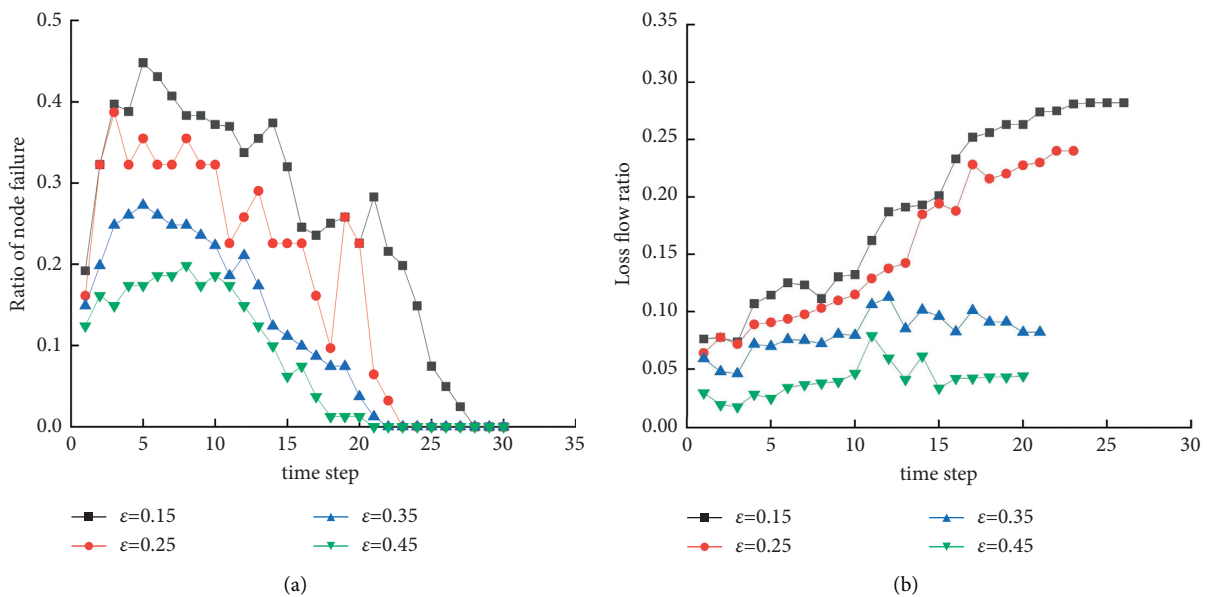
FIGURE 11: The vulnerability under the station with the largest degree interruption. (a) Ratio of node failure. (b) Loss flow ratio.

4.3.3. The Vulnerability Level of the BRTN. Finally, this study attacks each station in the network and obtains the weighted vulnerability index of the network to more clearly get the impact of station interruption on the vulnerability of BRTN. Taking the capacity coefficient $\varepsilon = 0.25$, the weight $\omega_1 = \omega_2 = 0.5$ and the estimation interruption time of the attacking station $T_d = 30$ min as an example. In the risk control system of the station, it is necessary to pay different attention to the stations with different vulnerabilities. Different vulnerability stations are divided into several levels for differentiated control to facilitate the management of the station. In order to more clearly reflect the vulnerability differences of different stations according to the calculation

results of BRTN, the levels of vulnerability are graded as shown in Table 2. The results are shown in Figure 13.

It can be seen that most of the vulnerability stations in BRTN are located on the loop line or at the important transfer stations. The distribution of vulnerability stations is relatively concentrated. The station with the largest degree, the station with the largest intensity, and the station with the largest betweenness are at an extremely high level of vulnerability.

At the same time, this article has made statistics on the number of stations under different vulnerability levels, as shown in Figure 14. It can be seen that the number of stations in each vulnerability level is 3, 3, 34, 49, and 173,

(a)   (b)
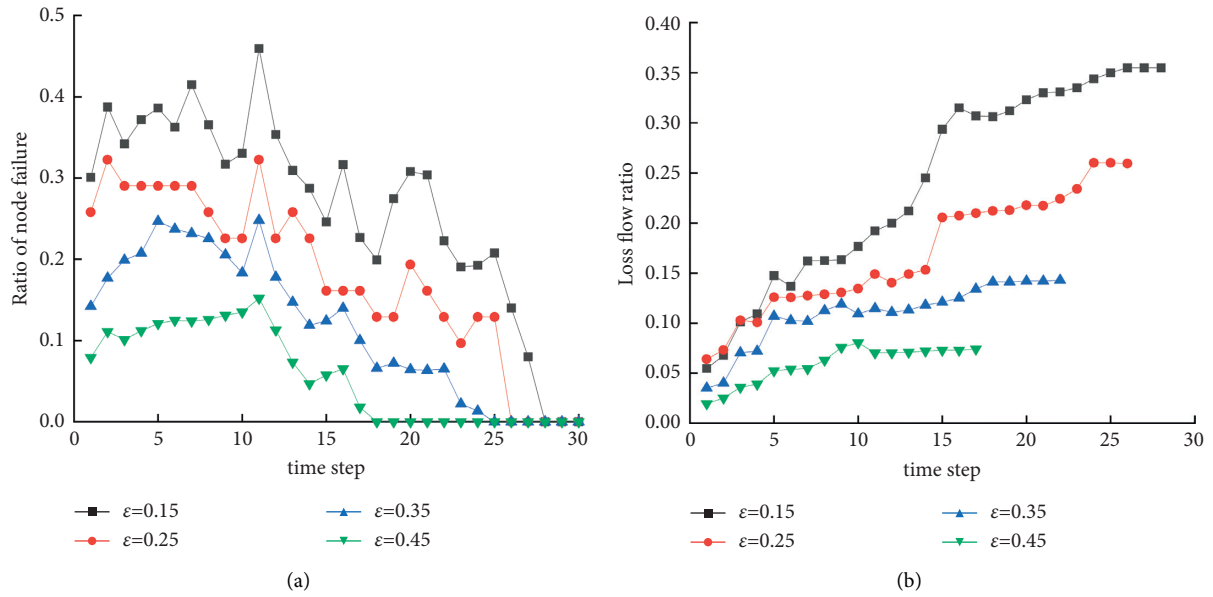
Figure 12: The vulnerability under the station with the largest intensity interruption. (a) Ratio of node failure. (b) Loss flow ratio.

Table 2: Vulnerability index range.

Vulnerability level	Vulnerability index range	Meaning
Extremely high level	[0.3, 1]	Station too prone to failure
High level	[0.25, 0.3]	Stations with a high probability of failure
Medium level	[0.15, 0.25]	Stations with a medium probability of failure
Low level	[0.1, 0.15]	Stations with a low probability of failure
Extremely low level	[0, 0.1]	Stations in robust state

Distribution of vulnerable stations
- 0.0 - 0.1
- 0.1 - 0.15
- 0.15 - 0.25
- 0.25 - 0.3
- 0.3 - 1.0

Figure 13: The distribution of vulnerability under cascading failure.
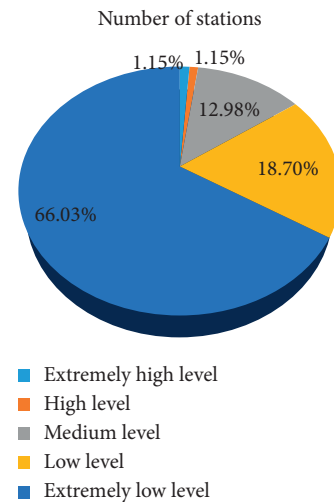
Figure 14: The levels of vulnerability under cascading failure.

accounted for 1.15%, 1.15%, 12.98%, 18.70%, and 66.03% of the total, respectively. Thus, we can draw that most stations in the operation failure cannot cause large-scale paralysis. The interruption of only a few stations with critical structures and functions may seriously reduce the efficiency of BRTN.

5. Conclusion

When a station in the urban rail transit network breaks down due to an emergency, the strong temporal and spatial correlation between stations is easy to cause the propagation of passenger congestion, further leading to station chain failure. In this scenario, the travel efficiency (such as travel time, and travel cost) of passengers is seriously affected. Hence, this article considers the changes in passenger travel modes in the case of abnormal disruption of the rail transit stations and designs travel plans for multiple modes of transportation (including road bus and rail transit). An algorithm for delay redistribution of passenger flows is proposed, and the dynamic vulnerability of the URTN under the phenomenon of cascading failure is finally evaluated.

By quantitatively analyzing the statistical characteristics of the structure and function of BRTN, this article finds that Xizhimen and Dongdan are important stations of the network structure, and GuoMao is an important station of the network function. The results of the dynamic simulation showed that the cascading failure of the URTN is closely related to the spatial-temporal distribution of passenger flow. Besides, compared with other important stations of the structure, the operational interruption of the station that has functional importance has a greater impact on the stability of the URTN operation. We also realize that decreasing the capacity coefficient of the stations can intensify the vulnerability of the URTN, especially in situations of interruption the largest intensity station. Therefore, increasing the capacity coefficient of stations in BRTN can improve the robustness of the network and reduce the impact of operation interruption on network traffic efficiency.

These conclusions provided guidance to operation managers. However, in order to simplify the model and reduce the complexity of calculation, this article only considers two travel modes: rail transit and bus travel. Reasonably speaking, the vulnerability of urban rail transit network must also be affected by other travel mode choices of passengers, such as private cars, taxis, and shared bicycles. When the station operation is interrupted, passengers may choose other travel modes, resulting in the loss of some passengers. However, at the same time, the lost passenger flow will not be redistributed to other rail transit feasible paths, thus reducing the failure risk of other stations. In addition, some scholars, such as Liu and Wang [39], have studied how dispatchers can quickly find a feasible and efficient operation table through the advanced dispatching decision support system to control passenger flow in case of traffic interruption. Therefore, in the future, we plan to further study the complex and diverse passenger travel choice behavior under emergencies and relevant passenger flow control measures, so as to more accurately describe the vulnerability change law of urban rail transit network under emergencies.

Data Availability

The Beijing Metro Corporation provided the operation data of Beijing Metro network from January 2016 to January 2017. However, because the data involves public security and personal privacy, they are not released to the public. If necessary, some sample data can be provided.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] P. Zhang, B. Cheng, Z. Zhao et al., "The robustness of interdependent transportation networks under targeted attack," *EPL (Europhysics Letters)*, vol. 103, no. 6, Article ID 68005, 2013.

[2] J. Tang, L. Xu, C. Luo, and T. Sheng Adam, "Multi-disruption resilience assessment of rail transit systems with optimized commuter flows," *Reliability Engineering & System Safety*, vol. 214, 2021.

[3] L.-G. Mattsson and E. Jenelius, "Vulnerability and resilience of transport systems-a discussion of recent research," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 16–34, 2015.

[4] L. H. Fink and K. Carlsen, "Operating under stress and strain [electrical power systems control under emergency conditions]," *IEEE spectrum*, vol. 15, no. 3, pp. 48–53, 1978.

[5] K. K. Kumari, A. K. Singh, and S. Mahto, "Mathematical models for stability in computer network," *Journal of Computer and Mathematical Sciences*, vol. 10, no. 1, pp. 92–98, 2019.

[6] Y.-H. Hao, J.-H. Han, Y. Lin, and L. Liu, "Vulnerability of complex networks under three-level-tree attacks," *Physica A: Statistical Mechanics and Its Applications*, vol. 462, pp. 674–683, 2016.

[7] S. L. Buglass, J. F. Binder, L. R. Betts, and J. D. M. Underwood, "When "friends" collide: social heterogeneity and user vulnerability on social network sites," *Computers in Human Behavior*, vol. 54, pp. 62–72, 2016.

[8] V. Rampurkar, P. Pentayya, H. A. Mangalvedekar, and F. Kazi, "Cascading failure analysis for Indian power grid," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1951–1960, 2016.

[9] L. Sun, Y. Huang, Y. Chen, and L. Yao, "Vulnerability assessment of urban rail transit based on multi-static weighted method in Beijing, China," *Transportation Research Part A: Policy and Practice*, vol. 108, pp. 12–24, 2018.

[10] S. Nozhati, B. R. Ellingwood, H. Mahmoud, and J. W. Van De Lindt, "Identifying and analyzing interdependent critical infrastructure in post-earthquake urban reconstruction," in *Proceedings of the 11th US National Conference on Earthquake Engineering: Integrating Science Engineering and Policy*, Los Angeles, CA, USA, June 2018.

[11] Q. Ye, *Assessing network vulnerability of rail transport networks–nodal connectivity, partial node failure, and shortest path network problems*, PhD dissertation, University of Tennessee, Knoxville. TN. USA, 2018.

[12] R. C. B. Bucar and Y. M. Hayeri, "Quantitative assessment of the impacts of disruptive precipitation on surface transportation," *Reliability Engineering & System Safety*, vol. 203, Article ID 107105, 2020.

[13] J. Liu, P. M. Schonfeld, Y. Yin, and Q. Peng, "Effects of link capacity reductions on the reliability of an urban rail transit network," *Journal of Advanced Transportation*, vol. 2020, no. 6, 15 pages, Article ID 9020574, 2020.

[14] Z. Liu and R. Song, "Reliability analysis of Guangzhou rail transit with complex network theory," *Journal of Transportation Systems Engineering and Information Technology*, vol. 10, no. 5, pp. 194–197, 2010.

[15] J. Zhang, X. Xu, L. Hong, and S. Wang, "Networked analysis of the Shanghai subway network, in China," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 23-24, pp. 4562–4570, 2011.

[16] Q. C. Lu, "Modeling network resilience of rail transit under operational incidents," *Transportation Research Part A: Policy and Practice*, vol. 117, pp. 227–237, 2018.

[17] Y. Yang, Y. Liu, M. Zhou, F. Li, and C. Sun, "Robustness assessment of urban rail transit based on complex network theory: a case study of the Beijing Subway," *Safety Science*, vol. 79, pp. 149–162, 2015.

[18] O. Cats and P. Krishnakumari, "Metropolitan rail network robustness," *Physica A: Statistical Mechanics and Its Applications*, vol. 549, Article ID 124317, 2020.

[19] J. Chen, J. Liu, Q. Peng, and Y. Yin, "Resilience assessment of an urban rail transit network: a case study of Chengdu subway," *Physica A: Statistical Mechanics and Its Applications*, vol. 586, Article ID 126517, 2022.

[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," *International Journal of Environmental Research and Public Health*, vol. 16, no. 3, p. 329, 2019.

[21] O. Cats and E. Jenelius, "Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information," *Networks and Spatial Economics*, vol. 14, no. 3-4, pp. 435–463, 2014.

[22] A. Tejedor, A. Longjas, I. Zaliapin, and E. Foufoula-Georgiou, "Delta channel networks: 2. Metrics of topologic and dynamic complexity for delta comparison, physical inference, and vulnerability assessment," *Water Resources Research*, vol. 51, no. 6, pp. 4019–4045, 2015.

[23] Z. Y. Liu, Y. B. Lv, B. S. Liu, Q. Li, and W. J. Lv, "Cascading failure resistance of urban rail transit network," *Journal of Transportation Systems Engineering and Information Technology*, vol. 18, no. 5, pp. 82–87, 2018.

[24] R. R. Xing, Q. F. Yang, and L. Zheng, "Research on cascading failure model of urban regional traffic network under random attacks," *Discrete Dynamics in Nature and Society*, vol. 2018, Article ID 1915695, 14 pages, 2018.

[25] A. E. Motter and Y. C. Lai, "Cascade-based attacks on complex networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 66, no. 6, Article ID 065102, 2002.

[26] W. X. Wang and G. Chen, "Universal robustness characteristic of weighted networks against cascading failure," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 77, no. 2, Article ID 026101, 2008.

[27] H. Y. Han, R. N. Yang, H. L. Li, and R. Fan, "Study on cascading failure of two-layer dependent command and control network," *Journal of Central South University: Science and Technology*, vol. 46, no. 12, pp. 4542–4547, 2015.

[28] T. He, N. Zhu, Z. Hou, and G. Xiong, "A novel cascading failure model on city transit network," in *Proceedings of the 2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer*, San Francisco, CA, USA, June 2016.

[29] J. J. Wu, H. J. Sun, and Z. Y. Gao, "Cascading failures on weighted urban traffic equilibrium networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 386, no. 1, pp. 407–413, 2007.

[30] J. Zhang and M. Wang, "Transportation functionality vulnerability of urban rail transit networks based on moving block: the case of Nanjing metro," *Physica A: Statistical Mechanics and Its Applications*, vol. 535, Article ID 122367, 2019.

[31] C. Szymula and N. Bešinović, "Passenger-centered vulnerability assessment of railway networks," *Transportation Research Part B: Methodological*, vol. 136, pp. 30–61, 2020.

[32] M. Kyriakidis, R. Hirsch, and A. Majumdar, "Metro railway safety: an analysis of accident precursors," *Safety Science*, vol. 50, no. 7, pp. 1535–1548, 2012.

[33] Z. Dai, X. C. Liu, Z. Chen, R. Guo, and X. Ma, "A predictive headway-based bus-holding strategy with dynamic control point selection: a cooperative game theory approach," *Transportation Research Part B: Methodological*, vol. 125, pp. 29–51, 2019.

[34] C. Von Ferber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Network harness: metropolis public transport," *Physica A: Statistical Mechanics and Its Applications*, vol. 380, pp. 585–591, 2007.

[35] A. A. De Bona, K. V. O. Fonseca, M. O. Rosa, and M. R. B. S. Delgado, "Analysis of public bus transportation of a Brazilian city based on the theory of complex networks using the P-space," *Mathematical Problems in Engineering*, vol. 2016, Article ID 3898762, 12 pages, 2016.

[36] Y. Z. Chen, C. H. Fu, H. Chang, L. Nan, and H. Da-Ren, "Connectivity correlations in three topological spaces of urban bus-transport networks in China," *Chinese Physics B*, vol. 17, no. 10, p. 3580, 2008.

[37] X. Liu, X. Qu, and X. Ma, "Improving flex-route transit services with modular autonomous vehicles," *Transportation Research Part E: Logistics and Transportation Review*, vol. 149, Article ID 102331, 2021.

[38] M. Nogal, A. O'Connor, B. Caulfield, and B. Martinez-Pastor, "Resilience of traffic networks: from perturbation to recovery via a dynamic restricted equilibrium model," *Reliability Engineering & System Safety*, vol. 156, pp. 84–96, 2016.

[39] H. Liu and J. Wang, "Vulnerability assessment for cascading failure in the highway traffic system," *Sustainability*, vol. 10, no. 7, p. 2333, 2018.

[40] B. Y. Chen, X. W. Chen, H. P. Chen, and W. H. K. Lam, "A fast algorithm for finding K shortest paths using generalized spur path reuse technique," *Transactions in GIS*, vol. 25, no. 1, pp. 516–533, 2020.

[41] F. Corman and A. D'ariano, "Assessment of advanced dispatching measures for recovering disrupted railway traffic situations," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2289, no. 1, pp. 1–9, 2012.

WILEY | Hindawi

Research Article

Analysis of Airport Risk Propagation in Chinese Air Transport Network

Xuejun Zhang [iD],[1,2,3] Shuaizhe Zhao,[1,2,3] and Hao Mei[1,2,3]

[1]*National Key Laboratory of CNS/ATM, School of Electronic and Information Engineering, Beihang University, Beijing 100191, China*
[2]*Beijing Key Laboratory for Network-based Cooperative Air Traffic Management, Beihang University, Beijing 100191, China*
[3]*Beijing Laboratory for General Aviation Technology, Beihang University, Beijing 100191, China*

Correspondence should be addressed to Xuejun Zhang; zhxj@buaa.edu.cn

In recent years, due to the close coupling between airports, airport risk propagation has become a huge challenge. However, it has not been fully understood on the network level. Airport risk can be transferred through other airports owing to connected resources. In this study, we consider two risk factors including airport delay and saturation and propose a risk coupling model based on a clustering algorithm to fit the index and form risk series. To understand the risk propagation mechanism, we build risk propagation networks based on the Granger Causality test, and we apply complex network theory to analyze the evolution of the risk propagation network. We study the regular pattern of risk propagation from perspectives of time and space. Through network analysis, we find four time stages in the risk propagation process and the participation of airports in risk propagation has a positive correlation with airport sizes. In addition, more large airports tend to prevent risk propagation in unoccupied and normal situations, while small airports perform better than large airports in busy situations. Via the conclusion, our work can assist airlines or air traffic managers in controlling the scale of risk propagation before its key time turning point. By identifying the critical airport level and related factors in risk propagation, they can also reduce single airport risk and risk participation through corresponding risk control measures, finally avoiding the large-scale spread of risk and reducing delay or cancellation of more flights.

1. Introduction

With the rapid development of economic globalization, the world's civil aviation industry has grown fast [1]. The scale of China's civil aviation transportation industry is expanding sharply in the meantime. In the past 10 years, China's total air transport turnover has increased by 300%. Nevertheless, the total mileage of China's domestic air routes has only increased by 30%, which leads to increasingly intensified conflicts between limited air transport network resources and the rapid development of the civil aviation industry. The conflicts will cause the appearance of airport risk and aggravate the airport risk propagation. In this paper, the Air transport network (ATN) refers to the airport network. According to [2, 3], the airport network takes airports as vertices and direct flights between airports as edges. Like

above, the Chinese air transport network (CATN) refers to the Chinese airport network(CAN); in this paper, the edges of the Chinese airport network means there is a direct flight between two airports which is constrained by the prescribed airline, and the two airports will be added to the network node set, so the daily structure of CAN will change with the flights in China on that day. And the risk in this paper means the poor operation of the airport, for example, the airport operation efficiency is reduced and the turnover is difficult. The risk is defined considering two risk factors including airport delay and congestion in this paper.

In a complex network, the occurrence of node risk will cause the load of the node to be transferred to other nodes, which will lead to the paralysis of the entire network [4]. Similarly, ATN involves a large number of flights, complex airline structures, uncertain disturbances, and other factors.

Uncertain disturbance factors can be transferred through ATN among airports, which will cause widespread ripple effects. In ATN, the primary risk of an airport results from various disturbances, including inclement weather conditions, flight delays, huge flow, air traffic accidents. Similar to flight delay [5], airport risk in ATN can also be transferred and be amplified through other airports and uncertain disturbance factors because of connected resources.

For example, on July 21, 2012, the rainstorm in Beijing had a great impact on the safety of a wide range of flights across the country. The traffic capacity of PEK airport dropped by 50%, causing thousands of flights to be canceled or delayed. As a result, the risk of PEK propagates to other airports and contributes to their risks. It further leads to the decline of the overall operation efficiency, capacity loss, large-scale flight delay, or cancellation of the aviation network, and finally resulting in huge economic losses or the decline of airline reputation. Therefore, airport risk propagation in CATN has become a huge challenge and also needs attention. Accordingly, if we study the risk propagation regular patterns of CATN, such as the spatial and temporal characteristics of risk propagation, we can effectively control the scale of risk propagation and reduce the global network risk caused by a local risk which can finally reduce losses.

Therefore, in this study, we built a risk propagation network (RPN) based on the Granger causality (GC) test and investigated the direction and range of risk propagation at the network-level. We try to analyze the law of RPN from the perspectives of time and space. From the time angle, we use the size of the largest connected component of RPN to analyze the evolution process of risk propagation scale with time, try to reveal the time characteristics of its diffusion and dissipation in a day, and get its key time turning point. The results reveal that there were 4 stages of risk propagation of a day: dissipation of remaining risks, generation, fluctuation, and dissipation of intraday risks. We should pay attention to several key periods to better control risk propagation. From the spatial angle, we analyze the role of airports of different categories in risk propagation and identify the key airport category and related factors in the process of risk diffusion and absorption. The results show that the participation of airports had a positive correlation with their sizes and number of flights, and participation of large airports is higher than middle airports and small airports, so large airports are more noteworthy in risk propagation control.

Our results provide support to air traffic managers (ATM) on decision-making. According to the findings, the air traffic managers can develop effective countermeasures to prevent the risk propagation in particular links before key time points to alleviate network-wide flight delay or congestion and finally reduce the global risk of ATN. Moreover, through the causality test, if one variable acts as the cause for another one, actively intervening on the first would lead to the changes in the second. Thus, policy makers could potentially adopt the proposed method to analyze the interaction of airports in order to identify critical ones. It will help them to make decisions on resource allocation for improving airport capacity. They can better control risk propagation in

the future according to the key periods in a day, high participation airports, and vital flow range in our analysis.

Beyond this introduction, this work is organized as follows. Section 2 introduces some relevant previous studies and the differences between this work and previous work. Section 3 presents the materials and methods here applied, including the risk coupling model (Section 3.1), a description of Granger Causality test (Section 3.2), and some topology parameters of network analysis (Section 3.3). Section 4 introduces the construction of RPNs and analysis of the mechanism of risk propagation in CATN. Finally, Section 5 summarizes our major conclusions.

2. Literature Review

In recent years, researchers around the world have largely investigated the risk assessment of the civil aviation industry [6] as well as the propagation of single risk factors such as flight delay [7–11], congestion [12–14], and aircraft conflicts [15]. Although there has been rich research on assessment and propagation of risk in the civil aviation domain, they mainly focused on the civil aviation macroindustry with less consideration on the network level.

Complex network theory has achieved mature results in many fields such as identification and evaluation of influential nodes [16, 17], network risk assessment [18, 19], and decision-making based on the safety risk suggestions [20, 21] in recent years. These studies have important theoretical value for mining risk propagation mechanisms and safety risk assessment. Specifically, complex network theory is a powerful tool in the study of transportation systems [22, 23], with many network models defined for air transport system [24].

There is also some research that analyzes flight delay propagation on the network level; [25] developed the maximum connected subgraph of congested airports for assessing the level of delays across the entire system. Afterward, via establishing a delay propagation network based on a causal model, [26] used two models to simulate flight delay propagation and assessed the effect of disruptions in the US and European aviation networks. Reference [27] created a multilayer delay propagation network of 50 busiest airports and 20 airlines in Europe in which each airline acts as a layer for comparing the differences between 20 single layer networks and 3 projection networks. Reference [28] created a delay causality network based on the Granger causality test to understand the mechanism of flight delay propagation at the system level. Reference [29] built a delay propagation network based on the Bayesian Network approach to study the complex phenomenon of delay propagation within the network consisting of the 100 busiest airports in the United States. Reference [30] proposed an adaptation of machine learning's clustering analysis to air transport delay propagation network. By using Granger tests, they can assign each airport to the corresponding cluster, which includes mostly forcing node, intermediaries node, and mostly forced node. To the best of our knowledge, research on risk propagation mechanisms at the level of ATN is timely yet challenging and has not attracted much attention. In addition, the above previous studies mainly focused on single local risk factors

such as delay and congestion without full consideration of overall airport risk. There is also no research on comprehensive risk propagation.

Different from previous studies, we investigate the overall airport risk propagation with consideration of combining different local risk factors. Besides, applying complex network theory, we study the mechanism of risk propagation on the network level instead of civil aviation macro industry or local areas. In this paper, we propose the use of complex network theory to understand the mechanism of risk propagation in CATN with empirical analysis. For this purpose, we firstly propose a risk coupling model to obtain airport risk time series through multiple local risk factors. Then, we apply a risk propagation identification model based on the well-known Granger Causality (GC) test [31, 32] due to its primary advance on the relevance problem [33]. GC is a powerful tool to determine if two time series are related. Finally, we construct a risk propagation network (RPN) based on GC for each hour of a day, then use complex network theory to investigate how airport risks appear, propagate, and dissipate in CATN with empirical analysis.

3. Materials and Methods

3.1. Risk Coupling Model. Comparing with studies on simple risk factor ranking, we consider the impact of different risk factors and form risk coefficients. Instead of a rough assessment that only gives the risk level, we provide specific numeric results of risk, which is more intuitive and convincing. According to the flight statistics of the Civil Aviation Administration of China (CAAC), the number of air traffic accidents has increased and the rate of regular flights has decreased in recent years, which results in the risk of security and efficiency of CATN operation. Therefore, considering temporal and spatial properties, we select two local risk factors to quantify airport risk. In this paper, the risk means the poor operation of the airport, for example, the airport operation efficiency is reduced, and the turnover is difficult. The risk in this paper is obtained by coupling the delay weight and congestion weight. The delay factor reflects airport efficiency from a temporal perspective and the congestion factor reflects airport security from a spatial perspective. These two risk factors can reflect the operation efficiency change and safety situation of airports and reveal its comprehensive risk.

In recent years, average departure delay has also attracted much attention. Research on departure delay optimization [34, 35] and prediction [36, 37] has achieved prominent results. For temporal local risk factors, we adopt average departure delay due to its significant position in air traffic risk assessment. In addition, the average departure delay of airports represents the detention of flights and focuses on measuring the operation efficiency of CATN in the time dimension. We focus on hourly time series as we intend to investigate changes in risk propagation of a day with time series of each hour. For airport i, we construct its departure delay time series D_i by splitting one hour into 12 time intervals. Each time interval contains 5 minutes. The

value of each time interval represents the average departure delay. The average departure delay for airport i of time interval t is defined as follows [28]:

$$d_i(t) = \frac{y_i(t) + c_i(t) * b}{n_i(t)}, \quad t \in \{0, 1, 2 \dots 11\}, \quad (1)$$

where $y_i(t)$ represents the total departure delay of airport i during. $(t, t + 1)$. $c_i(t)$ and $n_i(t)$ represent the number of cancelled flight and the total number of flights including canceled flights and scheduled flights during $[t, t + 1]$. Cancellation of flights should be taken into consideration in delay assessment [38]. b is the equivalent delay of canceled flights and generally equals 3 hours.

As for spatial local risk factors, congestion is a representative index in ATN and has attracted more and more attention in recent years [12–14]. The airport saturation describes the load situation of airports and focuses on the operation safety of ATN in the spatial dimension. As a result, we utilize airport saturation as a spatial factor to describe the congestion in CATN. For airport i, we construct its saturation time series S_i by splitting one hour into 12 time intervals. Each time interval contains 5 minutes. The value of each time interval represents the saturation. The saturation of airport i of time interval t is defined as follows:

$$s_i(t) = \frac{Q_i(t)}{C_i(t)}, \quad t \in \{0, 1, 2 \dots 11\}, \quad (2)$$

where $Q_i(t)$ represents the total arrival flights number of airport i during $[t, t + 1]$. $C_i(t)$ represents the capacity of airport i. In our work, airport capacity refers to the number of arrival aircraft that the airport can handle in an hour. With this method, we can construct delay and saturation as hourly time series.

After modeling risk factors, the next work is to couple them into the overall risk time series. For risk factor coupling, Li [14] utilized a grey clustering model to divide several local risk factors into 4 levels and weight them with a linear mapping function to obtain the final risk level. However, specific risk quantification results were not given. We propose a risk coupling model based on a clustering algorithm to obtain the overall risk coefficient. The specific method is defined as follows.

Firstly, we utilize (1) and (2) to obtain delay time series and saturation time series of each hour of a day. As delay and saturation have different dimensions; for both of them, we normalize each value of time series with the hyperbolic tangent function defined as follows:

$$d'_i(t) = \frac{e^{d_i(t)} - e^{-d_i(t)}}{e^{d_i(t)} + e^{-d_i(t)}},$$

$$s'_i(t) = \frac{e^{s_i(t)} - e^{-s_i(t)}}{e^{s_i(t)} + e^{-s_i(t)}}, \quad (3)$$

where $d'_i(t)$ is the normalized value of delay and $s'_i(t)$ is the normalized value of saturation of airport i in time interval t.

Secondly, in order to take advantage of internal relevance among airports and flights, we apply a K-means clustering algorithm [39] for delays of all airports in the same time

interval of an hour of a day and obtain 4 delay intervals. The same as delay, we adopt the same method to saturation and obtain 4 saturation intervals.

Then, according to the corresponding delay interval and saturation interval of airport i, we transform delay and saturation of airport i into delay weight $W_d(i,t)$ and saturation weight $W_s(i,t)$. We refer to the radial basis function in machine learning to define the delay weight mapping function and saturation weight mapping function in this paper, as shown in the following equations:

$$W_d(i,t) = \lambda_{d,c}^{0.5} * e^{d_i(t) - \lambda_{d,c}/\lambda_{d,c}} + \sum_{j=1}^{c-1} \lambda_{d,j}^{0.5}, \quad (4)$$

$$W_s(i,t) = \lambda_{s,c}^{0.5} * e^{s_i'(t) - \lambda_{s,c}/\lambda_{s,c}} + \sum_{j=1}^{c-1} \lambda_{s,j}^{0.5}, \quad (5)$$

$W_d(i,t)$ and $W_s(i,t)$ corresponds, respectively, to the delay weight and saturation weight of airport i in time interval $[t, t+1]$ $W_d(i,t)$. c is the index of delay and saturation interval. $\lambda_{d,c}$ and $\lambda_{s,c}$ correspond, respectively, to the right boundary of delay interval c and saturation interval c, on the left of (4) and (5), the value of c depends on which cluster interval $d_i'(t)$ and $s_i'(t)$ is in. $d_i'(t)$ and $s_i'(t)$ represent, respectively, the normalized delay and saturation of airport i in time interval t. n_s is the number of intervals and n_s in our paper and $c \in \{1, 2, 3, 4\}$, which represents the first to fourth intervals.

Finally, we obtain the overall risk coefficient $R_i(t)$ airport i during $[t, t+1]$ according to the following equation:

$$R_i(t) = \sqrt{W_d(i,t)^2 + W_s(i,t)^2}. \quad (6)$$

As a result, we apply this model to airport k for each time interval to obtain risk time series (R_i) in an hour. Then, we calculate all risk time series of 24 hours in a day for all airports with this method. In this way, we have 24 groups of risk time series. Each group has the same number of risk time series as the number of airports.

3.2. Risk Propagation Identification Model.
For risk propagation identification, we adopt the well-known Granger Causality (GC) test to test the causality between risk time series of 2 airports and identify the risk propagation according to the test result. GC test is proposed by the economy Nobel Prize winner Clive Granger. It is a statistical hypothesis test for determining whether one-time series is useful in forecasting another. For risk propagation, if the risks observed at one airport can explain the risks appearing at a second one, there exists a causality relationship, and we can identify the risk propagation from this airport to the second airport. Comparing with traditional methods, GC test is independent of airlines and air routes. GC test can identify risk propagation relation between any airports, ignoring their airlines and air routes. As a result, GC test can discover some undirect risk propagation relationships. In addition, GC test reveals propagation relationship from perspective of causality, which is more profound.

GC test is sensitive to the stationarity of time series. As a result, we here apply a Z-Score detrend procedure to reduce the nonstationarity of the time series, which can result in a biased evaluation of the Granger Causality metric. In detail, the detrended risk time series for one airport is calculated as follows:

$$R'(h,t) = \frac{R(h,t) - R(.,t)}{\sigma(R(.,t))}. \quad (7)$$

$R'(h,t)$ represents the detrended risk for hour h at time interval t. $R(h,t)$ is the original risk. $\langle R(.,t) \rangle$ and $\sigma(R(.,t))$ are the average and standard deviation of the risk time series, respectively. Then, an augmented Dickey–Fuller test (ADF) [40] is applied to verify whether the risk time series is stationary. The risk time series difference procedure is applied until the time series is stationary. The time series difference is presented as follows:

$$R'(h,t) = R(h, t+1) - R(h,t), \quad (8)$$

where $R(h,t)$ and $R(h, t+1)$ are the risk for hour h at time interval $[t, t+1]$, respectively. And $R'(h,t)$ is the risk after difference procedures.

In our research, causality reveals the impacts between airport pairs and reflects the interaction of airport risks. Here, GC will help understand the existence and direction of the risk propagation between two airports based on the risk time series. For a pair of airports, an airport j "Granger-causes" another airport i if the use of past values of the risk time series of j helps improving the prediction of risk time series of i. In mathematical terms, suppose that the risks of two airports i and j can be expressed as two stationary time series Yi and Yj. Then, Yj "Granger-causes" Yi if the following equation can be satisfied.

$$\sigma^2(Y_i|U^-) < \sigma^2\left(\frac{Y_i|U^-}{Y_j}\right). \quad (9)$$

$\sigma^2(Y_i|U^-)$ is the square of residual in forecasting the time series Yi using the past information of the entire universe U, and $\sigma^2(Y_i|U^-/Y_j^-)$ stands for the square of residual when the information about time series Yj is discarded. The specific process of the GC test is presented as follows.

Firstly, the Granger causality test adopts an unrestricted regression equation to obtain its residual sum of squares RSSr in the following equation:

$$y_i(t) = \sum_{m=1}^{P} a_m * y_i(t-m) + \sum_{m=1}^{P} b_m * y_j(t-m) + e(t). \quad (10)$$

$y_i(t)$ is the current value of time series Yi, $y_i(t-m)$ is the past value of time series Y_j, $e(t)$ is the error term, and am and bm are the regression coefficients. In addition, P stands for the lag, indicating that the current value should be regressed with the past P values. Statistical analysis shows that the average flying time between all airport pairs in CAN is around two hours; thus we adopt the lag value $P = 2$ in this paper. Then, the null hypothesis that j does not cause i is defined as follows:

$$b_1 = b_2 = \cdots = b_P = 0. \tag{11}$$

Secondly, we apply a restricted regression equation to obtain its residual sum of squares RSS_{ur} in the following equation:

$$y_i(t) = \sum_{m=1}^{P} a_m * y_i(t-m) + e(t). \tag{12}$$

Finally, we apply F-test to the model. F-statistic and p-value are adopted to test the null hypothesis as follows:

$$F = \frac{(RSS_r - RSS_{ur})/P}{RSS_{ur}/(n-P)}. \tag{13}$$

where n is the sample size of each time series. When the p-value is less than the chosen significance level ∂ (5% by default) of F-test and $>F_\partial$, it can be considered to pass F-test. In addition, we here apply the likelihood-ratio (LR) test [41] to the null hypothesis owing to insufficient length of risk time series. As a short length of time series may cause a fake causality relationship, LR can test the fitness of the model from the perspective of likelihood to increase accuracy of causality test. LR test is adopted as follows:

$$G = 2 * (\ln(L_{ur}) - \ln(L_r)), \tag{14}$$

where L_{ur} is the maximum likelihood value of the unconstrained regression model, and L_r is the maximum likelihood value of the constrained model. G is the test result of LR and approximately obeys the chi-square (χ^2) distribution. When the p-value is less than the chosen significance level ∂ (5% by default) of the chi-square test and $G > G_\partial$, it can be considered to pass the LR test.

When it passes both of F-test and LR test, the null hypothesis is rejected. If item Y_j belongs to this regression, Y_i is the cause Y_j. As a result, the value in j is partly attributed to i and we can consider that the risk of airport j propagates to airport i. In RPN, there is a directed link from airport j to airport i. For each pair of the airport which passed GC test, the link between them has no specific weight, and all the weight defaults to 1.

3.3. Network Analysis Metrics.

Due to a large number of airports and complex interactions, the features of risk propagation cannot be understood from the information at the individual airport level alone. Complex network theory and its associated metrics and tools present an opposite approach to study ATN beyond what is offered by classical techniques [42]. Thus, a network-level analysis is adopted to capture the characteristics of risk propagation in CATN. Some topology parameters with practical significance are introduced here to help analyze network-level risk propagation.

It is specified that $A_{i,j} = (a_{i,j})_{N*N}$ is the adjacency matrix of the network, where N is the number of network nodes. $a_{i,j} = 1$ if there is a directed edge from node i to node j; otherwise, $a_{i,j} = 0$. The number of nodes is the number of individual points in the network, and the number of edges is the number of links between nodes. In the undirected network, $a_{i,j} = a_{j,i}$.

(1) Degree of node corresponds to the number of links with it. The node i has two types of out-degree (k_i^{out}) and in-degree (k_i^{in}) links, representing the number of nodes pointing or pointed by this node. The mathematical form is listed as follows:

$$k_i^{out} = \sum_{j=1}^{N} a_{i,j}, \tag{15}$$

$$k_i^{in} = \sum_{j=1}^{N} a_{j,i}. \tag{16}$$

(2) Largest connected component [43] (LCC) is the largest group in which nodes are connected by existing edges with each other. It is generally introduced to reflect the scale of the network. In our research, we can assess the extent and seriousness of risk propagation through the size of LCC (N_{LCC}), which denotes the number of nodes in LCC.

(3) Clustering coefficient (CD) [44] is used to qualify the inherent cluster tendency of nodes. The clustering coefficient of a node is the fraction of pairs of its neighbor nodes that have a link (i.e., the number of triangles in the network). For a network, the overall clustering coefficient is calculated as follows:

$$C_D = \frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_j \sum_h (a_{i,j} + a_{j,i})(a_{i,h} + a_{h,i})(a_{h,j} + a_{j,h})}{d_{single} * (d_{single} - 1) - 2 * d_{double}}, \tag{17}$$

where $d_{single} = \sum_{j \neq i} a_{i,j} + \sum_{j \neq i} a_{j,i}$ and $d_{double} = \sum_{j \neq i} a_{i,j} * a_{j,i}$.

(4) Reciprocity (RC) [45, 46] of the network reflects the bidirectional nature of links between airport pairs. The reciprocity means that node i affects airport j, whereas airport j also affects airport i. It is defined as follows:

$$R_C = \frac{\sum_{i \neq j}^{N} (a_{i,j} - \overline{a})(a_{j,i} - \overline{a})}{\sum_{i \neq j}^{N} (a_{i,j} - \overline{a})^2}, \tag{18}$$

where $\overline{a} = \sum_{i \neq j}^{n} a_{i,j}/N(N-1)$ and n is the number of nodes.

(5) Link density (LD) [47] describes the density of edges in the network. In the risk propagation network, it describes the density of risk propagation links between airports. The link density is relative to the number of nodes. It is defined as follows:

$$L_D = \frac{\sum_{i,j} a_{i,j}}{N(N-1)}, \tag{19}$$

where $a_{i,j}$ is the existing links in the network and N is the total number of nodes.

(6) Efficiency (E) [48] of a network represents the ease for one node to reach other nodes (i.e., how many nodes have to pass by in order to reach the destination). The efficiency is defined by considering the inverse of the harmonic mean of the distances between pairs of nodes as follows:

$$E = \frac{1}{N(N-1)} \sum_{i,j \neq i} \frac{1}{d_{i,j}}, \tag{20}$$

where $d_{i,j}$ stands for the distance between node i and node j.

4. Results and Discussions

The flight data analyzed in this paper involves all domestic flight information in February, March, and May 2019 in China. The data of 3 months contained a total of 795973 domestic scheduled flights connecting 223 airports. The average number of daily flights is 8844. As a traditional Chinese holiday, May 1, 2019, was the busiest day with 9421 domestic flights. In contrast, as the Spring Festival of 2019, February 5 was the most unoccupied day with 6952 domestic flights. March 2, 2019, was a normal day with 8886 flights which is close to the average level. The weather is known to be one of the major factors influencing flight delays. By querying the air traffic website of the Civil Aviation Administration of China (CAAC) that publishes flight delay information, we can know the weather condition for the selected three days. On May 1, 2019 and March 2, 2019, there was no flight delay and airport capacity decline caused by weather. On February 5, 2019, only Wuhan Tianhe International Airport(WUH) was affected by the low visibility weather, the capacity of WUH was reduced by 30%, and other airports in CAN were not affected by the weather. Therefore, the selected three days in this paper can basically ignore the flight delay caused by weather factors and more effectively analyze the risk propagation of CATN under different dates. Details of the data set are shown in Table 1.

In order to perform a network-level analysis of risk propagation, we build an RPN using the pairwise GC test based on the flight data described in Table 1 for each hour of a day. For process and temporal characteristics of risk propagation, we focus on the hourly time series on March 2, 2019, due to its normal level of flight flow. For spatial characteristics, we focus on the features of airports of different sizes and their behaviors in risk propagation. We adopt the hourly time series on February 5, 2019, March 2, 2019, and May 1, 2019, to compare the features of different airports in different situations.

4.1. Process of Risk Propagation.
The propagation process is a significant item to consider as it intuitively reveals how risk propagates in CATN. Among all parameters of a complex network, the size of LCC (N_{LCC}) can reflect the extent of risk propagation. We adopt N_{LCC} as it can directly help understand the generation, fluctuation, and dissipation of risk propagation during the propagation process. In order to

TABLE 1: Available information for each flight.

Information for each flight	
Airline	Flight code
Scheduled departure time	Actual departure time
Scheduled arrival time	Actual arrival time
Flight departure airport	Flight state
Flight arrival airport	Flight date

reveal the general situation of the risk propagation process, we focus on the flight data on March 2, 2019, and construct an RPN for each hour of 24 hours. There are a total of 223 airports. For one RPN of each hour, 49729 (223*223) times GC tests are performed. After removing airports with no connections, we generate 6 snapshots of RPN of different hours on March 2, 2019, in Figure 1 to present an obvious and visualized process of risk propagation.

During the process of risk propagation in a day, risk propagation exhibits cyclical fluctuations due to daily schedules. We investigate the appearance, fluctuation, and dissipation of risks in the process of risk transmission by observing the changes of NLCC. There are 4 stages in the process of propagation shown in Figure 2 of a day: dissipation of remaining risks, generation, fluctuation, and dissipation of intraday risks. In the stage of dissipation of the remaining risks of yesterday, the impact of risks of yesterday has not disappeared. The extent of risk propagation decreases until 5 : 00 with fewer scheduled flights. In the generation stage, the extent of risk propagation increases rapidly owing to the sharp growth of scheduled flights. Most busy airports generate initial delays due to the "Morning Rush" phenomenon. In the fluctuation stage, through the connection of resources, risk propagation of airports becomes serious with the increasing scheduled flights. In addition, the scale of risk propagation fluctuates at a high level. In the stage of dissipation of intraday risks, the extent of risk propagation decreases smoothly. Risk propagation will disappear the next day. The Air transport network will recover from the risk propagation due to the removal of flights from the system. Therefore, we need to pay attention to several periods as 5 : 00–6:00, 9 : 00–10 : 00 and 20 : 00–21 : 00, which are the key periods of risk propagation in a day. By improving the risks in these periods, we can better control risk propagation in a day.

In addition, we compared the evolutionary relationship between NLCC and delay, flow, and risk, respectively, as in Figure 3. Delay corresponds to the hourly average departure delay on March 2, 2019, and flow stands for the total number of scheduled flights for each hour on March 2. Risk represents the hourly average risk coefficient of airports which we obtained in Section 2.1. As Figure 3(b) shows, NLCC has a strong correlation with the flow. They have a similar evolutionary trend. However, as Figure 3(a) shows, there are some deviations between NLCC and delay due to inertia of delay. Flow can decrease in time with the decreasing scheduled flights. However, delay can accumulate gradually with the increasing flights and last for a long period, which is the "Cumulative Effect" of delay. Although the number of flights decreases, it still needs time to recover and decrease, which is the reason that delay continues to maintain at a high
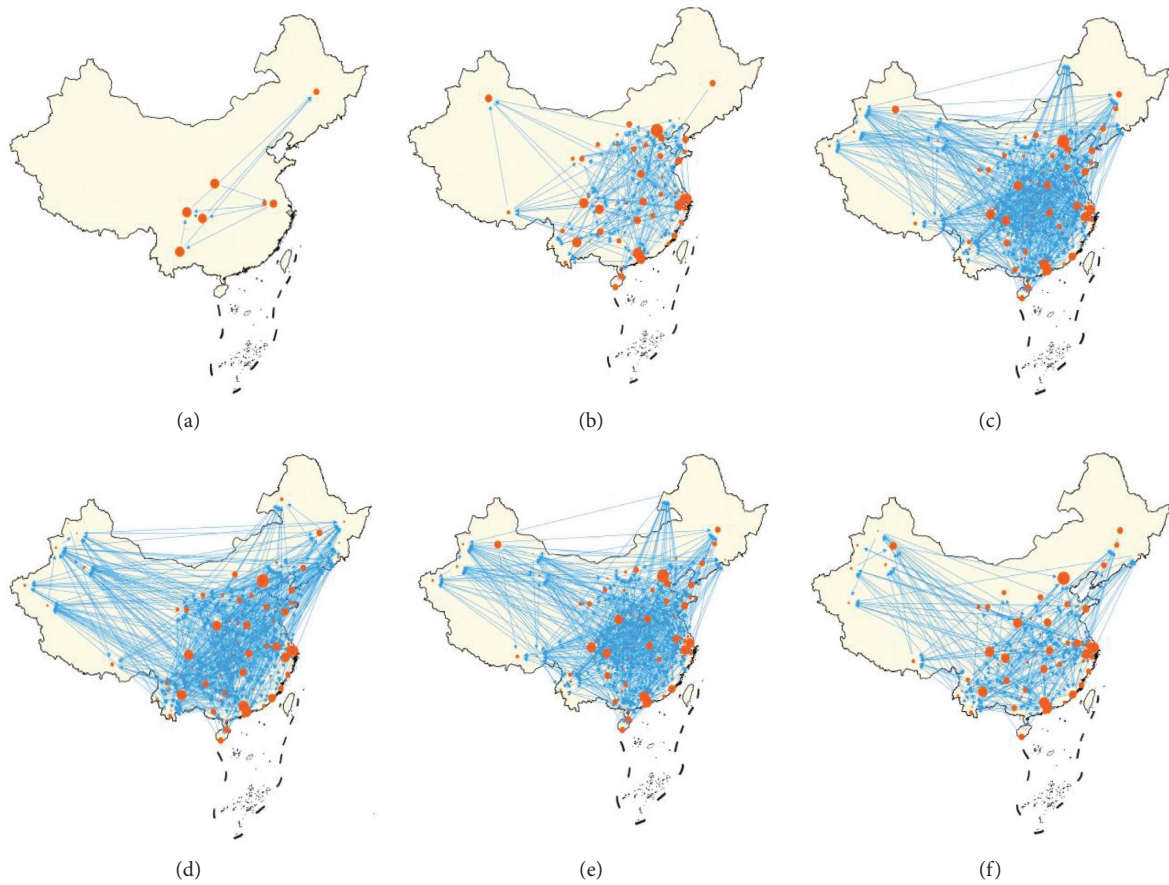
FIGURE 1: Snapshots of RPN on March 2, 2019. (a) 03 : 00. (b) 07 : 00. (c) 11 : 00. (d) 15 : 00. (e) 19 : 00. (f) 23 : 00.
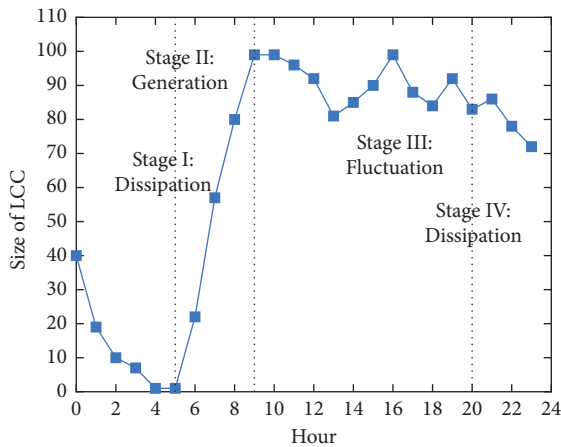
FIGURE 2: Stages of risk propagation process.

level at the end of the day. Risk is influenced by delay and flow and it is the result of them. As a result, risk combines their characteristics and has a similar evolutionary trend with *NLCC*, as Figure 3(c) shows. Results also reveal the close relationship between risk factors and *NLCC*.

