Fault Spread and Recovery Strategy of Urban Rail Transit System Based on Complex Network

Wei Yang¹, Liping Mu¹, Hua Jiang², Xiaoping Li² and Kun Hou²*¹

¹College of Computer Science and Technology, Changchun Normal University, Changchun 130032, China
²School of Information Science and Technology, Northeast Normal University, Changchun 130117, China

*Corresponding author e-mail: houk431@nenu.edu.cn

Abstract. Urban rail transit is an important existence in the urban transportation system. In recent years, a large number of people will take it as their preferred travel mode, but the station failure caused by emergencies is inevitable. The impact on the whole system makes people put forward higher requirements for the operation of rail transit. Therefore, how to suppress the fault spread and how to quickly and effectively recover the urban rail transit system after the fault spread is an urgent problem to be solved. In this paper, based on the metro network data obtained from Gaode map, the network topology model was established by using the complex network theory. Then, the fault spread law of the metro network without recovery and the fault spread law with recovery were studied. Finally, the four strategies were simulated and compared.

Keywords: Urban Rail Transit System, Complex Network Theory, Fault Spread, Recovery Strategy

1. Introduction
In recent years, China's rapid economic development, followed by people's demand for urban transportation is also growing. The emergence of urban rail transit systems such as light rail and underground lines has greatly eased the pressure on urban traffic. More and more passengers choose rail transit travel, the problem is that the rail traffic load is becoming more and more serious. Large passenger flow is the determining factor in the fault [1]. Many station nodes and rail transit networks formed by lines are characterized by complex systems. In recent years, many scholars have also combined it with complex network theory for research and analysis [2].

Complex network theory is a new theoretical method, which is generally used to analyze the structure and evolution of complex systems. It abstracts complex systems into networks of points and lines, and analyzes and studies abstract networks based on graph theory.

The station in the rail transit system is equivalent to the network node, and the connected lines are equivalent to the link between nodes [3]. Firstly, an abstract and complex network model can accurately describe the basic characteristics of urban rail transit system. Using the structural generation method of complex network, obtaining the parameter values of its characteristics, referring
to the complex network model and evolution generation model. Establishing a complex network model that can comprehensively describe the characteristics of the rail transit system. Laying the foundation for the next step to study the fault spread and recovery strategy. Secondly, the law of fault spread of urban rail transit system is studied. By simulating the process of fault spreading, the relevant parameters of the system network after the fault spread can be obtained, and the fault spread law of the urban rail transit system can be obtained through the change of these parameters. Finally, a set of efficient post-fault recovery scheme is given. After the spread of the fault, the system machine can be recovered quickly by using reasonable and efficient recovery means.

The presented research is organized as follows. Section 