4.2. Temporal Features. In addition to the process of risk propagation, we will investigate the evolution process of RPN to understand the temporal properties of risk propagation. We adopt clustering coefficient, reciprocity, link density and efficiency of the network to analyze the evolution characteristics of RPN.

As results are shown as Figure 4, network clustering, reciprocity and link density have a similar evolutionary trend as *NLCC*, which means that they have a strong correlation with each other. High link density indicates a large number of edges in RPN, which causes a high level of network clustering and efficiency. Particularly, they are all temporarily at a lower level in a short moment from 13 : 00 to 14 : 00 due to fewer flights after morning rush hours. Nevertheless, reciprocity is special, as Figure 4(b) shows. The low reciprocity in the generation and fluctuation stage reveals that the single-direction propagation plays a significant role. In contrast, reciprocity sharply increases at 2 : 00 owing to the little number of flights and the significant role of double-direction links. Since other network metrics are more sensitive to changes in the number of network nodes and edges, they have similar evolution processes. However, reciprocity is independent of the number of network nodes and edges and reveals the bidirectionality of the network. This is the reason why reciprocity is different from other metrics.

4.3. Spatial Features. The global features of risk propagation are analyzed by the RPNs' properties, while risk propagation cannot be fully understood without analyzing the spatial
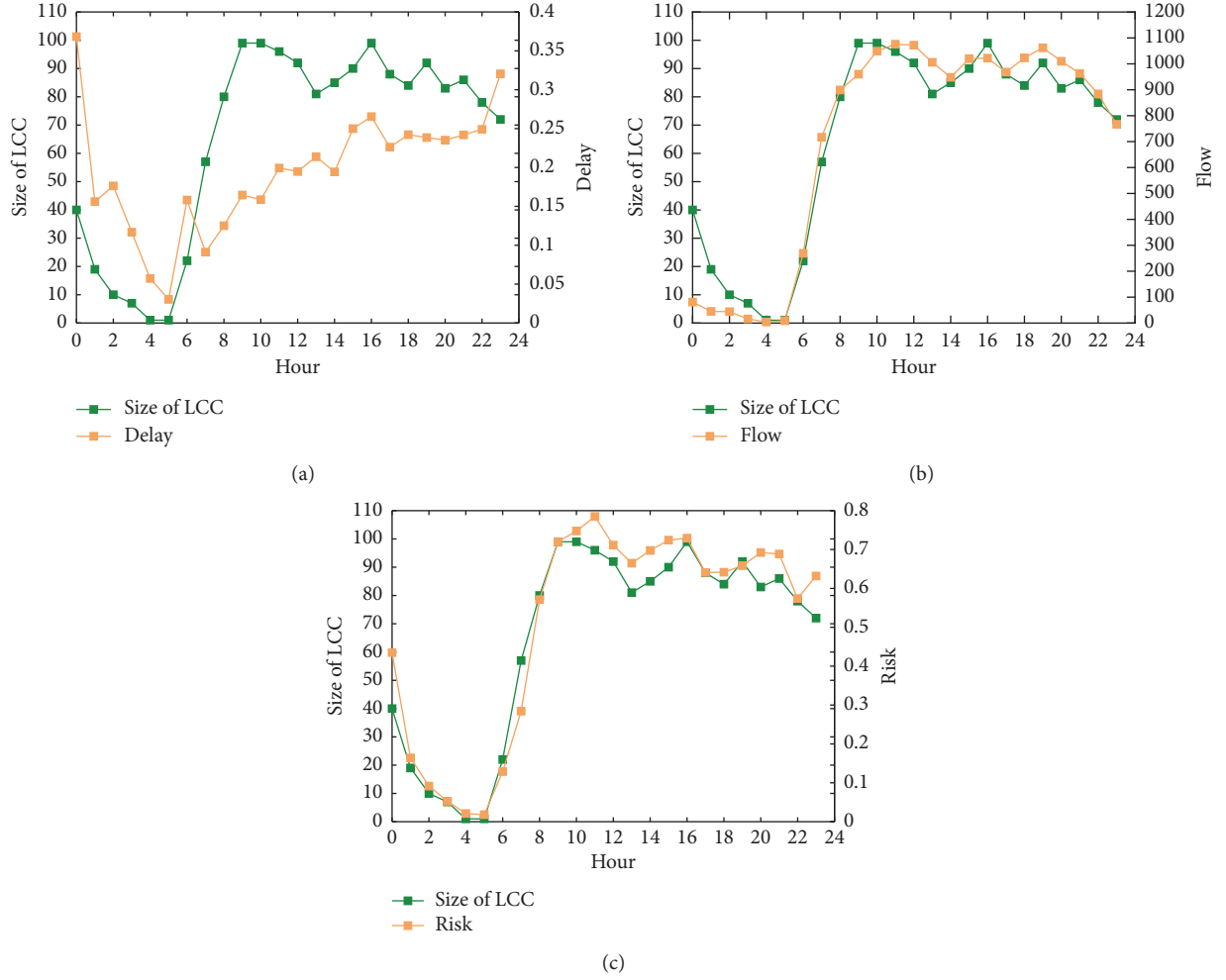
FIGURE 3: (a) Evolutionary relationship between (*N*)LCC and delay. (b) Evolutionary relationship between (N)LCC and flow. (c) Evolutionary relationship between (*N*)LCC and risk.

features. We investigate the spatial features from the point of view of behaviors of airports with different sizes in risk propagation. We divide airports into three categories according to their average departure and arrival flow within a day in Table 2.

The "Amount" in Table 2 means airport amount and the "Flow" means the number of flights. To understand the behaviors of airports in risk propagation, we will analyze the airports' participation in different categories and the impact of airports on risk propagation. We have three different categories of airports in this paper including "Large airports," "Middle airports," and "Small airports". The collective participation $P_c(t)$ of category c of airports is defined as the percentage of the number of airports involved in risk propagation belonging to category c to the number of total airports in category c. It is shown as follows:

$$P_c(t) = \frac{n_c(t)}{N_c}, \tag{21}$$

where $n_c(t)$ is the number of airports involved in risk propagation belonging to category c during $[t, t+1]$, and N_c is the total number of airports in category c.

We focus on the flight data on February 5, 2019, March 2, 2019, and May 1, 2019, to compare the participation of different categories in different situations. As Figure 5 shows, participation is related to the category of airports. Participation of large airports is the highest in all of the situations, which indicates that most large airports are involved in risk propagation regardless of situations. Although small airports represent the majority of the system, they are rarely involved in the spread of risks. In addition, the hourly average participation $\overline{P(t)}$ of different airports increases with increasing flights, as Table 3 shows.

In addition, we will also analyze the impact of airports on risk propagation. To explore how airports affect risk propagation, we define the absorption coefficient r_i to assess the absorption ability of airport i for risk propagation in

$$r_i = \frac{\overline{k_i^{\text{in}}} + 1}{\overline{k_i^{\text{out}}} + 1}, \tag{22}$$

where $\overline{k_i^{\text{out}}}$ ($\overline{k_i^{\text{in}}}$) is the hourly average out-degree (in-degree) of airport i in risk propagation network over an entire day. k_i^{in} is the number of airports, each of which partly results in

(a)

(b)

(c)

(d)
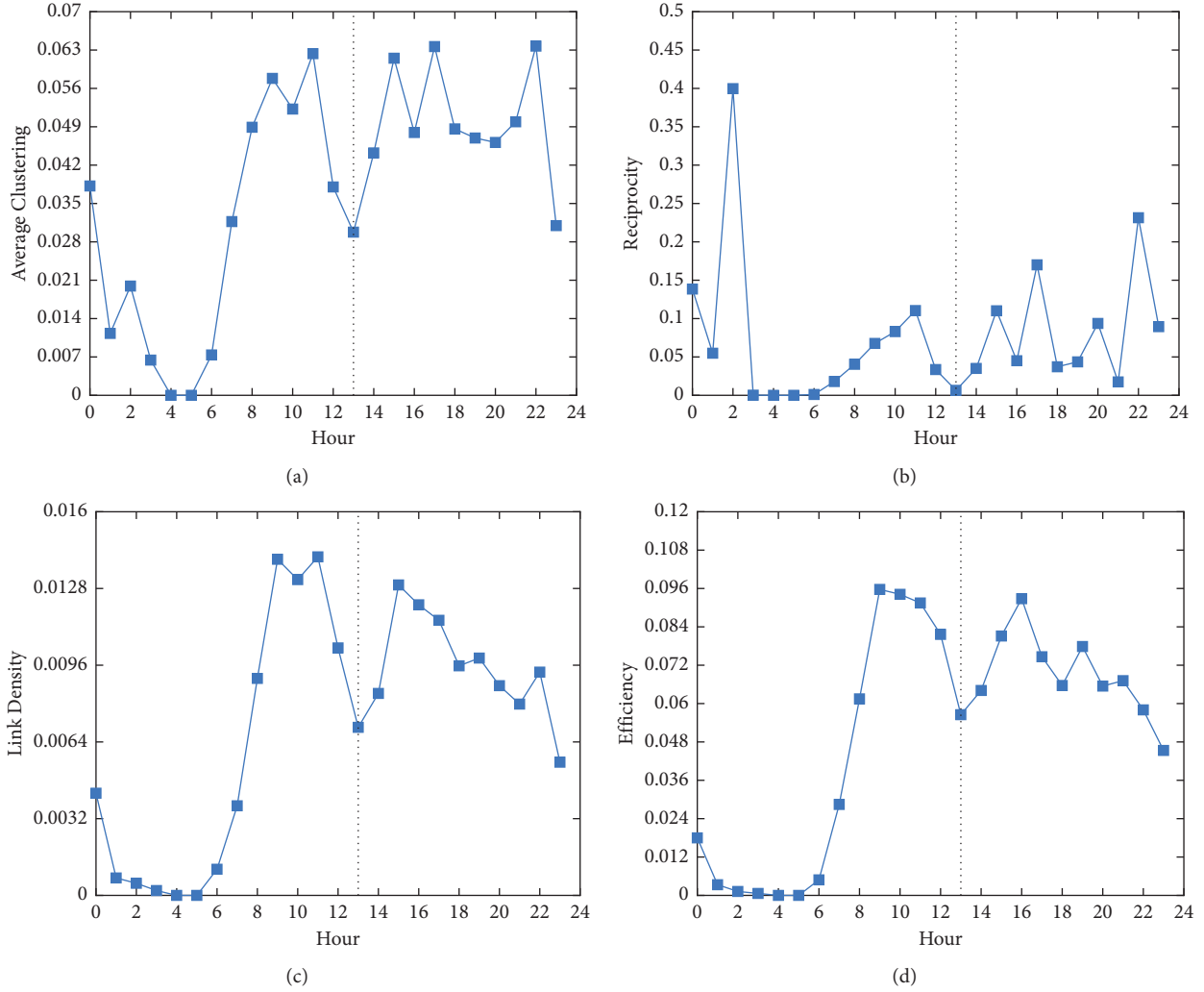
FIGURE 4: (a) Evolution of average clustering (16). (b) Evolution of network reciprocity (17). (c) Evolution of link density (18). (d) Evolution of network efficiency (19).

TABLE 2: Airport information.

Airport size	Amount	Flow (x)
Large	**35**	$x \geq 150$
Medium	**78**	$10 \leq x < 150$
Small	**110**	$x < 10$

risk in airport i, while k_i^{out} is the number of airports whose risk is partly caused by airport i. For airport i, ri > 1 means that outward links of airport i are fewer than its inward links, indicating that airports whose risks affect airport i are more than those whose risks are partly transferred by airport i, while ri < 1 is the opposite. Therefore, we consider that airports with ri > 1 have absorption ability for risk propagation and can mitigate the propagation, while those with ri < 1 have a diffusion effect and aggravate the propagation. To understand the impact of each category of airports, we focus on the situations mentioned above. To analyze the effect of each kind of airports in risk propagation, we calculate the percentage of absorption airports (ri > 1) a to the

number of airports that participated in risk propagation in the same category as follows:

$$a = \frac{n(r_i > 1)}{n_p},\qquad(23)$$

where $n(r_i > 1)$ represents the number of absorption airports, n_p is the total number of airports participating in risk propagation in the same category. High percentage of absorption means a significant role of airports in the absorption of risk propagation.

The final results are shown as Table 4. For February 5, 2019, and March 2, 2019, the majority of small airports exhibit the poor ability and low percentage of absorption of risk propagation due to their lack of resources and high vulnerability. As a result, more small airports tend to propagate risks under the influence of random factors. In contrast, large airports play a significant role in propagation absorption with a high percentage of absorption, indicating that the majority of large airports are affected by many upstream airports, but they impact fewer downstream

(a)

(b)

(c)
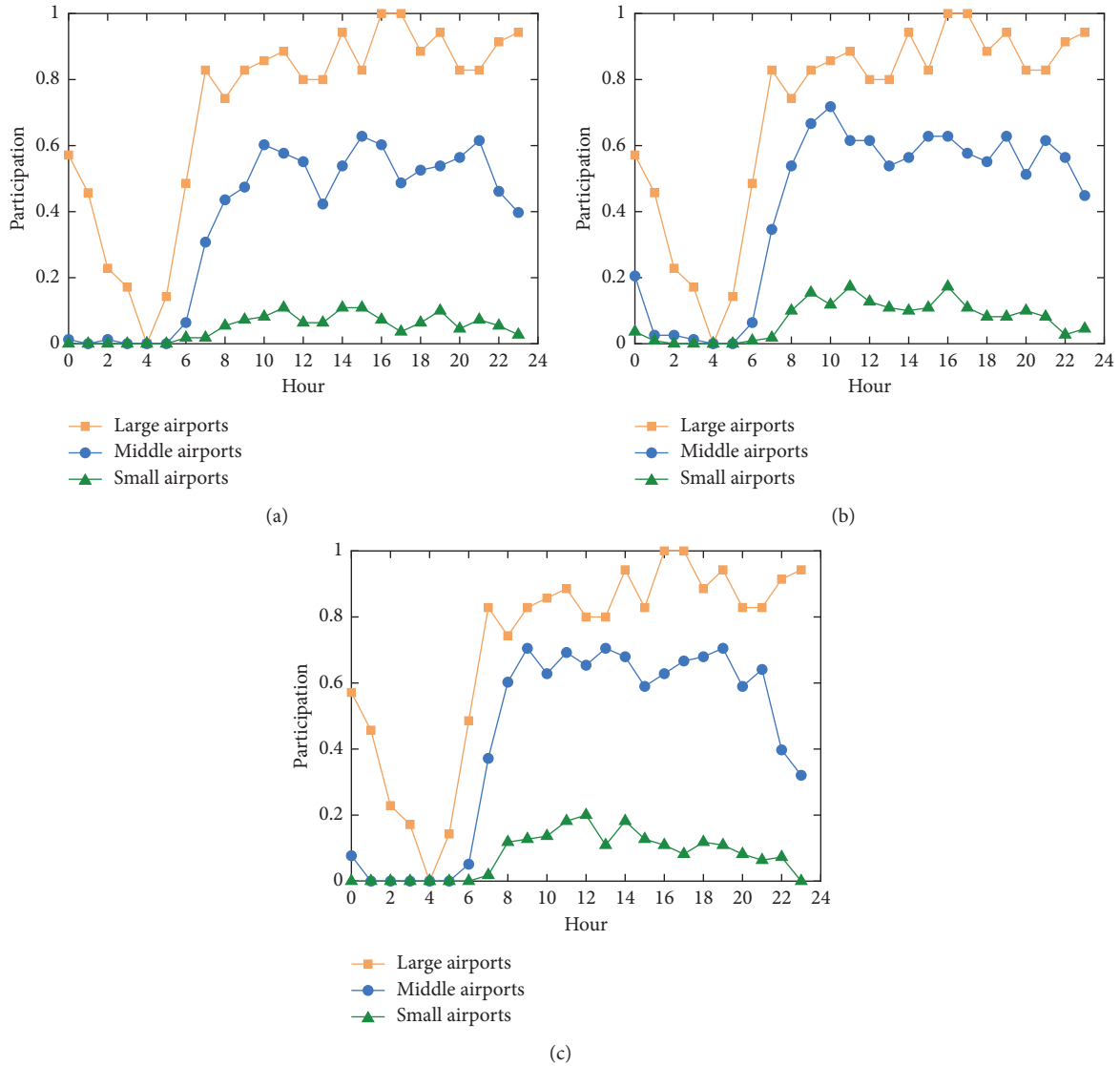
FIGURE 5: (a) Participation of airports on February 5, 2019. (b) Participation of airports on March 2, 2019. (c) Participation of airports on May 1, 2019.

TABLE 3: Average participation of airports ($\overline{P(t)}$) in risk propagation.

Airport size	February 5, 2019	March 2, 2019	May 1, 2019
Large	**0.656**	**0.662**	**0.705**
Medium	**0.368**	**0.420**	**0.433**
Small	**0.049**	**0.073**	**0.077**
Flow	**6452**	**8886**	**9421**

TABLE 4: Percentage of absorption airports (a) for risk propagation.

Airport size	February 5, 2019	March 2, 2019	May 1, 2019
Large	**0.571**	**0.629**	**0.429**
Medium	**0.506**	**0.474**	**0.423**
Small	**0.456**	**0.459**	**0.500**
Flow	**6452**	**8886**	**9421**

airports. In our expectation, the large airports are located in developed cities and bear the pressure of heavy traffic, as well as its large volume of flights with downstream airports, it should have a great impact on the downstream airports, but the findings are contrary to expectations. It may be because they also have sufficient resources and high robustness to partly resist the impact of delays and flow in the meantime. Therefore, they perform best in unoccupied and normal situations. The percentage of medium airports is at a medium level. Nevertheless, the situation was just the opposite on May 1, 2019. Small airports performed best, while edthe majority of large and medium airports tended to propagate risks. Due to a large number of flights and heavy air traffic pressure on May 1, 2019, large and medium airports bear most of the traffic pressure. Since large airports can partly resist the impact of delays and flow, why did they perform worse than small airports? This is mainly due to the overload of parts of large and medium airports. When a load of some airports is high enough to cause overload, the load

will be transferred to other airports according to the links in spite of their high robustness. As a result, small airports receive more risks transferred from large airports due to their overload, which is the reason why small airports performed best on May 1, 2019. It is worth noting that there is a critical flow between 8886 and 9421 where the situation will become opposite for large and small airports. To better control risk propagation, we need to pay attention to the flow between 8886 and 9421 and investigate the critical flow, which is meaningful for CATN risk control and safety improvement.

5. Conclusions

In this paper, we investigated the mechanism of risk propagation among airports in the Chinese air transport network (CATN) from a new perspective, i.e., constructing a risk propagation network (RPN) based on Granger Causality (GC) for each hour of a day and applying network analysis tools to reveal the macroscopic appearance of risk propagation. To realize this model, we firstly built a risk coupling model to obtain the risk time series of airports from delay and flow. Then, we built hourly RPNs based on risk time series and GC test.

We firstly investigated the process of risk propagation in a normal situation and analyzed the properties of the propagation process. There were 4 stages of risk propagation of a day: dissipation of remaining risks, generation, fluctuation, and dissipation of intraday risks. We should pay attention to several key periods as 5 : 00–6 : 00, 9 : 00–10 : 00 and 20 : 00–21 : 00 to better control risk propagation. Then, for temporal properties, the evolution of reciprocity is relatively special due to its independence of the number of network nodes and edges. Finally, we investigated the spatial properties of risk propagation. We found that the participation of airports had a positive correlation with their sizes and the number of flights. Generally, participation in large airports is more than 60%, while participation in small airports is less than 10%. The difference is significant. As a result, large airports are more noteworthy in risk propagation control. For absorption ability, i.e., the ability to prevent risk propagation, we calculated the percentage of absorption airports and found that the largest airports performed best on February 5, 2019, and March 2, 2019, because of their sufficient resources and high robustness to normal situations, while the opposite on May 1, 2019, due to the overload of some large airports causing more risk transferred to small airports. Therefore, it is necessary to focus on the situations where flow is between 8886 and 9421 to prevent the overload of large airports.

The results of the analysis provide a theoretical basis for the optimization and management of air transport network safety and efficiency. In addition, the results provide some useful suggestions for reducing air transport network risks and enhancing network load capacity in order to establish an advanced intelligent air traffic system. We can better control risk propagation in the future according to the key periods in a day, high participation airports, and vital flow range in our analysis. Nonetheless, the study of using RPN to study

airport risk propagation could be extended further. For example, we found the phenomenon of overload of airports. It is interesting to compare more situations to obtain the critical value of flow to cause an overload of large airports. According to critical flow, we can better control the risk propagation and optimize the overall risks of airports. In addition, it is also meaningful to study the relationship between the number of daily flights and the absorption percentage of airports in detail.

Data Availability

The flight data used in this paper are available at http://www.variflight.com/en/.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] P. Belobaba, A. Odoni, and C. Barnhart, *The Global Airline Industry*, John Wiley & Sons, Hoboken, NJ, USA, 2009.

[2] K. Cai, J. Zhang, W. Du, and X. Cao, "Analysis of the Chinese air route network as a complex network," *Chinese Physics B*, vol. 21, no. 2, Article ID 028903, 2012.

[3] X. Sun, V. Gollnick, and S. Wandelt, "Robustness analysis metrics for worldwide airport network: a comprehensive study," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 500–512, 2017.

[4] A. E. Motter and Y. C. Lai, "Cascade-based attacks on complex networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 66, Article ID 065102, 2002.

[5] N. Kafle and B. Zou, "Modeling flight delay propagation: a new analytical-econometric approach," *Transportation Research Part B: Methodological*, vol. 93, no. partA, pp. 520–542, 2016.

[6] J. Li, Y. Chen, and P. He, "Research on risk monitoring and simulation of China's civil aviation industry," *China Safety Science Journal*, vol. 1920 pages, 2009.

[7] N. Pyrgiotis, K. M. Malone, and A. Odoni, "Modelling delay propagation within an airport network," *Transportation Research Part C: Emerging Technologies*, vol. 27, pp. 60–75, 2013.

[8] M. Zanin, S. Belkoura, and Y. Zhu, "Network analysis of Chinese air transport delay propagation," *Chinese Journal of Aeronautics*, vol. 30, no. 2, pp. 491–499, 2017.

[9] S. Belkoura, J. M. Peña, and M. Zanin, "Beyond linear delay multipliers in air transport," *Journal of Advanced Transportation*, vol. 2017, Article ID 8139215, 11 pages, 2017.

[10] C.-L. Wu and K. Law, "Modelling the delay propagation effects of multiple resource connections in an airline network using a Bayesian network model," *Transportation Research Part E: Logistics and Transportation Review*, vol. 122, pp. 62–77, 2019.

[11] B. Baspinar and E. Koyuncu, "A data-driven air transportation delay propagation model using epidemic process

models," *International Journal of Aerospace Engineering*, vol. 2016, Article ID 4836260, 11 pages, 2016.

[12] M. Bernardo, V. D. P. Servedi, and L. Vittorio, "Congestion transition in air traffic networks," *PLoS One*, vol. 10, no. 5, Article ID e0125546, 2015.

[13] X. Dai, M. Hu, and W. Tian, "Modeling congestion propagation in multistage schedule within an airport network," *Journal of Advanced Transportation*, vol. 2018, pp. 1–11, Article ID 6814348, 2018.

[14] S. Li, *Research on Identification and Prediction Methods of Air Traffic Congestion*, Tianjin University, Tianjin, China, 2013.

[15] M. Zhang, B. Liang, S. Wang, M. Perc, and W. Du, "Analysis of flight conflicts in the Chinese air route network," *Chaos, Solitons & Fractals*, vol. 112, pp. 97–102, 2018.

[16] B. Song, G. Jiang, Y. Song, and L. Xia, "Rapid identifying high-influence nodes in complex networks," *Chinese Physics B*, vol. 24, no. 10, Article ID 100101, 2015.

[17] X. Yan, Y. Cui, and S. Ni, "Identifying influential spreaders in complex networks based on entropy weight method and gravity law," *Chinese Physics B*, vol. 29, no. 4, Article ID 048902, 2020.

[18] T. Bjerga, T. Aven, and E. Zio, "Uncertainty treatment in risk analysis of complex systems: the cases of STAMP and FRAM," *Reliability Engineering & System Safety*, vol. 156, pp. 203–209, 2016.

[19] Z. Mohaghegh, R. Kazemi, and A. Mosleh, "Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: a hybrid technique formalization," *Reliability Engineering & System Safety*, vol. 94, no. 5, pp. 1000–1018, 2009.

[20] W. Li, L. Zhang, and W. Liang, "An Accident Causation Analysis and Taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives," *Safety Science*, vol. 92, pp. 94–103, 2017.

[21] A. Nguyen, P. Do, and A. Grall, "Multi-level predictive maintenance for multi-component systems," *Reliability Engineering & System Safety*, vol. 144, pp. 83–94, 2015.

[22] M. Zanin, X. Sun, and S. Wandelt, "Studying the topology of transportation systems through complex networks: handle with care," *Journal of Advanced Transportation*, Article ID 3156137, 2018.

[23] X. Sun, S. Wandelt, and F. Linke, "On the topology of air navigation route systems," *Proceedings of the Institution of Civil Engineers - Transport*, vol. 170, no. 1, pp. 46–59, 2017.

[24] M. Zanin and F. Lillo, "Modelling the air transport with complex networks: a short review," *The European Physical Journal - Special Topics*, vol. 215, no. 1, pp. 5–21, 2013.

[25] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Systemic delay propagation in the US airport network," *Scientific Reports*, vol. 3, no. 3, p. 1159, 2013.

[26] B. Campanelli, P. Fleurquin, A. Arranz et al., "Comparing the modeling of delay propagation in the US and European air traffic networks," *Journal of Air Transport Management*, vol. 56, pp. 12–18, 2016.

[27] M. Zanin, "Can we neglect the multi-layer structure of functional networks?" *Physica A: Statistical Mechanics and Its Applications*, vol. 430, pp. 184–192, 2015.

[28] W.-B. Du, M.-Y. Zhang, and Y. Zhang, "Delay causality network in air transport systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 118, pp. 466–476, 2018.

[29] Q. Li and R. Jing, "Characterization of delay propagation in the air traffic network," *Journal of Air Transport Management*, vol. 94, 2021.

[30] M. Zanin, "Simplifying functional network representation and interpretation through causality clustering," *Scientific Reports*, vol. 11, no. 1, 2021.

[31] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[32] C. W. J. Granger, "Some recent development in a concept of causality," *Journal of Econometrics*, vol. 39, no. 1-2, pp. 199–211, 1988.

[33] M. R. Frank, N. Obradovich, and L. Sun, "Detecting reciprocity at a global scale," *Science Advances*, vol. 4, no. 1, Article ID eaao5348, 2018.

[34] A. Montlaur and L. Delgado, "Flight and passenger delay assignment optimization strategies," *Transportation Research Part C: Emerging Technologies*, vol. 81, pp. 99–117, 2017.

[35] T. Bolić, L. Castelli, L. Corolli, and D. Rigonat, "Reducing ATFM delays through strategic flight planning," *Transportation Research Part E: Logistics and Transportation Review*, vol. 98, pp. 42–59, 2017.

[36] G. Gui, F. Liu, J. Sun, J. Yang, Z. Ziqi, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2019.

[37] M. Zanin, Y. Zhu, R. Yan, P. Dong, X. Sun, and S. Wandelt, "Characterization and prediction of air transport delays in China," *Applied Sciences*, vol. 10, no. 18, Article ID 6165, 2020.

[38] J. Xiong and M. Hansen, "Value of flight cancellation and cancellation decision modeling," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2106, no. 1, pp. 83–89, 2009.

[39] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.

[40] W. Cheung and K. S. Lai, "Lag order and critical values of the augmented dickey-fuller test," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 277–280, 1995.

[41] W. Vandaele, "Wald, likelihood ratio, and Lagrange multiplier tests as an F test," *Economics Letters*, vol. 8, no. 4, pp. 361–365, 1981.

[42] A. Cook, H. A. P. Blom, F. Lillo, and R. N. Zanin, "Applying complexity science to air traffic management," *Journal of Air Transport Management*, vol. 42, pp. 149–158, 2015.

[43] A. Broder, R. Kumar, F. Maghoul et al., "Graph structure in the web," *Computer Networks*, vol. 32, no. 1-6, pp. 309–320, 2000.

[44] G. Fagiolo, "Clustering in complex directed networks," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 76, no. 2, Article ID 026107, 2007.

[45] L. da Rocha, "The International School for Advanced Studies (SISSA), find out more Structural evolution of the Brazilian airport network," *Journal of Statistical Mechanics: Theory and Experiment*, Article ID P04020, 2009.

[46] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[47] M. D. Laat, V. Lally, L. Lipponen, and R. J. Simons, "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: a role for Social Network Analysis," *International Journal of Computer-Supported Collaborative Learning*, vol. 2, no. 1, pp. 87–103, 2007.

[48] L. Vito and M. Massimo, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, Article ID 198701, 2001.

WILEY | Hindawi

Research Article

Deep Learning-Based Dynamic Stable Cluster Head Selection in VANET

Muhammad Asim Saleem [ID],[1] Zhou Shijie,[1] Muhammad Umer Sarwar,[2] Tanveer Ahmad [ID],[3] Amarah Maqbool,[4] Casper Shikali Shivachi [ID],[5] and Maham Tariq[4]

[1]*School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[2]*Department of Computer Science, Government College University, Faisalabad, Pakistan*
[3]*Department of Electrical and Electronic Engineering, Auckland University of Technology, Auckland 1010, New Zealand*
[4]*Department of Computer Science, Government College Women University, Faisalabad, Pakistan*
[5]*South Eastern Kenya University, Kitui, Kenya*

Correspondence should be addressed to Casper Shikali Shivachi; cshikali@seku.ac.ke

VANET is the spontaneous evolving creation of a wireless network, and clustering in these networks is a challenging task due to rapidly changing topology and frequent disconnection in networks. The cluster head (CH) stability plays a prominent role in robustness and scalability in the network. The stable CH ensures minimum intra- and intercluster communication, thereby reducing the overhead. These challenges lead the authors to search for a CH selection method based on a weighted amalgamation of four metrics: befit factor, community neighborhood, eccentricity, and trust. The stability of CH depends on the vehicle's speed, distance, velocity, and change in acceleration. These all are included in the befit factor. Also, the accurate location of the vehicle in changing the model is very vital. Thus, the predicted location with the Kalman filter's help is used to evaluate CH stability. The results have shown better performance than the existing state of the art for the befit factor. The change in dynamics and frequent disconnection in communication links due to the vehicle's high speed are inevitable. To comprehend this problem, a graphing approach is used to evaluate the eccentricity and the community neighborhood. The link reliability is calculated using the eigengap heuristic. The last metric is trust; this is one of the concepts that has not been included in the weighted approach to date as per the literature. An adaptive spectrum sensing is designed for evaluating the trust values specifically for the primary users. A deep recurrent learning network, commonly known as long short-term memory (LSTM), is trained for the probability of detection with various signals and noise conditions. The false rate has drastically reduced with the usage of LSTM. The proposed scheme is tested on the real map of Chengdu, southwestern China's Sichuan province, with different vehicular mobilities. The comparative study with the individual and weighted metric has shown significant improvement in the cluster head stability during high vehicular density. Also, there is a considerable increase in network performance in energy, packet delay, packet delay ratio, and throughput.

1. Introduction

A vehicular ad hoc network, popularly known as VANET, is a particular type of Mobile Ad Hoc Network (MANET), where mobile nodes are considered vehicles moving on the road [1]. The advent of developments in VANET has paved a way for the growth of Intelligent Transportation System (ITS) applications. These are broadly classified into safety-oriented applications that intended to increase safety and reduce fatal road accidents. The other is nonsafety that aims to provide additional services to passengers like traffic management, information sharing, and so on [2–4]. The vehicles communicate through the On-Board Units (OBUs), known as Vehicle to Vehicle (V2V). This becomes essential for most ITS application owing to low cost and availability [5]. Also, the vehicles can communicate with the auxiliary facilities (or installed infrastructure) like Road-Side Units (RSUs), using Vehicle-to-Infrastructure

(V2I) communications. The complete model design for VANET communication is shown in Figure 1.

Typically, in the V2V communication protocol, two approaches are used: flooding and relaying for data dissemination through the network. In flooding, each node broadcasts the received data packet to its neighbor received from the source. This process is repeated with the motive to reach the data packet to the source. In a flat, dense network, this approach leads to a storm problem [6]. In the relaying approach, the message broadcast from the source is forwarded to all the neighboring vehicles, and then a few are selected to forward it. The probability of success for data delivery is increased in the relaying approach but with high overhead and delay. To tackle such a problem, clustering in VANET is one of the prominent solutions. Some notations are listed in Table 1, which will be used for further gradation.

Through clustering in VANETs, there is information gathering, aggregation, and dissemination. Clustering is used to partition the network into smaller groups of moving vehicles. Typically, the dynamic clustering is done, which groups the vehicles on the fly as there is no physical connection among them, and they all are moving. There are several benefits like efficient bandwidth, proper distribution of resources, and scalability that the clustering approach offers [7]. There are many methods employed for clustering in VANETs [8, 9]. The urban roads are complicated as vehicle positions change frequently; they have unevenly distributed; thus, their routing and the forwarding capabilities change with their position. These issues result in an unstable network and a need for a cluster model that offers high stability in the dynamic VANET scenario. The methods especially designed for the stability of the cluster are compared in Table 2.

Through this analysis, we can find that most of the schemes conceived for stability are formulated on the effect of a stability parameter. The weighted scheme has also considered the parameters like speed, distance, acceleration, or link time. These schemes involve no information about the type of vehicle present in the traffic scenario. The connectivity with the neighbor is also not explored. These have been designed for highway where the vehicle's speed is too high, so the static cluster formation concerning speed and position will change dynamic, which will affect the stability of the cluster head. With all these literature analyses, the authors have designed a weighted cluster head selection based on multimetric. The befit factor, community neighborhood, eccentricity, and trust value are introduced. Also, the concept of trust has not to be utilized to date. The following are the contributions made by the authors in this paper:

(i) The befit factor designed by tuthorshe authors in [19] is based on T_{leave}, the time required by the vehicle to complete the remaining segment of the lane. This is calculated based on the speed of the vehicle. The speed of any vehicle is a variable quantity; in such a scenario, obtaining the accurate speed is essential for cluster head stability. This issue has been addressed in this paper; the authors have predicted each vehicle's current location, and this position has been used to calculate the befit factor. The authors showcase the validation through the comparative results.

(ii) The authors also intended to improve the VANET performance using an auxiliary facility, Road-Side Units (RSUs), improving the network's performance. However, the trade-off between the number of RSUs for better coverage and the high installation and maintenance costs is resolved in this paper by considering the angle suspended by the lanes. The effect of crossroads is also included to avoid ambiguity in the direction of the lane.

(iii) The authors have presented spectrum sensing as a classification problem and proposed the sensing method based on the recurrent neural network by employing the LSTM. The signal's power spectrum is given as input to the LSTM and uses various signal noise data types to train the network. The decision is made according to the confidence of the noise class. As the designed method is based on deep learning, the network automatically learns the parameters and can adapt well to different noise levels. This concept has been further extended to segregate the vehicles into two categories: the primary (PU) and secondary users. The primary includes vehicles like ambulance, a civic service that needs immediate assistance. The PU are given access to the network bandwidth and improve the stability of the VANET.

(iv) VANET has a dynamic topology where the vehicles travel at high speed and have frequent arrival with an irregular interval between them. To have a stable cluster, they have to be evolving in nature. The authors have designed the cluster-based evolving graph model that calculates the metrics that could select a stable cluster head to address this use.

The authors designed a new multimetric weighted CH selection scheme that considers different selection metrics that increase the cluster stability and efficiency. The designed scheme is well tested on the real data map, taken for the region Chengdu, southwestern China's Sichuan province. The clustering efficiency is tested for various vehicle densities, and also the network performance is evaluated using the parameters. The rest of the paper is organized as follows: In Section 2, the network model and the position prediction using the Kalman filter and the RSU placement are presented. In Section 3, the multimetric weighted CH selection scheme is discussed with all its metrics. Section 5 presents the simulation work and compares the proposed scheme with the existing state of the art. In the last section, the authors conclude the work with the prospects of the work.

2. Network Model with Prediction and RSU Deployment

The proposed scheme is mainly concentrated on the highway environment. Here, the model is specially designed for vehicles that are well equipped with a positioning system
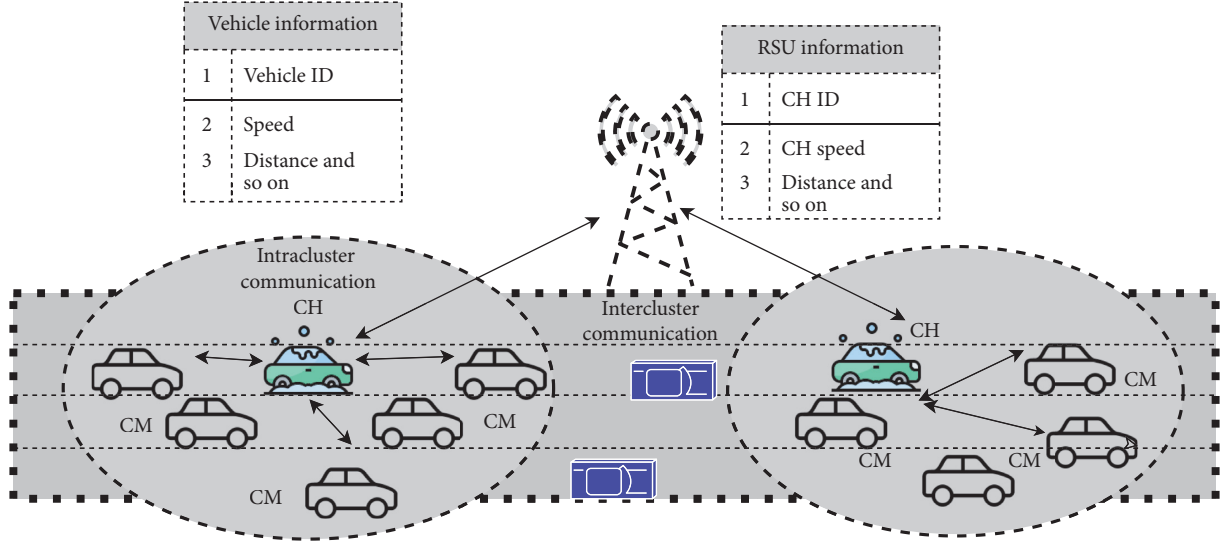
Figure 1: The complete system model.

Table 1: List of symbols with the nomenclature.

Symbol	Nomenclature
A	Affinity matrix
BF	Befit factor
C	Cluster
CN	Community neighbors
D	Distance
Dig	Diagonal matrix
d_{ij}	Diagonal elements
ΔD_{thr}	Threshold distance
b_u, b_f, b_o	The bias of each gate in LSTM
NH_D	Neighborhood
ψ_{vehi}	Relative velocity
ID	Vehicular ID
L	Length of the lane
Ecc	Eccentricity
$G(M, E, rl)$	Graph tuple
E	Energy
M	Total number of vehicles
N	Total number of clusters
NCC	Neighbor connection centrality
h	Channel
$\mathcal{H}_0$	Hypothesis
P_d	Probability of detection
t_p	Delay in p th path
P_f	Probability of false alarm
ϱ	Wavelength
rl	Link reliability
$\text{RSU}_{\text{tran}}^{L-ID}$	RSU
$\mathbb{Q}(\cdot)$	Complementary distribution function
S_{thr}	Speed threshold
T_{leave}	Time to leave
TRF	Transmission range factor
T	Total simulation time
TR	Transmission range
χ^2	Chi-square probability distribution function
t	An instant of time
tanh	Activation function
V_{vehi}	Velocity of the vehicle
ΔV_{thr}	Velocity threshold

TABLE 1: Continued.

Symbol	Nomenclature
V_{avg}	Average velocity
w	A random number $[0, 1]$
σ_x^2	Standard deviation
γ	SNR
$\hat{x}$	Predicted coordinates
$[x, y]$	Coordinates of vehicle
$\varnothing$	Angle of the vehicle
no_lane	Total number of lanes
δ	Penalty
λ	Eigenvalues
$\sigma_f, \sigma_u, \sigma_0$	Gates in LSTM
A_p	Channel gain
P_r	Probability of receiving
G	Antenna gain
$\odot$	Hamdard function
$T(Y)$	Test statistics

TABLE 2: Comparison of various state-of-the-art schemes designed for CH stability.

Year	Parameters for CH selection							Clustering scheme		Network and utility scenarios
	Speed	ID	Signal strength	Relative speed	Neighbor count	Link lifetime	Position	Individual	Weighted	
2011 [10]	Yes	Yes	Yes	—	—	—	—	Yes	—	Multihop with random walk mobility
2012 [11]	—	—	—	Yes	—	—	—	PDR	—	Multihop with manhattan mobility scenario
2013 [12]	—	—	—	Yes	—	—	—	Least mobility	—	Multihop with two lanes
2015 [13]	—	—	—	Yes	Yes	—	—	Neighbor count	—	Multihop with a dynamic scenario
2018 [14]	Yes	—	—	—	Yes	Yes	Yes	Priority	—	Multihop with two lanes
2018 [15]	Yes	—	—	—	—	Yes	Yes	The conditional probability of link reliability	—	Singel-hop with an urban scenario
2019 [16]	Yes	—	Yes	Yes	—	Yes	—	—	Yes	Singel-hop with an urban scenario and highway
2019 [17]	Yes	—	—	—	—	—	Yes	—	Yes	Singel-hop with highway
2019 [18]	Yes	—	—	Yes	—	—	Yes	—	Yes	Singel-hop with highway

(e.g., a GPS), which acquires information about the current location and an IEEE 802.11p-compliant radio transceiver that enables the communication between the other vehicles and the RSU (Road-Side Unit). The cars typically share different geographical proximity information, which can be harnessed for effective traffic management and organization. The designed weighted, dynamic, adaptive, and fuzzy clustering algorithm is manifested on the same ideology [20–22]. Here, the vehicles are grouped in small clusters based on their proximity concerning the RSU deployed near the lanes. A cluster head (CH) is selected based on the amalgamation of four components befit, trust, community neighborhood, and eccentricity. The rest are cluster members (CM). There is an exchange of information among the members using Vehicle-to-Vehicle (V2V) communication.

The complete knowledge is like the parameter speed, acceleration, distance travelled, location, and so on, which are maintained in each CM. The CH plays its role in broadcasting important information to the in-range RSU and the CM.

In this section, a detailed discussion of the designed model is done. Each parameter for the CH selection is examined and modeled without increasing the network's complexity and overhead delay.

2.1. Network Model. The network model considered is shown in Figure 2(a) that shows the nodes in the VANET in an unclustered fashion. In Figure 2(b), the nodes are now under each cluster head in continuous communication

(a)

(b)
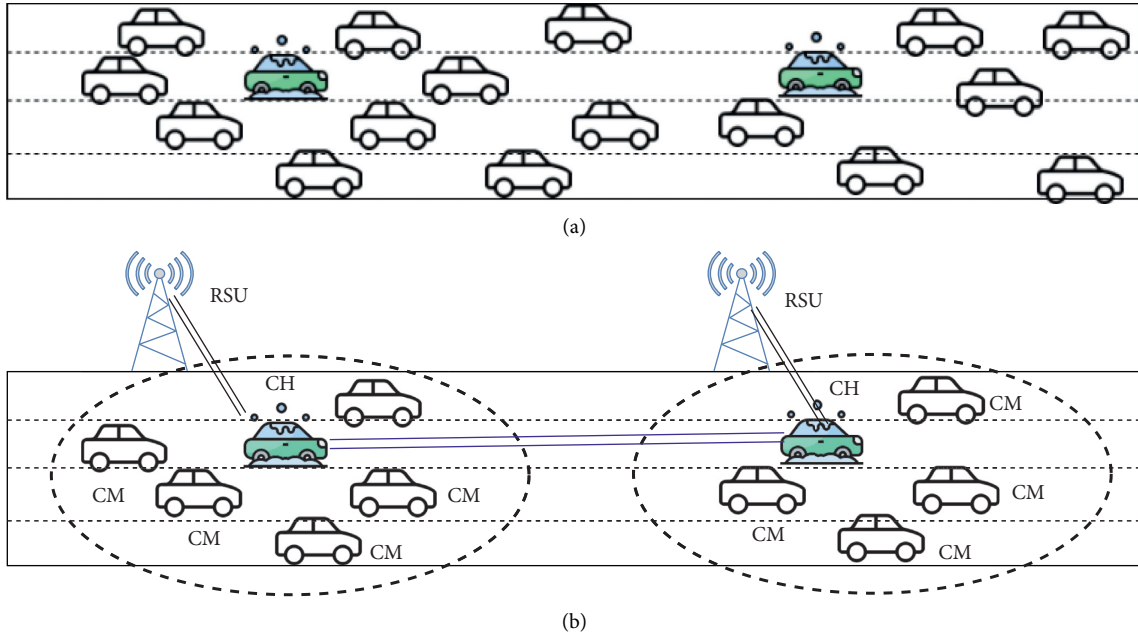
FIGURE 2: The proposed analogy for VANET clustering: (a) flat structure and (b) cluster concept with CH and RSU.

with the RSU. The network model's consideration is as follows: (1) The real-time road map is taken for the study. (2) The vehicle mobility is considered for a fixed time interval for study. (3) Each vehicle is considered as a node that is variable with time, and each car possesses a unique ID. (4) RSUs are effectively placed in each lane with proper analysis of the road with an identical communication range. (5) Each vehicle and the RSU are equipped with buffer memory and a GPS device for real-time data collection and synchronization. (6) The communication among the CM and CH-CM is enabled through V2V. (7) Each vehicle's position is estimated through the Kalman filter that intimates the RSU about the range of vehicles and ensures noninterruptible communication with smooth transitions among the RSU. (8) The network over time evolves into several dynamic clusters; the maximum hops between the cluster head and its CM are maxing hop. (9) Packet data of size 64 bits is taken.

2.2. RSU Deployment. To overcome the shortcomings of overhead delay and fewer storage facilities, the vehicles need an auxiliary facility like Road-Side Units (RSUs) to improve network performance. There is always a trade-off between the areas' total coverage with a sufficient number of RSUs with minimum installation and maintenance cost. This leads to the search for optimal locations where the RSU can maintain an efficient communication channel at a low price. The location affects communication as any void area will cause a drop of packets transferred among the vehicle and RSU. In this deployment, the lane coordinates are obtained from the SUMO simulation. Each RSU is assumed with a fixed circular transmission range RSU_{tran}^{L-ID}. There are placed on either of the roadsides. This ensures no interference in the coverage range of RSU and each lane as sufficient coverage

without any void area [23]. The deployment coordinates are calculated using the particular direction of the lane through the geometric methodology. At the beginning of any lane, an RSU is placed to get the next location lane curvature $\cos(\varnothing)$ which is used. Let the location of any point on the lane be expressed as $[x, y]$. The next location is calculated as

$$x_{next} = x_{old} + RSU_{tran}^{L-ID} * \cos(\varnothing),$$
$$y_{next} = y_{old} + RSU_{tran}^{L-ID} * \sin(\varnothing). \tag{1}$$

That takes into consideration the RSU transmission range. With the introduction of the angle factor, the cross-lane will not suffice to formulate such conditions. There are few checkpoints designed that could ensure that cross-lane will be utilized to place the RSU continuously. The deployment of the RSU is shown in Figure 3.

2.3. Location Prediction. After optimally placing the RSU on the desired area map, the next task is to reduce the drop rate in the communication between the RSU and CH. This will ensure better overall network performance and increase the reliability of the VANET. To make this possible, the RSU must be aware of the vehicle's next stamp position beforehand. In this work, the authors have utilized the work proposed by Kalman [24, 25] that predicts the vehicle's location with geographical routing. The prediction of the vehicle's location depends on the direction and the velocity. Also, it has been observed that the vehicle's angle plays a crucial role in predicting the position as it varies over time. The prediction is made using the Kalman filter. Each vehicle runs a Kalman filter prediction, which predicts the vehicles' position, velocity, and direction. The position vector at the moment t is $x(t) = [x(t), y(t), V_{vehi}(t), \phi(t)]^{T}$ and the predicted values for each parameter at the next instant are
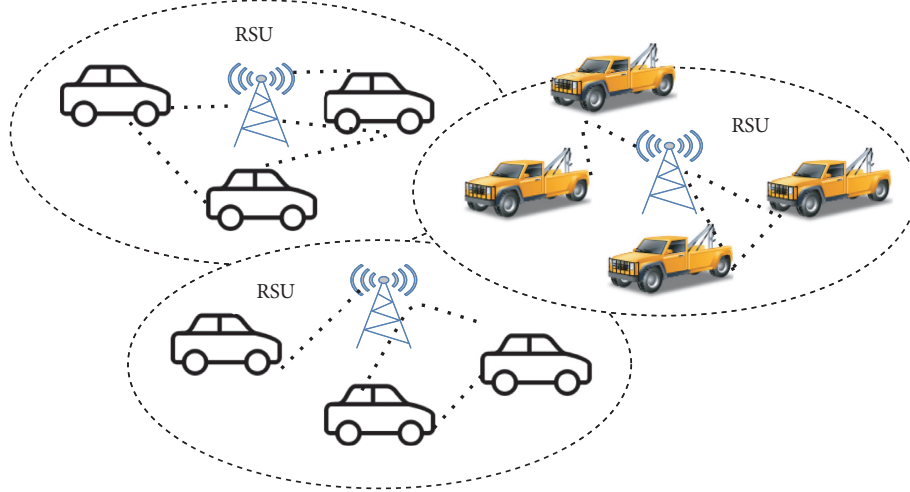
FIGURE 3: The deployment of RSU on the lane for the circular transmission range.

$$\check{x}(t) = \left| \widehat{x}(t), \widehat{y}(t), \widehat{V_{\text{vehi}}}(t), \widehat{\phi}(t) \right|^{\widehat{}} T,$$

$$\widehat{x}(t+1) = \left| \widehat{x}(t) + \widehat{V_{\text{vehi}}}(t) \times \cos(\widehat{\varnothing}(t)) \times \Delta T \right|, \quad (2)$$

$$\widehat{y}(t+1) = \left| \widehat{y}(t) + \widehat{V_{\text{vehi}}}(t) \times \sin(\widehat{\varnothing}(t)) \times \Delta T \right|.$$

The individual RSU scans for vehicles in its range; then using the following judgment equation, new predicted positions are derived. The difference between the past and the predicted position is used to get the new update value with the information on the angle suspended by the lane.

3. Dynamic Weighted Cluster Head Stability Algorithm

The designed clustering model has been broadly classified into two parts: (1) cluster formation and (2) CH selection. In the cluster formation, the vehicles are segregated into small groups and then based on a weighted approach, a weighted formulation of four metric befit factor (BF), community neighborhood (CN), eccentricity (Ecc), and trust (T) is designed to select the cluster head, and also these parameters ensure the stability of the cluster head over the time. All of these metrics are normalized to avoid the overriding effect.

3.1. Cluster Formation.

The RSU guides the cluster formation. Algorithm 1 encapsulates the cluster formation where each vehicle enters the lane for the first timestamp broadcast and communicates with all the RSU in its communication range. To form a stable cluster, two conditions are observed for that timestamp, the distance D and relative change in the vehicle's speed V_{vehi} in that timestamp. A distance between the RSU and that car is calculated. If this distance is less than a threshold ΔD_{thr}, the vehicle is attached to that RSU temporarily. This process is carried out for all the vehicles present in the vicinity and moving in the same direction. In the real-time scenario,

each vehicle's speed level may be different, and this variation can seriously affect the cluster formation. For stable cluster formation, the change in relative speed is also observed if it is less than a threshold ΔV_{thr} and then that vehicle gets permission to get permanently attached to that RSU. Now, the RSU stores the ID of that vehicle and becomes a cluster member (CM).

3.2. Cluster Head Selection.

The next step is selecting the cluster head, and it is a node in VANET that coordinates or heads the cluster. It takes the responsibility of broadcasting, discovery, and maintenance of the routing path. It remains in the contact of the RSU to sustainably maintain the intra- and intercommunication channels. The main motive of designing any clustering algorithm is the stability of the cluster head in VANET. The vehicles' high dynamic mobility can lead to frequent reclustering and eventually decreases the cluster stability. The methodology designed in this paper for the cluster head selection is a weighted combination of four factors, as mentioned above. All these parameters direct towards the search for a stable CH. All these and their amalgamation for selecting a suitable CH are discussed in the following.

3.2.1. Befit Factor (BF).

This parameter is defined to maximize the stability of the cluster structure. To ensure this, the elected CH is expected to stay connected with all the cluster members for a longer duration. Thus, the BF is derived from the three metrics as designed in [19]

$$\text{BF} = w_1 \times T_{\text{leave}} + w_2 \times \psi_{\text{vehi}} + w_3 \times \text{NH}_D, \quad (3)$$

where $w_1, w_2,$ and w_3 are the corresponding weights that vary in the range $[0, 1]$ satisfying the condition that ($w_1 + w_2 + w_3 = 1$) and can be reformulated by the local authority based on the road conditions and the cluster member behavior. The first metric is T_{leave}, which is time to leave; this is the time required for a vehicle to complete the lane's

```
Input: Velocity V_vehi and location [x, y]; the number of the lanes in a map (no_lane); time_span
Output: Cluster (C), N no of clusters
For t = 1:time_span
For i = 1: no_lane
    If any vehicle detected in RSU_tran^(L–ID)
        D: RSU location [x_RSU, y_RSU] and vehicles location [x_vehi, y_vehi]
        V_vehi: Speed of the vehicle
        If D < ΔD_thr ‖ V_vehi < ΔV_thr
            RSU ⟵ ID
        Endif
            C ⟵ RSU, ∀ ID
        Endif
        End
End
N ⟵ ∀ C; C ≠ Null.
```

ALGORITHM 1: Cluster formation.

remaining segment. This factor ensures selecting a CH with enough time left to complete the lane that makes it head for a longer time. It is calculated using L, the length of the lane, and D, the distance covered by a vehicle on the road segment, and t, the time consumed by the vehicle to drive that D section of the lane.

$$T_{\text{leave}} = \frac{L - D}{t}. \tag{4}$$

The second metric is ψ_{vehi} defined as relative average speed; this parameter determines how close a vehicle's velocity is to its neighbors. A reward function is conceived that takes into account the velocity of vehicles over the long term. Speed of each vehicle (V_{vehi}) is evaluated. Accordingly, their speed is either rewarded or penalized with an absolute value (δ); correspondingly, the relative average speed is incremented or decremented, as shown in

$$\psi_{\text{vehi}}(t + 1) = \psi_{\text{vehi}}(t) + \delta \; ; \left| V_{\text{vehi}} - V_{\text{avg}} \right| \leq S_{\text{thr}},$$
$$\psi_{\text{vehi}}(t + 1) = \psi_{\text{vehi}}(t) - \delta \; ; \left| V_{\text{vehi}} - V_{\text{avg}} \right| > S_{\text{thr}}, \tag{5}$$

where S_{thr} is the parameter that ensures that the vehicle moving with the velocity V_{vehi} is almost travelling at the same speed as that of the neighbors. The initial value of ψ_{vehi} is calculated using the TraCI parameters, and δ is taken as 0.01. The last metric is neighborhood degree (NH_D), and it is defined as the number of the neighbors whose speed difference with the vehicle is less than a threshold S_{thr}.

3.2.2. Eccentricity (Ecc). The next parameter is eccentricity; in real time, communication links break more frequently due to the vehicles' high speed. To maintain the link, there is a requirement for an evolving cluster model. Usually, reclsutering will become inevitable once the CH resigns or loses its suitability to continue as a CH. To ensure stability, the concept of eccentricity is introduced. Here, an evolving graph-based model is designed by using spectral clustering. A vehicular graph topology is intended to be $G(M, E, rl)$, where M is the number of vehicles present in the timestamp t, E is the ordered pair of the links among the vehicles, and rl is the link reliability. The affinity matrix is constructed to represent the graph topology for dimensions $M \times M$.

$$A = \begin{cases} r_{ij}, & \text{if } (D_i, D_j) \in E, \\ 0, & \text{if } i = j, \\ \infty, & \text{otherwise}, \end{cases}$$
$$A = \begin{bmatrix} r_{11} & \cdots & r_{1M} \\ \vdots & r_{22} & \vdots \\ r_{M1} & \cdots & r_{MM} \end{bmatrix}. \tag{6}$$

The tool employed for the spectral clustering is the Laplacian graph. The Laplacian graph is calculated for the affinity matrix:

$$G_L = \text{Dig} - A, \tag{7}$$

where Dig is the diagonal matrix with elements. Also,

$$d_{ij} = \sum A_{ij}. \tag{8}$$

In spectral clustering, eigenvectors of a similarity/affinity matrix are derived from the original dataset. The eigendecomposition of the graph will be serving as the model for dimensionality reduction of mobile vehicles. The optimal number of clusters is calculated by using the eigenvalue of the Laplacian graph. Based on the eigenmap heuristic [26], the eigenvalues $\lambda_i; i = \{1, 2, \ldots, M\}$ get sorted in the ascending order out of which k is picked that serves as the clusters in that timestamp.

$$k = \arg\max(\lambda_{i+1} - \lambda_i). \tag{9}$$

After the number of the clusters is obtained, then k eigenvectors are extracted in a matrix with dimensionality $M \times k$. In the last, K-means is applied to get the optimal number of clusters. Ecc is the mean/average eigenscore of each group that is calculated as in [27]

$$\text{Ecc}_i = \frac{1}{|N_i|} \sum_{\lambda_i \varepsilon N_i} \lambda_i. \tag{10}$$

The maximum value of Ecc ensures a stable cluster head selection designed based on the evolving graph.

3.2.3. Community Neighborhood (CN).

The evolving Laplacian graph also provides information about neighbors. The importance of neighbor ensures the CH's stability as the cluster member will not change for a given timestamp that will establish a reliable link among them. There is designing of the CN using the transmission factor (TRF); it represents the reliability of the connection between two vehicles that satisfy the following condition:

$$\text{TRF}\left(r_{ij}\right) = \begin{cases} 0, & if \ r_{ij} < \text{TR}, \\ \dfrac{\text{TR}^2 - r_{ij}}{\text{TR}^2}, & if \ 0 < r_{ij} \leq 1, \end{cases} \tag{11}$$

where TRF is the maximum transmission range of the vehicle and r_{ij} is the distance between those two vehicles at the timestamp t. There is a negative correlation between the distance and transmission range. This enables that if two vehicles are closer, then a more reliable connection is bound. The neighbor nodes are defined as those vehicles that satisfy the condition TRF $(r_{ij}) > 0$. Then, the next step is to count neighbor connection centrality that is defined as follows:

$$NCC = \sum_{r_{ij} \in NE} \text{TRF}(r_{ij}). \tag{12}$$

The last step is to get CN, the weighted average of NCC over the timestamp t [28].

$$CN = w_i \times NCC, \tag{13}$$

where w_i is the weighted associated at each timestamp.

3.2.4. Trust (T).

The type of vehicle also plays an essential role in cluster stability; this is an efficient technique for dealing with malicious and compromised nodes [29]. To address this problem, the authors have included the notion of spectrum sensing. Here, concepts are addressed that could remarkably enhance the cluster head's stability in the network. The first is the spectrum sensing technique that helps to utilize the spectrum efficiently. Here, the spectrum sensing is taken as a classification problem using the long short-term memory (LSTM). The network is trained using various types of signal and noise data. Data can be handled by using the latest technology called big data handling [30, 31], though this decision is made on the confidence of the noise class. Since the method includes a recurrent neural network, it automatically learns the energy features and adapts to any untrained noise or signal in a real and dynamic environment. This enables the detection of the primary users' (PU), like medical vans, police vans, or any other type of civil service, to use the spectrum as a priority if there is an emergency. The rest is considered the secondary users (SU).

In the following section, the formulation to address the above-stated issues is discussed.

LSTM Model Design. The internal structure of LSTM is shown in Figure 4. Here, $Y^{\langle t \rangle}$ is the input to the cell structure, and output is denoted by $a^{\langle t \rangle}$. The previous cell input is taken as $a^{\langle t-1 \rangle}$, current and previous cell states are represented by $c^{\langle t \rangle}$ and $c^{\langle t-1 \rangle}$. σ_f, σ_u, and σ_0 show the value of three gates forget gate, update gate, and output gate, respectively. The Hadamard function is denoted by $\odot$, and tanh shows the activation function with elementwise addition $\oplus$.

Any LSTM cell constitutes three main gates: update gate, forget gate, and output gate. The function of each gate is stated as follows:

(i) Update gate (σ_u): decide when to update the current cell state

(ii) Forget gate (σ_f): discard function of the current cell

(iii) Output gate (σ_0): provided to the output through the output gate

The updated vector for each state is obtained through

$$\tilde{c}^{\langle t \rangle} = \tanh\left(w_c\left[a^{\langle t-1 \rangle}, Y^{\langle t \rangle}\right] + b_c, \tag{14}\right)$$

where b_c is the bias term and tanh is the activation function. All the three gates get updated using the individual bias and the sigmoid function.

$$\begin{aligned} \Gamma_\mu &= \sigma\left(w_\mu\left[a^{\langle t-1 \rangle}, Y^{\langle t \rangle}\right] + b_\mu\right), \\ \Gamma_f &= \sigma\left(w_f\left[a^{\langle t-1 \rangle}, Y^{\langle t \rangle}\right] + b_f\right), \\ \Gamma_0 &= \sigma\left(w_0\left[a^{\langle t-1 \rangle}, Y^{\langle t \rangle}\right] + b_0\right), \\ \sigma_m &= \frac{1}{1 + e^{-m}}, \end{aligned} \tag{15}$$

where w_μ, w_f, and w_0 are the weight matrices. The bias terms are denoted by b_μ, b_f, and b_o. An elementwise product is taken among the previous cell state $c^{\langle t-1 \rangle}$ and forget gate Γ_f and among the update gate and candidate vector updating $\tilde{c}^{\langle t \rangle}$. The elementwise product among the output gate Γ_0 and hyperbolic tangent vector $c^{\langle t \rangle}$ is as follows:

$$\begin{aligned} c^{\langle t \rangle} &= \Gamma_\mu \odot \tilde{c}^{\langle t \rangle} + \Gamma_f \odot c^{\langle t-1 \rangle}, \\ a^{\langle t \rangle} &= \Gamma_0 \odot \tanh\left(c^{\langle t \rangle}\right). \end{aligned} \tag{16}$$

The architecture designed in this study has two-bit layer with every 100 nodes and a fully connected layer followed by a SoftMax layer as the decision is no binary bases. The network is trained for 500 epochs, having a learning rate of 0.01, and the batch size of the data chunk at a time is 500.

Data Modeling. The data is captured through the ideal probability of the primary user. We consider that each vehicle takes part in the Cognitive Radio (CR) networking to transfer the information. The bandwidth allotted to the
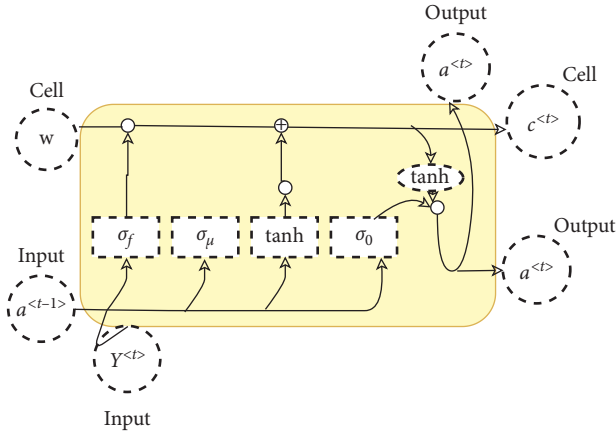
FIGURE 4: Internal structure LSTM network.

network is limited, so limited spectrum uses CR. The free spectrum of the PU can be utilized by secondary users, reducing the overhead problem. In PU's presence, the spectrum will be released for that vehicle if it receives a signal with an energy higher than the threshold signal energy. The factor trust (T) is increased if it a primary user and belongs to that neighborhood; else, the trust value is decremented. The behavior of the PU is mapped through the quantized energy vector called the sensing report. The actual status of the PU is estimated through the acknowledgment signal along with the fusion of reliable local decision of CR. The clean PU signal is acquired, and its power is measured as σ_x^2. The Gaussian noise is added to the computed power signal of PU to achieve signal-to-noise ratio (SNR) γ and evaluate by using the relation $\sigma_w^2 = \sigma_x^2/\gamma$.

Adaptive sensing at the individual vehicle level is as follows:

$$
\begin{aligned}
\mathcal{H}_0 &: Y^{\langle t \rangle} = w^{\langle t \rangle}, \\
\mathcal{H}_1 &: Y^{\langle t \rangle} = h^{\langle t \rangle} X^{\langle t \rangle} + w^{\langle t \rangle},
\end{aligned}
\tag{17}
$$

where $Y^{\langle t \rangle}$ is the signal envelop received at the t-th time instant by the sensing vehicle, the PSU signal is distorted using $w^{\langle t \rangle}$ an Additive White Gaussian noise with zero mean and variance σ^2, $X^{\langle t \rangle}$ is the SNR signal transmitted from the PU transmitter, and $h^{\langle t \rangle}$ is the channel gain [32]. This hypothesis can be viewed as a classification problem where there is a signal from the primary user or the presence of noise. The signal received at any instant is represented in general form with the previous sensing event of sample size M being fed to the current sensing event:

$$
Y = \left[Y^{\langle 1 \rangle} Y^{\langle 2 \rangle}, \ldots, Y^{\langle M \rangle} Y^{\langle M+1 \rangle} Y^{\langle M+2 \rangle}, \ldots, Y^{\langle 2M \rangle} \right]^T,
\tag{18}
$$

where M is the sample size taking into consideration these number of vehicles at timestamp t and the transpose of a vector is denoted by $[.]^T$. Here, the probability of detection and the probability of false alarm are used to detect the performance of the spectrum sensing algorithm, which are defined as

$$
\begin{aligned}
P_d &= P_r\{H_1 \,|H_1\}, \\
P_f &= P_r\{H_1|H_0\}.
\end{aligned}
\tag{19}
$$

The presence of the PU signal is denoted by H_1 while absence by H_0 in that case, the only noise is detected. At a given M sensing sample, the energy test statistic is represented by the energy that calculates test statistics $T(Y)$ in the time domain to compare with, which follows the Neyman–Pearson criteria.

$$
T(Y) = \frac{1}{2N} \sum_{i=1}^{N} \left| Y_i^{re} + Y_i^{im} \right|^2.
\tag{20}
$$

The test statistic $T(Y)$ is the random variable with a chi-square probability distribution function (χ^2) with k degrees of freedom. It can also be represented as $Q = \sum_{i=1}^{k} |z_i|^2$. Here, $k = 2\,N$ for the complex-valued case and $k = N$ for absolute values. The threshold ε can be defined using the central limit theorem. The detection probability can be defined as [33]

$$
P_d(\varepsilon, t) = \mathbb{Q}\left(\left(\frac{\varepsilon}{\sigma_u^2} - \gamma - 1 \right) \sqrt{\frac{t f_s}{2\gamma + 1}} \right),
\tag{21}
$$

where $\mathbb{Q}(.)$ is the complementary distribution function and is Gaussian in nature; that is,

$$
\mathbb{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\frac{t^2}{2} \right) dt.
\tag{22}
$$

If we take the inverse of equation (21), then the threshold for probability detection can be calculated as

$$
\mathbb{Q}^{-1}(P_d) = \left(\frac{\varepsilon}{\sigma_u^2} - \gamma - 1 \right) \sqrt{\frac{t f_s}{2\gamma + 1}}.
\tag{23}
$$

The false detection P_f is calculated as

$$
\mathbb{Q}^{-1}(P_f) = \left(\frac{\varepsilon}{\sigma_u^2} - 1 \right) \sqrt{t f_s}.
\tag{24}
$$

In the Rayleigh transmission channel, the converted carrier band signal is passed:

$$
h(t) = \sum_{p=1}^{N_p} A_p \delta(t - t_p),
\tag{25}
$$

where N_p are several paths, A_p is channel gain, and t_p is the delay in p th path. The signal moves in the line of sight (LOS) considered by a free-space path loss propagation model and receives power calculated as follows:

$$
P_r = \frac{G_l \times \varrho^2}{(4\pi d)^2} \times P_t.
\tag{26}
$$

The received and transmitted power are notated as P_r and P_t with wavelength ϱ. The LOS distance is denoted by d, and the factor dependent defines field radiation of PU and SU upon the antenna G_l. The square of the distance between PU and SU is inversely proportional to the received power. The noise variance is added to the received power and

updates the actual power value. The noisy channel is modeled as the convolution of Rayleigh channel response and noise variance where $n(t)$ is noise variance.

$$g(t) = h(t) * n(t). \tag{27}$$

This completes the design of the spectrum sensing structure.

In the training of LSTM, the behavior of CR users to the changing activity of PU in the operating environment is learned. The sensing report is generated by the CR user and makes a local decision based on its energy. Based on the outcomes, the acknowledgment signal status, the report is assigned to a sensing class. The information of the primary user is forwarded to the data center that decides spectrum sensing. The bandwidth of the transmitted signal is divided into N_s subcarriers and transmitted in chunks. These subcarriers are frequency spaced by $\Delta f = 1/T_d$ where T_d is time to transmit a signal. The signals are multiplexed using inverse discrete Fourier transform (IDFT) as

$$y_b(t) = \sum_{k=1}^{N_s} Y_k e^{j2\pi k \Delta f t}. \tag{28}$$

Every secondary node senses the energy of the transmitted signal by PU, and based on the comparison with a threshold $\mathbb{Q}^{-1}$ from equation (23), it decides whether it is from the primary user and to vacate the channel. This energy received enables the calculation of the trust value. Figure 4 portrays the complete model for the training using the concept of spectrum sensing and then obtaining the trust value. P_d and P_f are evaluated in the proposed scheme LSTM-based spectrum sensing detection. The signals collected are fed to the LSTM network one by one, and the corresponding values of the detection and false alarm probability are computed. The primary signal vector of each SNR value is processed to the LSTM network. The number of times it correctly classified the signal H_1 divided by the total number of primary user signals fed to the network determines P_d. Similarly, the noise sequence is forwarded to the LSTM network and calculates the false alarm probability P_f. The hypothesis H_0 is divided by the total number of noise sequences used in the prediction. The complete model designed for predicting the vehicle's trust value using LSTM is shown in Figure 5.

As discussed above, adaptive spectrum sensing is implemented for forming trust based on the energy detector's threshold. The threshold is set to the target that is the desired constant probability of detection.

The limits are as obtained from equation (23). The energy of the vehicle received is given to the trained LSTM network. The network decides whether the vehicle is a primary user or not, and in turn, the value of the trust is assigned if it successfully vacates the spectrum. This adaptive spectrum sensing is done using the trained data using the LSTM. In Figure 6, the training curve of LSTM is elucidated, where the loss function, along with the accuracy, is plotted for 3,500 epochs. The loss defines the difference between the actual and predicted values of the primary users' signals in various data and noise signals.

The analysis of trained LSTM is depicted in terms of the confusion matrix, as shown in Figure 7. Here, the adaptive spectrum sensing problem is formulated as a classification problem illustrated in equation (17). This prediction is made for a total of 4,000 vehicles. This binary classifier's accuracy is 89% for the correct detection of the absence of any PU and 83.5% for any PU presence in the network. This designed LSTM model is efficient and can predict with sufficient accuracy, even in various noise conditions and mixed signals. A comparison between P_d generated from the theoretical analysis as designed in [34] and P_d for the trained LSTM is shown in Figure 8. Since $P_{d,\max} = 1$, the higher the probability, the better the trust score, and accordingly, the vehicle is considered trustworthy. This completes the design and evaluation of the trust metric. Now, the proposed weighted CH selection is discussed in the following in detail.

The complete algorithm for the CH selection is explained in Algorithm 2. All the above parameters discussed are incorporated, and a weighted metric is formulated for selecting a stable cluster head for a maximum period [35]. The number of vehicles in a time instant is firstly clustered using the RSU as discussed in Algorithm 1. Then, for all the members (CM), the above four parameters are calculated, and then a weighted CH_{score} is calculated to select the cluster head as discussed in Algorithm 2.

3.3. Computational Complexity of the Designed Complete Algorithm.

In this section, the computational complexity of the dynamic weighted algorithm is discussed. The clustering algorithm is divided into two parts, the cluster formation and cluster head selection. Thus, the total time complexity of the algorithm can be stated as

$$O_{\text{tot}} = O_{\text{CF}} + O_{\text{CHS}}, \tag{29}$$

where O_{CF} is the time complexity of cluster formation and O_{CHS} is the cluster head selection.

In cluster formation, as already discussed in Section 3.1, the authors find only the distance between the vehicles and the RSU. M is the maximum number of vehicles taken in the worst-case analysis. Thus, the worst-case time complexity for this is

$$O_{\text{CF}} = O\left((\log M)^2\right). \tag{30}$$

In the cluster head selection, there are four parameters taken into consideration. The total time complexity for the cluster head selection is

$$O_{\text{CHS}} = O_{\text{BF}} + O_{\text{Ecc}} + O_{\text{CN}} + O_T. \tag{31}$$

The complexity of BF is a linear equation, where three things have been calculated, the time to leave, relative average speed, and the neighborhood degree. The time to leave and the relative average speed are only the constant values fetched and calculated. In contrast, in the neighborhood degree, the distance between the vehicles in the vicinity is calculated through their coordinates. Thus,
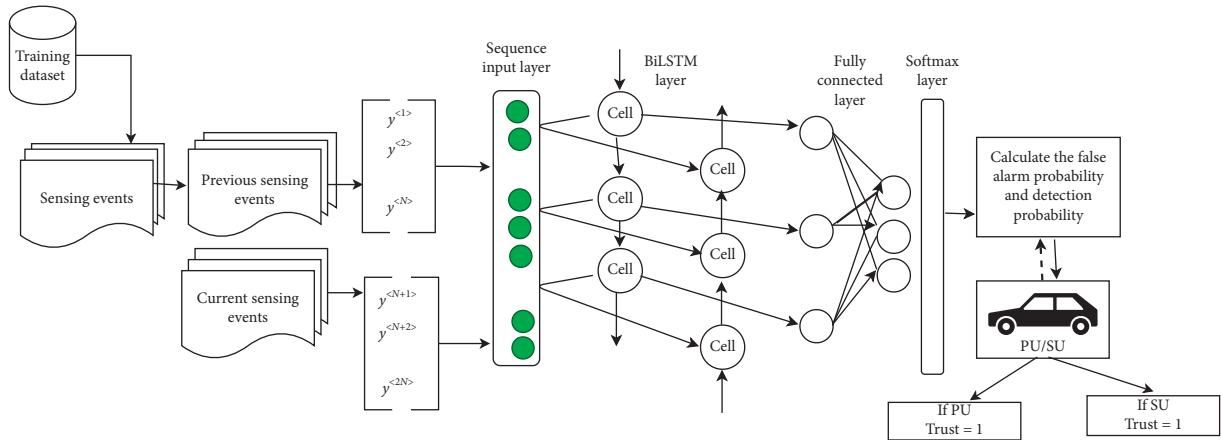
FIGURE 5: Adaptive spectrum sensing using LSTM.
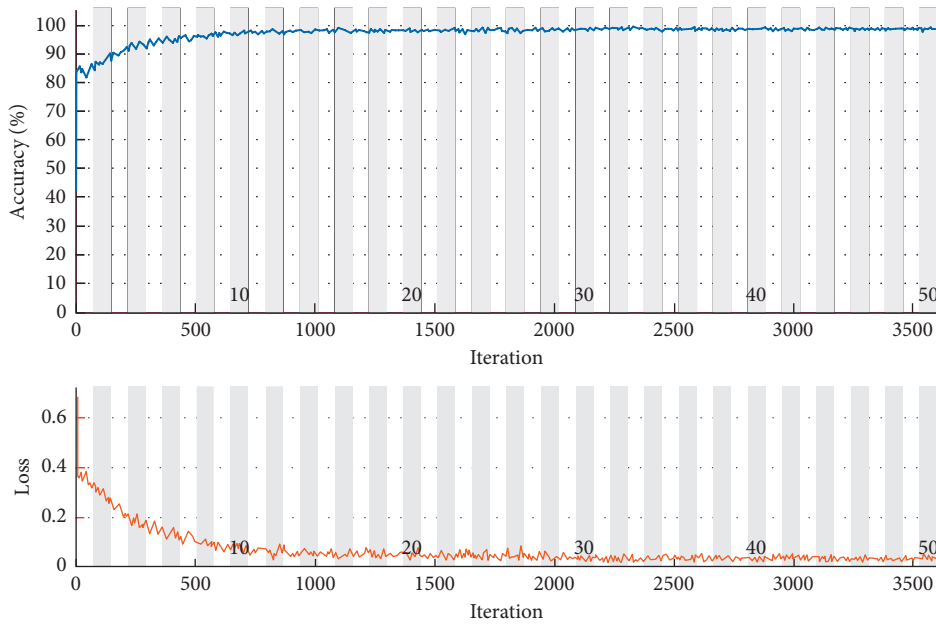
FIGURE 6: Training progress of designed LSTM for 3,500 iterations.
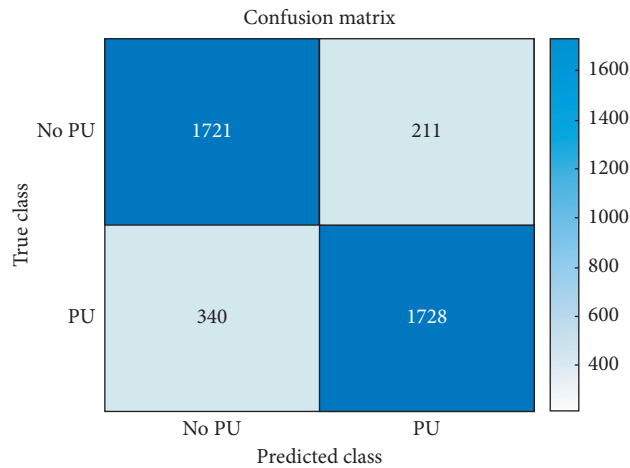
FIGURE 7: Confusion matrix for performance analysis of the classifier designed using LSTM.
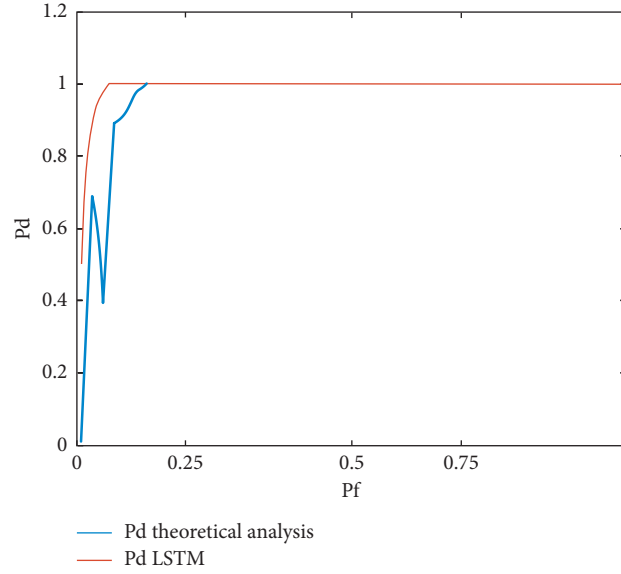
FIGURE 8: Comparison between the probability of detection and probability of false alarm using the proposed spectrum sensing LSTM method.

Input: Relative Speed V_{vehi} and Location $[x, y]$; Number of the lane in a map (no_lane); time_span; Cluster (C), N no of clusters.
Output: CH
For $t = 1$:time_span
For $i = 1$: N
    For $j = 1$: CM
    Obtain the BF using the equation (3)
    Calculate the r_{ij}; affinity matrix
    Fabricate the evolving graph G for each cluster
    Find the maximum eigenvalues λ ; calculate Ecc_j using equation (10)
    Determine the neighbor and find the CN using equation (12)
    From the LSTM trained network, calculate the T
    Obtain the CH_score for each CM
        $CH_{\text{score}} = w_1 \times \text{BF} + w_2 \times \text{Ecc} + w_3 \times \text{CN} + w_4 \times T$
    End $(w_1 + w_2 + w_3 + w_4 = 1)$
        CH = **max**$(\text{CH}_{\text{score}}) \forall j$
    End
End

ALGORITHM 2: CH selection.

$$O_{\text{BF}} = O(1) + O(1) + O\left((\log M)^2\right). \tag{32}$$

The next is the eccentricity, calculated using the spectral clustering methods that involve the affinity matrix and the eigenvalue decomposition. The complete complexity of spectral clustering is

$$O_{\text{Ecc}} = O(M)^2 + O(M)^3. \tag{33}$$

In the community neighborhood, an affinity matrix is generated for the near adjacent vehicles. Thus, the complexity for this is given as

$$O_{\text{CN}} = O\left((\log M)^2\right). \tag{34}$$

In the metric trust, the main role is played by the LSTM for spectrum sensing; the theoretical time complexity of LSTM is given as

$$O_{TH} = 4IH + 4H^2 + 3H + HK, \tag{35}$$

where I is the number of inputs, K is the number of outputs, and H is the number of hidden layers. In this study, as the model is trained once and then for a given vehicles signal, the LSTM, through its spectrum sensing, senses the vehicle either being a primary or secondary user. Thus, the time complexity bottles down to

$$O_T = O\left(4H^2\right). \tag{36}$$

Thus, the complete time complexity is reduced to removing all the terms with less complexity than cubic and quadratic terms.

$$O_{\text{CH}} = 2O\left((\log M)^2\right) + O(M)^3. \tag{37}$$

Thus, the total time complexity is

$$O_{\text{tot}} = 3\,O\!\left((\log M)^2\right) + O(M)^3. \tag{38}$$

4. Simulation Results and Discussion

This section discusses the results achieved at different stages of the cluster head model designing. It is bifurcated into the following sections: (1) Simulation Environment and Tool; (2) Network Performance Evaluation; (3) Experimental Evaluation with the existing state of the art and similar other cases.

4.1. Simulation Environment and Tool. All the simulation experiments are performed using MATLAB (R2020a), with processor Intel® Core TM i3, 1.98 GHz, Simulation of Urban Mobility [36] (SUMO 0.25.0; SUMO 0.12.0), and the TraCI [37]. SUMO is an open-source microscopic road traffic simulator licensed under the General Public License (GPL). It was developed through a collaboration between the Center for Applied Informatics Cologne (ZAIK) and the Institute of Transportation Systems (ITS) at the German Aerospace Center (DLR). While TraCI stands for the Traffic Control Interface specifically designed to get access to the traffic running on the road simulated, the embedded feature of extracting simulated objects' values. The network performance metrics like throughput, energy, packet delay, and packet delay ratio are evaluated through MATLAB. The simulation area is Chengdu, the capital of southwestern China's Sichuan province. The area taken for the simulation has the latitude $= 30.6598628°$N and longitude $= 104.0633717°$E. The area is busiest as there is a tourist place Chairman Mao statue. The total traffic environment summary is provided in Table 3.

The simulated section of the original region is shown in Figure 9; it is a vast area with a high urban and highway mobility model. The region also consists of the famous tourist spot that ensures dense vehicle movement around the peak working hours. The authors have deployed RSU, the additional facilities for network stability as per the proposed

algorithm discussed in Algorithm 3, given in Section 2.2. The map after the deployment is shown in Figure 10. It is evident from the figure that the total area has been covered efficiently. Also, enough RSU are concentrated at the cross-lane to provide sufficient coverage without any overhead delay or congestion problem, which can eventually lead to less drop in the packets and affect the network's stability.

4.2. Network Performance Evaluation. The proposed scheme's importance is tested using the four network performance metrics discussed in the following. The communication between vehicles is modeled through the Nakagmi channel, representing the obstacle medium to match real-life data transmission. Here a two-hop model is taken, where the packet transfer is initiated through any randomly selected vehicle that acts as a source to the CH and the CH sent the packet to the intended designation. The delay of this packet through the network enables the authors to calculate the desired network parameters. The size and data rate are as specified in the following table. The time taken for the packet transfer is measured using the MATLAB internal clock. Packet contains a random sequence of 1's and 0's. The communication network parameters are listed in Table 4.

The evaluation parameters for the proposed VANET stability performance assessment are as follows:

(1) Energy (E): the amount of energy consumed at each node for communication is measured in Joules. The consumption of energy is directly proportional to the distance between the hops.

$$E = \alpha_1 + \alpha_2 \times D + \alpha_3. \tag{39}$$

(2) Packet delivery ratio (PDR): PDR is the average ratio of successfully received packets at the destination vehicle over the total generated packets on the source vehicle. The delivery ratio decreases with increasing data rates.

$$\text{PDR} = \frac{\sum_{i=1}^{N} \text{packet received} \times (\text{datarate} \times \text{packet size})}{\text{packet generated at source} \times (\text{datarate} \times \text{packet size})}. \tag{40}$$

(3) Packet delay (PD): it refers to the time taken for a packet to be sent through the transmission media from the source to the destination vehicle. The delay

in packet delivery depends on network congestion, noise, and hop travel distance.

$$\text{PD} = \frac{\sum_{i=1}^{N} \text{packet transmitted} \times (\text{datarate} \times \text{packet size}) - \text{packet received} \times (\text{datarate} \times \text{packet size})}{t}. \tag{41}$$

(4) Throughput: it represents the amount of data successfully transferred from source vehicle to destination vehicle in a given period, typically

measured in Kilobits per second (kbps). The higher throughput can be achieved with less hop count and network stability.

TABLE 3: Simulation specification.

Parameter	Value
Scenario	Urban
Lanes	302
Average speed	12.71 m/sec
Simulation time	1500 sec
RSU placement	500 m
RSU transmission range	350
Vehicle transmission range	200
Vehicle length	5 m and 10 m
Vehicle speed	0–30 m/sec
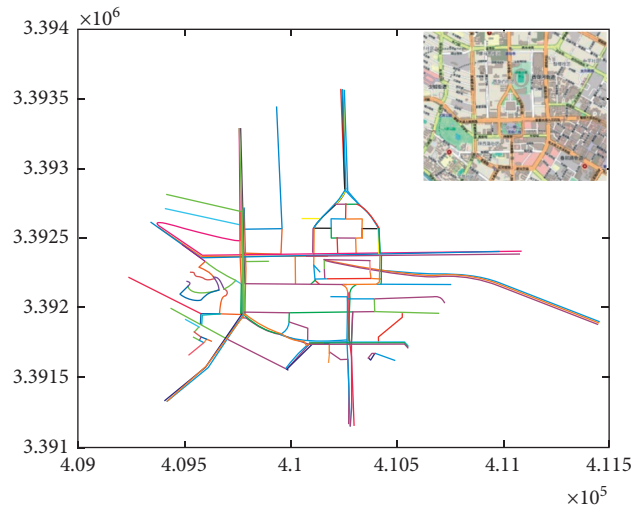
FIGURE 9: The simulated map area along with the original Chengdu, capital of southwestern China's Sichuan province.

```
Input: Lane coordinates [x, y]; angle (∅); number of the lanes in a map (no_lane)
Output: Change in angle as per Crossroad
Place the RSU in each lane.
For i = 1: no_lane
   If ([x_old, y_old] < [x_next, y_next])
      Lane is in [+, +]
   Elseif ([x_old, y_old] > [x_next, y_next])
      Lane is in [+, -]
   Elseif ([x_old, y_old] > [x_next, y_next]‖∅)
      Lane is in [-, +]
      Change the ∅
      If ∅ < 90°
         ∅ = ∅ + 180°
      Elseif ∅ > 90°‖∅ < 270°
         ∅ = ∅ + 90°
      Elseif ∅ > 270°‖∅ < 360°
         ∅ = ∅ + 180°
      Endif
   Else
         Lane is [-, -]
   Endif
End
```

ALGORITHM 3: RSU deployment in the presence of crossroad.
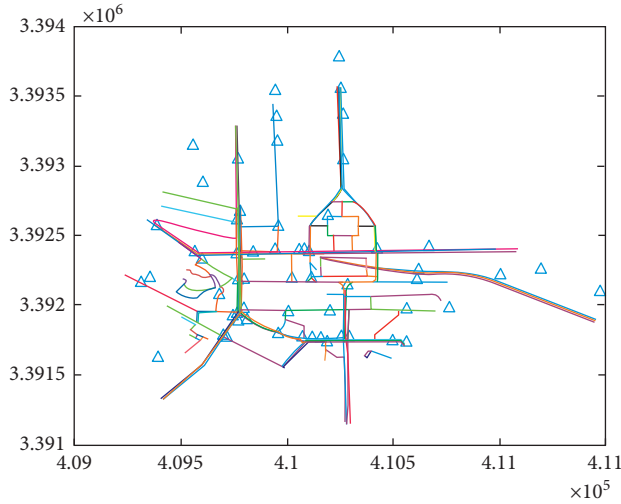
Figure 10: The deployment of RSU.

Table 4: Integrated network parameters.

Parameters	Value
Channel	Wireless
Propagation model	Propagation/Nagakami ($m = 3$)
MAC	Mac/802_11Ext
Data rate	4, 8, 10, 12, and 14 kbps
IEEE 802.11p transmission range	350 m
Packet sending interval	0.2 sec
Packet size	64 bytes

$$\text{Throughput} = \frac{\sum_{i=1}^{N} \text{packet received} \times (\text{datarate} \times \text{packet size})}{t}. \tag{42}$$

(5) Cluster head stability: it represents the number of times the same vehicle is chosen as cluster head in the total solution's time span.

$$\text{CHS} = \text{mode}\left(\sum_{i=1}^{t} \text{ID} \right). \tag{43}$$

4.3. Experimental Evaluation. The proposed scheme is tested with the following simulation environment settings:

(i) Different vehicles' densities

(ii) The vehicle position in a cluster is estimated by Kalman filter and calculated based on its current location

The complete evaluation for different vehicular densities is tabulated in Table 5. This study is done to evaluate the cluster stability in the different vehicular densities and the dynamic scenario. Vehicles are clustered on the ground of several RSUs. Each RSU covered area is considered as a cluster, and vehicles under that area are cluster members. The RSUs with empty clusters are omitted from the evaluation of the parameters. The dissipation of the energy

has a similar pattern for vehicular density. This ensures the stability of the cluster head. Also, the clusters with more number of cluster members or denser clusters pose a challenge as cluster head stability frequently changes in these conditions. The throughput is analyzed, two two-hop models. The throughput will be more for denser clusters as compared to the sparsely populated cluster. The other two parameters, packet delay and packet delay ratio, solemnly depend on the density and the distance of the nodes selected.

The authors have compared the evaluation of the befit factor in two ways. First, the authors have used the vehicles' current location to calculate the befit factor as suggested in the [19] and in the other way predicted and corrected location of the vehicles using Kalman filter as discussed in Section 2.3 employed. Figure 11 showcases the current location and the vehicles' corrected location for a single timestamp in the total simulation. We can analyze that, with dynamic evolving networks and frequent changing vehicle velocity, vehicles' exact location is trivial to know for the stability of the VANET. Also, the estimation of the vehicle is dependent on the direction of movement and current velocity. It has been observed that the vehicle's moving angle may vary instead of being constant at some points. These factors have significantly affected the befit factor analysis for selecting the cluster head and its stability. This assessment can be done from Figure 12, where 1,000 vehicle densities and 11 clusters are formed. The cluster head's stability is counted as the count of continuous timestamps for which any vehicle ID is constantly served as cluster head. Here, more stability is provided by the predicted method as we can observe that the frequency of a single vehicle in becoming the cluster head is about 180 times for a single period evaluation.

4.4. Baseline Comparisons. The baseline algorithms are the algorithms designed in the literature for the individual metric. In this analysis, three baseline algorithms are included as mentioned and discussed in Section 3. The analysis is carried out for the 100 vehicular density. The results are shown in Figure 13 at an instant of the total simulation time; in this, we can observe that CH selection is highest for the proposed scheme compared to the individual one as designed by the different arts state. All the simulation has been conducted on the same platform and done as described as in the literature for the comparison. As the model is urban, the vehicle's speed and density eventually lead to dynamic network changes. The cumulative effect of all the three metrics is evident compared to the individual as just change in velocity, community neighborhood, or eccentricity standalone cannot reflect the changes that a weighted approach can do. Thus, there is a need for a weighted approach that can evaluate the evolving changes in the network from time to time and make the transition of CH less.

The different network metrics are also tested on the 1,000 vehicular densities for the proposed scheme. The data packet is transmitted from a cluster member to the cluster head, and

TABLE 5: Comparison of various network parameters for different vehicular densities.

Network parameters	Clusters										
	1	2	3	4	5	6	7	8	9	10	11
Vehicular density: 200	*12*	*17*	*1*	*5*	*19*	*40*	*65*	*10*	*5*	*8*	*7*
Energy (J)	3.59	3.59	—	3.59	2.62	3.59	3.59	3.59	4.46	3.59	3.59
PDR	2	3	4	3	4	3	4	5	4	5	2
PD (s)	5.82	2.19	4.36	2.85	2.32	1.14	2.17	4.91	4.60	1.91	4.90
Throughput (kbps)	1.65	2.25	0.20	1.54	2.35	2.24	1.72	1.07	1.40	2.11	1.18
Vehicular density: 400	*15*	*18*	*141*	*80*	*50*	*19*	*12*	*71*	*17*	*33*	*12*
Energy (J)	3.59	3.59	3.59	5.36	2.62	1.15	3.59	3.59	1.77	3.59	3.59
PDR	1	3	2	5	2	1	3	4	2	3	2
PD (s)	1.51	1.36	4.89	2.31	3.09	3.76	1.82	6.64	1.24	1.36	1.69
Throughput (kbps)	2.02	1.93	8.26	4.81	2.78	2.95	1.32	1.51	1.11	1.55	1.32
Vehicular density: 600	*21*	*19*	*21*	*30*	*85*	*100*	*29*	*212*	*45*	*10*	*28*
Energy (J)	3.59	3.59	3.59	8.78	2.62	1.75	3.59	3.59	1.77	3.59	3.59
PDR	4.04	5.29	2.18	1.30	2.27	3.50	2.04	1.67	2.66	6.35	2.29
PD (s)	2	4	3	2	4	3	3	2	2	2	2
Throughput (kbps)	2.11	2.51	2.53	2.73	3.25	3.99	2.32	7.47	4.38	1.42	2.42
Vehicular density: 800	*35*	*135*	*40*	*30*	*50*	*145*	*45*	*96*	*110*	*62*	*52*
Energy (J)	3.59	3.59	3.59	4.90	2.62	6.16	3.59	3.59	6.82	3.59	3.59
PDR	3	1	2	5	4	3	1	3	3	3	3
PD (s)	8.16	2.05	2.21	3.67	2.48	2.95	1.38	2.34	7.12	6.27	1.13
Throughput (kbps)	1.05	7.59	1.82	1.25	2.00	7.07	2.19	6.13	4.45	2.33	2.34
Vehicular density: 1,000	*25*	*45*	*42*	*45*	*152*	*22*	*170*	*145*	*89*	*170*	*95*
Energy (J)	14.37	1.39	4.307	204.07	0.91	140.74	77.86	1.39	1.39	79.95	0.65
PDR	1	0.31	1.41	0.70	1.26	1.04	1.33	0.66	0.9	1	0
PD (s)	0.87	1.86	2.70	1.33	0.05	1.96	4.12	3.06	3.35	1.88	2.54
Throughput (kbps)	0.13	1.25	1.35	1.35	2.05	0.57	2.27	2.57	1.93	2.45	1.93
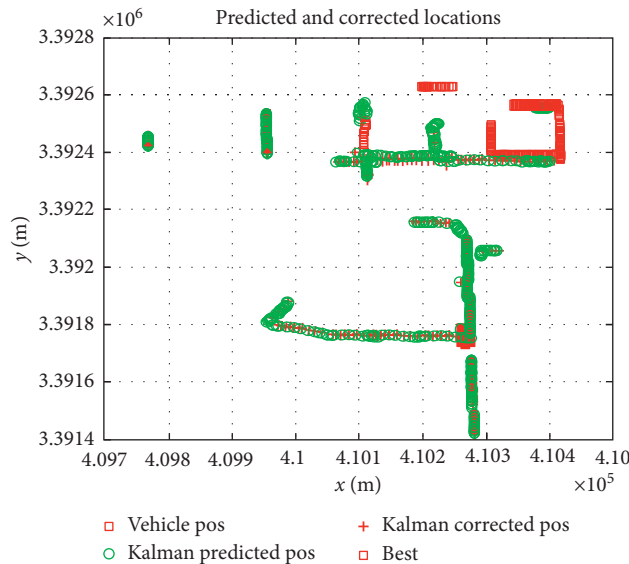
FIGURE 11: The corrected and the original location of the 1,000 vehicles for a single timestamp in the total simulation time.

the CH transmits it to the destination cluster member. The average amount of energy consumed at each node in a cluster for various data rates is shown in Figure 14. The consumption is observed to decrease or be constant for specific clusters for each data rate; this can be as the cluster topology has not changed, or the CH has not changed. The increase in energy is where the number of clusters is with fewer members and a larger distance, which requires more dissipation of the energy.

The packet delay ratio and the packet delay are demonstrated in Figures 15 and 16, respectively. The delay in a packet can be associated with the scant cluster formation; the distance among the chosen nodes is far apart, the network's congestion due to the dense
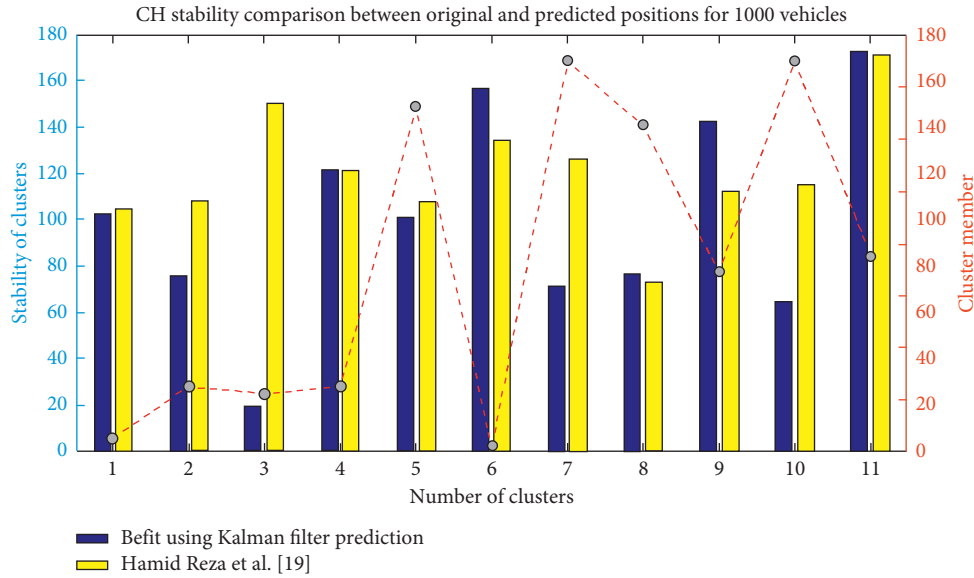
FIGURE 12: The comparison of the cluster stability for the befit factor using original and predicted 1,000 vehicles' position.
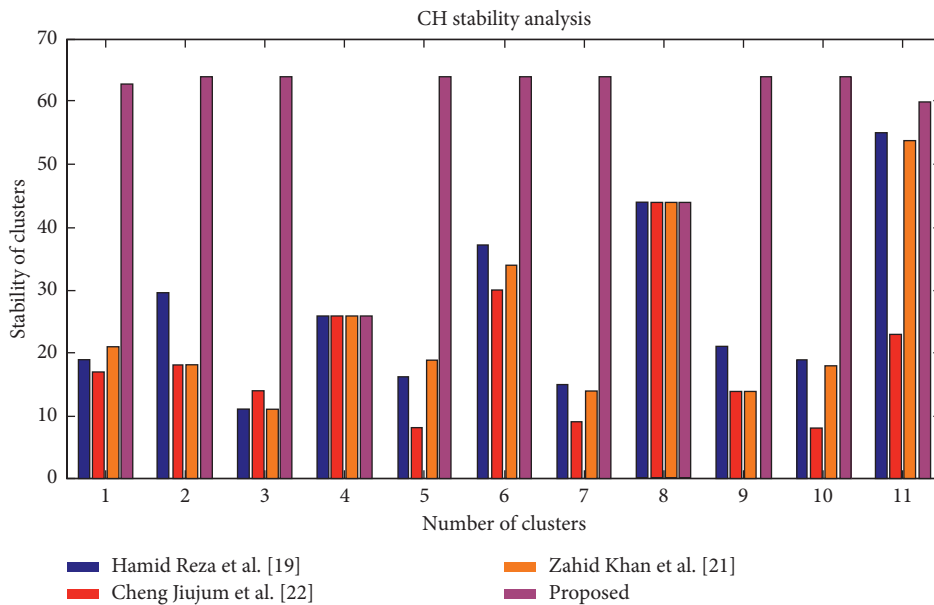
FIGURE 13: Comparative analysis of the CH stability for 1,000 vehicular densities.

population. The throughput is also analyzed, as shown in Figure 17. The throughput is the metric for the efficiency of the designed network. The high value of throughput resembles better performance with better communication among the cluster members. The drop and loss of packets eventually affect the throughput of the network. The clusters with only one cluster member ought to drop the packet.
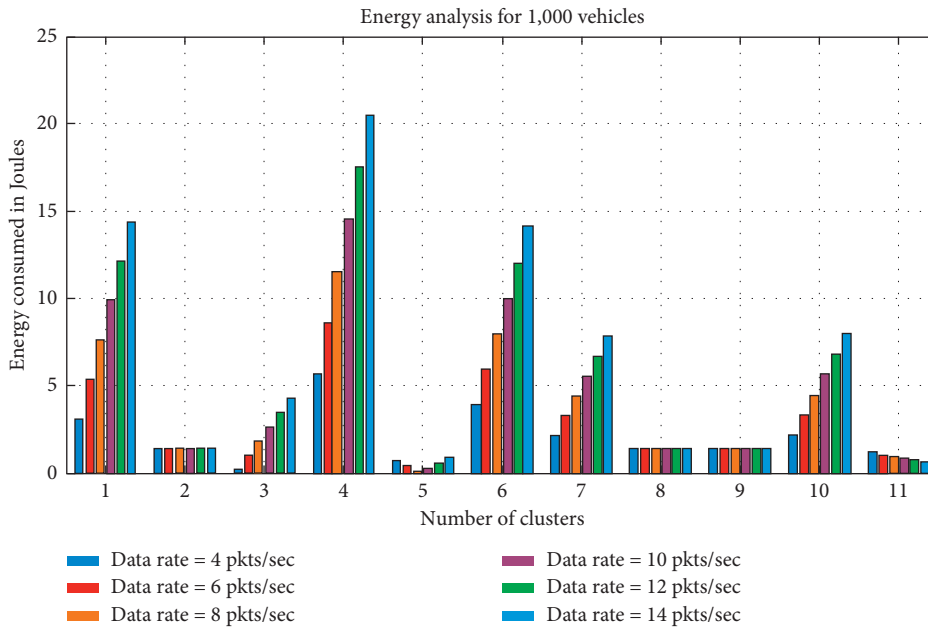
Figure 14: The energy consumption for various data rates at 1,000 vehicular densities.
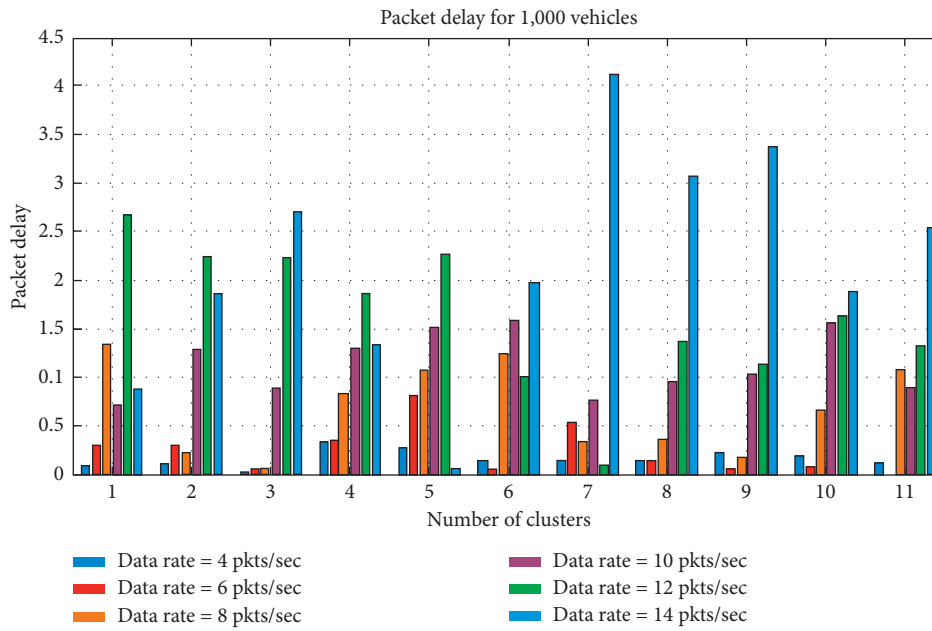
Figure 15: The packet delay for various data rates at 1,000 vehicular densities.

Packet delay ratio for 1,000 vehicles
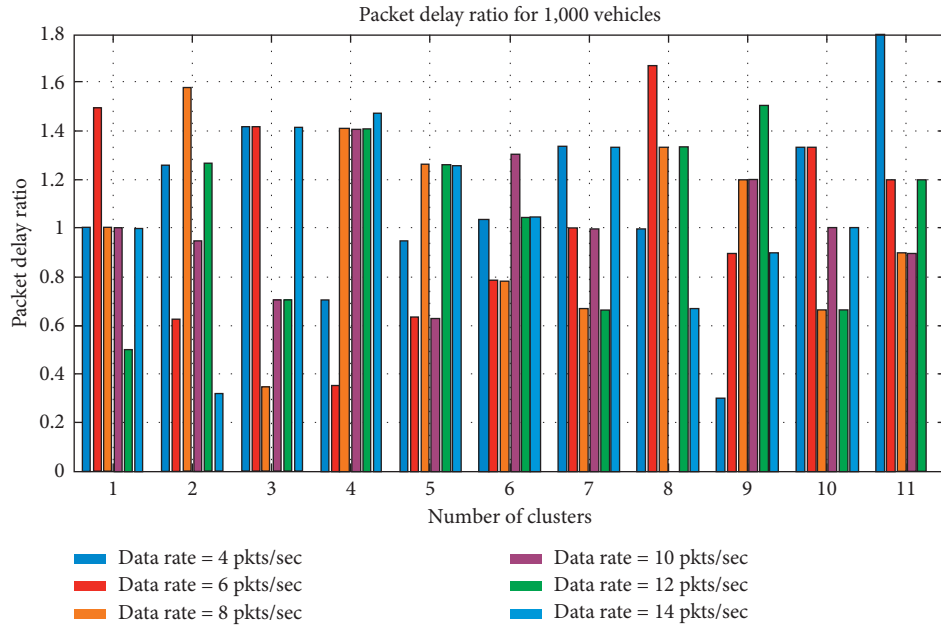
FIGURE 16: The packet delay ratio for various data rates at 1,000 vehicular densities.

Throughput analysis for 1,000 vehicles
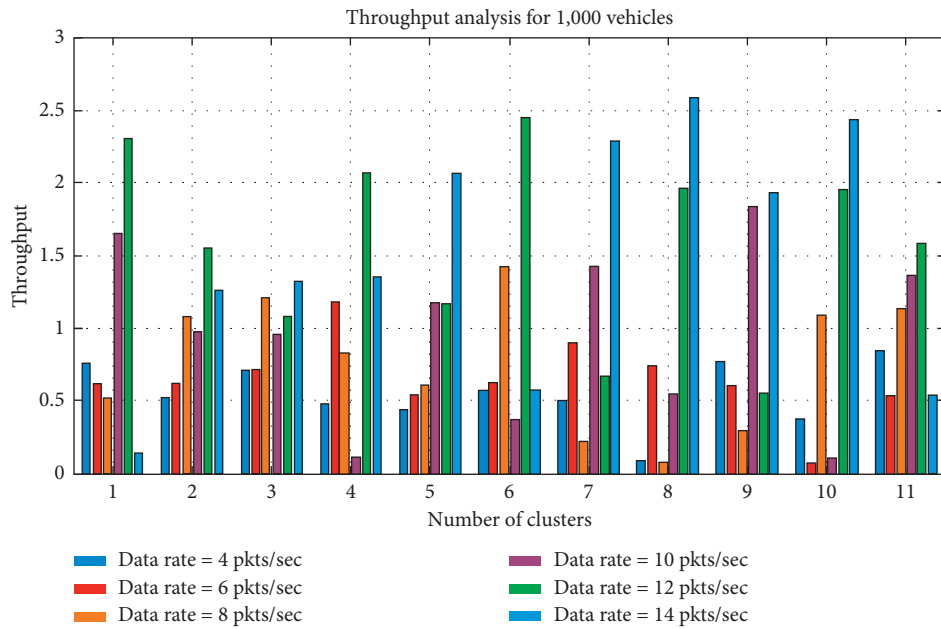
FIGURE 17: The throughput for various data rates at 1,000 vehicular densities.

5. Conclusion

In this paper, the authors have designed a new scheme for selecting a stable cluster head using the weighted approach. This formulation is crafted by including the four different metrics included to address all the parameters needed to enhance the dynamic network's stability. The authors have designed a practical methodology for the deployment of the RSU working as an additional facility to enhance the network based on the angle of the lane. The clusters are formed concerning RSU. These clusters are further improved to select a stable cluster head using metrics: befit factor,

eccentricity, community neighborhood, and trust. The befit factor is dependent on the speed of the vehicle. VANET is a network with high-speed vehicles and dynamic topology. The precision in the exact location and speed in such networks is trivial to combat such a situation. The authors have employed the Kalman filter to predict the vehicle's location at the next instant, which can improve the befit calculation and, in turn, suffice for the cluster head's stability. The results have demonstrated superior results as compared to the original formulated befit factor. The next two metrics are designed on the evolving graph structure as they are scalable and eliminate the need for recalculating in case of a change

in the network's topology. The last is the trust value included for the primary users for multiple reasons; the primary users get hidden in the network. The detection of the primary users' energy is resolved using an LSTM, deep learning trained for different signals and noises. The accuracy is around 80%, with a misclassification rate of around 14%. The cluster head's stability is measured in terms of mode, where the number of times a vehicle is selected as cluster head is noted. The results show the weighted approach's supremacy compared to the cluster head stability achieved through a single metric. The designed method is tested on the real map for the region of Chengdu, southwestern China's Sichuan province, for the different vehicle mobility densities. Also, the scheme is tested for various integrated network parametric analyses with a two-hop structure. The results regarding throughput, packet delay, energy, and packet delay ratio have shown the proposed scheme's domination for different data rates.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] S. I. Boucetta, C. Z. Johanyák, and L. Károly Pokorádi., "Survey on software defined VANETs," *Gradus*, vol. 4, no. 1, pp. 272–283, 2017.

[2] A. Paranjothi, M. S. Khan, and S. Zeadally, "A survey on congestion detection and control in connected vehicles," *Ad Hoc Networks*, vol. 108, Article ID 102277, 2020.

[3] M. Akhtar and S. Moridpour, "A review of traffic congestion prediction using artificial intelligence," *Journal of Advanced Transportation*, vol. 2021, Article ID 8878011, 2021.

[4] A. Ghansiyal, M. Mittal, and A. K. Kar, "Information management challenges in autonomous vehicles," *Journal of Cases on Information Technology*, vol. 23, no. 3, pp. 58–77, 2021.

[5] S. Vodopivec, J. Bešter, and A. Kos, "A multihoming clustering algorithm for vehicular ad hoc networks," *International Journal of Distributed Sensor Networks*, vol. 10, no. 3, Article ID 107085, 2014.

[6] S. Latif, S. Mahfooz, B. Jan, N. Ahmad, Y. Cao, and M. Asif, "A comparative study of scenario-driven multi-hop broadcast protocols for VANETs," *Vehicular Communications*, vol. 12, pp. 88–109, 2018.

[7] R. T. Goonewardene, F. H. Ali, and E. Stipidis, "Robust mobility adaptive clustering scheme with support for geographic routing for vehicular ad hoc networks," *IET Intelligent Transport Systems*, vol. 3, no. 2, pp. 148–158, 2009.

[8] P. Thakur and A. Ganpati, "A comparative study of cluster-head selection algorithms in VANET," in *Proceedings of*
International Conference on IoT Inclusive Life (ICIIL 2019), Chandigarh, India, December 2019.

[9] A. Katiyar, D. Singh, and R. S. Yadav, "State-of-the-art approach to clustering protocols in VANET: a survey," *Wireless Networks*, vol. 26, no. 7, pp. 5307–5336, 2020.

[10] Z. Zhang, A. Boukerche, and R. Pazzi, "A novel multi-hop clustering scheme for vehicular ad-hoc networks," in *Proceedings of the 9th ACM International Symposium on Mobility Management and Wireless Access*, Miami Beach, FL, USA, October 2011.

[11] L. Zhang and H. El-Sayed, "A novel cluster-based protocol for topology discovery in vehicular ad hoc network," *Procedia Computer Science*, vol. 10, pp. 525–534, 2012.

[12] S. Ucar, S. C. Ergen, and O. Ozkasap, "VMaSC: vehicular multi-hop algorithm for stable clustering in vehicular ad hoc networks," in *Proceedings of the 2013 Ieee Wireless Communications and Networking Conference (Wcnc)*, April 2013.

[13] Y. Chen, M. Fang, S. Shi, W. Guo, and X. Zheng, "Distributed multi-hop clustering algorithm for VANETs based on neighborhood follow," *EURASIP Journal on Wireless Communications and Networking*, vol. 2015, no. 1, p. 98, 2015.

[14] D. Zhang, H. Ge, T. Zhang, Y.-Y. Cui, X. Liu, and G. Mao, "New multi-hop clustering algorithm for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 4, pp. 1517–1530, 2019.

[15] X. Ji, H. Yu, G. Fan, H. Sun, and L. Chen, "Efficient and reliable cluster-based data transmission for vehicular ad hoc networks," *Mobile Information Systems*, vol. 2018, Article ID 9826782, 2018.

[16] G. H. Alsuhli, A. Khattab, and Y. A. Fahmy, "Double-head clustering for resilient VANETs," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 2917238, 2019.

[17] A. Bello Tambawal, R. Md Noor, R. Salleh, C. Chembe, and M. Oche, "Enhanced weight-based clustering algorithm to provide reliable delivery for VANET safety applications," *PLoS One*, vol. 14, no. 4, Article ID e0214664, 2019.

[18] X. Cheng and B. Huang, "A center-based secure and stable clustering algorithm for VANETs on highways," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 8415234, 2019.

[19] H. R. Arkian, R. Ebrahimi Atani, A. Pourkhalili, and S. Kamali, "Cluster-based traffic information generalization in vehicular ad-hoc networks," *Vehicular Communications*, vol. 1, no. 4, pp. 197–207, 2014.

[20] M. Mittal, R. K. Sharma, and V. P. Singh, "Modified single pass clustering with variable threshold approach," *International Journal of Innovative Computing Information and Control*, vol. 11, no. 1, pp. 375–386, 2015.

[21] M. Mittal, R. K. Sharma, and V. P. Singh, "Validation of k-means and threshold based clustering method," *International Journal of Advancements in Technology*, vol. 5.2, pp. 153–160, 2014.

[22] S. Kaur, R. K. Bansal, M. Mittal et al., "Mixed pixel decomposition based on extended fuzzy clustering for single spectral value remote sensing images," *Journal of the Indian Society of Remote Sensing*, vol. 47, no. 3, pp. 427–437, 2019.

[23] H. Yang, Z. Jia, and G. Xie, "Delay-bounded and cost-limited RSU deployment in urban vehicular ad hoc networks," *Sensors*, vol. 18, no. 9, p. 2764, 2018.

[24] Y. Mo, D. Yu, J. Song, K. Zheng, and Y. Guo, "Vehicle position updating strategy based on Kalman filter prediction in VANET environment," *Discrete Dynamics in Nature and Society*, vol. 2016, Article ID 1404396, 2016.

[25] J. Liu, X. Wang, Y. Li, X. Kang, and L. Gao, "Method of evaluating and predicting traffic state of highway network based on deep learning," *Journal of Advanced Transportation*, vol. 2021, Article ID 8878494, 2021.

[26] H. Zhou, D. Zhang, and K. Xie, "Accurate traffic matrix completion based on multi-Gaussian models," *Computer Communications*, vol. 102, pp. 165–176, 2017.

[27] Z. Khan, P. Fan, S. Fang, and F. Abbas, "An unsupervised cluster-based VANET-oriented evolving graph (CVoEG) model and associated reliable routing scheme," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3844–3859, 2019.

[28] J. Cheng, Q. Ma, R. Yu et al., "Research on the prediction-based clustering method in the community of medical vehicles for connected health," *IEEE Access*, vol. 7, pp. 71884–71896, 2019.

[29] A. Saidi, K. Benahmed, and N. Seddiki, "Secure cluster head election algorithm and misbehavior detection approach based on trust management technique for clustered wireless sensor networks," *Ad Hoc Networks*, vol. 106, Article ID 102215, 2020.

[30] M. Mittal, V. E. Balas, L. M. Goyal, and R. Kumar, *Big Data Processing Using Spark in Cloud*, Springer, Berlin, Germany, 2019, https://www.springer.com/gp/book/9789811305498.

[31] A. Srivastava, A. Prakash, and R. Tripathi, "Fuzzy-based beaconless probabilistic broadcasting for information dissemination in urban VANET," *Ad Hoc Networks*, vol. 108, Article ID 102285, 2020.

[32] X. Qian and H. Li, "On the performance of spectrum sensing in cognitive vehicular networks," in *Proceedings of the 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, August 2015.

[33] M. A. Saleem, S. Zhou, A. Sharif et al., "Expansion of cluster head stability using fuzzy in cognitive radio CR-VANET," *IEEE Access*, vol. 7, pp. 173185–173195, 2019.

[34] X. You, C.-X. Wang, and Y.-C. Liang, "Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts," *Science China Information Sciences*, vol. 64, no. 1, pp. 1–74, 2020.

[35] M. S. Kakkasageri and S. S. Manvi, "Multiagent driven dynamic clustering of vehicles in VANETs," *Journal of Network and Computer Applications*, vol. 35, no. 6, pp. 1771–1780, 2012.

[36] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO—simulation of urban mobility: an overview," in *Proceedings of the Third International Conference on Advances in System Simulation. SIMUL 2011*, pp. 1–6, Barcelona, Spain, October 2011.

[37] A. Wegener, M. Piórkowski, M. Raya, H. Hellbrück, S. Fischer, and J.-P. Hubaux, "TraCI: an interface for coupling road traffic and network simulators," in *Proceedings of the 11th Communications and Networking Simulation Symposium*, pp. 155–163, Ottawa Canada, April 2008.

Research Article

Dynamic Assessment of Road Network Vulnerability Based on Cell Transmission Model

Yu Sun ⓘ, Binglei Xie ⓘ, Shan Wang ⓘ, and Dazhuang Wu ⓘ

School of Architecture, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

Correspondence should be addressed to Binglei Xie; xiebingleihit@163.com

The road network maintaining stability is critical for guaranteeing urban traffic function. Therefore, the vulnerable links need to be identified accurately. Previous vulnerability research under static condition compared the operating states of the old equilibrium before the event and the new equilibrium after the event to assess vulnerability ignoring the dynamic variation process. Does road network vulnerability change over time? This paper combines the vulnerability assessment with the traffic flow evolution process, exploring the road network vulnerability evaluation from the perspective of time dimension. More accurate identification and evaluation of vulnerable nodes and links can help to strengthen the ability of road network resisting disturbances. A modified dynamic traffic assignment (DTA) model is established for dynamic path selection (reselect the shortest path at the end of each link) based on the dynamic user optimal (DUO) principle. A modified cell transmission model is established to simulate the traffic flow evolution processes. The cumulative and time-varying index of vulnerability assessment is established from the viewpoint of traveler's time loss. Then the road network vulnerability assessment combined the traffic flow model with the vulnerability index. The road network vulnerability assessment of Bao'an Central District of Shenzhen, China, reveals that road network vulnerability does contain a dynamic process, and vulnerable links in each phase can be exactly identified by the model. Results showed that the road network would have a large vulnerability during the disordered phase when the main road fails. Therefore, prioritizing the smooth flow of main roads can weaken the impact of road network vulnerability exposure.

1. Introduction

Road traffic network is the material basis for the existence and effectiveness of traffic activities. Once the road system suffers from natural disasters, terrorist attacks, traffic accidents, and other disturbances, it can lead to traffic congestion and road disruption, resulting in huge losses to the production and life of residents. Most interference events are uncontrollable. Therefore, it is particularly important to accurately identify the vulnerable links and repair them in a targeted manner. Improving the ability of the road network to resist interference events has attracted increasing attention from scholars in the transportation research field and continuously been applied in the traffic system emergency management field.

Most previous studies evaluated road network vulnerability under static conditions. The key links were identified by comparing the vulnerability assessment indicators of the two equilibrium operating states before and after the event. However, the analysis ignoring the process of evolving from the old to the new equilibrium cannot accurately reflect the changes in traffic service capacity during the sudden event evolution. Will it be more accurate to assess vulnerability from a dynamic perspective?

The goal of the paper can be decomposed into four aspects: First, the dynamic characteristics of road network vulnerability and the factors contributing to vulnerability assessment have been analyzed. Second, the previous static indicators are dynamically adjusted to establish dynamic evaluation indicators for road network vulnerability. Third, a comprehensive road network vulnerability dynamic evaluation model based on the improved traffic flow simulation model combining with the dynamic vulnerability evaluation index is established. Finally, taking the road network in

Bao'an Central Area, Bao'an District, Shenzhen City, as an example to evaluate its vulnerability, the dynamic law of vulnerability changes is summarized, and some suggestions are put forward for decreasing vulnerability.

The structure of the rest paper is as follows. Section 2 reviews existing research of urban road network vulnerability. Section 3 extracts a dynamic evaluation model of road network vulnerability. Section 4 applies the proposed improved model to evaluate the network vulnerability of the surrounding area of Bao'an Center in Bao'an District of Shenzhen City in China and then proposes the analysis of the case results. Section 5 concludes this paper.

2. Literature Review

The two mainstream definitions of vulnerability originate from Berdica [1] and D'Este as well as Taylor [2]. Both types of mainstream definitions take the consequences of road link failure into account, and the difference is whether to consider the probability of failure. Mattsson and Jenelius summarized the current research as a network-based topology and a transportation-based approach [3]. Studies on transport vulnerability have developed over time [4, 5]. The definition of road network vulnerability in this paper is as follows: vulnerability is the nature of a road link with full or partial failure that can significantly reduce the service capacity of the road network or the accessibility of some specific nodes in the road network.

The quantitative analysis of road network vulnerability has gradually developed to the three connotations of vulnerability: firstly, analysis from the perspective of reliability, represented by Bell in this respect, based on pure mathematical graph theory (such as complex network theory) for research [5–13]; secondly, analysis from the aspect of accessibility, represented by Chen, Taylor, and Jenelius, the research based on the cost of road network users and the importance of road sections [2, 3, 14–24]; thirdly, analysis from serviceability, the research based on the road network user cost and road network service level changes [25, 26].

The method of identifying the vulnerable connection through the whole network scanning has two much calculations in the traditional vulnerability analysis; therefore, some scholars have proposed a method for improving the computing efficiency [27, 28]. As there are many types of events that cause road network vulnerability, Jenelius and Mattsson classified the cause of the event and proposed a grid-based method to analyze the combined effects of multiple connections under regional-wide failures [19]. Xiangdong et al. proposed a new redundancy model that can find the best benefit point among travelers and planners effectively, which can also be used to improve and evaluate the vulnerability of urban road network [29].

Cats and Jenelius introduced a dynamic stochastic view into the study of public transport network vulnerability, taking the cumulative effects of interference events on system performance into account [30]. The dynamic evolution of traffic flow and the impact of traffic information on user behavior has gradually been introduced into vulnerability research [31]. Knoop et al. calculated the actual delay of the specific road network after the road link closed by dynamic traffic assignment (DTA) [32]. Cats and Jenelius analyzed the real-time information to promote the transfer of failure effects when the connection fails [30]. Kim and Yeo used a macroscopic basic diagram to observe changes in the traffic image of the road network to identify key links [33].

It can be inferred that using exposure to measure vulnerability under specific conditions has a good effect [16, 34]. Daganzo used the concept of Cell Automata (CA) to discretize the fluid dynamic traffic flow model Lighthill-Whitham-Richards (LWR) and proposed to capture the formation, propagation, and dissipation of queues effectively [35, 36]. Szeto and Wai used the Variational Inequality (VI) to construct the DTA model based on the dynamic user optimal (DUO) criterion and calculate the actual impedance of the path using the cell transmission model (CTM) simulation [37, 38]. Jiang et al. developed a probabilistic approach for assessing transport network vulnerability based on the clonal selection algorithm (CSA) [39].

Most of the existing vulnerability researches are under static conditions, ignoring the evolution process of traffic flow in the disordered phase, and the congestion phenomenon in the evolution process. There is still a lack of research on the dynamic evolution process of traffic flow, and few people have conducted vulnerability assessment from the perspective of the evolution process of traffic flow. Dynamic changes in vulnerability cannot be accurately reflected.

3. Methods

This paper divides dynamically the evolution stage of traffic flow so as to expand the dynamic evaluation results of vulnerability from the time dimension. A new indicator of vulnerability assessment from the perspective of traveler's time loss is established. The DTA model as well as the cellular transmission model is improved to simulate the dynamic evolution process of traffic flow. And the dynamic evaluation model of road network vulnerability is formed.

3.1. Path Selection Model for Vulnerability Assessment. Due to the advantage of Variational Inequality (VI) in dealing with asymmetric problems, it has been widely used in solving DTA problems. Under completely rational assumptions, users will always make the optimal travel choice; that is, the user's travel choice satisfies the DUO condition. According to the first principle of Wardrop equilibrium, the optimal condition of the basic model of DTA can be expressed as

$$f_r^{ij}(k)\left[\eta_r^{ij}(k) - \pi^{ij}(k)\right] = 0,$$

$$\sum_{r \in P_{ij}} f_r^{ij}(k) = q^{ij}(k), \eta_r^{ij}(k) - \pi^{ij}(k) \ge 0, f_r^{ij}(k) \ge 0, \pi^{ij}(k) \ge 0, \forall r \in P_{ij}, \tag{1}$$

where $\pi^{ij}(k)$ is the shortest travel time of OD pair $i \longrightarrow j$ in time period k, $\eta_r^{ij}(k)$ is the travel time of the feasible path r of OD pair $i \longrightarrow j$ in time period k, $f_r^{ij}(k)$ is the distribution flow of the feasible path r of OD pair $i \longrightarrow j$ in time period k, and $q^{ij}(k)$ is travel demand for OD pair $i \longrightarrow j$ in time period k.

However, the conventional DTA model cannot accurately reflect the changes of various traffic operating parameters in the road network under the influence of the event. Factors such as vehicle retention, queue length limitation, and maximum traffic capacity of the road section need to be considered. Therefore, in view of the impact of the event, this paper makes appropriate adjustments to the DTA model.

3.1.1. Demand Conservation Constraint considering Demand Retention.
Affected by the incident, some road links may have travel demand retention. Therefore, the traffic demand in a certain period of time should include not only the traffic flow successfully assigned to the road network but also the retention traffic demand.

$$q_{ij}(k) = \sum_{r \in p_{ij}} f_r^{ij}(k) + q_{ij}'(k), \tag{2}$$

where $q_{ij}'(k)$ is the retention traffic demand for OD pair $i \longrightarrow j$ in time period k.

3.1.2. Association Constraint between Link and Path.
The link traffic should equal the sum of all feasible path flows containing this link.

$$x_a(k) = \sum_{i \in I} \sum_{j \in J} \sum_{r \in p_{ij}} f_r^{ij}(k) \delta_{ra}^{ij}(k), \tag{3}$$

where $x_a(k)$ is the flow of road link a in time period k and $\delta_{ra}^{ij}(k)$ determines whether road link a is included in path r.

3.1.3. Road Capacity Constraint and Queue Capacity Constraint.
Considering the congestion conditions under the influence of the event, restrictions are imposed on the traffic flow and queuing capabilities of the road link. In the disordered phase, the traffic flow gradually increases but cannot exceed the capacity of the section. With the further increase of vehicles, excessive vehicle density will cause traffic congestion, and the running speed and the actual outflow of the road links decrease, so the vehicle would be queued, but the number of queued vehicles cannot exceed the maximum queue capacity of the road link.

$$x_a(k) \le C_a(k),$$
$$n_a(k) \le N_a^{\max}(k), \tag{4}$$

where $C_a(k)$ is the capacity of road link a in time period k, $n_a(k)$ is the actual number of vehicles on road link a in time period k, and $N_a^{\max}(k)$ is the maximum queuing capacity of link a in time period k.

3.1.4. Traffic Demand Time Delivery Constraint.
Traffic flow has a dynamic evolution process. In the case of road failure and traffic supply capacity decline, traffic distribution should consider both the traffic demand staying in the previous period and the new traffic demand in the current time. The dynamic traffic allocation method can be used to shift the retained traffic demand to the next time period. This improvement is especially important for the road network vulnerability analysis in the disordered phase. The dynamic assessment of vulnerability needs to meet the constraint shown as

$$q_{ij}(k+1) = q_{ij}^0(k+1) + q_{ij}'(k), \tag{5}$$

where $q_{ij}(k+1)$ is actual traffic demand in time period $k+1$, $q_{ij}^0(k+1)$ is original traffic demand during the time period of $k+1$, and $q_{ij}'(k)$ is unsatisfied traffic demand in time period k.

3.1.5. Traffic Flow Space Propagation Constraint.
It takes time for the vehicle to be assigned to the road network to complete the trip in the context of dynamic analysis. During each analysis time period, the vehicle is possible to complete the trip or may only flow into subsequent sections. Therefore, it is necessary to clarify the inflow and outflow relationship between the vehicles and each road link to determine the actual number of vehicles affected in the current period corresponding to the affected degree.

For a certain road link, the number of vehicles in period k should be determined by the number of vehicles in the previous period and the inflow and outflow of the current period.

$$n_a(k+1) = n_a(k) + u_a(k+1) - v_a(k+1), \tag{6}$$

where $u_a(k+1)$ is the inflow of vehicles on link a during the time period $k+1$ and $v_a(k+1)$ is the outflow of vehicles on link a during the time period $k+1$.

The outflow of link a of the path r shall be the same as the sum of the number of vehicles in all subsequent links of the path and the number of vehicles that have completed the trip.

$$v_a^{rij}(k) = \sum_{b \in r'} \left[n_b^{rij}(k) - n_b^{rij}(k-1) \right] + E_r^{ij}(k), \qquad (7)$$

where $E_r^{ij}(k)$ is the amount of traffic that completes $i \longrightarrow j$ travel along path r in time period k and r' is the subsequent path of link a in path r.

3.1.6. Other Constraints.
The model also needs to satisfy the basic nonnegative constraints and set affiliation and other related constraints.

$$
\begin{aligned}
f_r^{ij}(k) &\geq 0, \\
n_a(k) &\geq 0, \\
x_a(k) &\geq 0, \\
u_a(k) &\geq 0, \\
v_a(k) &\geq 0, \quad r \in P_{ij}, i \in I, j \in J, a \in A.
\end{aligned}
\qquad (8)
$$

Travel route selection needs to follow certain traffic distribution rules, as the first-principle requirement of Wardrop equilibrium: all the paths adopted must have the shortest travel time, and the travel time is equal. In this model, it can be expressed as follows: if there is a new vehicle inflow in a road link, the road segment must be included in one of the shortest paths; otherwise, the optimal path selection condition for vehicle inflow will not be added.

$$
\begin{aligned}
\left[\pi_{hj}^*(k) + \tau_a^*(k) - \pi_{gj}^*(k) \right] u_a^*(k) &= 0, \\
u_a^*(k) &\geq 0, \\
\pi_{hj}^*(k) + \tau_a^*(k) &\geq \pi_{gj}^*(k),
\end{aligned}
\qquad (9)
$$

where $\tau_a^*(k)$ is the travel time of link a, $\pi_{hj}^*(k)$ is the shortest travel time of node $h \longrightarrow j$, $\pi_{gj}^*(k)$ is the shortest travel time of node $g \longrightarrow j$, and $u_a^*(k)$ is the vehicle inflow of link a in time period k.

The model needs to be carried out for traffic allocation separately when the vehicle is loaded and at the end of each road link in order to realize the flexible adjustment of the travel path.

3.2. CTM for Vulnerability Assessment.
The road network topology has changed during the continuous impact of the road link failure, and travelers go through the process from unknowing to knowing. For the traffic flow in the road network, there were still vehicles flowing into the incident road link in the early stage of the incident, then no vehicles flowed out of the incident road link as time passed, and the upstream link of the incident site became congested or even overflowed. Travelers confirm the fact that the road network has failed, and the directly affected vehicles need to leave the incident road link and reselect the route to travel. Vehicles that have not flowed into the incident road link will directly change the travel path. Thereby, the evacuation of vehicles on the accident road link and the reevolution process of the traffic flow of the road network are formed.

As shown in Figure 1, the cellular transmission model can well simulate the evolution of the traffic flow under the
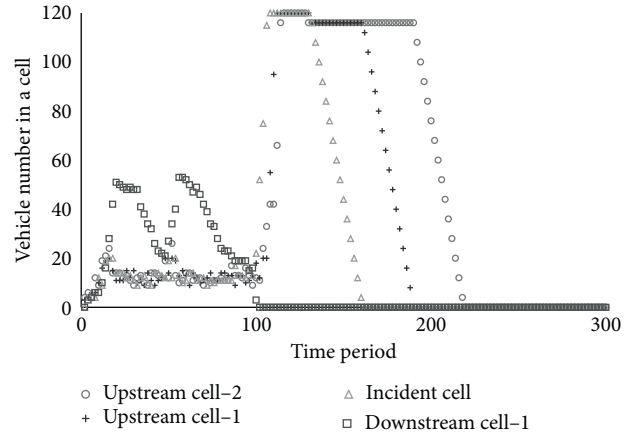
FIGURE 1: Schematic diagram of the formation and dissipation of congestion on the incident road link.

influence of the event. Congestion has gradually formed upstream of the event point since the event occurs and dissipates as the vehicle gradually changes the travel path.

3.2.1. Cell Transmission Method within the Road Link.
When the traffic flow propagates in the road section, CTM expresses the basic propagation relationship of the traffic flow as follows:

$$
\begin{aligned}
n_i(k+1) &= n_i(k) + y_i(k) - y_{i+1}(k), \\
y_i(k) &= \min\{n_{i-1}(k), Q_i(k), N_i(k) - n_i(k)\},
\end{aligned}
\qquad (10)
$$

where $n_i(k)$ is the number of vehicles in the i-th cell in the time period k, $y_i(k)$ is the outflow of vehicles in the i-th cell in the time period k, $Q_i(k)$ is the capacity of the i-th cell in the time period k, and $N_i(k)$ is the maximum number of vehicles that the i-th cell can hold in the time period k.

Since the LWR model assumes that the traffic flow is a continuous fluid, its solution is relatively complicated. In order to simplify the solution of the model, CTM assumes that the traffic flow satisfies the flow-density relationship shown in Figure 2. Formula (11) gives the calculation method of flow under this relationship.

$$q = \min\{v\rho, q_{\max}, w(\rho_j - \rho)\}, \quad 0 \leq \rho \leq \rho_j, \qquad (11)$$

where v is the free speed, w is the speed of backward shock wave propagation, and ρ_j is the blocking density.

At this time, the amount of traffic spread between cells satisfies

$$y_i(k) = \min\left\{ n_{i-1}(k), Q_i(k), \left(\frac{w}{v}\right)[N_i(k) - n_i(k)] \right\}. \qquad (12)$$

In order to further clarify the propagation relationship of the traffic flow, the maximum output capacity S and the maximum receiving capacity R of the cell are, respectively, given by equations (13) and (14):

$$S_i(k) = \min\{n_i(k), Q_i(k)\}, \qquad (13)$$

$$R_i(k) = \min\left\{ Q_i(k), \left(\frac{w}{v}\right)[N_i(k) - n_i(k)] \right\}, \qquad (14)$$
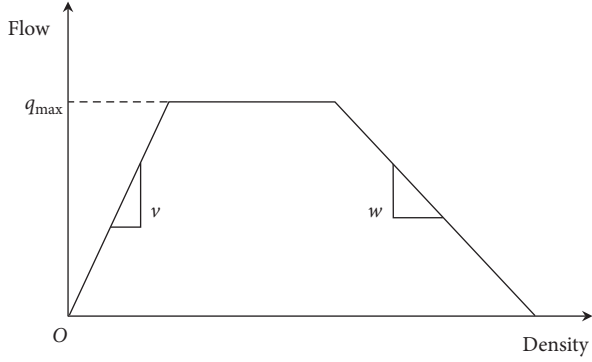
FIGURE 2: Basic flow-density relationship diagram of CTM.

where $S_i(k)$ is the maximum output capacity of cell i in time period k and $R_i(k)$ is the maximum receiving capacity of cell i in time period k.

The relationship between the flow and the amount of communication between the cells is transformed into

$$y_i(k) = \min\{S_i(k), R_{i+1}(k)\}. \tag{15}$$

3.2.2. Cell Transmission Method between Road Links. The transmission of vehicles between road links includes two forms: merging and diversion. The basic transmission principle is the same as the transmission process in the road link, but the output capacity constraints of the upstream cells and the receiving capacity constraints of the downstream cells must be satisfied at the same time; that is, the traffic flow propagation relationship should satisfy formulas (16) and (17), respectively.

$$y_{i,j}(k) \le S_{i,j}(k),$$
$$\sum_i y_{i,j}(k) \le R_j(k), \quad i = 1,2,3,\ldots, \tag{16}$$

$$\sum_j y_{i-1,j}(k) \le S_{i-1}(k), y_{i-1,j}(k) \le R_j(k), \quad j = 1,2,3,\ldots. \tag{17}$$

The cell capacity at the intersection cannot reach the same capacity as the cell in the road link. At this time, the maximum output capacity of the cell is calculated as

$$S_i^c(k) = \min\{n_i(k), Q_i(k), Q_i^c(k)\}, \tag{18}$$

where $Q_i^c(k)$ is the capacity of the i-th cell at the intersection in the time period k and $S_i^c(k)$ is the maximum output capacity of the i-th cell at the intersection in the time period k.

For the actual road network, formula (19) is always valid, and the cell output capacity can be calculated according to formula (20).

$$Q_i^c(k) \le Q_i(k), \tag{19}$$

$$S_i^c(k) = \min\{n_i(k), Q_i^c(k)\}. \tag{20}$$

The receiving capacity of the intersection cell is calculated according to

$$R_i^c(k) = \left(\frac{w}{v}\right)[N_i(k) - n_i(k)]. \tag{21}$$

Combining the above-mentioned transmission method during link and between links, the basic cell transmission relationship of the traffic flow in the road network is formed.

3.2.3. Vehicle Evacuation Methods on the Accident Road. This paper adds a vehicle evacuation method for interference events based on the basic CTM. For the situation that the road link capacity is greatly reduced, the CTM can naturally capture the increasing amount of vehicle queues in the upstream cell of the road link due to the existence of the road link transmission constraint, and the actual travel time of the road link is gradually extended. The vehicle not arriving at the event-occurred link gradually changed the path according to the DUO condition, and the queue is gradually dissipated because there is no new vehicle flowing into it. In the road link interruption situation, the vehicle upstream of the incident point cannot flow into the downstream cell and stay in the road section. In order to make the simulation results more realistic, it is necessary to add a method to make the vehicle know the link interruption and then turn around and leave the event-occurred link.

(1) The Method of User Knowing the Invalid Link and Reselecting the Path The cell transit ability of the cell is 0 at first. After the time of T, since the incident occurs at time k_a, the road Bottom Vertex (BV) detects that the vehicle outflow is zero, or the road Top Vertex (TV) meets that the number of vehicles in the cell (n) equals the number of vehicles that the cell can accommodate (N). At this time, the vehicle knows that the road link is interrupted, and the new path is selected according to the new road network connectivity.

(2) The Method of the Vehicle Turning around in the Upstream Cell. Transfer the reverse method on the upstream cell vehicle queue of the event-occurred link, and the vehicle queue is sorted in reverse order. The updated cell capacity is the capacity after turning around, and reverse transmission is performed according to formula (22). After reaching the top node of the incident link, the shortest path is selected again based on taking this node as a new starting point, and then the destination transmission is continued according to the normal transmission mode. The upstream cell vehicle U-turn driving method can be specified as

$$n_{i-1}(k+1) = n_{i-1}(k) + y_i(k) - y_{i-1}(k), \tag{22}$$

where $n_i(k)$ is the number of vehicles in i-th cell during k period and $y_i(k)$ is vehicle outflow in i-th cell during k period.

3.2.4. Path Travel Time Calculation Method. The path and link impedance can be calculated based on the actual travel time of the vehicle in this paper. Because the model uses a road link-based approach to traffic distribution (every

vehicle completes a road link), it will reselect the shortest path at the end of the road link and continue to travel according to the new shortest path), and then when calculating the actual travel time of the road link, it no longer distinguishes the origin and destination point of the vehicle in the road link from the driving path but counts the running time of all vehicles directly. The path travel time calculation method can be specified as

$$\tau_a(k) = \frac{\sum_{z \in M^a} \tau_z^a(k) \xi_z^a(k)}{\sum_{z \in M^a} \xi_z^a(k)},$$

$$T_{ij}^r(k) = \sum_{a \in A_{ij}^r} \tau_a(k), \tag{23}$$

where $\xi_z^a(k)$ is whether the z-th traveler has completed the travel on link a, $\tau_z^a(k)$ is the actual travel time of z-th traveler completes of the road link a, and $T_{ij}^r(k)$ is the travel time of the path r.

The actual travel time of all the vehicles traveling on link a is counted in the k period, and the average value is taken as the actual travel time on link a of this period. The actual travel time of each path is calculated according to the travel time of the link, which is used as the basis for the existing vehicle to reselect the path and for the new vehicle to select the path.

3.3. Transmission Efficiency Optimization Method.

There are two main problems when applying this model to a large-scale road network: the number of feasible paths and the number vehicles adjusting the path in real time are both large. It is necessary to optimize the computational efficiency of the model for the above two problems.

3.3.1. Search Method Based on Path Set Matrix.

A path matrix is established in order to reduce the number of calculations when searching the shortest path, and the total feasible paths between all the node pairs are obtained by traversing the entire road network at one time; then the path to the corresponding position is recorded in the path set. Next, the path search range is limited to the existing feasible path in the set when performing traffic assignment. For the case where the number of times of distribution is extremely large, the amount of calculation can be cut down by reducing the number of searches:

$$\mathbf{R} = \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{pmatrix}; \tag{24}$$

among them

$$P_{ij} = \begin{cases} \varnothing, & i = j, \\ \{r_{ij} | 0 < l(r_{ij}) < +\infty\}, & i \neq j, \end{cases} \tag{25}$$

where R is the road network path set matrix, P_{ij} is the set of feasible paths from node i to j, r_{ij} is the feasible path

from node i to j, and $l(r_{ij})$ is the full travel time of the path r_{ij}.

When searching for the shortest path, directly follow the formula (26) to traverse the path set of the corresponding starting and ending points to obtain the shortest path.

$$r_{ij}^*(k) = \left\{ r | l(r, k) = \min_{r \in P_{ij}} \{l(r, k)\} \right\}, \tag{26}$$

where $r_{ij}^*(k)$ is the shortest path from node i to j in time period k and $l(r, k)$ is the full travel time of the path r.

3.3.2. Feasible Path Node Number Limit.

It is important to set the upper limit of the number of nodes in the path reasonably, which can not only reduce the candidate path set, improving the efficiency of establishing the path set matrix, but also increase the traversal calculation amount of each search for the shortest path in the path set, improving model calculation efficiency comprehensively.

Here, the search method based on the path set matrix is adjusted (the node number limit is introduced) and the path set is changed as

$$P_{ij} = \begin{cases} \varnothing, & i = j, \\ \{r_{ij} | 0 < l(r_{ij}) < +\infty, 0 < m(r_{ij}) \leq M_{ij}\}, & i \neq j, \end{cases} \tag{27}$$

where P_{ij} is the feasible path sets from nodes i to j, r_{ij} is the feasible path from nodes i to j, $l(r_{ij})$ is the travel time of path r_{ij}, M_{ij} is the upper limit of the path nodes number from node i to j, and $m(r_{ij})$ is the number of nodes included in path r_{ij}.

3.3.3. Node Group Division.

It is necessary to group these more concentrated nodes reasonably for the large-scale road network, in order to avoid searching for a more tortuous but not feasible path and then select the paths within the group and between groups to form the final travel path.

3.4. Vulnerability Index Calculation Method.

Vulnerability evaluation indicators are mainly concentrated in two categories: one is an indicator that starts from the network topology and ignores the characteristics of traffic flow and is suitable for analyzing networks with relatively certain traffic volumes and small changes in traffic; the other is under noncrowded conditions. The traffic performance index reflects the operating status of the composite transportation system and therefore gradually becomes the mainstream of the road network vulnerability evaluation index.

Most of the indicators considering the elements of traffic flow use the number of individuals affected by the event and the degree of individuals affected by the event as basic factors and then combine them in a certain expression form to establish vulnerability evaluation indicators.

$$V = F(n, c), \tag{28}$$

where V is the vulnerability evaluation index, n is the number of individuals affected by the event, and c is the extent to which the individual is affected by the event.

In vulnerability evaluation, it is a commonly used evaluation method to rank the criticality of road sections by measuring the severity of the consequences caused by the incident. Take the traffic flow, traffic demand, population, and other factors of the evaluated area as the basic number of affected individuals, calculate the degree of individual impact from the generalized travel cost, travel resistance, or accessibility loss, and use the product of the two as the basis data to measure the consequences of vulnerability exposure has been widely adopted by scholars.

Therefore, referring to the vehicle inflow and outflow curve shown in Figure 3, vulnerability evaluation indicators are constructed considering two basic factors in this paper, road traffic flow and traveling time consumption, and then discretize them based on their continuous expression. Taking the product of the instant vehicle number and the instant duration dt of a certain link at time t as the instantaneous travel time consumption of the traveler in the road link at that time, its integral is calculated during the whole research period as the traveler's total travel time consumption on the link a in the study period. where a means the occurrence of event a in the road network. When a is 0, it indicates that no interference event occurs, $u_a(t)$ is the cumulative vehicle inflow of the road network at time t, and $v_a(t)$ is the cumulative vehicle outflow of the road network at time t.

$$B_a = \int_0^t [u_a(t) - v_a(t)] dt, \tag{29}$$

Then discretize the overall travel time consumption of travelers on link a during the study period based on their continuous expression, and bring in the vulnerability calculation:

$$V_B = \frac{B_a - B_0}{B_0}, \tag{30}$$

where B_a is the travel time consumed of the evaluated area when the interference event a occurs, B_0 is the travel time consumed of the evaluated area when no interference event occurs, and V_B is the cumulative vulnerability index of the evaluated area in the research time.
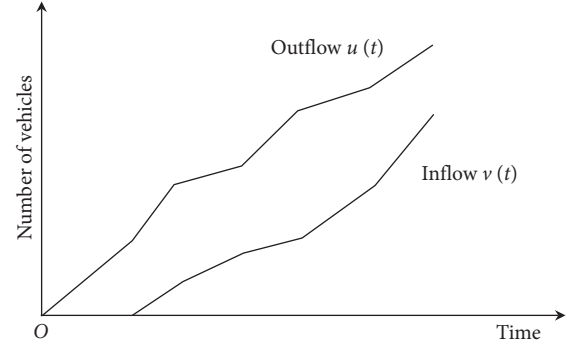
FIGURE 3: Schematic diagram of the inflow and outflow of vehicles on the road section.

The cumulative travel time of the road network relative to the incremental proportion of the undamaged road network in the study period is used as the cumulative evaluation index of the road network vulnerability, and the ratio of the travel time consumption of the road network under normal and abnormal conditions in the same period is used as the time-varying evaluation index of the road network vulnerability in this paper. The vulnerability assessment index is combined with the traffic flow evolution model to give a calculation method for the vulnerability dynamic indicator value in the model.

Under static conditions, the evaluation indicators of the two equilibrium operating states before and after the event are compared, but the analysis of the static equilibrium conditions implies the assumption that the equilibrium state will be realized, so its conclusion cannot make an accurate evaluation of the road network vulnerability if the equilibrium state has not been realized in the end, and the process of evolving from the old equilibrium to the new equilibrium is ignored. Correspondingly, during the evolution process, the phenomenon that caused traffic congestion was also ignored; therefore, the evaluation results have certain limitations. Therefore, this paper uses dynamic indicators to explore the cumulative and time-varying effects of vulnerability and evaluate the vulnerability of the road network more accurately and comprehensively.

3.4.1. Individual Link Vulnerability Index Calculation Method. The cumulative and time-varying indicators of link vulnerability are shown, respectively, as

$$B_a(k) = T_k [u_a(k) - v_a(k)] = T_k \sum_{i \in I_a} n_i^a(k), V_a = \frac{\sum_k [B_a(k) - B_a^0(k)]}{\sum_k B_a^0(k)}, V_a(k) = \frac{B_a(k) - B_a^0(k)}{B_a^0(k)}, \tag{31}$$

where T_k is the duration of the k-th time period, $n_i^a(k)$ is the number of vehicles in the i-th cell in link a during k period actual, $B_a(k)$ is the total time consumption of the road link a

in the k period, where 0 corresponds to the value when no event is affected, V_a is the cumulative vulnerability index value of link a as of k period, and $V_a(k)$ is the time-

varying vulnerability index value of road link a during k period.

3.4.2. Road Network Vulnerability Index Calculation Method.

According to the road network structure and the calculation method of the vulnerability index of the road link, the cumulative index and the time-varying index of the road network vulnerability are shown as

$$V_{\text{net}} = \frac{\sum_{a \in A} \sum_k \left[B_a(k) - B_a^0(k) \right]}{\sum_{a \in A} \sum_k B_a^0(k)},$$

$$V_{\text{net}}(k) = \frac{\sum_{a \in A} \left[B_a(k) - B_a^0(k) \right]}{\sum_{a \in A} B_a^0(k)},$$

$$(32)$$

where A is a set consisting of all links in the road network, V_{net} is the cumulative vulnerability index value of road network as of k period, and $V_{\text{net}}(k)$ is the time-varying vulnerability index value of road network during k period.

4. Case Study

4.1. Regional Overview and Parameter Design.

The surrounding area of Bao'an Center in Bao'an District of Shenzhen City is selected as an example study object. The boundary of the area is mainly composed of Guangshen Highway, Xixiang Avenue, Haibin Avenue, and Hubin Road, with Bao'an Avenue passing through and the dense branch network matching inside. The area is about 16 square kilometers and the planned permanent population is 380,000. The internal land use types are mainly commercial land near the Qianhai Bay (Binhai area), a large number of residential lands scattered within the area (Bihai area), and parts of school land.

The traffic flow of the study area is mainly composed of a large number of transit traffic flows and some internal traffic trips on Guangshen Highway and Bao'an Avenue. According to the detailed level of OD data, the road network in the area is collated and simplified appropriately based on the actual layout of the road network in Bao'an Central Area, and then the abstraction is the road network structure shown in Figure 4.

The time range is 6:30–9:30 pm, and a blocking density is 133 pcu/ln•km. The remaining parameters are shown in Table 1. Some roads do not have the same section throughout the entire road. The same value is used here to simplify the road network. In addition, for the road links where the road capacity and the free flow speed can be queried, the value is taken according to the actual situation; for the link where the relevant parameters cannot be queried, the value is taken according to the "City Planning Design Code CJJ37-2012." Main interference situation setting is shown in Table 1.

The details of the change process cannot be reflected when the time interval is too large, and it is difficult for the model to achieve the required accuracy if it is too small. This paper has conducted multiple tests based on the road length and speed limit of the studied area based on

previous article [37]. According to the principle of speed limit × time interval = cell length, for example, two nodes are separated by 1000 m and are divided into 5 cells with a length of 200 m. The driving distance can be exactly the length of 1 cell in each cycle according to the speed limit; thus, it can be realized that every cycle moves one cell to the downstream cell when there is no congestion. Meanwhile, the length of all the cell is controlled to a similar level and uses an integer number of cells. Finally, based on the results of multiple trial calculations, the simulation takes a time period T of 20 s.

4.2. Road Network Vulnerability Evaluation Results

4.2.1. Vulnerability Evaluation Results under Complete Failure.

Figure 5(a) shows the cumulative index changes of the road network vulnerability when the main section of the Bao'an Central Area completely fails, reflecting the overall time loss ratio of all travelers in the region since the event occurred relative to the undisturbed situation.

It can be seen from Figure 5(a) that the cumulative vulnerability curve shows two kinds of trends. One as in situation 1, its cumulative vulnerability value gradually increases to about 0.10 since the event occurs and then tends to be stable. In the others as in situations 2, 3, and 4, the cumulative vulnerability values have increased to a greater extent and produced more obvious signs of recovery. The cumulative vulnerability value of situation 4 is the largest, and its peak value reaches 0.28.

During the period of continuous impact of the event, the vulnerability index value is significantly higher than other situations. The recovery efficiency of situation 2 is relatively high, and the slope of the cumulative vulnerability curve in the recovery phase is significantly higher than in other situations. Corresponding the invalid road link of each situation to the road network, it can be found that when the contact link of Bao'an Avenue and Guangshen Highway fails, more overall time loss is generated and the recovery speed is slow.

Figure 5(b) shows the time-varying index of road network vulnerability when the main link of Bao'an Central Area fails. This index indicates the traveler's time loss in the road network at each time period since the event occurrence. This index eliminates the cumulative effect since the occurrence of the event and directly reflects the consequences of the event at the current time.

A similar conclusion to Figure 5(b) can be obtained from (a). It is worth noting that the initial impact of situation 1 is less than situation 2 at the peak of each curve, but the subsequent sustained impact is basically maintained above situation 2. It is indicated that the main impact periods of different road links are different, and the criticality of the two links needs to be discussed differently based on the continuous impact time of the events.

At the maximum peak of each curve, the initial impact of situation 1 is less than that of situation 2, but the subsequent continuous impact is basically maintained above situation 2. Thus, it may neglect the operational state of the road

(a)

(b)
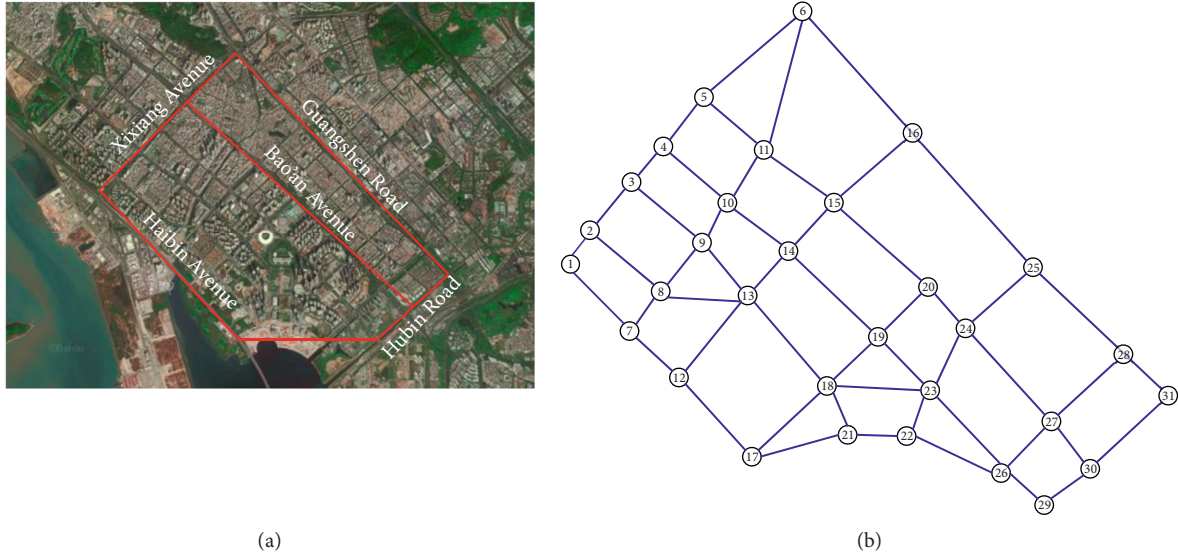
FIGURE 4: Example road network structure diagram. The left figure shows the actual layout of the road network in Bao'an Central Area, and the right figure shows that the abstraction is the road network structure.

TABLE 1: Main interference situation setting table.

Interference situation	Links	Nodes	Number of lanes	Degree of interference
Situation 1	Guangshen Highway (Xixiang Avenue-Chuangye 1st Road)	6⟶25	4 lanes	Complete failure
Situation 2	Bao'an Avenue (Xin'an 6th Road-Yu'an Road)	15⟶20	4 lanes	Complete failure
Situation 3	Bao'an Avenue (Chuangye 1st Road-Xin'an 1st Road)	24⟶27	4 lanes	Complete failure
Situation 4	Chuangye 1st Road (Guangshen Highway-Bao'an Avenue)	24⟶25	3 lanes	Complete failure
Situation 5	Chuangye 1st Road (Guangshen Highway-Bao'an Avenue)	24⟶25	3 lanes	Two-lane failure
Situation 6	Chuangye 1st Road (Guangshen Highway-Bao'an Avenue)	24⟶25	3 lanes	One-lane failure

(a)

(b)
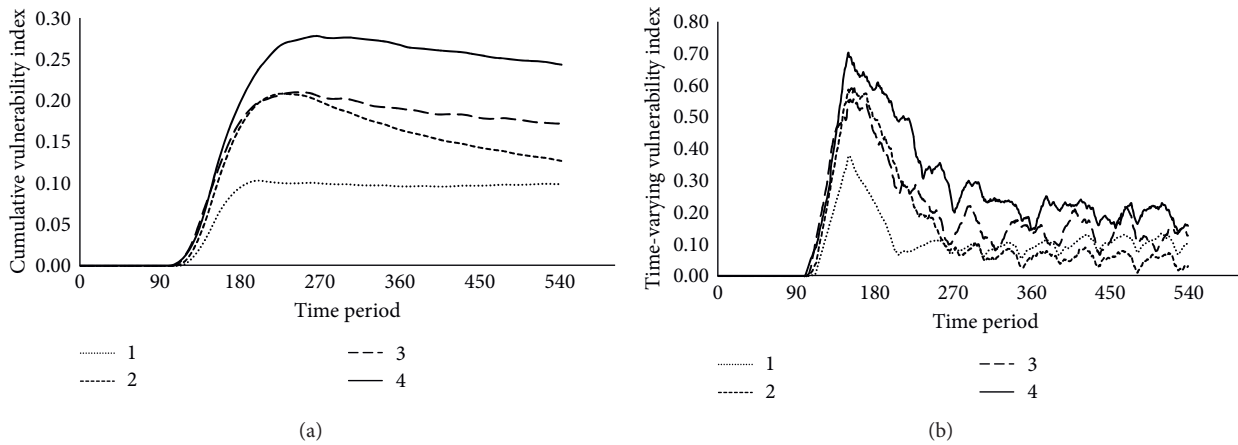
FIGURE 5: Vulnerability variation diagram under link complete failure. (a) shows the cumulative vulnerability when the main section of the Bao'an Central Area completely fails, and (b) shows the time-varying vulnerability. (a) Cumulative vulnerability. (b) Time-varying vulnerability.

network in the disordered phase initially affected by the event if applying static indicators only.

The vehicle arrival rate is used to analyze the road network flow variation. It can be visually observed from Figure 6 that the vehicle arrival rate in the disordered phase has dropped significantly in various interference situations, significantly in situations 3 and 4; see Figures 6(a) and 6(b). This corresponds to the obvious rising phase of the time-

(a)

(b)

(c)

(d)
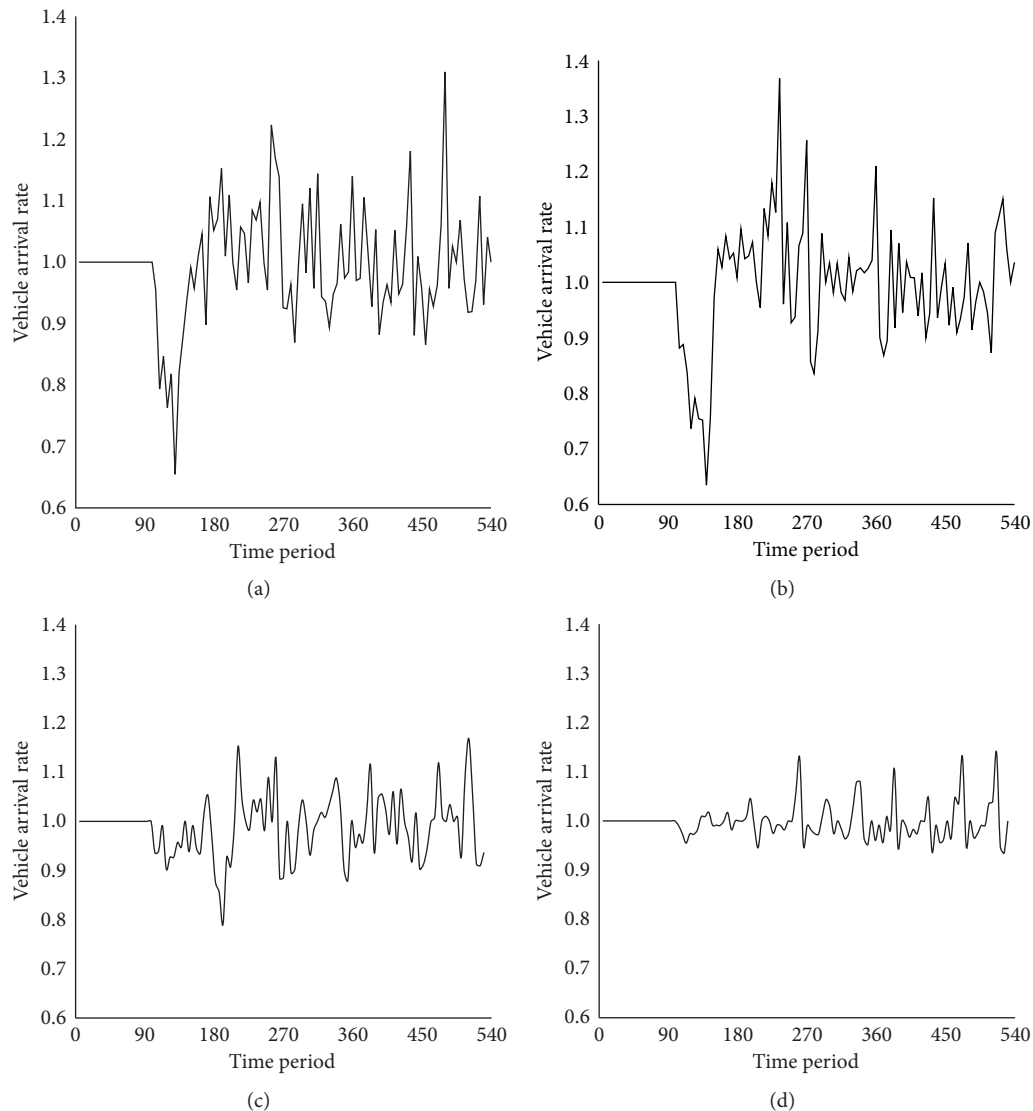
Figure 6: Vehicle arrival rate diagram under complete or partial failure. (a) Situation 3 and (b) situation 4 of complete failure show significant fluctuations. (c) Situation 5 and (d) situation 6 of partial failure situation show flat fluctuations. (a) Situation 3. (b) Situation 4. (c) Situation 5. (d) Situation 6.

varying peak and cumulative vulnerability of each interference situation at the initial stage of the event. The vehicle arrival rate of situations 3 and 4 still maintains a large fluctuation as time passes; that is, the unstable fluctuation of traffic flow makes the time-varying network vulnerability also maintained at a relatively high level.

According to the partial failure situation (see Figures 6(c) and 6(d)), the vehicle arrival rate curves are different from that of complete failure, there is no obvious impact on the disordered phase, and the fluctuation of the period impact amount is small. Situation 6 is even maintained at a level that is comparable to the noninterference situation. The vehicle could still travel to the destination along the original route since the road link is not interrupted in situation 5, so that the vehicle arrival rate does not observe a significant drop in the early stage of the event. After the event occurred for about 100 time periods, the arrival rate

shows a large decline. This is because the formation of congestion needs to go through a process after the partial failure of the road. The impact on traffic flow is gradually shown up after the congestion is basically formed.

4.2.2. Vulnerability Evaluation Results of Partial Failure. A further discussion is made based on selecting situation 4 which has a relatively high vulnerability value throughout the simulation. By setting the road partial failure situation from the failed road link $24 \longrightarrow 25$ in situation 4 as situations 5 and 6, the road network vulnerability under different interference levels is discussed.

The single-lane failure of the road link can barely expose the vulnerability of the road network (Figure 7). Because the minimum travel time of the link $24 \longrightarrow 25$ increases due to the single-lane failure, the link will not become the shortest path
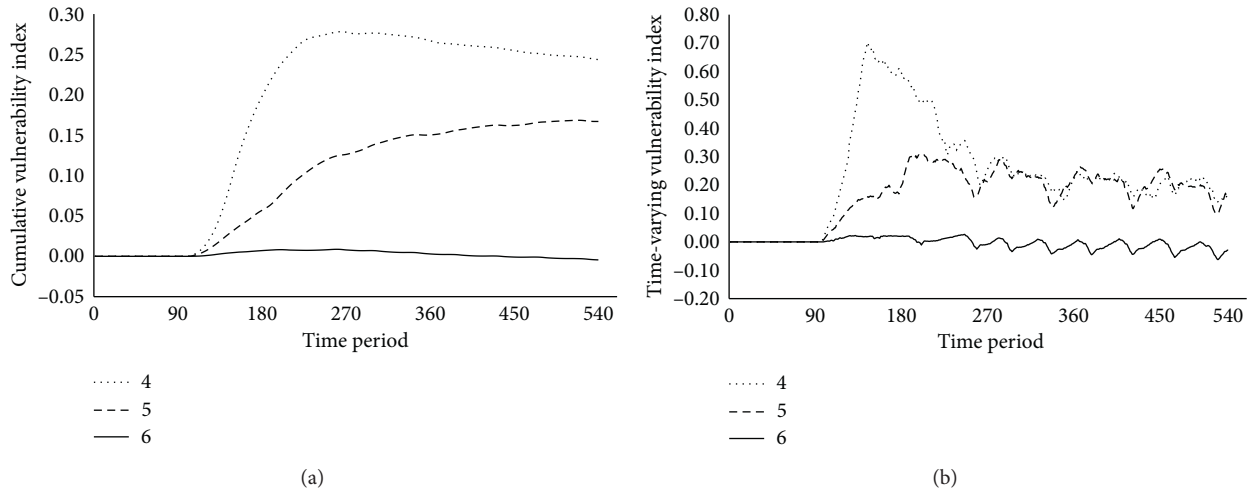
FIGURE 7: Vulnerability variation diagram under link partial failure. (a) Cumulative vulnerability when setting the road link 24⟶25 failed. (b) Time-varying vulnerability. (a) Cumulative vulnerability. (b) Time-varying vulnerability.
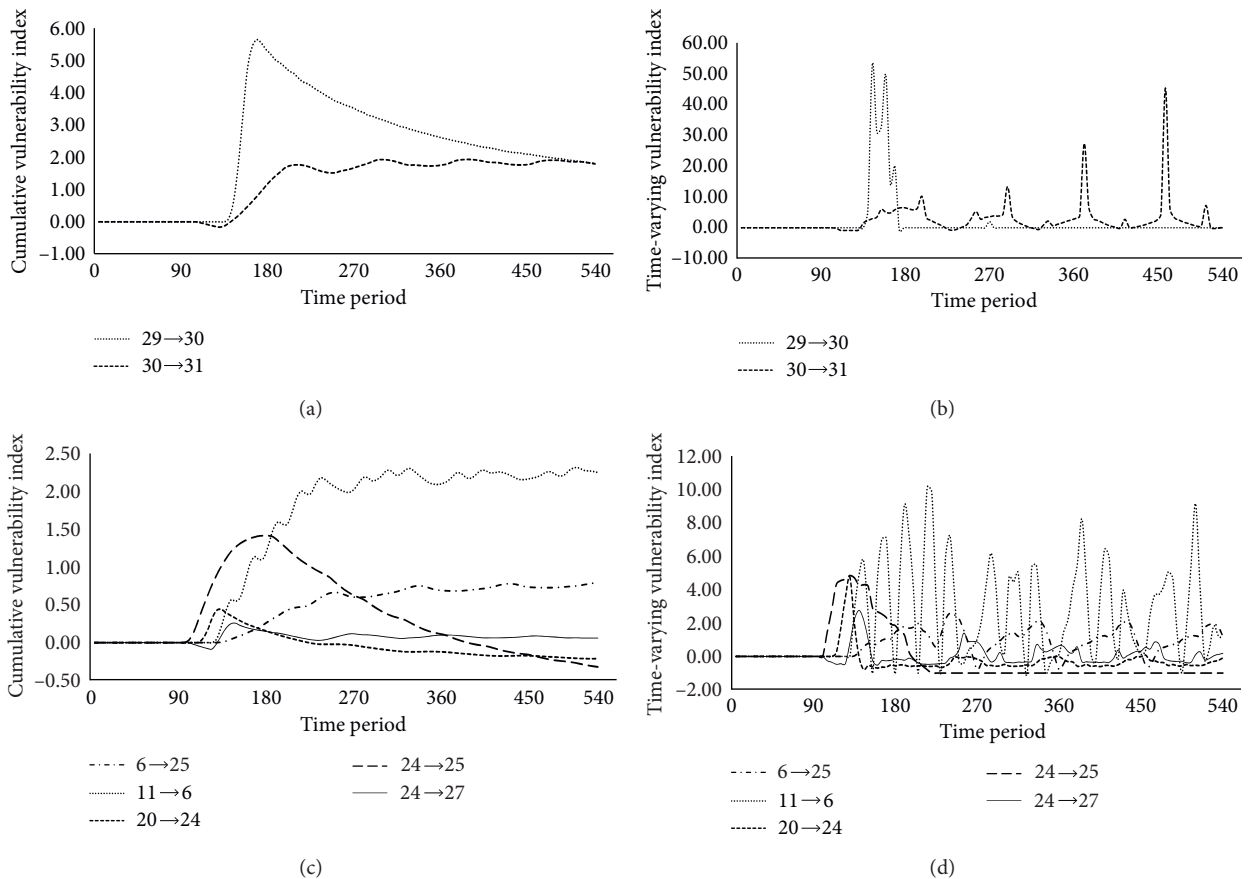
FIGURE 8: Individual link vulnerability variation. (a, b) Cumulative and time-varying vulnerability under severe impact. (c, d) Cumulative and time-varying vulnerability under general impact. (a) Cumulative vulnerability under severe impact. (b) Time-varying vulnerability under severe impact. (c) Cumulative vulnerability under general impact. (d) Time-varying vulnerability under general impact.

for some travel demand, so that the traffic volume of the link is more stable compared with the noninterference situation, and the road network vulnerability has also decreased.

As Figure 7(b) shows, the initial impact of two-lane failure (scenario 5) is much lower than that of road section complete failure (scenario 4), but the continuous impact

of the two is basically the same. Using static indicators to compare the vulnerability indicators of the two equilibrium states may lead to conclusions that the vulnerabilities of the two situations are similar, which is inaccurate.

4.3. Individual Link Vulnerability Evaluation Results. It is not convenient to list the evaluation results of all the links under all interference situations since there are many links in the example road network. Therefore, take the situation with the highest vulnerability value in the vulnerability evaluation of the former road network as the evaluation object, giving the evaluation result in which the link vulnerability changes significantly under this situation.

Figure 8 shows the vulnerability index variation of the road links with large vulnerability values in situation 4, depending on the degree of each link impact. Obviously, links 29⟶30 and 30⟶31 are severely impacted, and the impact is even higher than the direct event-occurred link 24⟶25. Among them, the influence of the link 29⟶30 is mainly in the disorder stage, and 30⟶31 is affected multiple times during the event impact, which is mainly because the travelers alternately select the shortest path and the second shortest path with the traffic flow evolution, causing link congestion to be formed several times and gradually dissipated.

The time-varying vulnerability of the road link 11⟶6 shows a large fluctuation, and the cumulative vulnerability also has risen with fluctuations. This repeated fluctuation of time-varying vulnerability is also caused by the traveler's alternately selection between the shortest path and the second shortest path.

The two vulnerability index values on the event-occurred link 24⟶25 rise sharply in the disordered phase and decline gradually in the recovery phase. In addition, the time-varying vulnerability of this link is maintained at a lower position after the vehicle is evacuated, and the cumulative vulnerability is also gradually reduced correspondingly. The two vulnerability index values of the links 20⟶24 and 24⟶27 have a small increase with fluctuations in the disordered phase and then quickly return to the low position and enter a new stable state. This change is mainly due to the fact that the two links connect the upstream nodes of the event-occurred link 24⟶25, undertaking the task of evacuating the link traffic flow. Therefore, the vulnerability value in the disordered phase has a large increase obviously and then gradually enters a new equilibrium state after the vehicle evacuation is completed.

5. Conclusions

The vulnerability evaluation of the example road network is carried out, and the dynamic change law of vulnerability is summarized. The model is used to evaluate the vulnerability of road network and link from the two aspects of complete failure and partial failure. The specific research results and conclusions obtained are as follows:

(1) The impact of interference events usually includes a large increase in the time-varying vulnerability value in the disordered phase and then gradually returns to the low level in the recovery phase. The change in cumulative vulnerability value is mainly reflected in the significant increase in the disordered phase and the gradual decline or substantial stability after the evolution into the recovery phase.

(2) Comparing the evaluation results of the road network and individual link, it is found that the road network vulnerability change is relatively flat, and the influence of individual link has the order of influence of event-occurred link, adjacent road links, and similar functional links. Therefore, the road clearing and traffic guidance should be carried out according to the specific conditions of each stage during the dynamic process, strengthening the traffic guidance for the direct event-occurred link and then the similar functional links to avoid repeated congestion.

(3) Comparing the results of complete failure and partial failure of the road, it is found that the interfering event will not have a significant impact on the road network vulnerability when the road has sufficient spare capacity. When the spare capacity of the road link is insufficient, the complete failure of the road causes a greater initial impact. Therefore, it is necessary to maintain a certain traffic capacity for the road links to avoid the vehicles from rushing into adjacent road links in a short period of time.

(4) From the perspective of road traffic flow, when the main road with large traffic flow fails, the road network has a large cumulative vulnerability and a greater time-varying vulnerability during the disordered phase. Therefore, prioritizing the smooth flow of major roads can weaken the impact of road network vulnerability exposure.

(5) From the perspective of network topology, the less the spare capacity or the number of alternative road links, the higher the vulnerability of the road link. Therefore, improving the density of the road network and improving the redundancy of the road network from the topological level can also improve the ability of the road network to resist interference events.

In terms of the universality, it should be noted that lessons from this study could not be generalized to regions like the US or Europe where urban traffic condition is different from that in China. However, for some developing countries, our findings have potential to provide some reference comparably, especially for rapid urbanizing and renewing cities in southeast Asia, with similar traffic condition and road structure to China. The dynamic road vulnerability is complex and the data from Shenzhen city are of limited size, which calls for more context-sensitive and place-based research to supplement the existing knowledge in the future. Our findings could contribute to guiding

sustainable development of integrated transport planning for these developing countries.

For future study, the CTM will be integrated with other transportation networks. Improved versions of this model will consider the elastic demand of traffic travel and the cognitive error of the user since only the complete failure or partial failure of an individual link was considered. The multilink combination failure caused by severe interference events can be studied in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] K. Berdica, "An introduction to road vulnerability: what has been done, is done and should be done," *Transport Policy*, vol. 9, no. 2, pp. 117–127, 2002.

[2] G. M. D'este and M. A. P. Taylor, "Network vulnerability: an approach to reliability analysis at the level of national strategic transport networks," in *Proceedings of the 1st International Symposium on Transportation Network Reliability (INSTR)*, May 2003.

[3] L.-G. Mattsson and E. Jenelius, "Vulnerability and resilience of transport systems - a discussion of recent research," *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 16–34, 2015.

[4] Y. Gu, X. Fu, Z. Liu et al., "Performance of transportation network under perturbations: reliability, vulnerability, and resilience," *Transportation Research Part E: Logistics and Transportation Review*, vol. 133, 2020.

[5] K. Sugishita and Y. Asakura, "Citation network analysis of vulnerability studies in the fields of transportation and complex networks," *Transportation Research Procedia*, vol. 47, pp. 369–376, 2020.

[6] M. G. H. Bell, "The use of game theory to measure the vulnerability of stochastic networks," *IEEE Transactions on Reliability*, vol. 52, no. 1, pp. 63–68, 2003.

[7] M. G. H. Bell, U. Kanturska, J. D. Schmocker et al., "Attacker-defender models and road network vulnerability," *Philosophical Transactions of The Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1872, pp. 1893–1906, 2008.

[8] M. G. H. Bell, A. Fonzone, and C. Polyzoni, "Depot location in degradable transport networks," *Transportation Research Part B: Methodological*, vol. 66, pp. 148–161, 2014.

[9] M. G. H. Bell, F. Kurauchi, S. Perera, and W. Wong, "Investigating transport network vulnerability by capacity weighted spectral analysis," *Transportation Research Part B: Methodological*, vol. 99, pp. 251–266, 2017.

[10] J. Pan, M. G. H. Bell, K. F. Cheung et al., "Identifying container shipping network bottlenecks along china's maritime silk road based on a spectral analysis," *Maritime Policy & Management*, vol. 48, no. 2, 2020.

[11] A. Candelieri, B. G. Galuzzi, I. Giordani, and F. Archetti, "Vulnerability of public transportation networks against directed attacks and cascading failures," *Public Transport*, vol. 11, no. 1, pp. 27–49, 2019.

[12] B. Abdulla and B. Birgisson, "Characterization of vulnerability of road networks to random and nonrandom disruptions using network percolation approach," *Journal of Computing in Civil Engineering*, vol. 35, no. 1, 2020.

[13] M. Chen and H. Lu, "Analysis of transportation network vulnerability and resilience within an urban agglomeration: case study of the greater Bay area, China," *Sustainability*, vol. 12, no. 18, p. 7410, 2020.

[14] M. A. P. Taylor, S. V. C. Sekhar, and G. M. D'Este, "Application of accessibility based methods for vulnerability analysis of strategic road networks," *Networks and Spatial Economics*, vol. 6, no. 3-4, pp. 267–291, 2006.

[15] A. Chen, C. Yang, S. Kongsomsaksakul, and M. Lee, "Network-based accessibility measures for vulnerability analysis of degradable transportation networks," *Networks and Spatial Economics*, vol. 7, no. 3, pp. 241–256, 2007.

[16] E. Jenelius, T. Petersen, and L.-G. Mattsson, "Importance and exposure in road network vulnerability analysis," *Transportation Research Part A: Policy and Practice*, vol. 40, no. 7, pp. 537–560, 2006.

[17] E. Jenelius, L.-G. Mattsson, and D. Levinson, "Traveler delay costs and value of time with trip chains, flexible activity scheduling and information," *Transportation Research Part B: Methodological*, vol. 45, no. 5, pp. 789–807, 2011.

[18] E. Jenelius, "Network structure and travel patterns: explaining the geographical disparities of road network vulnerability," *Journal of Transport Geography*, vol. 17, no. 3, pp. 234–244, 2009.

[19] E. Jenelius and L.-G. Mattsson, "Road network vulnerability analysis of area-covering disruptions: a grid-based approach with case study," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 746–760, 2012.

[20] M. A. P. Taylor and fnm Susilawati, "Remoteness and accessibility in the vulnerability analysis of regional road networks," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 761–771, 2012.

[21] E. Jenelius and L.-G. Mattsson, "Road network vulnerability analysis: conceptualization, implementation and application," *Computers, Environment and Urban Systems*, vol. 49, pp. 136–147, 2015.

[22] J. Ding, S. Gao, E. Jenelius et al., "Routing policy choice set generation in stochastic time-dependent networks, transportation research record," *Journal of the Transportation Research Board*, vol. 2466, no. 1, pp. 76–86, 2014.

[23] M. Bababeik, M. M. Nasiri, N. Khademi, and A. Chen, "Vulnerability evaluation of freight railway networks using a heuristic routing and scheduling optimization model," *Transportation*, vol. 46, no. 4, pp. 1143–1170, 2019.

[24] G. Gecchele, R. Ceccato, and M. Gastaldi, "Road network vulnerability analysis: case study considering travel demand and accessibility changes," *Journal of Transportation Engineering*, vol. 145, no. 7, pp. 05019004.1–05019004.10, 2019.

[25] D. M. Scott, D. C. Novak, and F. Guo, "Network robustness index: a new method for identifying critical links and evaluating the performance of transportation networks," *Journal of Transport Geography*, vol. 14, no. 3, pp. 215–227, 2006.

[26] C. Balijepalli and O. Oppong, "Measuring vulnerability of road network considering the extent of serviceability of critical road links in urban areas," *Journal of Transport Geography*, vol. 39, pp. 145–155, 2014.

[27] E. Alex, B. James, A. Kay et al., "Vulnerability assessment methodology for Swiss road network," *Transportation Research Record*, vol. 2137, no. 1, pp. 118–126, 2009.

[28] B. Y. Chen, W. H. K. Lam, A. Sumalee, Q. Li, and Z.-C. Li, "Vulnerability analysis for large-scale and congested road networks with demand uncertainty," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 3, pp. 501–516, 2012.

[29] X. Xiangdong, C. Anthony, J. Sarawut et al., "Transportation network redundancy: complementary measures and computational methods," *Transportation Research Part B Methodological*, vol. 114, pp. 68–85, 2018.

[30] O. Cats and E. Jenelius, "Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information," *Networks and Spatial Economics*, vol. 14, no. 3-4, pp. 435–463, 2014.

[31] E. Jenelius, "Incorporating dynamics and information in a consequence model for road network vulnerability analysis," in *Proceedings of the 3rd International Symposium on Transport Network Reliability (INSTR)*, The Hague, The Netherlands, July 2007.

[32] V. L. Knoop, M. Snelder, and S. P. Hoogendoorn, "Link-level vulnerability indicators for real-world networks," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 5, pp. 843–854, 2012.

[33] S. Kim and H. Yeo, "Evaluating link criticality of road network based on the concept of macroscopic fundamental diagram," *Transportmetrica A: Transport Science*, vol. 13, no. 2, pp. 162–193, 2017.

[34] E. Jenelius and L. G. Mattsson, "Developing a methodology for road network vulnerability analysis," *Nectar Cluster*, vol. 1, pp. 1–9, 2006.

[35] C. F. Daganzo, "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.

[36] C. F. Daganzo, "The cell transmission model, part II: network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.

[37] H. Szeto and Y. S. Wai, "A cell-based variational inequality formulation of the dynamic user optimal assignment problem," *Transportation Research Part B: Methodological*, vol. 36, no. 5, pp. 421–443, 2002.

[38] W. Y. Szeto and H. K. Lo, "A cell-based simultaneous route and departure time choice model with elastic demand," *Transportation Research Part B: Methodological*, vol. 38, no. 7, pp. 593–612, 2004.

[39] Y. Jiang, Y. Wang, W. Y. Szeto et al., "Probabilistic assessment of network vulnerability with equilibrium flows by metaheuristics," *International Journal of Sustainable Transportation*, vol. 3, 2020.

WILEY | Hindawi

Research Article

Constructing Scenarios' Network-of-Flight Conflict in Approach of Intersecting Runway

Ming Cheng [ID]**, Yixuan Li, and Xiaolian Han**

College of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China

Correspondence should be addressed to Ming Cheng; figocm@163.com

For studying the mechanism of flight conflict in approach of the intersecting runway, this paper applies the case study, scenario construction, and complex network, analyzes the operational risks of the intersecting runway, and researches the general rule of flight conflict. We constructed a network model of scenario evolution of flight conflict with selecting Beijing Daxing International Airport as the research object, which included 169 nodes and 263 edges. It proposed path evolution and verified the effectiveness of this network. We analyzed the degree centrality, median centrality, and closeness centrality of the network, and the results showed that the comprehensive value of 5 nodes is high, including go-around (V2), conflict resolution (C22), the warning light of aft cargo dJor was extinguished (F12), suspend subsequent take-off operations (F5), and keeping visual flying (C26). The results show that this method provides a new research way for the control strategy of chain breakage and the mechanism of scenario evolution of flight conflict.

1. Introduction

In recent years, in order to cope with the shortage of airport runway capacity in the development of civil aviation industry of China, airports are being rebuilt and expanded all over the country, which is a project with huge demands on land resources and environment. The airport with the intersecting runway has many advantages. On the one hand, it can adapt to the change of wind direction, realize omnidirectional take-off and landing, improve the efficiency and safety of the runway operation, and increase the flow. On the other hand, land resources can be greatly saved. The airports using the intersecting runway is becoming a superlarge and large airports, for example, Beijing Daxing International Airport [1], which operated in 2019.

There has been rich experience in the operation of the parallel runway in China. In 2005, Beijing Capital International Airport was the first airport in China to implement the independent operation of two runways. At the same time, the Civil Aviation Administration of China has continuously released the operation rules for multiple runways, such as CCAR-98TM [2] and CCAR93TM—R2 [3]. However, serious flight conflicts still occur in the operation of parallel runways in China. For example, the "10.11" flight conflict (serious accident) sign occurred in Shanghai Hongqiao Airport in 2016 [4].

2. Research Actuality

Because flight conflict is the direct cause of the plane crash [5], the research on the intersecting runway is relatively early abroad; most of the work progress to the prediction and prevention of the conflict stage, whose content involves the ground early warning system, the pilot alarm information, intersecting runway take-off location identification system, take-off and landing aircraft separation operation [6–8], etc.

However, due to the fact that the intersecting runway operation mode has not been officially used in China, current studies still focus on the wake interval, runway capacity, flight data analysis [9–11], and other related issues. The research in this area is relatively rare. In the relevant research on flight conflict, domestic scholars mostly take parallel runway as the basis and focus on conflict hotspot identification and the establishment and solution of the relief model [12–14].

Complex network has been applied to the risk scenarios' evolution process of the deep-water drilling platform [15] and amphibious seaplanes [16] and achieved remarkable results. At the same time, many hot problems in the field of civil aviation also applied to the single-layer and multi-layer complex network, such as the route network [17], the flight delays [18], traffic distribution strategy [19], and security vulnerability analysis [20], based on the performance of navigation (PBN) [21].

The research actuality at home and abroad is supplemented and modified as follows: at present, many research studies on the prediction and prevention of aviation safety incidents rely on aviation safety incident reports, but most of the reports have problems such as large content and non-standard language and writing style [22]. Therefore, it is vital to accurately identify why these incidents occurred in the aviation safety incident report [23]. Xu et al. studied Natural Language Process (NLP), text mining techniques, machine learning, and other aspects, which effectively improved the accuracy of information processing [24, 25].

At present, there are relatively few studies on the combination of the operation mode of the intersecting runway and the scenarios' evolution of flight conflict. In view of the actual operational requirements and potential safety problems of airports with superlarge intersecting runway in China, this paper adopts complex network theory and scenario analysis method to conduct scenario evolution and risk analysis of flight conflict of the intersecting runway, in order to explore a new way of describing and analyzing flight conflict risk.

3. Construction of Method and Model

3.1. Theoretical Method

3.1.1. Construction of Scenario Theory. A scenario is a collection of a large number of similar events and various risks that may occur in the future [26, 27]. Scenario elements are usually analyzed from three dimensions of disaster body, disaster-resistant body, and disaster body [28]. Situational elements are the key factors of situational construction, which can reflect the development state and trend of events.

3.1.2. Complex Network Theory. Qian Xuesen defined that the complex network is the network with a part or all of the properties of self-organization, self-similarity, attractor, small world, and scale-free [29]. Based on system theory, graph theory, and statistical theory, the complex network can represent intuitively connectivity between system structures by establishing accident scenarios [30]. The complex network is described by a weighted directed acyclic connected graph $G = (V, L, W)$ of the sparse matrix, which is suitable for the study of accidents with complex accident mechanism, numerous risk factors, and risk factors with complex relationships and complex accident models' components.

(1) The node degrees

Node degree is the set of the input degree and exit degree of the node and the number of edges connected by nodes. The degree of node t is denoted as

$$k_t = \sum_s a_{ts}. \tag{1}$$

In the type, k_t is the degree of the node and a_{ts} is the number of edges connected between nodes v_t and v_s. Node degrees can reflect the importance of nodes. The greater the degree of the node is, the more important the nodes in the network are.

(2) The degree centrality

The degree centrality of nodes is to measure how closely a node in the network is connected with all other nodes. The degree centrality of nodes is denoted as

$$C_D(v_t) = \frac{\sum_{t=1}^s a_{ts}}{s - 1}. \tag{2}$$

In the type, $C_D(v_t)$ is the degree centrality of nodes and a_{ts} is the number of edges connected between nodes v_t and v_s (excluding self-ring). The degree centrality of a node reflects the degree of association between a node and other nodes in the network. The greater the degree centrality of a node, the closer the connection between the node and other nodes.

(3) Median centrality

Median centrality of a node is the ratio of the number of shortest paths a node has passed to all shortest paths in the network. The median centrality of a node is noted as

$$C_B(v_t) = \sum_{s \neq v \neq t \epsilon V} \frac{\sigma_{st}(v)}{\sigma_{st}}. \tag{3}$$

In the type, $C_B(v_t)$ is the median centrality of nodes and $\sigma_{st}(v)$ is the number of shortest paths through nodes v. Median centrality of nodes is another index that reflects the importance of nodes in the network. The greater the median centrality of nodes, the higher the position of nodes in the network.

(4) Closeness centrality

The closeness centrality of a node is the ratio of the number of all nodes related to a node in the network to the number of all shortest paths passing through this node:

$$C(v_s) = \frac{1}{\sum_{v_s} d(v_s, v_t)}. \tag{4}$$

In the type, $C(v_s)$ is the closeness centrality of nodes and $d(v_s, v_t)$ is the distance between nodes v_s and v_t. Closeness centrality of nodes is a parameter that measures the importance of nodes by the average length of shortest paths between nodes. The greater the proximity centrality of nodes is, the more important the nodes in the network are.

3.2. Analysis of Risk Characteristics of Intersecting Runway
Operation. The flight conflicts of the intersecting runway

have varied causes and complex evolvement process, which is suitable for using the complex network to study. The complex network is between the regular grid and random network. The nodes are connected into edges in a self-organizing way, and the initial event evolves into the final event through different paths. The evolution of the intersecting runway flight conflict scenario has the following characteristics.

3.2.1. Complexity.

The risks of flying at low and medium altitudes, especially in tower control areas, eventually emerge in the operation. In the process of flight, facing the influence of turbulence, thunderstorm, wind shear, ice accumulation, and other bad weather, restrictions on airspace imposed by military aviation activities have randomness and variability, and aircraft and air traffic control equipment are prone to failure. These risk factors interact with each other in a complex way, projecting them into the network as nodes.

3.2.2. Small-World Character.

The small world of the evolution of the risk of intersecting runway flight conflict is reflected in that, although there are many risk factors affecting flight conflict events, it can occur in a few short nodes between the initial event and the resulting event.

3.2.3. Scale-Free Features.

The scale-free property of the complex network mainly describes the problem of the node degree. A few nodes in the network have a lot of connections, while most do not. In the evolution process of flight conflict scenarios for the intersecting runway, most risk factors revolve around the results of flight conflict and several major risk factors leading to flight conflict, such as aborted approach, go-around, re-approach, and so on, which reflect the scale-free characteristics in the evolution process.

3.2.4. Community Structure Characteristics.

Intersecting runway flight conflict situation evolution concerns the four types of risk factors of Human, Machine, Ring, and Tube. The complex network provides a model that can show the interrelationships between each type of the risk factor and the interaggregation of related risk factors. Categorizing these risk factors, we can identify the commonalities of these risk factors and the relationships between each type of factor.

3.3. Construction of Scenarios' Evolutionary Network Model of Flight Conflict in Approach Stage

3.3.1. Identification of Operational Risk Factors for Airport with Intersecting Runway Configuration.

This study was based on the real layout of the intersecting runway of Daxing Airport. Major operational risks in the approach phase are shown in Table 1.

3.3.2. Network Model Construction Procedures.

The process of constructing the evolution network model of flight conflict scenarios in the approach stage of Daxing Airport's cross-runway is as follows:

(1) *Data Processing.* Collected and sorted out laws and regulations related to intersecting runway operation of civil aviation as well as relevant data of Daxing Airport's operating and natural environment; a total of 906 flight conflict incidents were collected and summarized from 2010 to 2019. The cases were divided into 6 categories, including aborted approach, rejected take-off, runway unusable, ground activities, unmanned aerial vehicles, and certain consequences.

(2) *Case Study.* We defined relevant risk factors that could lead to flight conflict as a keyword library, applied the Chinese Word Segmentation technology of Python ieba function library to extract the keywords in the event, and counted the frequency of statistical keyword, logical relationship, and other parameters, and the keyword library is modified by the results of the Chinese Word Segmentation technology to make it closer to the case contents.

(3) *Construction of Scenario Group.* Constructed the logical link between the keywords in taking a single case as a unit, extracted the safety risk factors, scenario description, scenarios' elements and nodes of flight conflict occurring during aircraft approach in the terminal area of the airport with the intersecting runway, and constructed the scenario group and evolution network

(4) *The Construction of Complex Network.* Sorted out the public node of different scenarios and plotted complex network diagrams for flight conflict scenarios in the intersecting runway terminal area of Daxing Airport.

(5) *Analysis of Experimental Results.* Calculated network parameters including node, edge, and weight and analyzed their influence on flight safety.

4. The Empirical Analysis

4.1. The Experimental Background.

This paper is based on the actual layout of the intersecting runway of Daxing Airport and assumes that the flight conflict would occur after the 11L/29R runway was put into operation: the landing would be made on the 29R runway, the approaching aircraft would stop the approach and go-around, and its track would cross with the aircraft taking off and landing on other runways, resulting in flight conflict.

4.2. Complex Network Construction.

The risk factors of the abort approach event case set, namely, nodes of the complex network, were extracted, including 37 nodes. The risk factors (nodes) of the abort approach scenario are shown in Table 2.

By sorting and screening invalid edges in the network (go-around⟶abort approach), the obtained directed network graph contains 78 edges, as shown in Figure 1.

By the same token, the nodes of the other 5 cases were extracted and constructed to network. Because of the limit of

FIGURE 1: Directed network graph of the aborted approach cases.

3.143, indicating that each node can affect other nodes only through the average of 3.143 units. The network diameter of the network is 9, indicating that any of the nodes in the network may cause flight conflicts up to 9 steps. The average cluster coefficient is 0.097, which reflects that the interaction between nodes is low.

5. Experimental Results' Discussion

5.1. Node Degree and Degree Distribution. Table 5 shows that, in the flight conflict scenario evolution network model of Daxing Airport's intersecting runway, for the degree, the degrees and in-degree of the node (V1) are max. The degrees (degree value 38) are max, indicating that it is the most important node in the network, and the in-degrees (degree value 32) are max, indicating that the risk factors leading to the suspension of approach are the most, and it is difficult to control. V1 is a key risk factor and one of the necessary conditions leading to flight conflict. The result is completely consistent with the actual situation.

In terms of out-degree, the controller makes notification (V18), blind approach (V16), and check (V19), which are the three nodes with the largest out-degree. V18 is a process event and not a risk factor, so it can be ignored here. V16 is a node describing the approach state, which is a risk factor, and its large degree indicates that it is more

likely to cause subsequent risk events in the process of blind approach.

Nodes with higher degrees should be paid attention during the evolution of flight conflict scenarios at the Daxing Airport crossing runway.

5.2. Betweenness Centrality of Nodes. Table 6 shows that the betweenness centrality values of V19 and V1 are the largest, which indicates that the shortest paths V19 and V1 pass are the most, and V19 and V1 play the most important role in the risk transmission process of the whole network.

The analysis of the actual case shows that the factors causing V1 include the abnormal state of runway and all kinds of approach equipment. In this case, the controller will inform the relevant ground personnel to check and clear trouble at the first time, so the betweenness centrality value of V19 is the largest among all risk event nodes, which is completely in line with the reality.

V19 and V1 play an important role in the evolution of flight conflict scenarios on the intersecting runway.

5.3. Closeness Centrality of Nodes. The 12 nodes in Table 7 had the highest closeness centrality. The closeness centrality of nodes shows of the location of nodes in the network. The

TABLE 3: Risk factors (nodes).

Numbers	Risk factors	Degree
V1	Abort approach	38
V2	Go-around	9
V3	Rainfall	2
V4	Turbulence	3
V5	Reapproach	13
V6	Flap fault	2
V7	Bird strike	3
V8	Runway suspended	7
V9	Avoid	5
V10	Ground proximity warning	3
V11	Crossing the runway waiting line	4
V12	Conflict	10
V13	TCAS warning	1
V14	Conflict warning	2
V15	PTCAS	1
V16	Blind approach	14
V17	Waiting in place	3
V18	Controller briefing	33
V19	Inspect	18
V20	Safe landing	11
V21	Approach coordination	3
V22	Normal approach	6
V23	Continue approach	5
V24	Run off the runway	9
V25	Alarm elimination	2
V26	Suit of pavement	3
V27	Change runway to land	4
V28	Glide back	25
V29	Take-off interrupted	21
V30	Automatic pressurization system failure	1
A1	Wind shear	3
A2	Turbulent flow	3
A3	Thunderstorm	1
A4	Unable to see the runway	1
A5	Excessive tailwind	1
A6	Excessive gust	1
A7	Orders plane to slow down	12
A8	Causes of runway configuration	1
A9	Failed to get off the runway in time	1
A10	Unable to see the front plane	9
A11	Still has catch-up trend	3
A12	Under the wake interval	9
A13	Overweight	2
A14	Continue to catch up with the front plane	7
A15	Catch up with the front plane	7
A16	Dissatisfaction landing interval	2
A17	There is a trend of catching up	6
A18	Unstable approach	3
A19	Bias navigation	3
A20	Crosswind	1
A21	Frost fog	1
B1	Repair	2
B2	No foreign matter was found	1
B3	The equipment is normal	3
B4	Departure aircraft waiting	1
B5	The course signal is normal	3
B6	Confirm whether the blind drop signal is stable	2
B7	Course stability	2
B8	Fragments	1
B9	Plastic bag	1

TABLE 3: Continued.

Numbers	Risk factors	Degree
B10	Course signal instability	3
B11	Course instability	5
B12	The one minute vector line swings left and right	3
B13	The course signal is unstable	2
B14	Signal instability of glide path	2
B15	The radar signal swings left and right	1
B16	Radar track swing	2
C1	MSAW alarm	1
C2	Descent height	2
C3	Stop descent	2
C4	Alarm release	1
C5	Unidentified vehicle	2
C6	Controller call field service assistance handling	2
C7	Radio jamming channel	2
C8	Get off the runway	1
C9	Deviation taxiway	2
C10	Guided vehicle passes the waiting point without permission	1
C11	Runway intrusion warning	2
C12	Controller verification	2
C13	The guide car exits outside the waiting point	1
C14	Breaking the command height	1
C15	Keep going up	2
C16	Controller command descent	2
C17	The height setting is correct	1
C18	Flight procedure error	1
C19	Upwind not turning according to the procedure	2
C20	Deviation from procedure	2
C21	Rejoin the correct take-off procedure	1
C22	Conflict resolution	6
C23	No TCAS alarm	1
C24	Controller asked if it could be visualized	2
C25	Visualization	2
C26	Keep visualization	2
C27	Converging flight at the same altitude	1
C28	Converging flight	2
C29	Slow down	1
C30	Under the regular interval	1
C31	Waiting outside the runway	1
C32	Drive-bird car for road inspection	1
C33	Waiting on taxiway	2
C34	Delay	1
D1	Drone	5
D2	Tower verification to crew	4
D3	Not found by the crew	5
D4	The operation was not affected	2
D5	Departure aircraft affected	1
D6	Approach aircraft affected	1
D7	The moving direction is uncertain	1
D8	Balloon	4
D9	Crew visual balloon activity	1
D10	Kite	1
D11	Floater	1
D12	Laser irradiation	1
E1	Airforce activities	1
E2	Reasons for passengers	1
E3	Mechanical fault	1
E4	Aircraft fault	1
E5	Fuel leakage	1
E6	Pollute taxiway	2
E7	The reason of frontier defense	1

Numbers	Risk factors	Degree
E8	Crew timeout	1
E9	The visual range of runway is lower than its landing standard	1
E10	Pavement icing	1
E11	Flight control system fault	1
E12	Weather radar fault	1
E13	Aircraft technical reasons	1
E14	Front wheel turning fault	1
E15	Engine core de-icing component fault	1
E16	Departure time limit	1
E17	There are approaching planes on final	2
E18	Oil replenishment	1
E19	Abnormal front tire pressure display	1
E20	Hit by a special vehicle	1
F1	No impact on runway	1
F2	The tower asked if there was any hydraulic oil leakage	2
F3	No hydraulic oil leakage	2
F4	Need rescue service	2
F5	Suspension of subsequent take-off activities	2
F6	Further confirm the fault information	1
F7	Apply for glide back	2
F8	Uncertain whether there is any abandoning and scattering objects	2
F9	No service is required	1
F10	Recovery of runway	2
F11	Coordination of relevant airport departments	1
F12	The rear cargo door warning light extinguish	2
F13	RebJoting	2
F13	The rear cargo door is closed	1
F14	The push back light is on at the same time	2
F16	The computer shows that the hydraulic pressure is low	1
F17	Lost GPS signal	1
F18	Rear gate light on	1
F19	Rear cargo door warning light on	1
F20	Rear passenger compartment gate warning light on	1
F21	Fire engine in place	2
F21	Warehouse fire	1
F22	Fire emergency	2
F24	There is no smoke or fire outside the engine room	2
F25	The fire engine taxied behind	1
F26	There is no abnormal phenomenon in the fire report	1
F27	Right engine fault	1
F28	Oil leakage may occur	1
F29	Hatch open	1
F30	Cockpit voice recorder fault	1
F31	Computer fault	1
F32	Configuration alert	1
F33	Take-off configuration alert	1
F34	Front door of engine room not opened	1
F35	Air-brake fault	1
F36	Left side engine fault	1
—	—	—

closer the node is near the network center, the more important is the node.

Go-around (V2) and conflict resolution (C22) are the key risk factors related to the occurrence of flight conflict events at Daxing Airport, and their closeness centrality is high.

The warning light off (F12) of the rear cargo door and interphone card group channel (C7) are both related to equipment failure. The inspection (V19) node is the node with the greatest median centrality, while the equipment failure-related node is closely related to V19.

The description of deviation procedure (C20) in the case is that the pilot did not operate according to the prescribed procedure to cause deviation procedure, which is a human factor. The whole process of flight conflict in actual operation cannot be separated from human behavior.

Suspend of subsequent take-off activities (F5), maintain visual (C26), stop descent (C3), controller verification (C12),
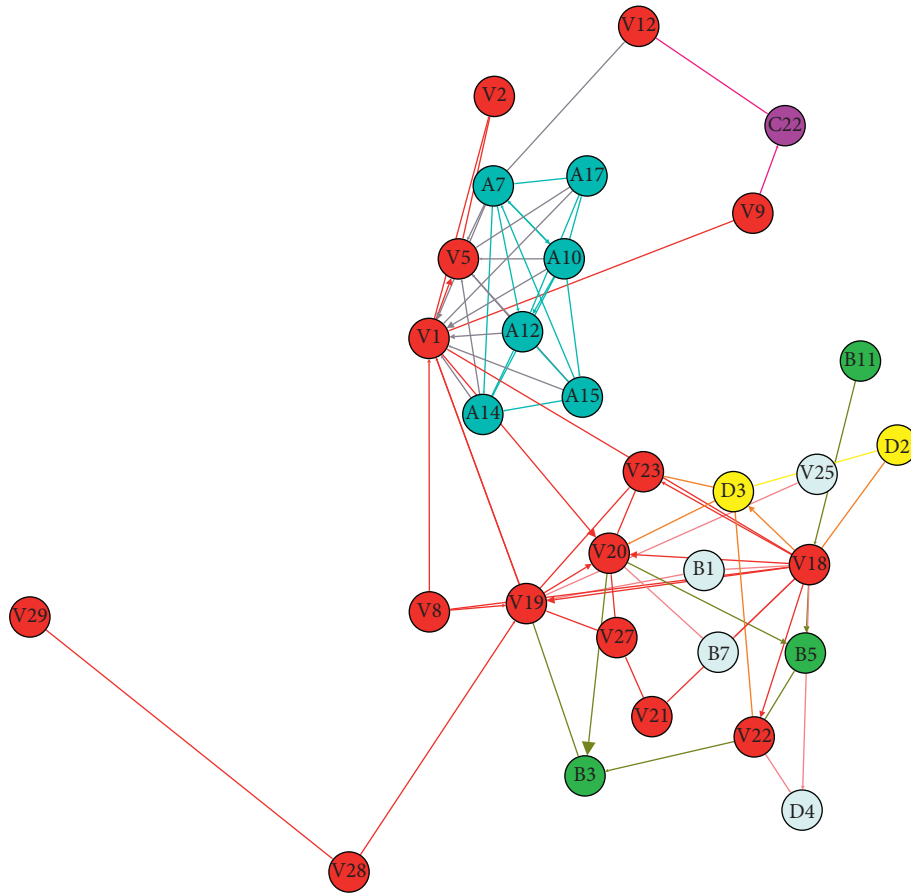
Figure 2: Complex networking graph of flight conflict in the intersecting runway. Note: only nodes with the node degree greater than or equal to 3 are retained in the figure, while nodes with the low connectivity degree are not shown in the figure.

Table 4: Parameters of the network.

Parameter name	Value
Number of nodes	169
Number of edges	263
Network density (directed)	0.009
Network average degree	1.556
Network average weighted degree	20.337
Network diameter	9
Network average clustering coefficient	0.097
Network average path length	3.143

Table 5: Distribution of nodes' degree values.

Number	Nodes	Degree	Degree centrality	In-degree	Out-degree
1	V1	38	0.2262	32	6
2	V18	33	0.0536	15	18
3	V28	25	0.0774	22	3
4	V29	21	0.0417	15	6
5	V19	18	0.0595	7	11
6	V16	14	0.0833	1	13
7	V5	13	0.1964	13	0
8	A7	12	0.1071	4	8
9	V20	11	0.0655	8	3
10	V12	10	0.0357	8	2

Note: only the top 10 nodes are shown in the table.

and controller instruction descent (C16) are all belong to the node that describes the state, not risk factors, which carry the transmission of risk factors.

Since there is no flight accident in the case set in this paper, the closeness centrality of safe landing (V20) is 1.

5.4. The Comprehensive Value of the Node. According to the relevant research results [15], the importance of nodes is determined to be described by comprehensively considering the relevant parameters. In this paper, the comprehensive value of nodes is defined as the average value of the sum of degree centrality, betweenness centrality, and closeness centrality. The importance of each node in the network is described by the comprehensive value, as shown in Table 8.

From the result of the comprehensive value, the case set of unsafe events adopted in this case did not cause serious consequences, so the safe landing (V20) can be eliminated. At this time, the comprehensive value of go-around flight (V2) is the largest, indicating that go-around flight is the most critical factor leading to flight conflict. Conflict relief (C22), rear cargo door warning light off (F12), suspension of subsequent take-off activities (F5), visual maintenance (C26), and other nodes have significant influences on flight conflict, and the conclusions obtained from the analysis are also consistent with the actual situation.

Table 6: Betweenness centrality distribution of nodes in networks.

Number	Nodes	Betweenness centrality
1	V19	0.0493
2	V1	0.0430
3	V28	0.0337
4	V18	0.0262
5	V29	0.0219
6	F21	0.0100
7	F24	0.0072
8	V20	0.0070
9	V24	0.0050
10	F7	0.0045

Note: only the top 10 nodes are shown in the table.

Table 7: Closeness centrality distribution of nodes in networks.

Number	Nodes	Closeness centrality
1	V2	1.0000
2	F5	1.0000
3	C3	1.0000
4	C7	1.0000
5	V20	1.0000
6	C12	1.0000
7	C16	1.0000
8	F12	1.0000
9	C20	1.0000
10	C22	1.0000
11	C26	1.0000
12	C33	1.0000

Table 8: Comprehensive value distribution of nodes in networks.

Number	Nodes	Comprehensive value
1	V20	0.3575
2	V2	0.3512
3	C22	0.3454
4	F12	0.3385
5	F5	0.3375
6	C26	0.3374
7	C3	0.3373
8	C7	0.3373
9	C12	0.3373
10	C16	0.3373

Note: only the top 10 nodes are shown in the table.

5.5. Break Chain Control Strategy. Perform scenario analysis by using a large number of cases, find the key nodes that affect the outcome of the event as well as the correlation and logic among the key nodes, and then find out the prevention and control strategies. In this paper, V2 is the most critical factor, which can effectively prevent the occurrence of flight conflict by controlling it. By controlling the risk factors that lead to resort (V2), it can reduce the occurrence or improve the safety of go-around. For the field of civil aviation, the most effective strategy includes enhancing the capability of small-scale weather forecast in the airport area, improving the conflict resolution ability of flight crews and controllers in the go-around scenario, and developing equipment to provide the capability of conflict prediction in the airport terminal area.

6. Conclusion

(1) In this paper, the operational risk factors of the airport with intersecting runway configuration are identified, and the visual model characteristics and key nodes of flight conflict scenario evolution of civil aircraft in the approach phase are described.

(2) The evolution network model of the flight conflict scenario of civil aircraft in the approach phase is constructed, which includes 169 nodes and 263 edges. By analyzing the parameters of the network, the risk evolution path is given and the effectiveness of the network is verified.

(3) The mechanism of flight conflict scenario evolution and propagation has not been analyzed in detail, which will be the next research focus.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Gao, X. He, and I. Zhan, "Research on the development of multi-runway airport Joperation in my country," *Jiaotong Qiye Guanli*, vol. 28, no. 6, pp. 60-61, 2013.

[2] Civil Aviation Administration of China, *Regulations on Simultaneous Instrument Operation of Parallel Runways Order No. 123*, Civil Aviation Administration of China, Beijing, China, 2004.

[3] Civil Aviation Administration of China, *Civil Aviation Air Traffic Management Rules CCAR-93TM-R5*, Civil Aviation Administration of China, Beijing, China, 2017.

[4] B. Yang, "Reflection on runway intrusion of Hongqiao airport," *Jetliner*, no. 6, pp. 75–79, 2016.

[5] J. Sun, Z. Zhang, and D. Yao, "On the handling of flight conflict from the runway incursion of Shanghai Hongqiao Airport," *Technology Innovation and Application*, vol. 34, no. 264, p. 292, 2016.

[6] K. Mcgarry, "Don't go there: using runway intersection lights to prevent runway incursions," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 52, no. 1, pp. 16–20, 2008.

[7] M. P. ensen, I. S. Wyatt, T. L. Feyereisen, and A. Gannon, *System and Method for Identifying Runway Position During an Intersection Takeoff*, Honeywell International Inc., Charlotte, NC, USA, 2014.

[8] D. G. Gellert, *Safe Runway Aircraft Arrival and Departure System Using Split Runway Design*, USA, 2009.

[9] Y. Yan, D. Zhu, and Q. Zhang, "Research on wake interval of intersecting runway," *China Science and Technology Information*, vol. 7, no. 12, pp. 43–47, 2016.

[10] D. Zhu, X. Zhou, T. Wu, and K. Chen, "Evaluation on intersecting runway capacity of Chengdu new airport based on MACAD model," *Command Information System and Technology*, vol. 7, no. 6, pp. 40–44, 2016.

[11] J. Wang, W. Deng, Z. Xia, and J. Wan, "Flight risk assessment method of transport aviation," *China Safety Science Journal*, vol. 29, no. 12, pp. 114–120, 2019.

[12] Y. Gao, X. Wang, and D. Zheng, "Research on method for predicting risk of aircraft flight conflict in airport terminal," *China Safety Science Journal*, vol. 1, no. 28, pp. 158–162, 2017.

[13] X. Kong, Z. Wu, and C. Yang, "Research on short-term flight conflict decision model," *Science Technology and Industry*, vol. 19, no. 4, pp. 112–116, 2019.

[14] Q. Wen and W. Chen, "The solving strategy of the flight conflict resolution based on the "Two-step method"" *High Technology Letters*, vol. 30, no. 3, pp. 268–273, 2020.

[15] X. Meng, G. Chen, C. Zheng, X. Wu, and G. Zhu, "Risk evaluation model of deepwater drilling blowout accident based on risk entropy and complex network," *CIESC Journal*, vol. 70, no. 1, pp. 388–397, 2019.

[16] Q. Xiao and L. Fan, "Safety risk evolution of amphibious seaplane during takeoff and landing-based on complex network," *Complex Systems and Complexity Science*, vol. 2, no. 3, pp. 20–30, 2019.

[17] W. Du, *Research on Complex Network and Network Dynamics for Air Transportation Systems*, University of Science and Technology of China, Hefei, China, 2010.

[18] J. Du, *Research on Analysis and Control Problems of Airline Network Stability*, Nanjing University of Aeronautics And Astronautics, Nanjing, China, 2015.

[19] Z. Zhou, *Vulnerability Analysis of Civil Aviation Land-Based Air Traffic Control Maintenance System Based on Complex Network Theory*, Civil Aviation Flight University of China, Deyang, China, 2018.

[20] K. Wang, *Research on Traffic Congestion Identification and Control Based on Complex Network*, Civil Aviation University of China, Tianjin, China, 2018.

[21] B. Ran, *Study on Complex Network Model of Air Traffic Control System for PBN Operation*, Civil Aviation Flight University of China, Guanghan, China, 2020.

[22] L. Tanguy, N. Tulechki, A. Urieli, E. Hermann, and C. Raynal, "Natural language processing for aviation safety reports: from classification to interactive analysis," *Computers in Industry*, vol. 78, pp. 80–95, 2016.

[23] M. A. Abedin, V. Ng, and L. Khan, "Cause identification from aviation safety incident reports via weakly supervised semantic lexicon construction," *Journal of Artificial Intelligence Research*, vol. 38, pp. 569–631, 2010.

[24] B. Xu and S. Kumar, "A text mining classification framework and its experiments using aviation datasets," in *Proceedings of the Workshop of Information Technology and Systems*, Dallas, TX, USA, 2015.

[25] S. Kierszbaum and L. Lapasset, "Applying distilled BERT for question answering on ASRS reports," in *Proceedings of the 2020 New Trends in Civil Aviation (NTCA)*, pp. 33–38, Prague, Czech Republic, 2020, November.

[26] T. Liu, "Theories and methods of building major accident disaster scenarios," *Fudan Public Administration Review*, vol. 9, no. 2, pp. 46–55, 2013.

[27] Y. Sheng, "Study on assessment method of emergency preparedness capability based on scenario construction technology," *Journal of Safety Science and Technology*, vol. 10, no. 7, pp. 44–47, 2017.

[28] J. Hui, "Research on scenario expression and relationship between scenarios in real-time emergency decision," *Journal of UESTC (Social Sciences Edition)*, vol. 14, no. 1, pp. 48–52, 2012.

[29] 2003 http://news.qdu.edu.cn/news.Aspx?%20new%20sid=1514.

[30] X. Meng, G. Chen, and H. Zhu, "Complex network analysis on risk evolution of submarine pipeline leakage," *Journal of Safety Science and Technology*, vol. 13, no. 4, pp. 26–31, 2017.

[31] Z. Yao, K. Shang, and X. Xu, "Fundamental statistics of weighted networks," *Journal of University of Shanghai for Science and Technology*, vol. 34, no. 1, pp. 19–26, 2012.

WILEY | Hindawi

Research Article

Urban Rail Transit System Network Reliability Analysis Based on a Coupled Map Lattice Model

Shaojie Wu [ID],[1] **Yan Zhu,**[1] **Ning Li,**[2] **Yizeng Wang** [ID],[3] **Xingju Wang,**[4] **and Daniel Jian Sun** [ID][5]

[1]*State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*Ulanqab Vocational College, Ulanqab 012000, Inner Mongolia, China*
[3]*School of Communication and Information Engineering, Shanghai University, Shanghai, 200444, China*
[4]*Transportation School, Shijiazhuang Tiedao University, Shijiazhuang 050043, Hebei, China*
[5]*Smart City and Intelligent Transportation (SCIT) Center, School of Design, Shanghai Jiao Tong University, Shanghai 200240, China*

Correspondence should be addressed to Daniel Jian Sun; danielsun@sjtu.edu.cn

During the last twenty years, the complex network modeling approach has been introduced to assess the reliability of rail transit networks, in which the dynamic performance involving passenger flows have attracted more attentions during operation stages recently. This paper proposes the passenger-flow-weighted network reliability evaluation indexes, to assess the impact of passenger flows on network reliability. The reliability performances of the rail transit network and passenger-flow-weighted one are analyzed from the perspective of a complex network. The actual passenger flow weight of urban transit network nodes was obtained from the Shanghai Metro public transportation card data, which were used to assess the reliability of the passenger-flow-weighted network. Furthermore, the dynamic model of the Shanghai urban rail transit network was constructed based on the coupled map lattice (CML) model. Then, the processes of cascading failure caused by network nodes under different destructive situations were simulated, to measure the changes of passenger-flow-weighted network reliability during the processes. The results indicate that when the scale of network damage attains 50%, the reliability of the passenger-flow-weighted network approaches zero. Consequently, taking countermeasures during the initial stage of network cascading may effectively prevent the disturbances from spreading in the network. The results of the paper could provide guidelines for operation management, as well as identify the unreliable stations within passenger-flow-weighted networks.

1. Introduction

With the rapid development of urbanization, rail transit lines of megacities have been extended into a network undertaking large-scale urban commuter passengers. Metro networks not only improve the efficiency of the transit system but also expand the risk of fault propagation. Although the topological network of the urban transit system is relatively simple, as one typical social network, the passenger-flow-weighted one is relatively complex [1]. Therefore, from the perspective of a complex network, we establish

the network dynamics model of the rail transit system, to analyze the cascading failure process and the changes of passenger-flow-weighted reliability will assist the metro management agency to improve the capacity of passenger transportation and effectively prevent large-scale burst accidents.

Many studies have been devoted to the cascading failure and passenger-flow-weighted reliability from different perspectives in transportation networks [2, 3]. Previous research mainly concentrated on the network vulnerability, accessibility, and other characteristics, while few studies

focus on the passenger-flow-weighted network reliability and cascading failure process of the subway network. Latora and Marchiori [4] investigated the Boston metro network and verified that the subway network has small world characteristics, from which the concepts of network efficiency and connectivity index were proposed. Scale-free networks behave differently under random and intentional attacks [5, 6]. Random attacks may merely cause the network to be slightly affected, while an intentional attack seriously affects the network which is the beginning of network robustness research. Sun and Guan [7] set up the Shanghai network model and proposed indicators to measure the vulnerability of the network from the line perspective. They found that circular lines usually have the highest value because of the specific topology. Jing et al. [8] proposed the travelers' route redundancy (or route diversity) index, defined as the number of behavioral effective routes between each origin-destination (O-D) pair in the network. The new index incorporates the travelers' route choice, assisting to evaluate the predisaster preparedness of the metro networks from the O-D level to network level.

In terms of rail transit network reliability studies, early scholars mainly applied unweighted indexes to measure the reliability of the rail transit network, such as network efficiency, degree distribution, or average path length. For example, the degree distributions and clustering coefficient indexes are utilized to measure the differences in reliability between scale-free networks and random networks [9]. However, most are weighted networks in reality, and research considering the factors of passenger flow weight has emerged during the last 10–20 years. Li et al. [10] pointed out that link weight is crucial in a weighted complex network, and the obtained results show that the change of weight distribution can cause some significant effects on the subtle structural features and functions of the given networks. It is also found that the heterogeneity and vulnerability of the Beijing subway network vary over time when passenger flow is taken into consideration [11]. From the perspective of passenger ridership, Chen et al. [12] utilized the binomial logit model (BNL) to estimate mode choices and distinguish the relationship between metro and taxi as substitutable, complementary, and extended types. The temporal and spatial characteristics of passenger trips and the unbalanced features of line entry and exit by mining the AFC card data of Shanghai Metro are explored as an application of passenger data [13]. The research evaluated the travel reliability of passengers based on travel time index. For the network cascading failure, the impact of vertex failures on the probability of trip failure in a number of transit topological networks is analyzed by an absorbing Markov chain model [14]. The absorbing Markov chain model also measures the unreliable probability of a missing transit line due to insufficient capacity. The results indicated that the hub and spoke graph is the most reliable type for random vertex failures. For link failure, a normative approach for transportation network reliability based on game theory is proposed to analyze the weak points of the network and the performance in case of link failure [15]. The reliability estimates were divided into two parts, one is the probability to

encounter a failure and the other is the probability to arrive at a destination within an acceptable travel time. The two methods analyze the network reliability from the perspective of probability.

The reliability evaluation methods above have rather good performance compared with traditional methods. However, the change of network reliability during dynamic processes has still not been mentioned. In addition, the network cascading failure behavior is another important issue. This study simulates the processes of cascading failures caused by network nodes based on a coupled map lattice (CML) model and measures the passenger-flow-weighted network reliability during cascading failure processes. The CML model is a common dynamic network simulation method, which describes the continuous changes of the chaotic state. Crucitti et al. [16] simulated the cascading failure of power grid and Internet, utilizing the efficiency index to measure the changes of network function. They found that the breakdown of a single node is sufficient to collapse the efficiency of the entire system if the node is among the ones with the largest loading. Cui et al. [17] applied the CML model to cascading failure of small-world networks and found that a larger mean node degree (referring to the number of edges directly connected with the node within the network) can delay the cascading failure process. Cui et al. [18] modified the original CML model and proposed a sequential cascading failure model with edge disturbance, in which the application of the CML model is extended to edge elements. Zhang et al. [19] proposed an improved CML model to simulate the urban road traffic network of Beijing, in which the cascading failures were tested using different attack strategies.

To sum up, the network cascading failure process has been investigated through different disrupted node selection strategies, measuring the reliability of the network through the change of these properties by network efficiency and invulnerability during the process. However, the passenger-flow-weighted reliability is still less involved, which has a large influence on the rail transit network. In this paper, based on the actual metro passenger flow data, improved reliability indexes involving passenger flow parameter were proposed, to better reflect the field network reliability situations. The remainder of the paper is structured as follows: Section 2 describes the definitions and methodology for the urban metro system network reliability analysis used, in which the dynamic state modeling of metro stations based on coupled map lattice is proposed. An empirical study on network cascading failure process of the Shanghai Metro system is conducted. The cascading failure processes under different conditions are investigated. Finally, conclusions and recommendations are provided in Section 4.

2. Definitions and Methodology

Reliability concept reflects the function of the network during the processes of cascading failure. As a metro network is generally composed of multiple nodes and connected edges between nodes, network reliability can be considered as the sum of the reliability of all nodes.

Therefore, measuring reliability changes need to utilize other function indicators. Network topology has a large impact on reliability, but in the real world, due to the different distribution and routes of passenger flows. Reliability analysis from the perspective of passenger flow weight has more practical significance [20]. We try to combine topological index indicators with passenger flow weight to obtain real metro network reliability.

2.1. Reliability Measure Indexes

2.1.1. Betweenness and Passenger-Flow-Weighted Betweenness. Betweenness refers to the number of shortest paths passing through the node for the shortest paths between all nodes within the network [21], which is often used to evaluate the centrality of nodes in an unweighted topological network. However, the node betweenness only considers the factor of network topology, ignoring the passenger flow factor. However, nodes with large betweenness values in an unweighted network may not be practically significant if incorporating the real passenger flow weight [3]. By combining the betweenness centrality of nodes with the actual passenger flow weight, the passenger-flow-weighted betweenness index of node passenger flow can be obtained with the mathematical expressions as follows:

$$
\begin{aligned}
C_B(v) &= \sum_{s,t \in V} \frac{2\theta(s,t|v)}{(N-1)(N-2)\theta(s,t)}, \\
C_B^W(v) &= W_v C_B(v) = W_v \cdot \sum_{s,t \in V} \frac{2\theta(s,t|v)}{(N-1)(N-2)\theta(s,t)}, \\
W_v &= \frac{N Q_v}{\sum_{i=1}^N Q_i}, \\
C(V) &= \frac{1}{N} \sum_{v=1}^N C_B^W(v),
\end{aligned}
\tag{1}
$$

where $C_B(v)$ denotes the betweenness value of node v after normalization; $C_B^W(v)$ denotes the passenger-flow-weighted betweenness value of node v; N denotes the number of all nodes in the network; $\theta(s,t)$ denotes the total number of shortest paths from nodes s to t; $\theta(s,t|v)$ denotes the number of shortest paths from s to t passing through node v; Q_i denotes the total amount of passengers passing through node i in one day; Q_v denotes the total amount of passengers passing through node v in one day; W_v denotes the ratio of passenger flow at node v and the average node passenger flow; and $C(V)$ denotes the passenger-flow-weighted betweenness value of network V.

2.1.2. Entropy and Passenger-Flow-Weighted Entropy. Entropy is applied for describing the chaotic degree of the system. The higher entropy value is related to the higher chaos degree of the system. A network topology entropy

index as calculated in equation (2) is introduced to measure the uniformity of the unweighted network as follows:

$$
\begin{aligned}
D_v &= \frac{k_v}{\sum_{i=1}^N k_i}, \\
E &= -\sum_{v=1}^N D_v \ln D_v,
\end{aligned}
\tag{2}
$$

where D_v represents the proportion of node degree v to all nodes; k_i represents the degree of node i; N denotes the number of all nodes in the network; and E represents the value of network topology entropy; Furthermore, the node passenger flow intensity was introduced to measure the equilibrium degree of the served passenger flow intensity of network nodes. Then, the entropies of network passenger flow intensity after normalization are as follows:

$$
\begin{aligned}
G_s(V) &= \frac{G - G_{\min}}{G_{\max} - G_{\min}} = \frac{-\sum_{v=1}^N I_v \ln I_v}{\ln N} = -\sum_{v=1}^N I_v \log_N^{I_v}, \\
G &= -\sum_{v=1}^N I_v \ln I_v, \\
I_v &= \frac{Q_v}{\sum_{i=1}^N Q_i},
\end{aligned}
\tag{3}
$$

where $G_{\min} = 0$ denotes the theoretical minimum entropy of the network; $G_{\max} = lnN$ denotes the theoretical maximum entropy of the network; G denotes the passenger-flow-weighted entropy of the network; N denotes the number of all nodes in the network; Q_i denotes the total amount of passengers passing through node i in one day; Q_v denotes the total amount of passengers passing through node v in one day; and I_v denotes the proportion of passenger flow passing through node v to the total passenger flow passing through all nodes.

Thus, based on the concept, the proposed passenger-flow-weighted reliability considers the passenger flow factor by indexes of betweenness and entropy.

2.2. Dynamic State Modeling of Metro Stations Based on CML.
Being widely used in modeling complex dynamic systems, the CML model defines the coupling relationship between adjacent network nodes and describes the spatiotemporal chaotic state changes of the nodes. This means that the CML state value of node is solely determined by the previous time state of the node and the previous time state of the adjacent nodes. The expressions are defined as follows:

$$
X_i(t+1) = \left| (1-\varepsilon) f(X_i(t)) + \varepsilon \sum_{j=1, j \neq i}^N \frac{a_{ij} f(X_j(t))}{k(i)} \right|, \tag{4}
$$

where $X_i(t)$ is the CML state of node i at time t, ε is the coupling coefficient, and the larger coupling coefficient is always related to the higher mutual influence of the nodes. a_{ij} is the corresponding row i and column j element within the

adjacency matrix $A = (a_{ij})_{N \times N}$, reflecting the network connection information, $a_{ij} = 1$ denotes that node i is directly connected to node j; otherwise, $a_{ij} = 0$. $f(x) = 4x(1-x)$ is the chaotic logistic mapping function ($x \in [0,1]$), and $X_i(t+1)$ is obtained from $X_i(t)$ through chaotic mapping function and other mathematical operations. When $0 < X_i(t) < 1$, node i is in the healthy state. During investigating the cascading failure processes of the network system caused by node i, exerting an external disturbance $R \geq 1$ at time step m, the node fails, which can be expressed as follows:

$$ X_i(t+1) = \left| (1-\varepsilon) f(X_i(t)) + \varepsilon \sum_{j=1, j \neq i}^{N} \frac{a_{ij} f(X_j(t))}{k(i)} \right| + R. \tag{5}$$

The failed node i will be removed from the network at the next time step $m+1$. Due to the coupling mechanism of the CML model, the CML state of adjacent nodes of i is affected and the cascading failure of the network is induced.

3. Empirical Study for the Network Cascading Failure Process

3.1. Passenger Flow Data Process and Analyses. As the necessary parameter of weighted reliability indexes, the actual passenger flow is an essential part extracted from the card record provided by Shanghai Public Transport Card Company Limited. We selected the data of August 29, 2016 (Monday). During the day, about 2 million passengers and 5 million trips were obtained after the data process. Considering that most metro passengers tend to choose the shortest routes and high cost of subway transfer time, the Dijkstra shortest path algorithm [22] is applied to assign 5 million trips to the Shanghai network according to the corresponding OD of each passenger. After assignment, the amounts of daily passenger flow of each network node were obtained and used as the parameter of reliability indexes.

3.2. CML State of Network Nodes. Although the states of nodes are decided by the CML model, the dynamic state of different nodes is different due to their respective connection relationship within the network. Therefore, each node has a special CML state value $X_i(t)$ at any time step.

3.2.1. Changes of CML State Values of Network Nodes without Disturbance R. As the CML model describes the relationship between adjacent nodes in the network, any single node is controlled by the model. At the initial time $t = 0$, the state value $X_i(0)$ of all nodes in the network is set to a random value between $(0, 1)$, and the states at subsequent time steps are calculated according to formula (4). If not disturbed, the node state value is always within the range of $(0, 1)$. Taking the Dongchuan Road subway station node of Metro Line 5 as an example, the CML states of each time step of the Dongchuan Road station are affected by the CML state of adjacent Jianchuan Road and Jiangchuan Road stations. The change of the CML state value without R interference is shown in Figure 1.
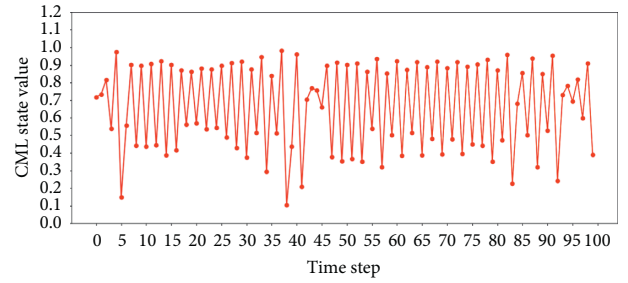
Figure 1: Without disturb R CML state value change of the Dongchuan Road station by time step.

Figure 1 demonstrates that the CML value of the Dongchuan Road Station is in a chaotic state without disturbance R, which means the station is in a healthy state between $(0, 1)$.

3.2.2. Changes of CML State Values of Network Nodes with Disturbance R. By selecting node S randomly and exerting disturbance $R \geq 1$ at time step m, the node will fail and be removed from the network at the next time step, and the node status value is set to 0 at the subsequent time step. Taking the node of the Dongchuan Road Station as an example, by exerting the interference $R = 1$ at the 51st time step, the change of the CML status is as shown in Figure 2.

As presented, after adding disturbance $R = 1$ at time step t, the network node fails and the CML state value of Dongchuan Road, $X(t)$, exceeds the critical value 1. Then, the CML value of the following step $X(t+1)$ was set to 0, and the rest of the steps are followed.

3.3. Cascading Failure Processes and Passenger-Flow-Weighted Reliability Analysis of the Shanghai Network. Under the conditions of different node degrees, coupling coefficients ε, and disturbances R, the cascading failure process of the rail transit network system behaves differently, and the corresponding network reliability state changes are also different. A small R disturbance cannot lead to cascading failure of the network, while the large coupling coefficient ε leads to too fast propagation speed, and the node degree mainly describes the impact of node importance on network cascading failure.

3.3.1. Influence of the Node Degree on Passenger-Flow-Weighted Reliability. The higher degree value of the node generally indicates higher importance within the network, which tends to be located in the central area of the network. To investigate the influences of nodes with different degrees on the cascading failure process of network, four stations, Pudong International Airport (degree value = 1), Jianchuan Road (degree = 2), Lancun Road (degree = 3), and South Xizang Road (degree = 4), were selected randomly. With the coupling coefficient $\varepsilon = 0.2$, by exerting the same disturbance $R = 4$, the cascading failure process caused by network nodes is as shown in Figure 3.

Figure 2: CML state value changes of the Dongchuan Road station with time (disturb $R = 1$).

Figure 3: Network cascading failure processes caused by stations of different degrees ($R = 4$, $\varepsilon = 0.2$).

Figure 3 compares the cascading failure processes of the metro network caused by nodes with different degrees, while Figures 4(a) and 4(b) measure the changes of passenger-flow-weighted network betweenness and entropy during the processes of different cascading failures. Under t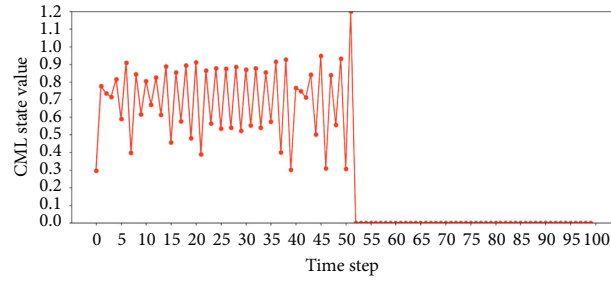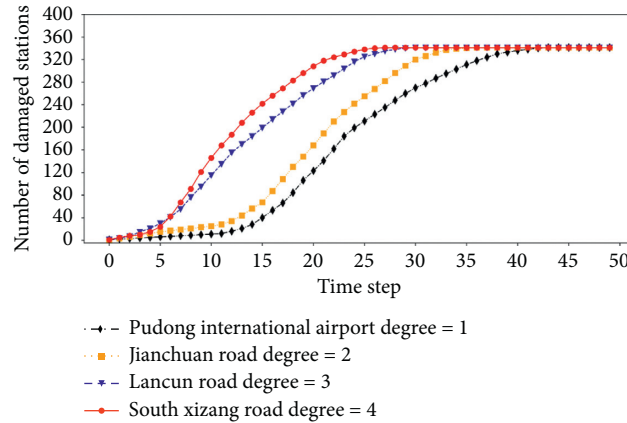he same disturbance R and coupling coefficient ε, nodes with larger degrees tend to cause cascading failure processes faster than nodes with smaller degrees.

When time step $t \leq 5$, the cascading failure processes are in the initial stage, and the disturbance R can only propagate to the adjacent nodes of the failure node in a relatively slow speed. During the initial stage, the passenger-flow-weighted betweenness and entropy of the network are slightly affected, which indicates that the overall function of the network was not affected. When $t > 5$, the speed of cascading failure begins to accelerate because the disturbance R spreads to the center area of the network and other lines through the transfer nodes. Eventually, all nodes were infected and failed, and the passenger-weighted betweenness and entropy indexes decrease to 0.

At the acceleration stage, the passenger-flow-weighted entropy declines constantly, indicating that the passenger flow chaos is declining. Different from the passenger-flow-weighted entropy, the weighted betweenness rapidly decreases to 0. As the disturbance R spread to nodes in the central area of the network, which undertakes large amounts of passenger flow, the connectivity function of the network is seriously affected. It was found that the node degree value mainly affects the speed of cascading failure processes.

3.3.2. Influence of Disturbance R on Passenger-Flow-Weighted Reliability.
Disturbance R mainly refers to the size of node failure, during which the network system shows different situations from reliability to collapse under different R values. The network will not cause cascading failure if R value is small, but when R value is large, the network will collapse continuously. To highlight the influence of different R values on the network cascading failure processes, the node with the largest degree value is selected. Consequently, the Century Avenue Station (degree = 6) is selected and exerts different disturbance values ($R = 2.6$ to 4.4), and the network cascading failure processes is shown in Figure 5.

Figure 5 describes the range and speed of cascading failure processes caused by the Century Avenue Station under different disturbances ($R = 2.6$–4.4). The disturbance R factor mainly affects the scale of the network. Under the same coupling coefficient ε, nodes with larger disturbance R tend to cause larger scale cascading failures. When $R \leq 3.6$, it is not enough to cause consecutive network collapse, only part nodes fail. However, when $R > 3.6$, diagrams of the cascaded failure processes of different R values merely coincide to each other, which indicates that the speed of network cascading failure attains the maximum. Nodes affected by disturbance R at each time step are all infected and become ineffectual.

Figures 6(a) and 6(b) present the changes of network passenger-flow-weighted betweenness and passenger-flow-weighted entropy reliability during the process of cascading failure, respectively. During the process of cascading failure

(a)

(b)
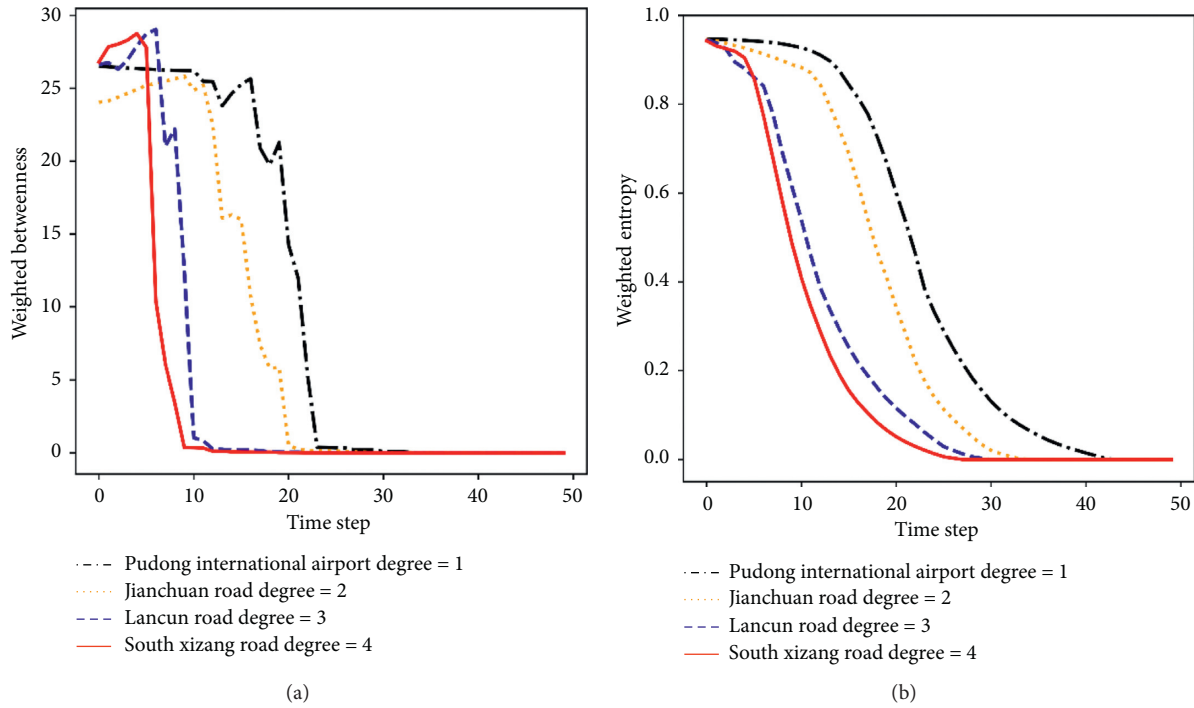
FIGURE 4: Changes of passenger-flow-weighted network reliability during cascading failure processes caused by stations of different degrees (a) for weighted betweenness and (b) for weighted entropy.
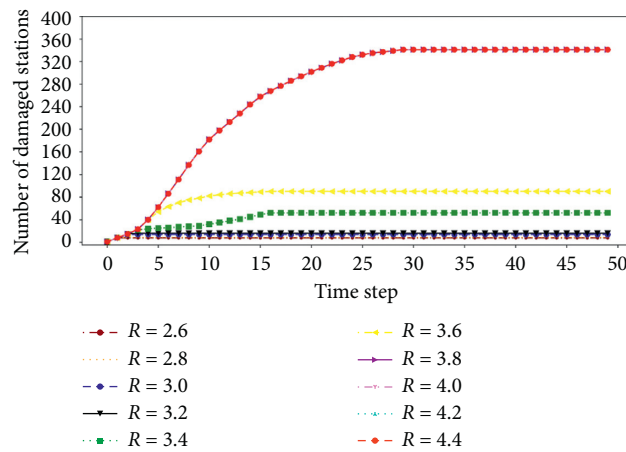
FIGURE 5: Network cascading failure processes caused by the Century Avenue Station under different R values of disturbance with $\varepsilon = 0.2$.

caused by different R values, the reliability of passenger-flow-weighted entropy decreases continuously. However, the passenger-flow-weighted betweenness declines rapidly. For the Century Avenue Station, disturbance $R = 3.6$ is the critical value. Once exceeding the critical value, the scale of network cascading failure attains the maximum. In general, R disturbance mainly affects the scale of network cascading failure, and the network has another critical value leading to the disturbance spread.

3.3.3. Influence of Coupling Coefficient ε on Passenger-Flow-Weighted Reliability. The coupling coefficient ε measures the strength of the interaction between adjacent

nodes. In general, the value of ε affects the disturbance spread ability, in which a large ε indicates that the interaction between nodes is strong and the disturbance spread ability is strong. Conversely, a smaller ε indicates weak interaction and disturbance spread ability. By randomly selecting the Pudong Avenue Station with a degree value of 2 and exerting $R = 4$ under different values of ε from 0.1 to 0.7, the cascading failure process and the changes of reliability are as shown in Figure 7.

Figure 7 describes the range and speed of the cascading failure process of the Pudong Avenue node under different ε coupling coefficients with $R = 4$. When $\varepsilon \leqq 0.2$, merely a part of the network nodes fail, which will not lead to continuous network collapse. When $\varepsilon > 0.3$, the cascading failure

(a)

(b)
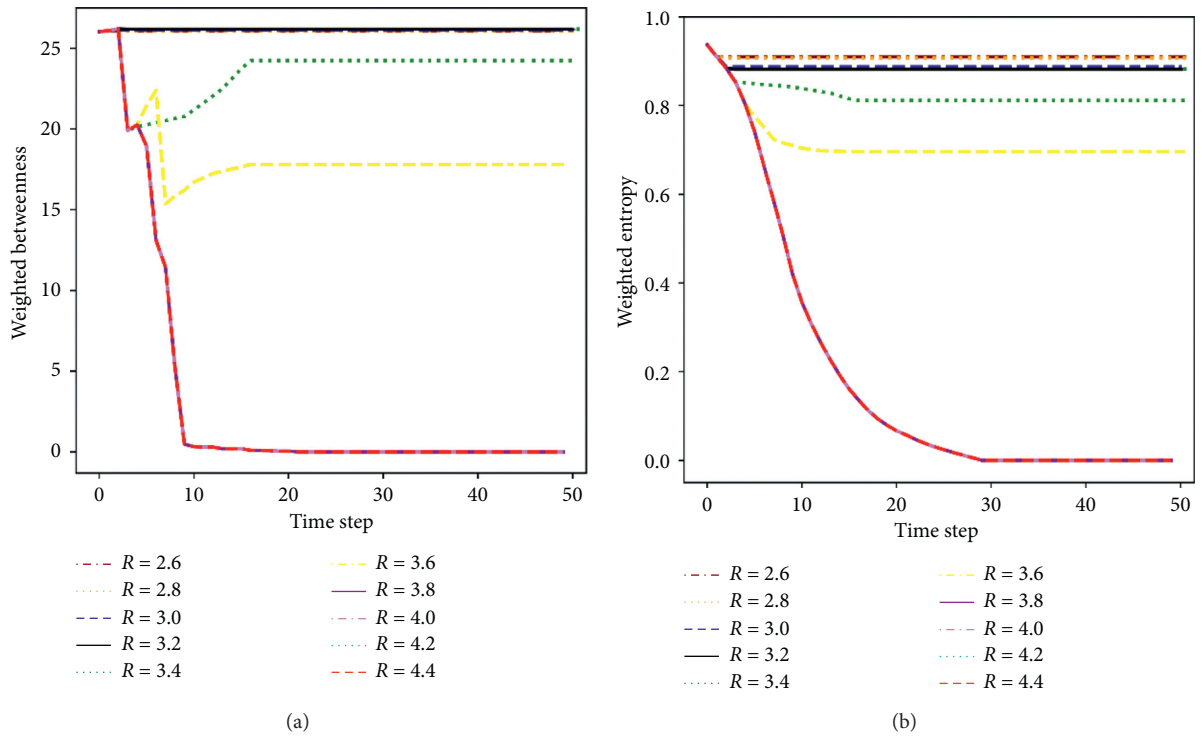
Figure 6: Changes of passenger-flow-weighted network reliability during cascading failure caused by the Century Avenue Station (a) for weighted betweenness and (b) for weighted entropy.
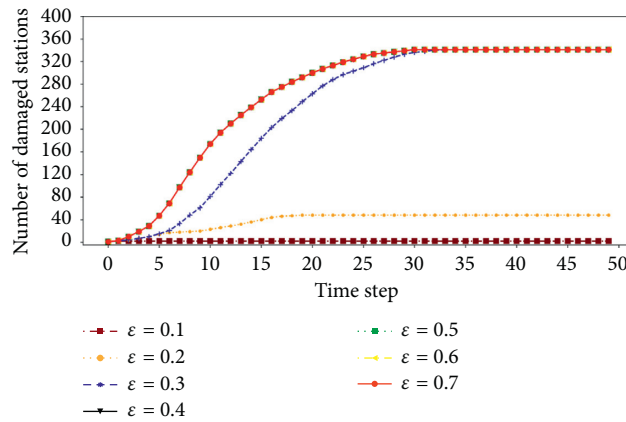
Figure 7: Network cascading failure processes caused by the Pudong Avenue Station under different ε values with $R = 4$.

diagrams in Figure 7 coincide with each other, indicating the network cascading failure process attains the maximum speed. When $\varepsilon = 0.3$, all networks are damaged, but the damage speed is different from the conditions of $\varepsilon > 0.3$.

Figures 8(a) and 8(b) present the changes of network passenger-flow-weighted betweenness and passenger-flow-weighted entropy reliability during the process of cascading failure, respectively. The passenger-flow-weighted betweenness decreases faster compared to the passenger-flow-weighted entropy, which demonstrates the overall connectivity of the network decreases faster.

In general, the coupling coefficient ε affects the scale and speed of network cascading failure by affecting the size of R in the disturbance spread processes.
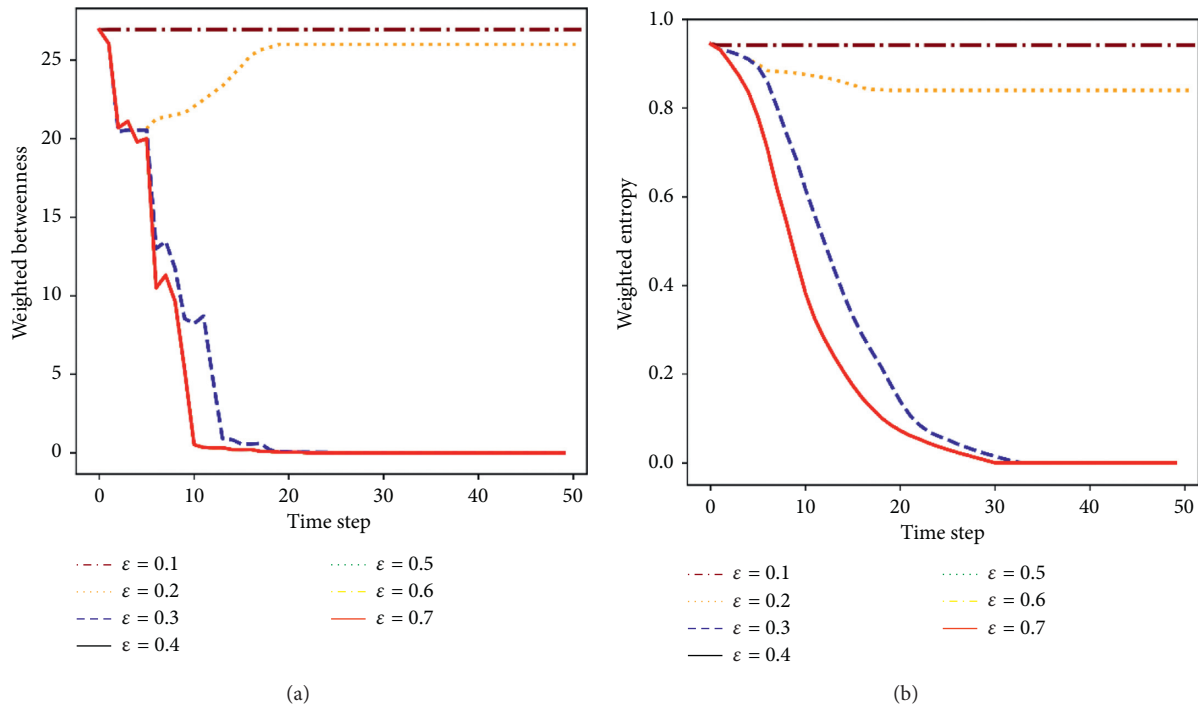
FIGURE 8: Changes of passenger-flow-weighted network reliability during cascading failure caused by the Pu-Dong Avenue Station under different ε values (a) for weighted betweenness and (b) for weighted entropy.

4. Conclusions and Recommendations

Although urban rail transit network reliability theory has been developed for many years, the passenger-flow-weighted reliability of the metro network system are relatively unexplored. The passenger-flow-weighted reliability index is essential for metro management to identify the unreliable nodes in the network. The cascading processes analysis also provides a new perspective for the planning of metro networks. This study assumed that passengers in an urban metro travel according to the shortest path, which is close to the real travel habits of the passenger under large-scale data. The main contributions of the paper are concluded as follows:

(i) Constructing the passenger-flow-weighted reliability evaluation index of the network: by utilizing the Shanghai public transport card data, the field actual passenger flow weight was taken as the influencing factor, considering the impact of the real passenger flow weight on the network reliability. Eventually, the passenger flow was proved to have a significant impact on the network reliability.

(ii) Establishing the dynamic model of the Shanghai Metro network based on CML model, thus to investigate the influence of different parameters on the cascading failure processes: the node degree and disturbance R, respectively, affect the speed and scale of cascading failure processes, in which the coupling coefficient affects both the scale and speed of cascading failure processes.

(iii) Measuring the changes of passenger-flow-weighted reliability of the Shanghai Metro network during cascading failure processes: the passenger-flow-weighted betweenness index decreases faster than the entropy index, which reaches zero when the scale of cascading failure is more than half, indicating rapid decline of the network connectivity.

Data Availability

The data will be made available upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

The authors confirm contribution to the paper as follows: Shaojie Wu, Yan Zhu, Ning Li, and Xingju Wang were involved in study conception and design; Yan Zhu, Yizeng Wang, and Daniel (Jian) Sun collected data; Yizeng Wang, Ning Li, and Xingju Wang carried out model simulation; Yan Zhu, Ning Li, and Daniel (Jian) Sun analysed and interpreted the simulation results; and Shaojie Wu, Yan Zhu, and Daniel (Jian) Sun prepared the draft manuscript. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

presented in the Transportation Research Board 100th Annual Meeting. The research was funded in part by the National Nature Science Foundation of China (71971138), the Shanghai "Science and Technology Innovation Action Plan"—"One Belt and One Road" International Cooperation Project (19210745600), and the Open Subject of Traffic Safety and Control Lab in Hebei Province (JTKY2019001). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] X.-Y. Ni and D. J. Sun, "Agent-based modelling and simulation to assess the impact of parking reservation system," *Journal of Advanced Transportation*, vol. 2017, Article ID 2576094, 10 pages, 2017.

[2] X. Ding, S. Guan, D. J. Sun, and L. Jia, "Short turning pattern for relieving metro congestion during peak hours: the substance coherence of Shanghai, China," *European Transport Research Review*, vol. 10, no. 2, pp. 1–11, 2018.

[3] G. Nian, F. Chen, Z. Li, Y. Zhu, and D. Sun, "Evaluating the alignment of new metro line considering network vulnerability with passenger ridership," *Transportmetrica A: Transport Science*, vol. 15, no. 2, pp. 1402–1418, 2019.

[4] V. Latora and M. Marchiori, "Is the Boston subway a small-world network?" *Physica A: Statistical Mechanics and Its Applications*, vol. 314, no. 1-4, pp. 109–113, 2002.

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

[6] P. Malighetti, G. Martini, R. Redondi, and D. Scotti, "Integrators' air transport networks in Europe," *Networks and Spatial Economics*, vol. 19, no. 2, pp. 557–581, 2019.

[7] D. Sun and S. Guan, "Measuring vulnerability of urban metro network from line operation perspective," *Transportation Research Part A: Policy and Practice*, vol. 94, pp. 348–359, 2016.

[8] W. Jing, X. Xu, and Y. Pu, "Route redundancy-based network topology measure of metro networks," *Journal of Advanced Transportation*, vol. 2019, Article ID 4576961, 12 pages, 2019.

[9] D. Sun, Y. Zhao, and Q.-C. Lu, "Vulnerability analysis of urban rail transit networks: a case study of Shanghai, China," *Sustainability*, vol. 7, no. 6, pp. 6919–6936, 2015.

[10] M. Li, Y. Fan, D. Wang et al., "Effects of weight on structure and dynamics in complex networks," arXiv preprint cond-mat/0601495, 2006.

[11] X.-M. Xiao, L.-M. Jia, and Y.-H. Wang, "Correlation between heterogeneity and vulnerability of subway networks based on passenger flow," *Journal of Rail Transport Planning & Management*, vol. 8, no. 2, pp. 145–157, 2018.

[12] F. Chen, Z. Yin, Y. Ye, and D. Sun, "Taxi hailing choice behavior and economic benefit analysis of emission reduction based on multi-mode travel big data," *Transport Policy*, vol. 97, pp. 73–84, 2020.

[13] Y. Sun, J. Shi, and P. M. Schonfeld, "Identifying passenger flow characteristics and evaluating travel time reliability by visualizing AFC data: a case study of shanghai metro," *Public Transport*, vol. 8, no. 3, pp. 1–23, 2016.

[14] M. G. H. Bell, J.-D. Schmoecker, Y. Iida, and W. H. K. Lam, "Transit network reliability: an application of absorbing Markov chains. Transportation & Traffic Theory in 21st Century," in *Proceedings of the 15th International Symposium on Transportation and Traffic Theory*, pp. 43–62, Adelaide, Australia, June 2002.

[15] C. Cassir, M. Bell, and J. Schmocker, "A normative assessment of transport network reliability based on game theory. Network Reliability of Transport," in *Proceedings of the 1st International Symposium on Transportation Network Reliability (INSTR)*, pp. 225–243, Kyoto, Japan, May 2003.

[16] P. Crucitti, V. Latora, and M. Marchiori, "Model for cascading failures in complex networks," *Physical Review E: Statistical, Nonlinear, & Soft Matter Physics*, vol. 69, no. 4 Pt 2, 2003.

[17] D. Cui, Z. Gao, and X. Zhao, "Cascades in small-world modular networks with cml's method," *Modern Physics Letters B*, vol. 21, no. 30, pp. 2055–2062, 2007.

[18] D. Cui, Z. Gao, and J.-F. Zheng, "Tolerance of edge cascades with coupled map lattices methods," *Chinese Physics B*, vol. 18, no. 3, pp. 992–996, 2009.

[19] Y. Zhang, Y. Lu, G. Lu, P. Chen, and C. Ding, "Analysis of road traffic network cascade failures with coupled map lattice method," *Mathematical Problems in Engineering*, vol. 2015, Article ID 101059, 8 pages, 2015.

[20] Z. Zhang, L. Jia, Y. Qin, and T. Yun, "Optimization-based feedback control of passenger flow in subway stations for improving level of service," *Transportation Letters*, vol. 11, no. 8, pp. 413–424, 2019.

[21] G. Chen, Z. Y. Dong, D. J. Hill, G. H. Zhang, and K. Q. Hua, "Attack structural vulnerability of power grids: a hybrid approach based on complex networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, no. 3, pp. 595–603, 2010.

[22] H. Wang, Y. Yu, and Q. Yuan, "Application of Dijkstra algorithm in robot path-planning," in *Proceedings of the 2nd IEEE International Conference on Mechanic Automation & Control Engineering*, Hohhot, Inner Mongolia, China, July 2011.

WILEY | Hindawi

Research Article
Analysing the Vulnerability of Public Transport Networks

Mary Luz Mouronte-López (ID)

Higher Polytechnic School of Universidad Francisco de Vitoria, Pozuelo de Alarcón, Madrid 28223, Spain

Correspondence should be addressed to Mary Luz Mouronte-López; maryluz.mouronte@ufv.es

This paper analyses the robustness of specific public transport networks. Common attributes and which of them have more influence on the networks' vulnerability are established. Initially, the structural properties of the networks in two graphical representations (L-Space and P-Space) are checked. Afterwards, the spread of problems (traffic jams, etc.) are simulated, employing a model based on a propagation and recovery mechanism, similar to those used in the epidemiological processes. Next, the size of the largest connected subset of stops of the network (giant component) is measured. What is shown is that the faults randomly happened at stops or links, also displaying that those that occurred in the highest weighted links spread slower than others. These others appear at stops with the largest level of betweenness, degree, or eigenvector centralities and PageRank. The modification of the giant component, when several stops and links are removed, proves that the removal of stops with the highest interactive betweenness, PageRank, and degree centralities has the most significant influence on the network's integrity. Some equivalences in the degree, betweenness, PageRank, and eigenvector centrality parameters have been found. All networks show high modularity with values of index Q close to 1. The networks with the highest assortativity and lowest average number of stops are the ones which a passenger can use to travel directly to their destination, without any change. The Molloy–Reed parameter is higher than 2 in all networks, demonstrating that high integrity exists in them. All stops were characterized by low k-cores ≤ 3.

1. Introduction

The question of vulnerability of a public transport network (PTN) against problems in nodes or links is a matter of high interest. All nodes and links may accidentally fail with a similar probability, but the severity of the damage caused is higher when it occurs in some nodes or links rather than in others. The PTN is a key issue in the main cities of the world, having to operate efficiently on a constant basis. They must display a high tolerance to problems, keep the possible delays to a minimum, and manage the duration of a journey in accordance with the trip planning. With the purpose of persuading passengers to make use of the PTN, some strategies should be implemented to ensure the continued running of networks, the prevention of problems before, and the mitigation of losses during and after the failure [1, 2].

This paper analyses the vulnerability of some PTNs. The study firstly focuses on the analysis of the structural properties of the PTN (betweenness centrality, degree, eigenvector centrality, PageRank, average path length, k-core, modularity, and assortativity). Next, the fault propagation is simulated (e.g., traffic jams caused by accidents or roadworks), employing different selection criteria of nodes and links, in order to detect the stops in whose immediate vicinity congestion is more prone to happening. A procedure based on epidemic spread algorithms is used. The effect of the structure of the network on the failure propagation is also examined. Finally, the size of the largest connected subset of stops of the PTN (giant component, GC) is estimated, which is calculated when several stops and links are disconnected according to certain criteria. Depending on the value of this magnitude, a serious issue could occur due to the existence of inaccessible routes (construction work or redesign activities in networks), which cause unavoidable stops or important delays during a trip. The analysis allows the identification of the most critical stops and links of a PTN and establishes a subnetwork of highly connected stops.

Ten networks are analysed. They are Auckland in New Zealand, which consists of 5223 stops and 318,896 links; EMT in Madrid, Spain, which has 4636 stops and 2,070,508 links; BC Transit in Vancouver, Canada, which consists of 3981 stops and 702,197 links; Kolumbus in Rogaland, Norway, which has 3828 stops and 487,432 links; AVL, CFL, RGTR, and TICE in Luxembourg which contains 1372 nodes and 340,684 links; STAR in Rennes, France, which consists of 1415 stops and 9,477,213 links; Thunder Bay Transit, in Ontario, Canada, which contains 825 nodes and 78,247 links; TransAntofagasta, in Chile, which has 650 nodes and 724,362 links; Linja-Karjala Oy, in Kuopio, Finland, which contains 551 nodes and 63,339 links; and finally, CIT Chambly-Richelieu-Carignan, in Quebec, Canada, which consists of 346 nodes and 9366 links. The reason for the selection was to have networks of various sizes (small, medium, and large) and topologies.

Specifically, the goals of this research are as follows:

(1) Analysis of the robustness of some PTNs: detection of commonalities and key aspects.

(2) Discovery of properties on a network's vulnerability that can be generalized.

Several investigations exist about PTN, some of them describing a comprehensive survey of statistical properties of PTN based on the data of several cities, and at the same time, presenting models that reproduce the best part of their properties [3]. Other research detects the critical nodes in PTN in order to enable contingency mechanics in them.

Berche et al. [4] studied the PTN's resilience applying different attack scenarios, which range from random failure to targeted destruction. Some nodes are removed according to certain operating characteristics. The candidate nodes to be eliminated were selected according to their highest degree, closeness, graph, stress, and betweenness centralities, as well as the largest clustering coefficient and next nearest neighbouring number. In [5], an explanation exists of a procedure for the detection of critical airports in the global air transport network (ATN), which is based on simulating an attack on specific airports using several adaptive selection criteria. Wandelt et al. [6] evaluated several strategies for dismantling of networks and identified large heterogeneities in their performance. The authors show that the use of the interactive betweenness creates far stronger attacks. Because of the high complexity of this algorithm, the authors recommend its use on small networks. After, in [7], the authors propose approximations to interactive betweenness in order to improve the computational cost of the algorithm. Other research analyses the effect that the addition of new links has on the robustness of the networks or design methodologies in order to detect the most vulnerable links in PTN [8].

The robustness of a network can be estimated in terms of passenger welfare [9]. Additionally, the accessibility of the PTN is evaluated by several types/examples of research. Albacete et al. [10] compared and applied two location-based methods to analyse PTN accessibility. The results had considerable implications for transport policy making. In [11], a thorough assessment is carried out, focusing on the possible applications of an optimization-based scheme for the building of several timetables and proactive railway traffic management. This is carried out over a large network, employing stochastic disturbances. Wandelt et al. [12] proposed a framework to evaluate and improve the robustness of transportation systems by exploiting the existence of communities. The method is applied to several real-world transportation systems. Sun et al. [13] studied the resilience of cities when disruptions in their airports happen.

Relationships between robustness and accessibility are also studied [14, 15]. Other authors suggested methods and spread models which analyse the performance of networks. He et al. [16] defined a mathematical model to evaluate the effect of the propagation of the risk of failure in multimodal transport networks. Baspinar and Koyuncu [17] proposed a new model for the air transportation network under stress and defined parameters in order to describe not only air sectors, but also airports, as well as flights. They used epidemic spreading processes by assuming that the characteristics of disease spreading and the delay propagation are similar. Akdere et al. [18] studied the reliable data dissemination in the context of wireless sensor network environments. It shows the applicability of epidemic spread algorithms in those environments and carries out a comparative performance analysis of several mechanisms in terms of message delivery rate, average message latency, and messaging overhead on the network. As other research suggests, Nekovee [19] presented a new epidemic algorithm for information dissemination in highly dynamic and intermittently connected vehicular ad hoc networks (VANET). It shows through realistic simulations of highway traffic that the proposed algorithm is suitable to achieve a reliable and efficient level of information transmission in a context with frequent network fragmentation and large density variations.

This paper is organized as follows: firstly, resources and methods used in this work are explained in detail. It focuses on the used data and the developed software programs, which correspond with the design of the network and the employed parameters to characterize the networks' topology. The propagation model and the study of evolution of the GC when certain elements are removed are also studied. Afterwards, the results, conclusions, and future projects are described and discussed.

2. Materials and Methods

2.1. Overview of Used Resources. Information of stops and routes of Auckland, EMT, BC Transit, Kolumbus, STAR, AVL, CFL, RGTR, TICE, Thunder Bay Transit, Trans-Antofagasta, Linja-Karjala Oy, and CIT networks was retrieved from public sources (see the Data Availability section for details). Several programs in R [20] and Python [21] were implemented, using the R.3.6.0 and 3.8.3 version, respectively. We use the Python networkx and pathpy packages and the igraph R package. The programs followed the typical development life cycle with stages of specification, design, and testing.

2.2. Overview of Used Methods. The structural properties of the networks are analysed. A new propagation algorithm, which was designed and undertaken in the networks in order to simulate the spread of a failure, is detailed. This is followed by a description of how the evolution of the GC is calculated, with the aim of studying the specific impact a problem might have on each network.

2.3. Study of Structural Properties. A PTN can be represented in two topological spaces L-Space [22, 23] and P-Space [2, 24]. In both spaces, the network is mapped as a graph $G = (N; L)$, where N is the set of nodes symbolizing the stops and L is the set of links established between them. In L-Space, one node symbolizes one stop, and one link means a union between two consecutive stops, which tells us there is a link between two stops, if one stop is the successor of the other on a route. This space aims to show the geographical proximity between stops. In the P-Space [2, 24], one node represents one stop, and one link joins a pair of stops, if at least one route provides direct service between them. A link means that passengers can take at least one route for a direct trip between two stops. If travellers have to exchange routes, then the pair of stops is joined by more than one link. This space aims to show the transfer relationship between routes.

In L-Space and P-Space, the average values of betweenness (<BC>) and eigenvector (<EC>) centralities, the PageRank <PR>, and degree (<k>) of nodes as well as the minimum distance between nodes distributions were estimated. Pagerank was measured considering a damping factor equal to 0.85. In order to determine if the degree distributions, similar in a way to what happens in other networks [3, 25], followed a power law function, this characteristic was also analysed. Next, all these magnitudes are defined:

(i) The degree of a node i, $k(i)$, for an undirected graph, G, such as a PTN, is [26]

$$k(i) = \sum_{j=1}^{N} A_{ij},\qquad(1)$$

where A_{ij} is the element ij of the adjacency matrix, for example, $A_{ij} = 1$ if the node i is linked to node j and 0 otherwise.

(ii) The minimum distance between two nodes i, j in G, l, is the length of the shortest path between them.

(iii) The betweenness centrality of a node i in G, BC(i) is [27]

$$BC(i) = \sum_{u \neq i \neq w} \frac{\sigma_{u,w}(i)}{\sigma_{u,w}},\qquad(2)$$

where $\sigma_{u,w}$ is the total number of shortest paths from node u to node w and $\sigma_{u,w}(i)$ is the number of those paths that pass through i.

(iv) Regarding the eigenvector centrality of a node i in G, EC(i) [28, 29]:

$\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_N$ are the eigenvalues of the adjacency matrix $A = \{A_{ij}\}$ of G. Then, the largest eigenvalue of matrix A is $\lambda_{\max}$ with an eigenvector $e = [e_1, e_2, \ldots, e_N]T$ such that $\lambda_{\max} * e_i = \sum_{j=1}^{N} A_{ij} * e_j$. The eigenvector centrality for node i denoted as EC(i) can be defined as

$$EC(i) = \frac{1}{\lambda_{\max}} \sum_{j=1}^{N} A_{ij} * e_j.\qquad(3)$$

(v) Pagerank, PR, of a node i in G is [28, 30]

$$PR(i) = \frac{q}{N} + (1 - q) \sum_{j:j \longrightarrow i} \frac{PR(j)}{k_{out}(j)}, \quad i = 1, 2, 3, \ldots, N,\qquad(4)$$

where N is the number of nodes in G, PR(j) is the PageRank of a node j, and $k_{out}(j)$ is the outdegree of node j. We add $(PR(j)/k_{out}(j))$ for all nodes with a link ending in i. In the case of the PTN, it is considered that G is an undirected graph; therefore, $k_{out}(j) = k(j)$. q is the damping parameter, $\in [0, 1]$.

Regarding the modularity of the network, in L-Space, this was calculated employing the generate overlapping cluster generator (OCG) method [31]. The clusters were originally hierarchically joined together, optimizing the modularity of the partition, resulting in overlapping clusters. Similarly, to what happens in the protein networks, the uniqueness of node categorization prevents from disclosing the implication of some stops in various trip processes. In L-Space, the assortativity parameter [32] which determines the tendency of nodes being connected to other similar ones was also estimated. Additionally, the Molloy–Reed parameter (MRP) [33] was calculated to know the structural integrity of the networks.

2.4. Simulation of Faults Propagation. In the L-Space, the propagation of failures in each network is simulated using an infection and recovery mechanism, which consists of the following steps:

Calculate the total number of nodes in the network (N),

Initialize the infection rate $(IR = 0)$,

Infect the first node,

Calculate the total number of infected nodes in the network (μ'),

Initialize time $(t = 1)$,

while $(IR < 0.60)$ and $(t < T)$:

{

Obtain all uninfected nodes in the network, $\{uin_i\}_{i=1}^{i=\mu}$, μ being the total number of uninfected nodes.

For $i = 1$ to μ

{

Obtain the neighbour nodes of $uin_i\{nnui$ $n_{i-j}\}_{j=1}^{j=\gamma_i}$, γ_i being the total number of neighbours of uin_i

For $j = 1$ to γ_i:

{

If RandomBinomial $(\mu', \quad pI) > 0$ $nnuin_{i-j}$ is still infected

Else

To disinfect $nnuin_{i-j}$

}

}

Calculate how many nodes are infected (δ),

For $i = 1$ to δ:

{

If RandomBinomial then $(1, pR) > 0$ then

To disinfect in_i, pR being the recuperation probability,

Else

To infect in_i

}

Calculate the total number of infected nodes in the network (μ')

Calculate IR as:

IR $= \mu'/N$

}

where RandomBinomial (n, p) is a generator of random numbers, which are calculated from the binomial distribution specified by the number of trials n and the probability of success for each trial p. The np.random.binomial function in python is used.

Several simulations were carried out until a time T and without reaching 60% percent of the total of infected nodes in each transport network. pI and pR were modified in the range 0 to 0.1 in steps of 0.0025 (values were considered within a range of 0-0.1, because with $pI > 0.1$, all networks very quickly reached 60% infection (they were saturated in a short time for all groups). It was checked if some differences existed when the propagation process started in nodes with specific characteristics, such as nodes with the highest betweenness (B) and eigenvector centralities (E), nodes with the largest PageRank (P) and degree (D), and nodes randomly selected (A). Pagerank was estimated considering a damping factor equal to 0.85. Once the simulation was concluded, the existing differences between groups (B, E, P, D, A) were analysed. The following procedure was carried out:

(1) The normality of distributions was tested using the D'Agostino test [34] with a significance level $\alpha = 0.05$. This test provides very effective results when it is applied to large size samples [34, 35]. The considered hypothesis was as follows:

(i) H_0: "the samples came from a normal distribution."
(ii) H_a: "the samples did not come from a normal distribution."
    If p value $\leq \alpha$, H_0 is rejected and H_a is taken, else H_0 is accepted.

(2) If normality existed, the homoscedasticity of variances would be studied in each distribution using the Breusch–Pagan test [36] with a significance level of $\alpha = 0.05$. The following hypothesis should be taken into account:

(i) H_0: "the variance was constant."
(ii) H_a: "the variance was not constant."

Again, if p value $\leq \alpha$, H_0 should be rejected and H_a should be accepted; otherwise, H_a should be taken.

(3) If there was no normality or homoscedasticity in the distributions, the Kruskal–Wallis test [37] should be executed in order to detect whether the population distributions were identical, or at least one differed from the rest. The used significance level was $\alpha = 0.05$. The hypothesis employed was as follows:

(i) Null hypothesis: "the groups were from identical populations."
(ii) Alternative hypothesis: "at least one of the groups comes from a different population than the others."

(4) In the next step, the Wilcoxon rank sum test [37] was executed with a significance level $\alpha = 0.05$ in order to realize a pairwise comparison between groups. The hypothesis was as follows:

(i) Null hypothesis: "both groups had the same distribution."
(ii) Alternative hypothesis: "both groups have a different distribution."

(5) If there are normality and homoscedasticity, the analysis of variance (ANOVA) method was realized.

2.5. Analysis of Evolution of the GC. The size of the GC describes the largest fraction of overall nodes such that any pair of them is linked through a path. It evaluates the largest extension that a route (or a union of them) provides in terms of the available stops a traveller can reach from an origin inside the GC.

In the L-Space, for each network, the size of GC was calculated as nodes and links were removed according to the following criteria: links (criterion 1) or nodes (criterion 2) randomly eliminated; nodes removed in descending order of betweenness centrality (criterion 3), degree (criterion 4), eigenvector centrality (criterion 5), and PageRank (considering a damping factor equal to 0.85) (criterion 6); and the highest weighted links removed (criterion 7) following a decreasing sequence. In this criterion, a weighted graph was built where the weight of each link, w_{ij}, represented the

number of links between the nodes i, j and the highest number of links between two nodes in the entire network. Finally, the interactive betweenness calculation was carried out (criterion 8). This criterion, instead of removing the nodes in decreasing order of the static betweenness, recomputes the betweenness after the removal of a node.

It is relevant to analyse which of the used criteria allow us to achieve an 80% reduction in the size of GC in less time. It is true that when the network is fragmented, the stops belonging to the GC are linking a fraction of the PTN, which can still be functional in certain cases. However, it is also relevant to observe the evolution of size over time in order to estimate the attack tolerance. Several commonalities and differences between PTN can be detected. The relevance of studying the evolution of PTN is also corroborated by the interdependencies between PTN and the accessibility to other infrastructures such as hospitals, business centers, schools, and airports. Therefore, the fluctuations in the availability of stops and routes might greatly impact urban life.

3. Results and Discussion

3.1. Study of Structural Properties. The PageRank (with a damping parameter equal to 0.85), betweenness and eigenvector centralities, degree, and minimum distance characteristics were analysed in the L-Space and P-Space. Figures 1 and 2 show the L-Space for all studied networks. The average values of degree, PageRank, eigenvector, and betweenness centralities for this space can be observed in Table 1. Those correspond to degrees, which are similar in all networks. The average eigenvector and PageRank centralities are lower than 0.002, having very small values in all networks. Therefore, the failures randomly happening in a node will have a low probability of arriving at any other. The same happens with respect to betweenness centrality, presenting small magnitudes (a mean value minor than 0.025), which denotes a slight average in how a node plays a bridging role in a network. Table 1 also shows the value of MRP in all networks, proving the existence in them of a GC (MRP >2).

In L-Space, an equivalence between the degree distributions of the PTN of some cities was pointed out by Von Feber et al. [3], where the authors showed that this distribution followed a power law. This characteristic was checked for the analysed networks, fitting their degree distribution to the function:

$$P(k) \sim k^{-\alpha}. \tag{5}$$

The obtained values for the fitting are shown in Table 2. It is observed, according to the R square value, that those obtained for the fitting were not good (R squared is not close to 1). Figure 3 depicts, in L-Space, the degree distribution for the analysed networks. It is clear that all of them have a low value of k ($k_{\max} \leq 17$) compared with the total links. Therefore, the failures occurring at random on a node will have a high probability of having a small impact. Even so, these networks might be vulnerable to disruptions since a very high amount of nodes is joined to very few other nodes (low degree).

It is clear from the above that the values of betweenness, eigenvector, and PageRank centralities in the analysed networks ensure that they are highly robust against random failures.

The lowest average minimum distance between nodes shows a relevant variability in all networks. Trans-Antofagasta and EMT as well as AVL, CFL, RGTR, and TICE networks have the lowest value. The highest value corresponds to Thunder Bay Transit network.

The modularity of the network is also evaluated in L-Space, employing the generate overlapping cluster generator (OCG) method [31]. It can be noted in Table 3 that all networks show a high value of Q with magnitudes near to 1, which demonstrates a strong community structure in these networks. Romano et al. [38] showed that network efficiency is modularity dependent, with the highest values of transmission occurring at intermediate levels and low values of modularity. High values tend to negatively influence transmission in networks [38]. Therefore, the high modularity in PTN could help to slow down the propagation of a failure in the network.

Table 4 depicts the values of the assortativity coefficient [32]. All networks show assortative higher than 0.4, demonstrating that a significant chance exists in which fractions of stops will randomly join with stations of the same degree. The nodes appear to join with a specific preference. A failure in a node with the highest degree will not have much negative impact, since other nodes with the highest degree are still connected to ones with a similar degree.

The highest k-core in all networks was 3 or 2, demonstrating that the largest subgraphs exist where every node is connected to at least 3 or 2 other nodes within the network. Most nodes show a small and similar k-core (≤ 2).

With respect to P-Space, Table 5 shows the average values of eigenvector and betweenness centralities, PageRank, degree, minimum distance between nodes, diameter, and density.

The degree distributions of the analysed networks were also fitted to a power law function according to equation (5). The results are shown in Table 6. It can be noted that, as in the case of L-Space, in accordance with the obtained R square value, the fitting is not good. P-Space allows us to analyse interconnections between routes and transfers in the network. In this space, k reveals how many stops a person can travel through by using only one route. The network with the highest value average k, $<k>$, is TransAntofagasta. The obtained values for eigenvector and PageRank centralities are very low and similar in all networks, reinforcing the idea exposed in the analysis of L-space that failures randomly occurring in a node have a low probability of quickly arriving at any other one. Many possible transfer nodes exist, making the values for betweenness centrality very small and achieving high robustness against random failures on a node. The PTNs with lower $<l>$ are the best connected in this space, having a minor level of vulnerability. Figure 4 shows, in P-Space, the degree distribution for the analysed networks.

(a)   (b)   (c)

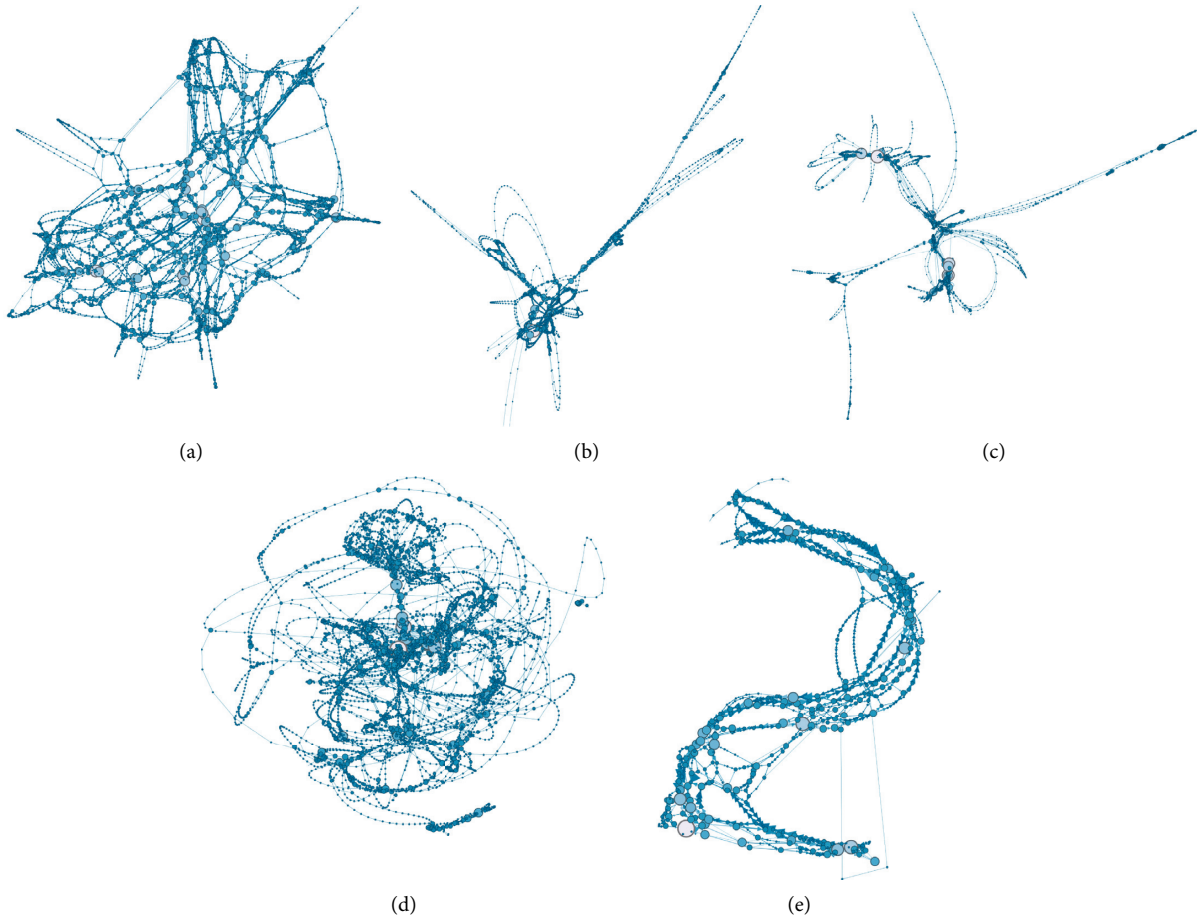(d)   (e)

FIGURE 1: In L-Space, for EMT (a), BC Transit (b), Kolumbus (c), Auckland (d), and TransAntofagasta (e) networks.
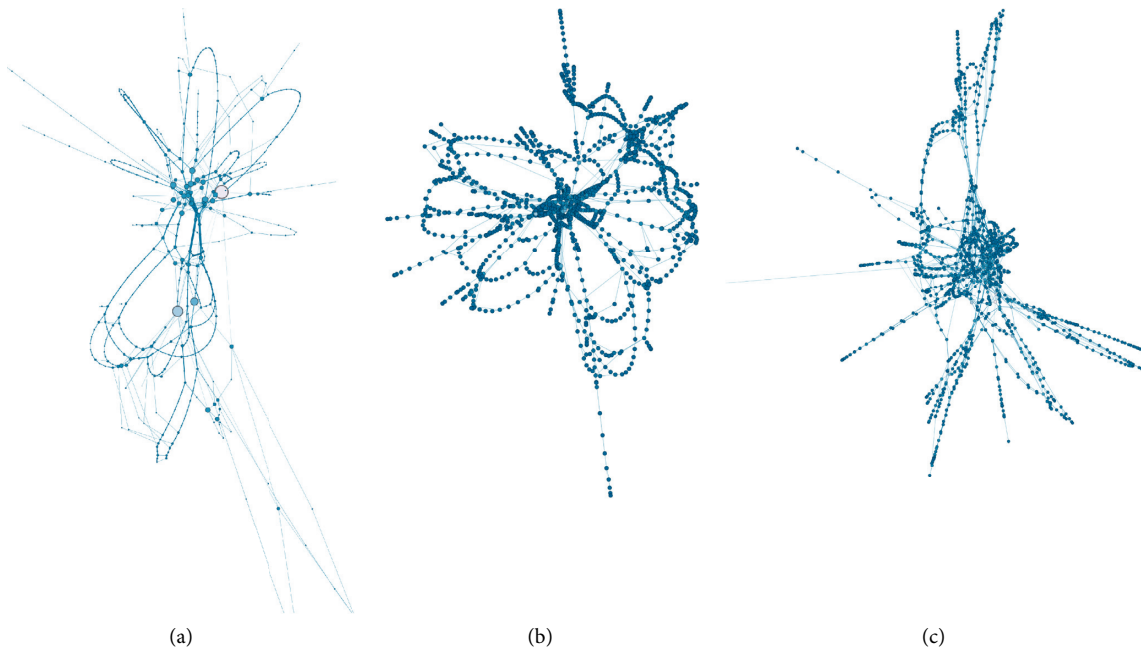

(a)   (b)   (c)

FIGURE 2: Continued.
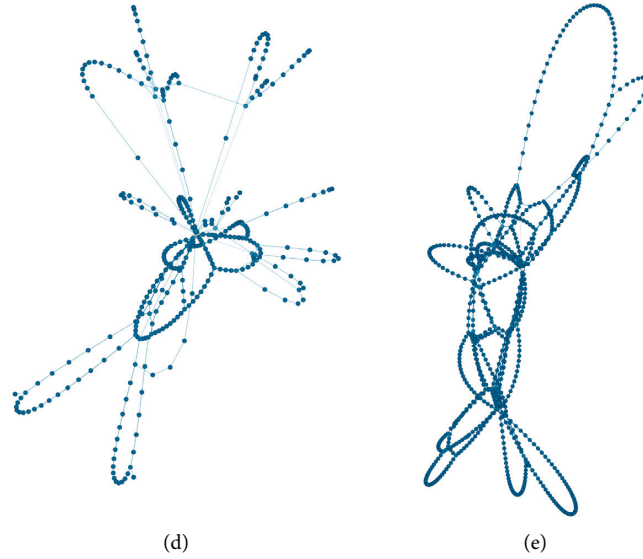
(d)   (e)

FIGURE 2: In L-Space, for Linja-Karjala Oy (a), AVL, CFL, RGTR, and TICE (b), STAR (c), CIT (d), and Thunder Bay Transit (e) networks.

TABLE 1: Structural properties in L-Space. Average Betweenness Centrality (<BC>), average degree (<k>), average Eigenvector Centrality (<EC>), average PageRank (with a damping factor = 0.85) (<PR>), average path Length (<l>), diameter (d), density, and MRP.

Network	Number of nodes	Number of links	<BC>	<k>	<EC>	<PR>	<l>	d	Density	MRP
Auckland	5,223	318,896	0.00655	2.36071	0.00217	0.00019	2.69523	151	0.00045	2.66878
EMT	4,636	2,070,508	0.00383	2.71140	0.00181	0.00022	1.00000	61	0.00058	3.21782
BC Transit	3,981	702,197	0.00233	2.28894	0.00167	0.00025	9.11427	83	0.00058	2.54653
Kolumbus	3,829	487,432	0.00258	2.59546	0.00216	0.00026	2.89389	98	0.00068	3.24314
STAR	1,415	661,150	0.00664	2.81696	0.00559	0.00071	10.0000	37	0.00199	3.54340
AVL, CFL, RGTR, and TICE	1,372	340,684	0.00706	2.81050	0.00564	0.00072	1.00000	34	0.00205	3.79201
Thunder Bay Transit	825	78,247	0.03499	2.17697	0.00560	0.00121	29.79	92	0.00264	2.35412
TransAntofagasta	650	724,362	0.02436	2.96308	0.00902	0.00154	16.78525	51	0.00456	3.46729
Linja-Karjala Oy	551	63,399	0.01611	2.55898	0.01497	0.00181	4.50476	29	0.00465	3.18865
CIT	346	9,366	0.02714	2.24855	0.01700	0.00289	2.33333	29	0.00651	2.75804

TABLE 2: Fitting in L-Space of the network degree distribution to a power law function $P(k) \sim k^{-\alpha}$.

Network	L-Space	
	α	R square
Linja-Karjala Oy	0.158	0.600
Kolumbus	0.459	0.330
EMT	0.622	0.239
Auckland	0.603	0.237
BC Transit	0.412	0.161
AVL, CFL, RGTR, and TICE	0.236	0.156
CIT	0.302	0.156
Thunder Bay Transit	0.406	0.068
TransAntofagasta	0.182	0.044
STAR	0.158	0.041

3.2. Simulation of Faults Propagation.
Failure propagation was simulated in L-Space using an infection and recovery algorithm.

(1) Firstly, a total of 100 experiments for each selection criterion and duple (pI, pR) were undertaken in each network. In each experiment, the simulation was implemented until $t = T$ or until the percentage of infected nodes was higher or equal to 60%. The values of t were modified in steps of 10 units.

(2) The second step was to check the normality of distributions of the infection rate (IR), in all groups (A, B, D, E, P) using the D'Agostino test for any T, which took different values until reaching a level propagation ≥60% of the network. The results show that p value was not higher than 0.05 in all cases. Since not all distributions presented normality, it was not necessary to study the homoscedasticity of variances applying the Breusch–Pagan Test. The Kruskal–Wallis test could be used for the detection of differences between groups.

(3) The third step was using the Kruskal–Wallis test to verify for each T, whether all groups (A, B, D, E, P) were identical or by contrast, one of the groups

(a)

(b)

(c)

(d)

(e)

(f)
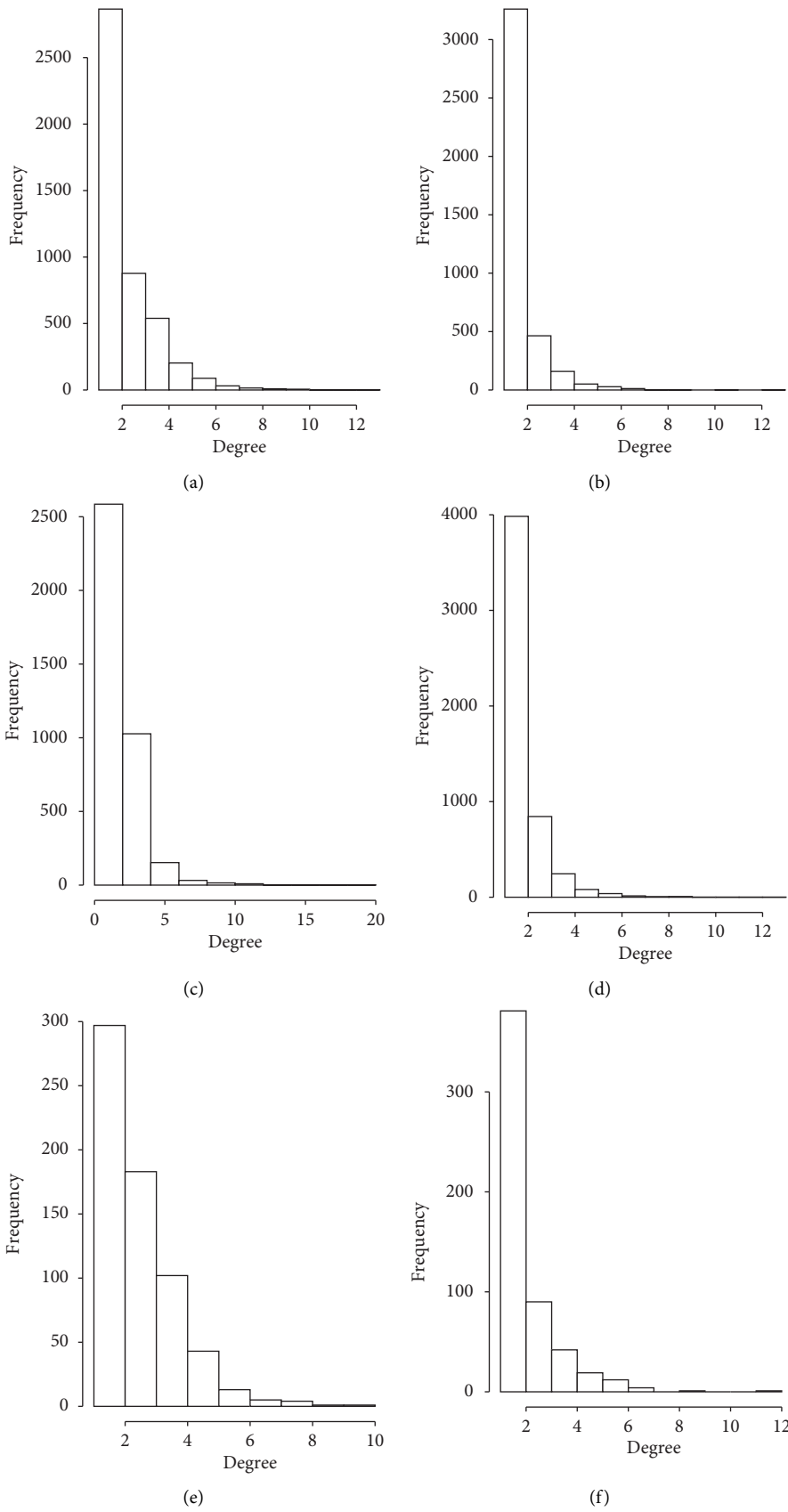
Figure 3: Continued.
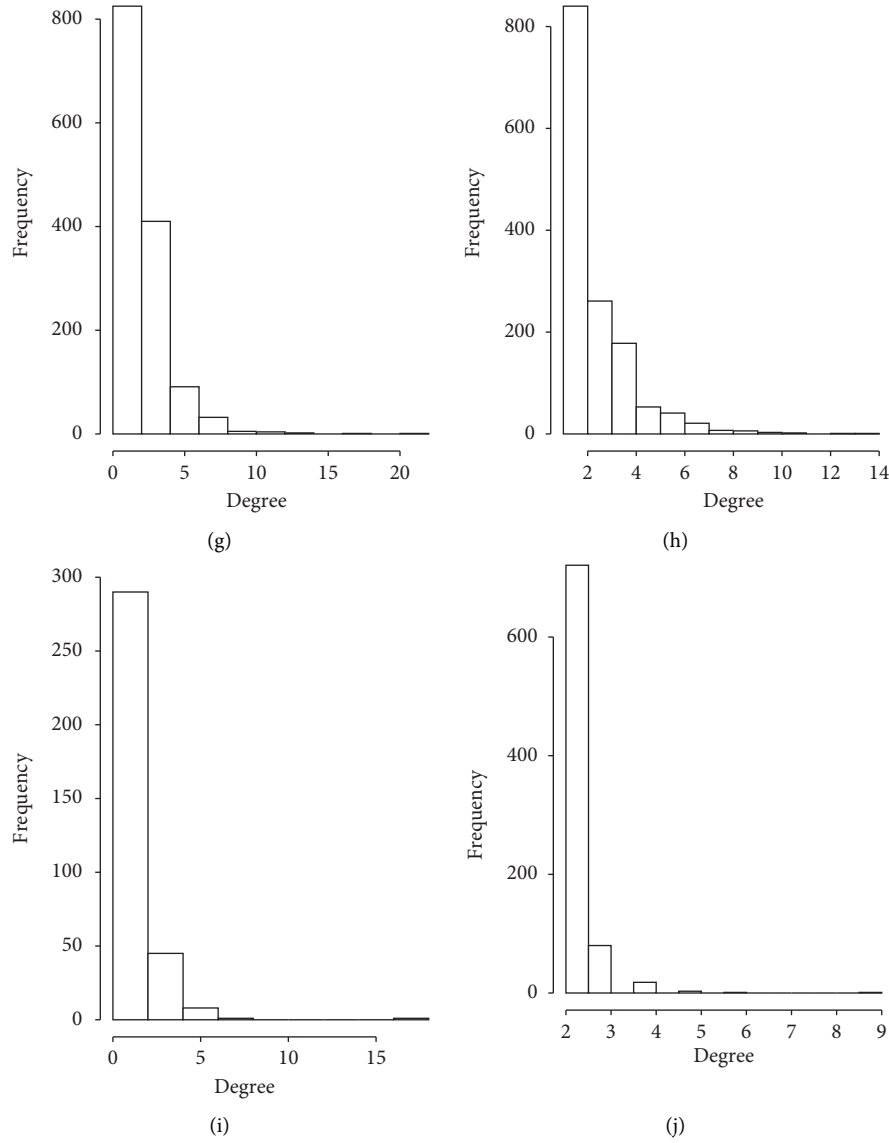
Figure 3: Degree distribution in linear (*x*-axis) and linear scale (*y*-axis) for the following networks: EMT (a), BC Transit (b), Kolumbus (c), Auckland (d), TransAntofagasta (e), Linja-Karjala Oy (f), AVL, CFL, RGTR, TICE (g), STAR (h), CIT (i), and Thunder Bay Transit (j).

Table 3: *Q* Modularity computed using the overlapping cluster generator (OCG) method in L-Space.

Network	*Q*
BC Transit	0.99750
Auckland	0.99610
EMT	0.98550
Thunder Bay Transit	0.99100
Kolumbus	0.98730
STAR	0.97020
AVL, CFL, RGTR, and TICE	0.96880
CIT	0.95880
TransAntofagasta	0.93910
Linja-Karjala Oy	0.93960

tended to give observations that were different from those of other groups. Again, the *p* value obtained was lower than 0.05. The null hypothesis had to be rejected because at least one group had a different distribution.

(4) The fourth step was carried out utilising the Wilcoxon rank sum test in order to compare paired groups; we also applied the Bonferroni correction, both with a significance level $\alpha = 0.05$.

In all analysed duos (*pI*, *pR*), for EMT, there was a significant difference (p value ≤ 0.05) between all groups except between groups D-E. With respect to BC Transit network, there were dissimilarities (p value ≤ 0.05) between some groups except between groups D-E, D-P, and E-P. Regarding the Kolumbus network, a relevant similarity is found between groups B-E and P-D. In the Auckland network, similarities happened between groups P-D, P-B, and D-B. For $pI = 0.05$ and $pR = 0.0025$, in $T = 100$, Figure 5 shows the results of the propagation

TABLE 4: Assortativity coefficients in L-Space.

Network	Assortativity
Kolumbus	0.90599
STAR	0.89463
BC Transit	0.86067
Auckland	0.84123
Linja-Karjala Oy	0.81336
AVL, CFL, RGTR, and TICE	0.72469
EMT	0.67211
TransAntofagasta	0.58252
Thunder Bay Transit	0.45575
CIT	0.42243

algorithm in EMT, BC Transit, Kolumbus, and Auckland networks. For the same values of pI and pR and $T = 65$, Figure 6 shows the outcomes for STAR and AVL, CFL, RGTR, and TICE networks. Figure 7 depicts the results for TransAntofagasta, Linja-Karjala Oy, CIT, and Thunder Bay Transit networks. For those values of pI, pR, and T, Tables 7–16 show the obtained p value in the Wilcoxon rank sum test in each network.

For STAR networks, in all studied duos (pI, pR) there was a similarity (p value > 0.05) between groups B-D, B-E, B-P, and D-P. In AVL, CFL, RGTR, and TICE network there is an equivalence between groups E-D and B-P.

With respect to the TransAntofagasta network, a similarity exists between groups D-P. In Linja-Karjala Oy and CIF networks, there is a difference between groups A-B, A-D, A-E, and E-P. Additionally, dissimilarities exist between B-E groups in the Thunder Bay Transit network.

In all analysed networks, according to the analysis carried out, the A category presents differences with respect to other groups, showing that those failures randomly happening on a node have a distinct behaviour to those originated by target attacks.

In all networks, for $pI = 0.05$ and $pR = 0.0025$, Tables 17 and 18 show the time in which IR reaches 60%.

It can be noted that in all analysed networks, random failures take the longest propagation time, demonstrating, as was previously covered in the section labelled study of structural properties, that these networks are less vulnerable to such failures. Depending on the network, some categories (D, E, P, or B) show similar propagation times. This seems to point out that in some cases, there is an existence of a certain amount of correlation between betweenness, degree, PageRank, and eigenvector centralities. Research shows that topology and density significantly affect the correlation between centrality measures and suggested that they could also affect the robustness [39].

The networks that showed reduced failure propagation times were AVL, CFL, RGTR, and TICE; STAR; Linja-Karjala Oy; and TransAntofagasta. However, those that presented the highest times were in the order of BC Transit, Kolumbus, Auckland, EMT, and finally Thunder Bay Transit. In the magnitude of these times, there is an influence of the diameter presented by the network in L-space.

3.3. Analysis of Evolution of the GC.

It is necessary to know the sensitivity of the network to the elimination of stops and links according to certain criteria in order to investigate which of them has a higher impact. For criteria 1–7, the nodes are removed in blocks of five elements and the percentage of the reduction in the GC is calculated. The applied criteria are the removal of random links (criterion 1) and the removal of random nodes (criterion 2). Other strategies such as the removal of nodes in the descending order of betweenness centrality (criterion 3), degree (criterion 4), eigenvector centrality (criterion 5), PageRank (with a damping factor equal to 0.85) (criterion 6), and highest weight links (criterion 7) are also applied. Using this criterion, a weighted graph was built, in which the weight of each link, w_{ij}, represented the portion of links taken from the total number of links between i and j. The interactive betweenness calculation (criterion 8) was also used. Figures 8 and 9 show, in all networks, the obtained results for the eight criteria.

In Table 19 and Figures 8 and 9, it can be observed that in all networks, the four criteria that most quickly achieve 80% of the reduction of the GC after removing a number of nodes in the range [0%, 20%] have the highest interactive betweenness, PageRank, degree, and static betweenness centrality. And among them, in concordance with [6], in most networks, the elimination of nodes with the highest betweenness interactively calculated obtained an 80% reduction in the GC more rapidly, with a minor % of nodes eliminated. The criteria in which 80% was reached more slowly were those with a removal of random links, random nodes, the highest weighted links, and eigenvector centrality.

The information above highlights the importance of taking special protective measures at some stops. The networks are more robust to failures in random links, random nodes, the highest weighted links, and eigenvector centrality than in the nodes with the highest interactive betweenness, PageRank, static betweenness centrality, and degree. However, when the faults occur randomly, those happening in links produce a higher impact. There is consistency with [40], which proves that links are more robust than nodes when the problems happen in highest degree nodes or in links, for instance, with a weight based on passengers' flow between stops.

TABLE 5: Structural properties in P-Space. Average Betweenness Centrality (<BC>), average degree (<k>), average Eigenvector Centrality (<EC>), average PageRank (with a damping factor = 0.85) (<PR>), average path Length (<l>), diameter (d), and density.

Network	Number of nodes	Number of links	< BC>	<k>	< EC>	< PR>	<l>	d	Density
Auckland	5,223	8,079,095	0.00039	101.28508	0.02386	0.00019	2.69143	7	0.01940
EMT	4,010	25,263,108	0.00065	47.66983	0.04685	0.00025	3.59238	7	0.01189
BC Transit	3,981	15,737,729	0.00017	70.76714	0.03415	0.00025	2.65481	6	0.01778
Kolumbus	3,829	11,317,714	0.00028	78.98355	0.02584	0.00026	2.96889	7	0.02063
STAR	1,415	9,477,213	0.00137	43.11236	0.01862	0.00071	1.48276	6	0.03049
AVL, CFL, RGTR, and TICE	1,372	4,254,656	0.00120	48.42570	0.01791	0.00073	1.00000	6	0.03532
Thunder Bay Transit	825	2,014,779	0.00133	75.51030	0.02197	0.00121	2.09	4	0.09164
TransAntofagasta	650	31,171,307	0.00108	199.74769	0.03666	0.00154	1.70123	3	0.30778
Linja-Karjala Oy	551	1,167,682	0.00180	64.26860	0.03500	0.00181	1.47619	4	0.11685
CIT	346	135,740	0.00274	33.93478	0.03445	0.00289	1.00000	4	0.103578

TABLE 6: Fitting in P-Space of the network degree distribution to a power law function.

Network	α	R square
EMT	0.513	0.180
Kolumbus	0.169	0.166
Auckland	0.204	0.077
BC Transit	0.177	0.068
Thunder Bay Transit	0.275	0.047
TransAntofagasta	0.163	0.044
STAR	0.118	0.037
CIT	0.109	0.036
AVL, CFL, RGTR, and TICE	0.061	0.022
Linja-Karjala Oy	0.015	0.001

(a)

(b)

(c)
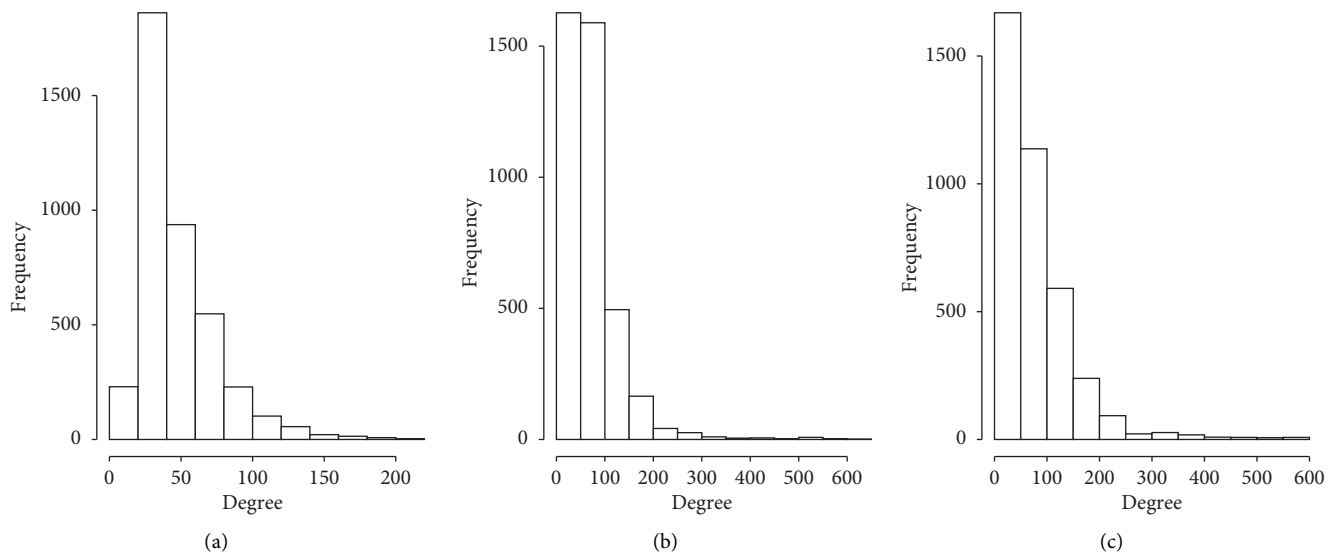
FIGURE 4: Continued.

(d)

(e)

(f)

(g)

(h)

(i)

(j)
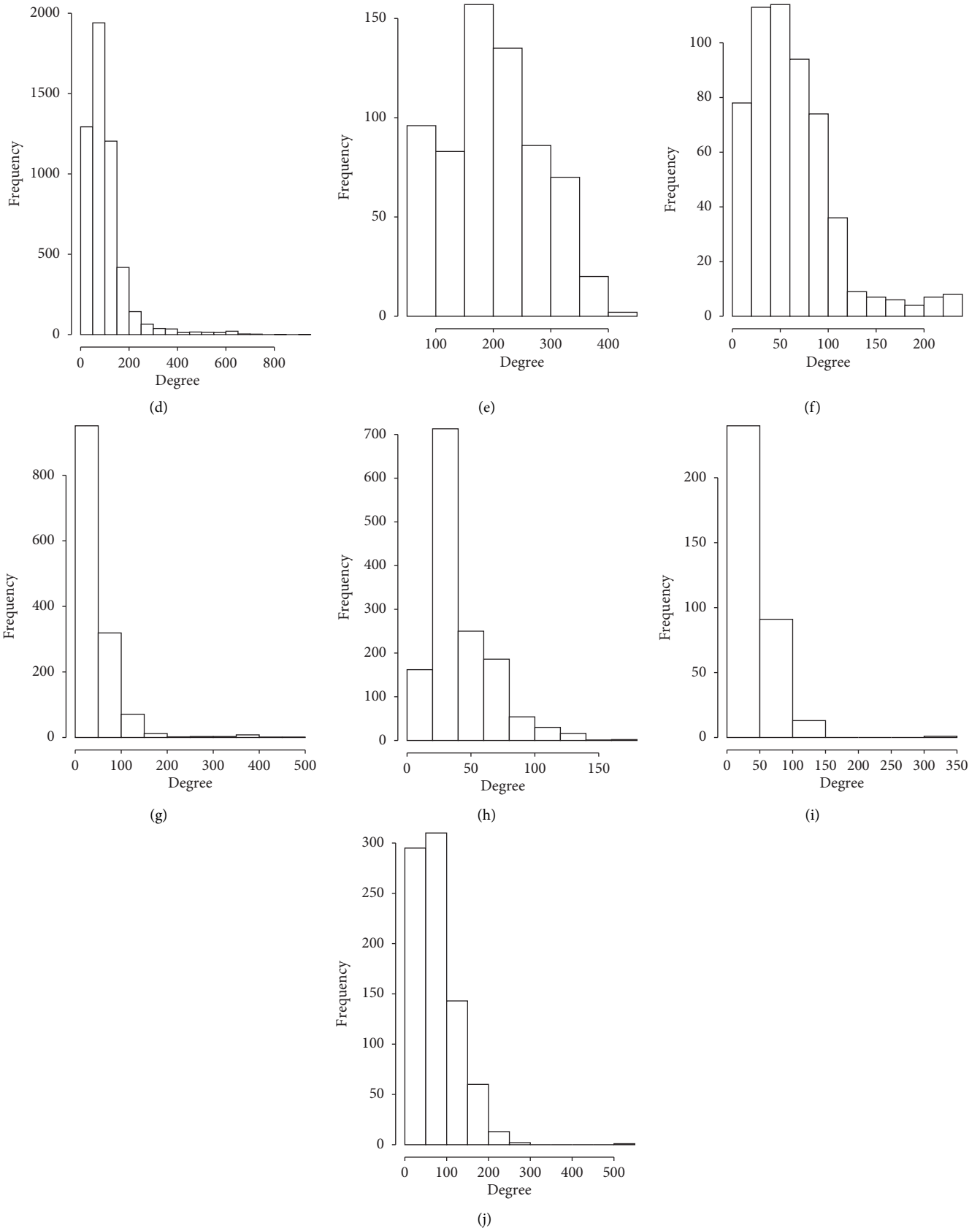
FIGURE 4: Degree distributions in P-Space in linear (*x*-axis) and linear scale (*y*-axis) for the following networks: EMT (a), BC Transit (b), Kolumbus (c), Auckland (d), TransAntofagasta (e), Linja-Karjala Oy (f), AVL, CFL, RGTR, and TICE (g), STAR (h), CIT (i), and Thunder Bay Transit (j)
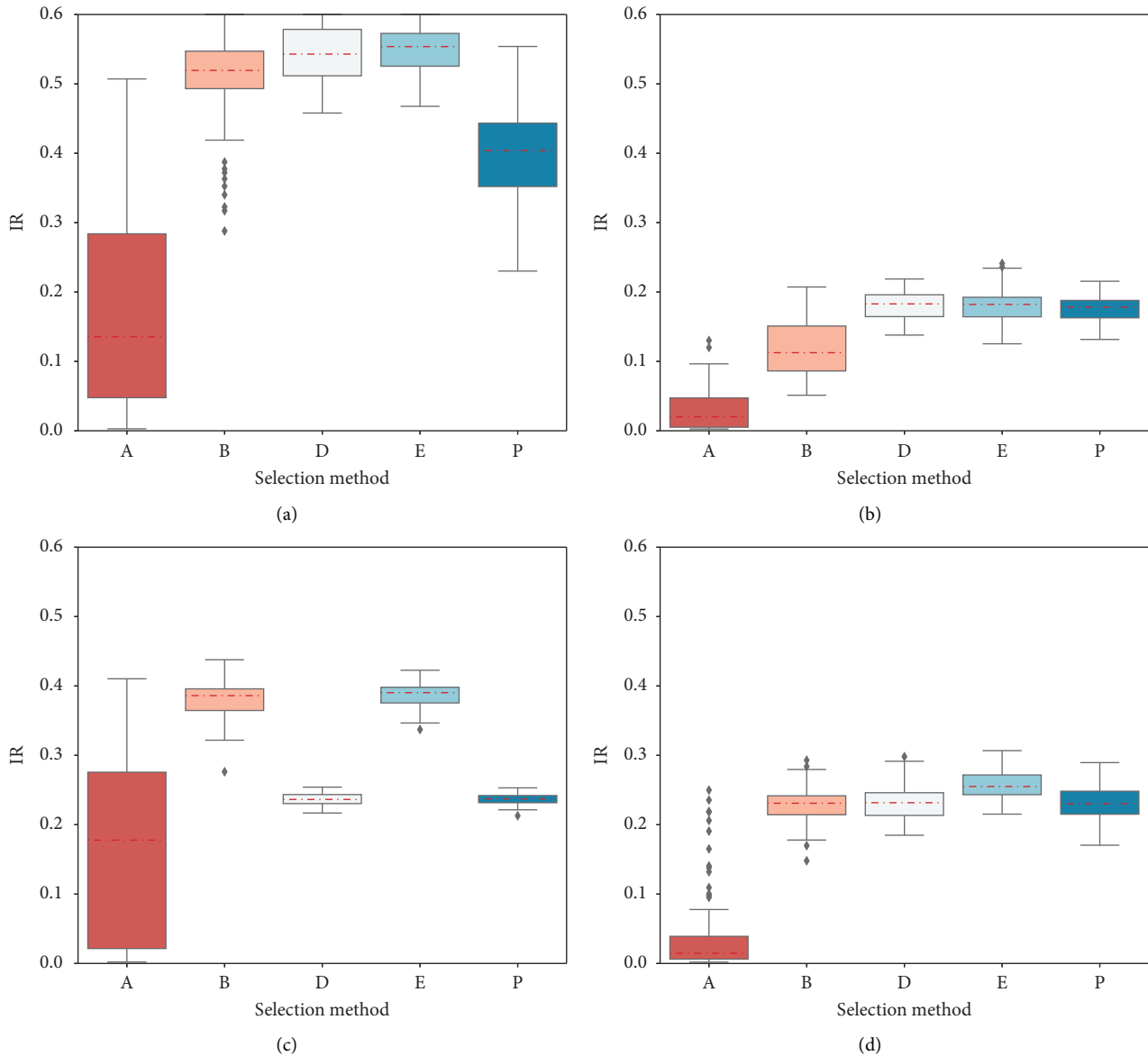
FIGURE 5: In L-Space for (a) EMT, (b) BC Transit, (c) Kolumbus, and (d) Auckland networks. $pI = 0.05$, $pR = 0.0025$, and $T = 150$ (boxplot diagram for A, B, D, E, and P groups).

Regarding k-core feature, it does not seem to influence the level of impact that a failure originating from a node will have on the network. This is in line with some research points, in reference to social networks, where in rumour spreading models, only the k-core of a node does not determine its propagation capacity [41].

Additionally, if some of the topological parameters in L-Space, which are depicted in Table 1, are seen in conjunction with the results shown in Tables 17 and 18, it can be noted that the networks with the highest value of density and lowest value of diameter are between those that obtain 60% in the IR more quickly.

(a)

(b)

Figure 6: In L-Space for (a) STAR and (b) AVL, CFL, RGTR, and TICE networks. $pI = 0.05$, $pR = 0.0025$, and $T = 65$ (boxplot diagram for A, B, D, E, and P groups).

(a)

(b)

Figure 7: Continued.

(c)

(d)

FIGURE 7: In L-Space for (a) TransAntofagasta, (b) Linja-Karjala Oy, (c) CIT, and (d) Thunder Bay Transit. $pI = 0.05$, $pR = 0.0025$, and $T = 65$ (boxplot diagram for A, B, D, E, and P groups).

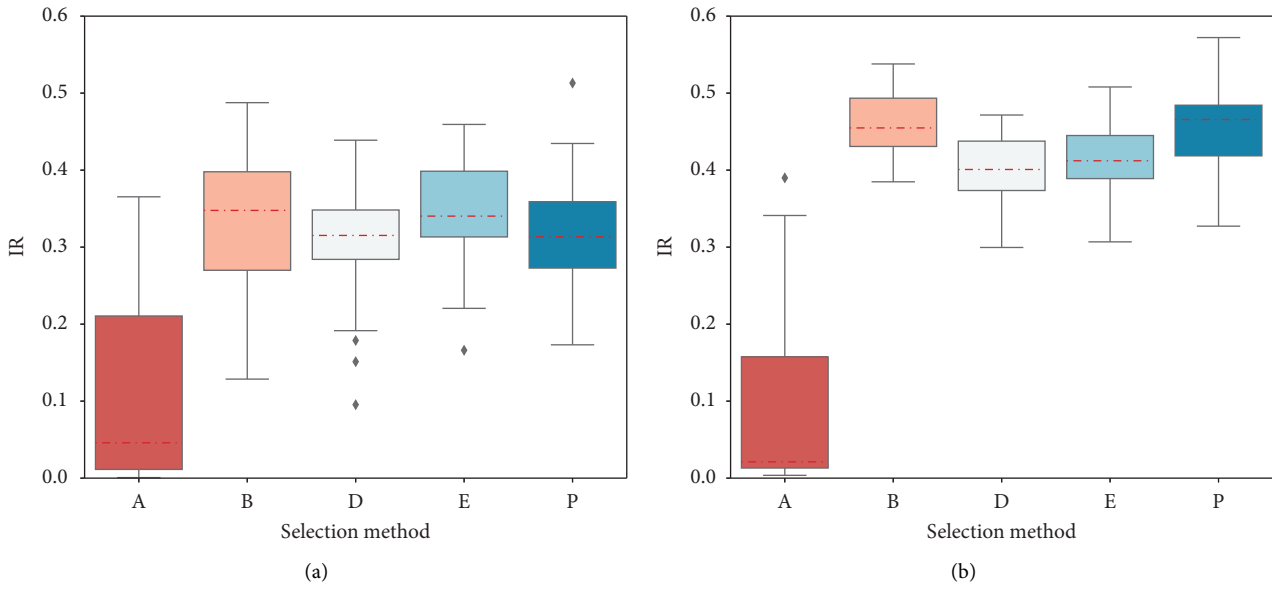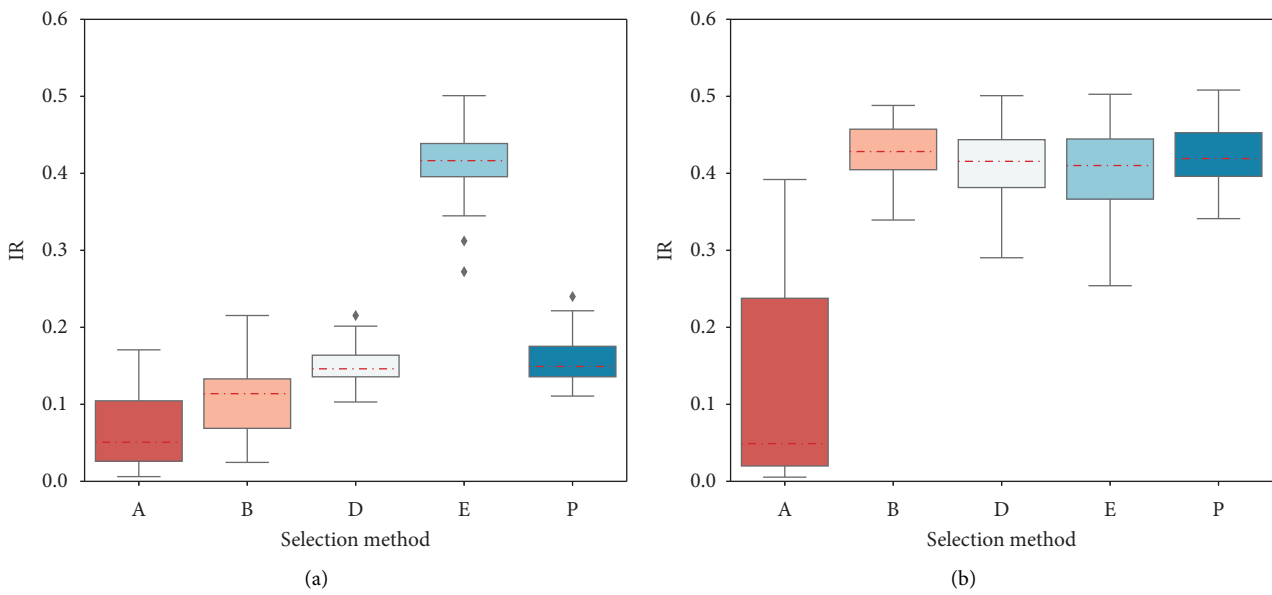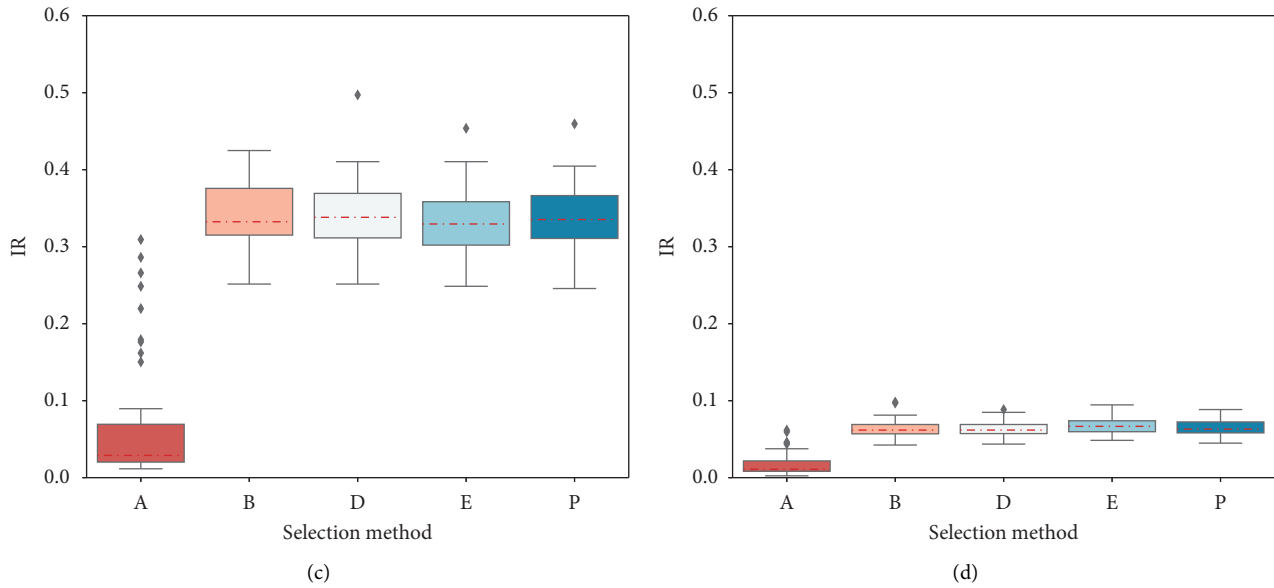TABLE 7: In L-Space, for BC Transit network, $pI = 0.05$, $pR = 0.0025$, and $T = 150$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$<5.70e-11$ ($8.60e-10$)			
D	$1.10e-14$ ($1.60e-13$)	$1.60e-09$ ($2.30e-08$)		
E	$2.40e-14$ ($3.50e-13$)	$3.70e-09$ ($5.50e-08$)	0.90000 (1.00000)	
P	$2.20e-14$ ($3.30e-13$)	$1.30e-08$ ($2.00e-07$)	0.27000 (1.00000)	0.34000 (1.00000)

TABLE 8: In L-Space, for Kolumbus network, $pI = 0.05$, $pR = 0.0025$, and $T = 150$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$<1.40e-13$ ($2.00e-12$)			
D	0.00473 (0.071)	$<2.00e-16$ ($1.10e-15$)		
E	$3.00e-14$ ($4.50e-13$)	0.44174 (1.000)	$<2.00e-16$ ($1.00e-15$)	
P	0.00554 (0.083)	$<2.00e-16$ ($1.10e-15$)	0.99123 (1.000)	$<2.00e-16$ ($1.00e-15$)

TABLE 9: In L-Space, for EMT network, $pI = 0.05$, $pR = 0.0025$, and $T = 150$, p-value of in the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$<2.00e-16$ ($<2.00e-16$)			
D	$<2.00e-16$ ($<2.00e-16$)	0.00027 (0.0041)		
E	$<2.00e-16$ ($<2.00e-16$)	$1.40e-06$ ($2.00e-05$)	0.30986 (1.00000)	
P	$<2.00e-16$ ($<2.00e-16$)	$<2.00e-16$ ($2.00e-16$)	$<2.00e-16$ ($<2.00e-16$)	$<2.00e-16$ ($<2.00e-16$)

TABLE 10: In L-Space, for Auckland network, $pI = 0.05$, $pR = 0.0025$, and $T = 150$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$<2.00e-16$ ($<2.00e-16$)			
D	$<2.00e-16$ ($<2.00e-16$)	0.49000 (1.00000)		
E	$<2.00e-16$ ($<2.00e-16$)	$7.40e-14$ ($1.10e-12$)	$3.60e-12$ ($5.30e-11$)	
P	$<2.00e-16$ ($<2.00e-16$)	0.54000 (1.00000)	0.96000 (1.00000)	$1.00e-12$ ($1.50e-11$)

TABLE 11: In L-Space, for STAR network, $pI = 0.05$, $pR = 0.0025$, and $T = 65$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$1.20e-10$ ($1.20e-09$)			
D	$5.70e-12$ ($5.70e-11$)	0.056 (0.56)		
E	$2.90e-14$ ($2.90e-13$)	0.752 (1.00000)	0.007 (0.07000)	
P	$1.20e-12$ ($1.20e-11$)	0.193 (1.00000)	0.689 (1.00000)	0.01900 (0.19000)

TABLE 12: In L-Space, for AVL, CFL, RGTR, and TICE networks, $pI = 0.05$, $pR = 0.0025$, and $T = 65$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$2.30e-14$ ($2.3e-13$)			
D	$2.00e-16$ ($4.8e-16$)	$1.70e-07$ ($1.7e-06$)		
E	$< 2.00e-16$ ($3.1e-16$)	$9.00e-06$ ($9.0e-05$)	0.19 (1.00000)	
P	$< 2.00e-16$ ($<2e-16$)	0.19 (1.00000)	$2.70e-07$ ($2.7e-06$)	$4.30e-05$ (0.00043)

TABLE 13: In L-Space, for TransAntofagasta network, $pI = 0.05$, $pR = 0.0025$, and $T = 65$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	0.0013 (0.008)			
D	$1.80e-11$ ($1.10e-10$)	$5.50e-07$ ($3.30e-06$)		
E	$1.70e-14$ ($1.70e-13$)	$2.90e-14$ ($2.90e-13$)	$<2.00e-16$ ($<2.00e-16$)	
P	$6.50e-12$ ($3.90e-11$)	$1.50e-07$ ($8.90e-07$)	0.5348 (1.00000)	$<2.00e-16$ ($<2.00e-16$)

TABLE 14: In L-Space, for Linja-Karjala Oy network, $pI = 0.05$, $pR = 0.0025$, and $T = 65$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$1.10e-13$ ($1.10e-12$)			
D	$2.40e-16$ ($2.40e-15$)	0.11600 (1.00000)		
E	$1.50e-15$ ($1.50e-14$)	0.11000 (1.00000)	0.59500 (1.00000)	
P	$<2.00e-16$ ($3.20e-16$)	0.65500 (1.00000)	0.16400 (1.00000)	0.09800 (0.98000)

TABLE 15: In L-Space, for CIT network, $pI = 0.05$, $pR = 0.0025$, and $T = 65$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$1.30e-13$ ($1.30e-12$)			
D	$<2.00e-16$ ($7.50e-16$)	1.00000 (1.00000)		
E	$<2.00e-16$ ($1.40e-15$)	0.29000 (1.00000)	0.20000 (1.00000)	
P	$<2.00e-16$ ($7.30e-16$)	0.63 (1.00000)	0.56 (1.00000)	0.41 (1.00000)

TABLE 16: In L-Space, for Thunder Bay Transit network, $pI = 0.05$, $pR = 0.0025$, and $T = 65$, p-value of the Wilcoxon rank sum test (resultant p-value per Bonferroni correction is indicated in brackets).

	A	B	D	E
B	$3.00e-13$ ($3.00e-12$)			
D	$<2.00e-16$ ($1.50e-15$)	0.47900 (1.00000)		
E	$<2.00e-16$ ($5.60e-16$)	0.02700 (0.27000)	0.08300 (0.83000)	
P	$<2.00e-16$ ($9.30e-16$)	0.16900 (1.00000)	0.44800 (1.00000)	0.27900 (1.00000)

TABLE 17: For $pI = 0.05$ and $pR = 0.0025$, average propagation time in which IR reached 60% in each group (A, B, D, E, P) and globally.

	A	B	D	E	P	All
Auckland	542.13	386.85	389.85	408.83	389.37	415.24
EMT	261.31	171.75	162.83	162.06	195.33	202.99
BC Transit	>700	>700	>700	>700	>700	>700
Kolumbus	552.99	458.23	>700	489.01	>700	>700
STAR	116.75	93.51	99.74	95.09	97.88	99.67
AVL, CFL, RGTR, and TICE	121.37	84.42	90.43	89.58	83.73	91.51
TransAntofagasta	>150	145	>150	101.16	>150	105.57
Linja-Karjala Oy	127.22	98.64	102.54	102.12	97.50	103.78
CIT	140.00	115.65	117.98	119.56	120.18	120.01
Thunder Bay Transit	>150	>150	>150	>150	>150	>150

TABLE 18: For $pI = 0.05$ and $pR = 0.0025$, median propagation time in which IR reached 60% in each group (A, B, D, E, P) and globally.

	A	B	D	E	P	All
Auckland	545	385	390	409	388	399
EMT	249.50	169	164	161	193	181
BC Transit	>700	>700	>700	>700	>700	>700
Kolumbus	547	455	>700	486	>700	486.50
STAR	112.50	94	98	95	99	99.67
AVL, CFL, RGTR, and TICE	116.00	84	89	90	81.50	88
TransAntofagasta	>150	149	>150	100	>150	101
Linja-Karjala Oy	128	97	103	103	97	102
CIT	138	113	117	119	120	119
Thunder Bay Transit	>150	>150	>150	>150	>150	>150

Evolution of GC

— Random links removal
— Random nodes removal
— Nodes removal according to their betweenness
— Nodes removal according to their degree
— Nodes removal according to their eigenvalue
— Nodes removal according to their PageRank
— Weighted links removal
— Nodes removal according to their interactive betweenness
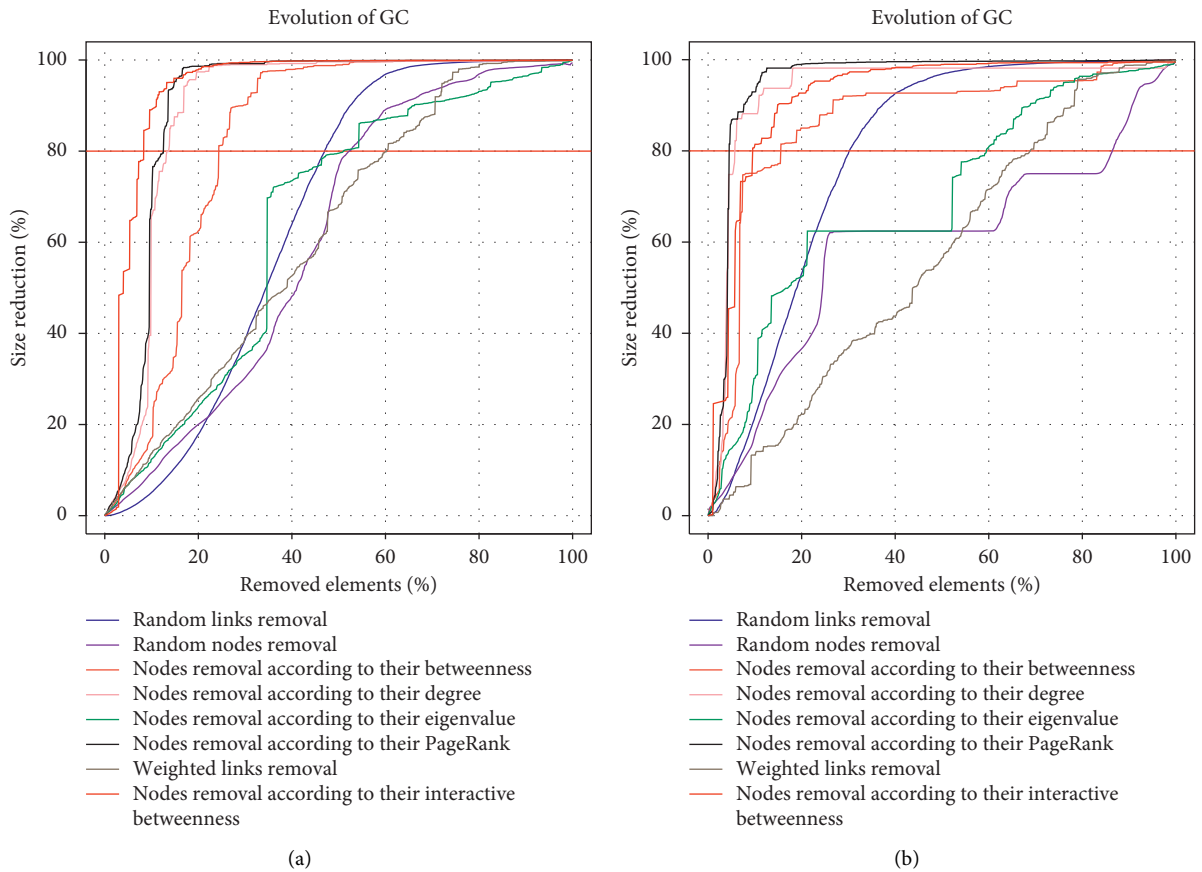
(a)
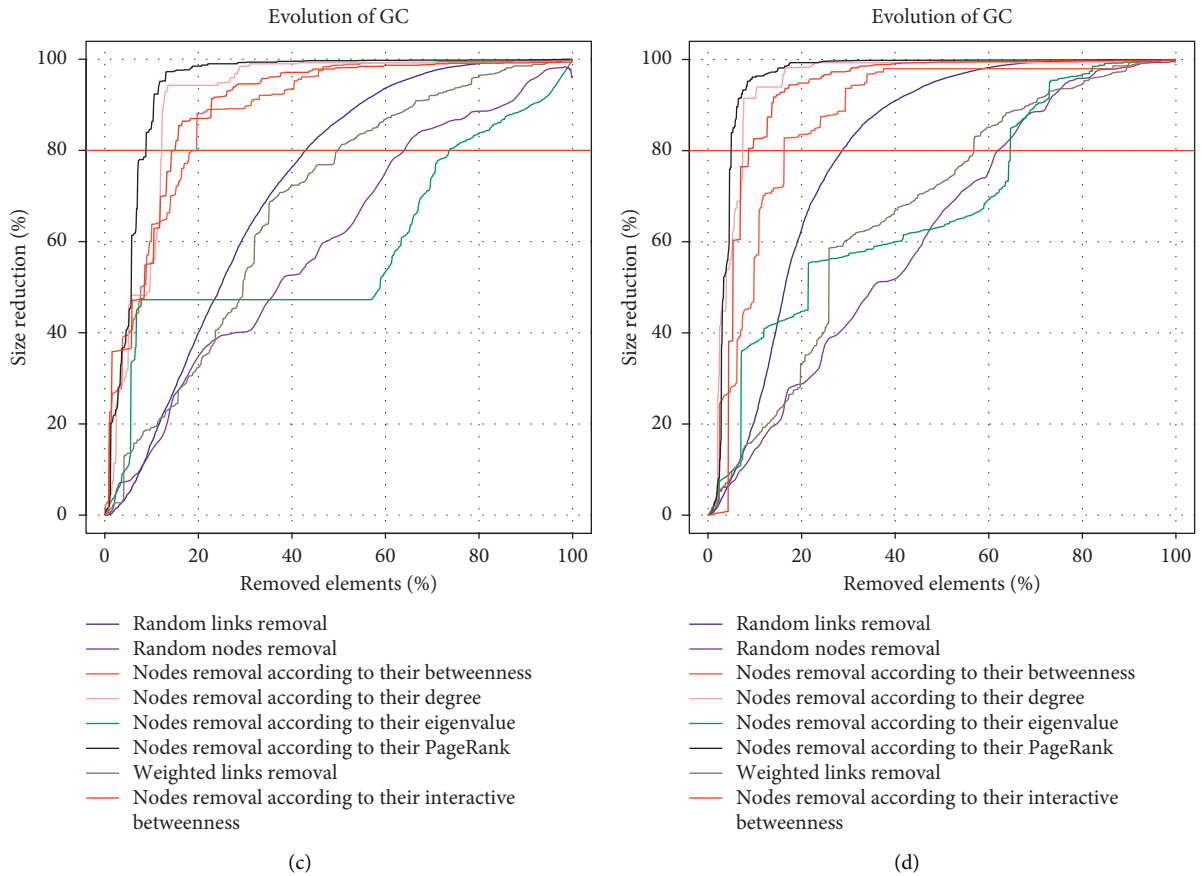
(b)

FIGURE 8: Continued.

(c)

(d)

FIGURE 8: Evolution of the GC in EMT (a), BC Transit (b), Kolumbus (c), and Auckland (d) networks.
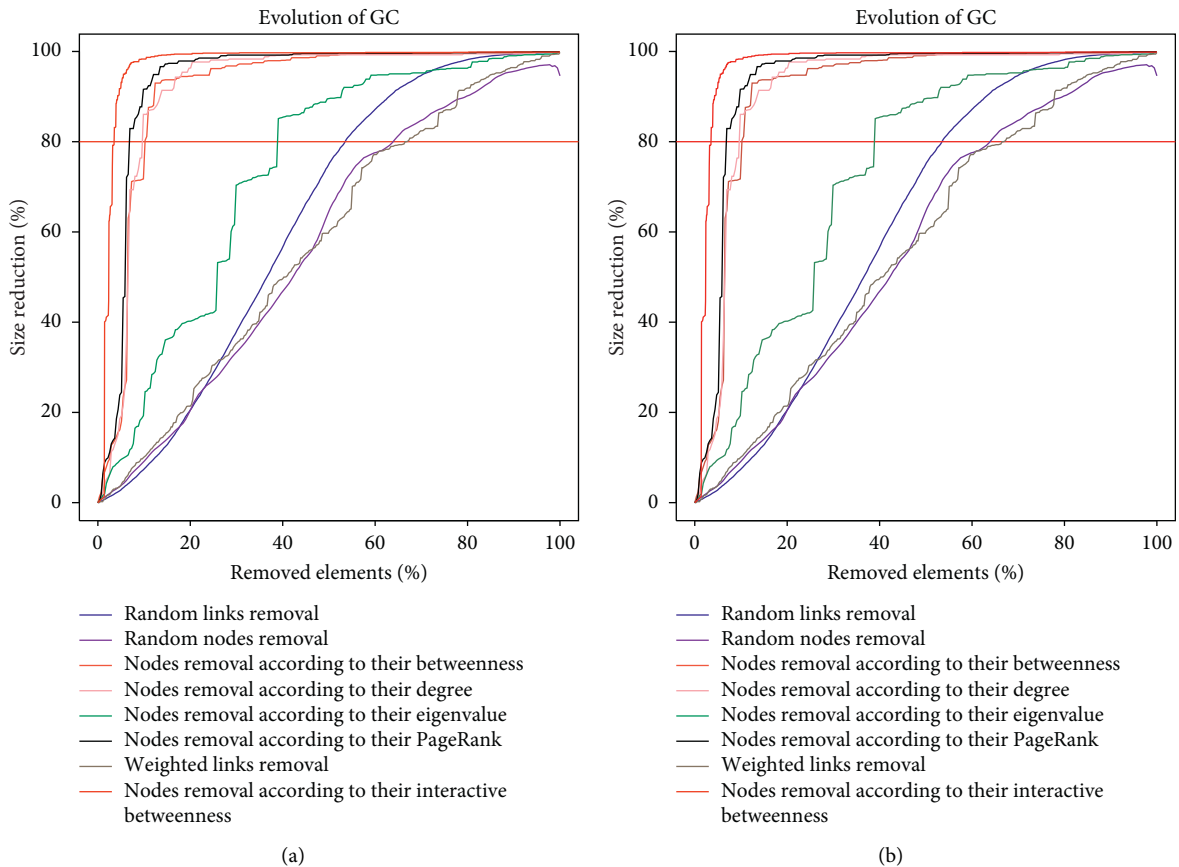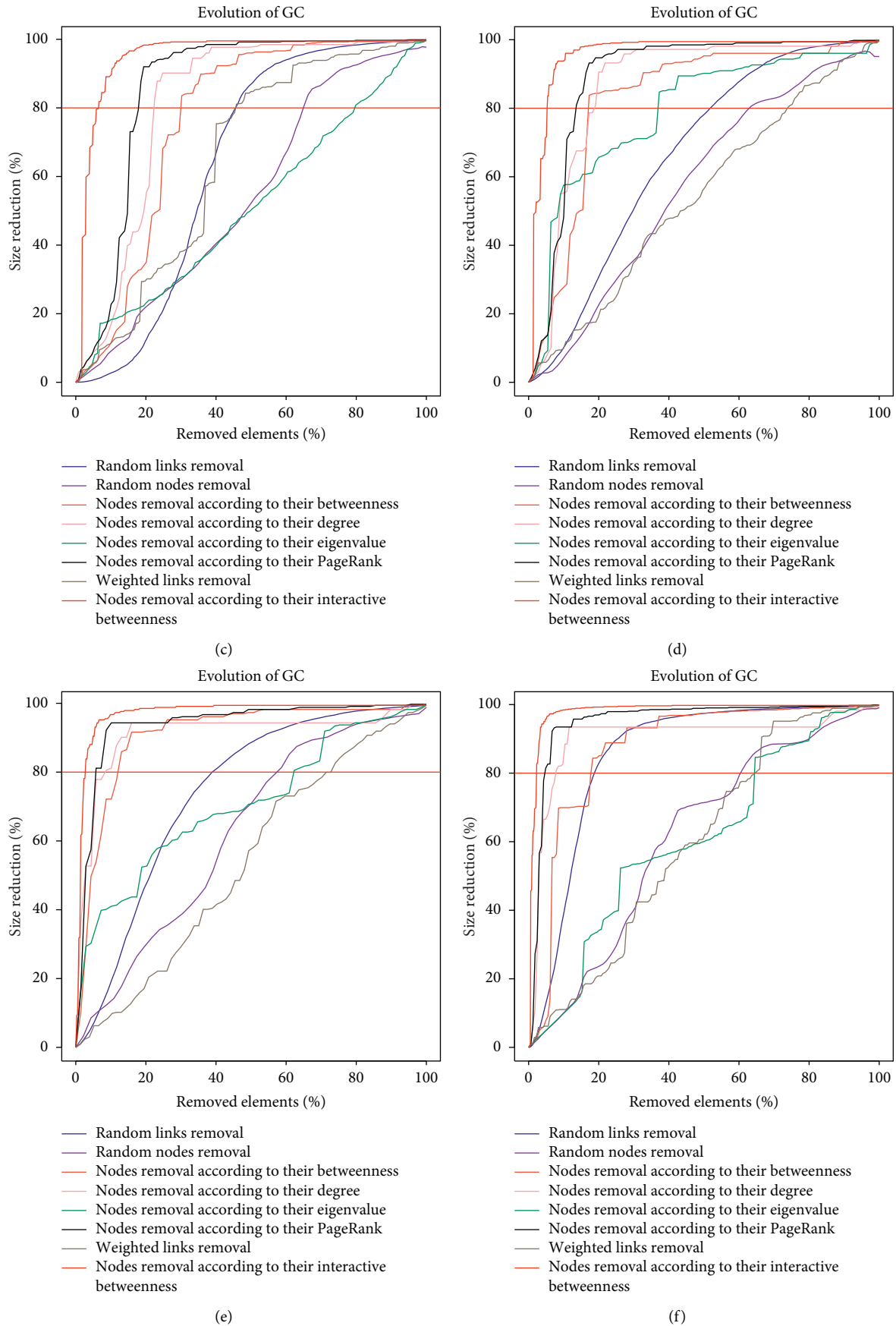

(a)

(b)

FIGURE 9: Continued.

Figure 9: Evolution of the GC in AVL, CFL, RGTR, and TICE (a), STAR (b), TransAntofagasta (c), Linja-Karjala Oy (d), CIT Chambly-Richelieu-Carignan network (e), and Thunder Bay Transit (f).

TABLE 19: Percentage of removed nodes with which the 80% of reduction in the GC is achieved.

Network	Method	% of removed elements	Network	Method	% of removed elements
Auckland	Pagerank centrality	4.98	EMT	Interactive betweenness centrality	8.35
	Degree	7.47		Pagerank centrality	12.62
	Interactive betweenness centrality	8.67		Degree	13.70
	Betweenness centrality	16.28		Betweenness centrality	24.38
	Random links	28.79		Random links	47.02
	Weighted links	56.85		Eigenvector centrality	51.02
	Random nodes	61.99		Random nodes	52.42
	Eigenvector centrality	64.66		Weighted links	60.22
BC Transit	Pagerank centrality	4.52	Kolumbus	Pagerank centrality	8.89
	Degree	5.79		Degree	12.29
	Interactive betweenness centrality	9.55		Interactive betweenness centrality	15.04
	Betweenness centrality	15.58		Betweenness centrality	19.74
	Random links	30.19		Random links	43
	Eigenvector centrality	59.55		Weighted links	49.75
	Weighted links	68.94		Random nodes	64.17
	Random nodes	86.50		Eigenvector centrality	73.86
STAR	Interactive betweenness centrality	4.53	AVL, CFL, RGTR, TICE	Interactive betweenness centrality	3.57
	Pagerank	14.18		pagerank	6.93
	Degree	15.60		Degree	9.85
	Betweenness centrality	17.73		Betweenness centrality	10.22
	Eigenvector centrality	45.74		Random links	53.51
	Random links	55.53		Eigenvector centrality	39.05
	Weighted links	64.07		Random nodes	64.25
	Random nodes	78.49		Weighted links	66.49
TransAntofagasta	Interactive betweenness centrality	6.63	Linja-Karjala Oy	Interactive betweenness centrality	5.27
	Pagerank	17.83		Pagerank	13.64
	Degree	22.48		Betweenness centrality	17.27
	Betweenness centrality	30.23		Degree	19.09
	Random links	45.31		Eigenvector centrality	37.27
	Weighted links	45.83		Random links	52.48
	Random nodes	65.12		Random nodes	63.51
	Eigenvector centrality	79.84		Weighted links	73.76
CIT	Interactive betweenness centrality	2.90	Thunder Bay Transit	Interactive betweenness centrality	2.31
	Pagerank	5.80		Pagerank	4.88
	Degree	8.70		Degree	7.93
	Betweenness centrality	13.04		Betweenness centrality	18.29
	Random links	38.96		Random links	18.99
	Random nodes	58.70		Random nodes	61.11
	Eigenvector centrality	62.32		Eigenvector centrality	64.63
	Weighted links	74.03		Weighted links	96.97

4. Conclusions

An efficient PTN must be robust to provide a good quality service, even when it faces faults or disaster. This study analyses several PTNs, applying a methodology based on the study of structural properties, simulations of faults propagation, and the examination of the evolution of the GC. The analysis aims to detect the most vulnerable nodes and links when some faults occurred in them. The influence of the topological properties of the networks over their robustness has also been checked. This could allow preventative action in certain nodes and links against traffic congestion and unattainable stops, occurring as a result of construction work and catastrophes.

Problems that happened at random nodes and over the strongest links spread slower than other types of failure. Those nodes whose failures spread more quickly depended on the network, but they were those with the highest betweenness centrality, degree, or PageRank. All networks show high modularity >0.8 and a Molloy–Reed parameter >2. Additionally, all nodes were characterized by low k-core.

The AVL, CFL, RGTR, and TICE networks were that presented the highest-level propagation in a time T, for all

analysed pairs (*pI*, *pR*), followed in descending order by STAR, Linja-Karjala Oy, and TransAntofagasta. In the majority of the networks, 80% reduction in the size of GC with the minor percentage of removed nodes is reached for those nodes with the highest betweenness calculated interactively. In a small number of networks, this was achieved with the removal of nodes with the highest PageRank. The networks that showed the highest amount of sensitivity were in sequential order of Thunder Bay Transit; CIT; and AVL, CFL, RGTR, and TICE, with 2.31%, 2.90%, and 3.57% of deleted nodes, respectively. In all networks, random failures took the longest time to spread across all networks, which proves, in conjunction with some structural properties (*k*, PageRank, and eigenvector in L-Space), that the networks are less vulnerable to such failures.

This research can be continued by analysing passenger flow in the PTN as a dynamically changing characteristic. Some models that take into account the individual passengers' behaviour, with respect to the choice of routes and connections during a trip, can be implemented in order to dynamically establish the possible consequences of the congestion. It is also possible to study how the passengers modified their travel behaviour when considering their previous experience. Additionally, the PTN vulnerability can be examined in other graph representations, as well as considering different link weights, such as trip time or financial costs. In addition to the above, in a catastrophic situation, the postdisaster circumstances can be examined, completing the analysis of the evolution of GC with a study of the access to critical facilities (i.e., emergency services).

Data Availability

Information of stops and routes of Auckland, EMT, BC Transit, Kolumbus, STAR, AVL, CFL, RGTR, TICE, Thunder Bay Transit, TransAntofagasta, Linja-Karjala Oy, and CIT networks were retrieved from the operating companies public web sites, the Deconet Public Transport Network Data and GTFS Data Exchange repositories

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] C. Wan, Z. Yang, D. Zhang, X. Yan, and S. Fan, "Resilience in transportation systems: a systematic review and future directions," *Transport Reviews*, vol. 38, no. 4, 2018.

[2] C. Wan, Z. Yang, Z. Xu, W. Zhang, and T. Zheng, "Space P-based empirical research on public transport complex networks in 330 cities of China," *Reliability Engineering and System Safety*, vol. 13, no. 1, 2013.

[3] F. C. Von Feber, T. Holovatch, Y. Holovatch, and V. Palchykov, "Public transport networks: empirical analysis and modelling," *European Physical Journal B*, vol. 68, 2009.

[4] B. Berche, C. von Ferber, T. Holovatch, and Y. Holovatch, "Resilience of public transport networks against attacks," *Physics of Condensed Matter*, vol. 71, 2009.

[5] O. Lordan, J. M. Sallan, P. Simo, and D. González-Prieto, "Robustness of the air transport network," *Transportation Research Part E: Logistics and Transportation Review*, vol. 68, 2014.

[6] S. Wandelt, X. Sun, D. Feng, M. Zanin, and S. Havlin, "A comparative analysis of approaches to network-dismantling," *Scientific Reports*, vol. 8, 2018.

[7] S. Wandelt, X. Shi, and X. Sun, "Approximation of interactive betweenness centrality in large complex networks," *Complexity*, vol. 2020, Article ID 4046027, 2020.

[8] M. D. Yap, N. Van Oort, R. Nes, and B. Arem, "Identification and quantification of link vulnerability in multi-level public transport networks: a passenger perspective"" *Transportation*, vol. 45, 2018.

[9] E. Jenelius and O. Cats, "The value of new public transport links for network robustness and redundancy," *Quantitative Approaches to Resilience in Transport Networks*, vol. 11, no. 9, 2015.

[10] X. Albacete, D. Olaru, V. Paül, and S. Biermann, "Measuring the accessibility of public transport: a critical comparison between methods in Helsinki," *Applied Spatial Analysis and Policy*, vol. 10, no. 2, 2017.

[11] F. Corman F, A. D'Ariano, and I. A. Hansen, "Evaluating disturbance robustness of railway schedules," *Journal of Intelligent Transportation Systems*, vol. 18, no. 1, 2014.

[12] S. Wandelt, X. Shi, and X. Sun, "Estimation and improvement of transportation network robustness by exploiting communities," *Reliability Engineering & System Safety*, vol. 206, 2021.

[13] X. Sun, S. Wandelt, and A. Zhang, "Resilience of cities towards airport disruptions at global scale," *Research in Transportation Business & Management*", vol. 34, 2020.

[14] A. Chen, C. Yang, S. Kongsomsaksakul, and M. Lee, "Network-based accessibility measures for vulnerability analysis of degradable transportation networks," *Networks and Spatial Economics*, vol. 7, 2007.

[15] F. Liao and B. Van Wee, "Accessibility measures for robustness of the transport system," *Transportation*, vol. 44, no. 5, 2017.

[16] Z. He, J. N. Guo, and J. X. Xu, "Cascade failure model in multimodal transport network risk propagation," *Mathematical Problems in Engineering*, vol. 12, 2019.

[17] E. Baspinar and B. Koyuncu, "A data-driven air transportation delay propagation model using epidemic process models"" *International Journal of Aerospace Engineering*, vol. 2016, 2016.

[18] M. Akdere, C. C. Bilgin, O. Gerdaneri, I. Korpeoglu, Ö. Ulusoy, and U. Çetintemel, "A comparison of epidemic algorithms in wireless sensor networks"" *Journal Computer Communications*, vol. 3, no. 2, 2006.

[19] M. Nekovee, "Epidemic algorithms for reliable and eficient information dissemination in vehicular ad hoc networks," *IET Intelligent Transport Systems*, vol. 3, no. 2, 2009.

[20] R Project, *R Project for Statistical Computing*, https://www.r-project.org/, 2020.

[21] Python, *Python Lenguaje de programación*, https://www.python.org/, 2020.

[22] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Physical Review Letters*, vol. 87, no. 19, 2001.

[23] P. Angeloudis and D. Fisk, "Large subway systems as complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 367, 2006.

[24] B. Hu, Y. Pei, J. Tang, and W. Gao, "Common network characteristics of four bus transport networks in Northeast China based on a perfect space P," *International Journal of Modern Physics B*, vol. 32, no. 21, 2018.

[25] K. Ma, Z. W. Wang, J. W. Jiang, G. X. Zhu, and W. Li, "Power law and small world properties in a comparison of traffic city networks," *Chinese Science Bulletin*, vol. 56, no. 34, 2011.

[26] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, 2003.

[27] E. Estrada, D. Higham, and N. Hatano, "Communicability betweenness in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 5, 2009.

[28] M. Dehmer, *Structural Analysis of Complex Networks"*, Birkhäuser, Vienna, Austria, 2011.

[29] C. Durón, "Heatmap centrality: a new measure to identify super-spreader nodes in scale-free networks," *PLoS One*, vol. 15, no. 7, 2020.

[30] N. Perra and F. Santo, "Spectral centrality measures in complex networks," *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 78, 2008.

[31] E. Becker, B. Robisson, C. Chapple, A. Guenoche, and C. Brun, "Multifunctional proteins revealed by overlapping clustering in protein interaction network," *Bioinformatics*, vol. 28, no. 1, 2012.

[32] M. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, 2002.

[33] A. L. Barabasi, *Network Science Network Robustness*, https://barabasi.com/f/619.pdf, 2014.

[34] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Wallingford, CT, USA., 2011.

[35] R. Khan and F. Ahmad, "Power comparison of various normality tests," *Pakistan Journal of Statistics and Operation Research*, vol. 11, no. 3, 2017.

[36] A. Halunga, C. D. Orme, and T. Yamagata, *Handbook of Parametric and Nonparametric Statistical Procedures*, School of Social Sciences, University of Manchester, Manchester, UK, 2011.

[37] M. Neuhauser, *Nonparametric Statistical Tests: A Computational Approach*, Chapman and Hall/CRC, New York, NY, USA, 2017.

[38] V. Romano, M. Shen, J. Pansanel et al., "Social transmission in networks: global efficiency peaks with intermediate levels of modularity," *Behav Ecol Sociobiol*, vol. 72, no. 154, 2018.

[39] I. McCulloh, "Network topology effects on correlation between centrality measures," *Connections*, vol. 30, no. 1, pp. 21–28, 2010.

[40] S. Ferreira, C. Rodrigues, and F. Fagundes, "Structure and robustness of Sao Paulo public transport network," *Physics and Society*, vol. 4, no. 3, 2018.

[41] J. Borge-Holthoefer and Y. Moreno, "Absence of influential spreaders in rumour dynamics," *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 85, no. 2, 2012.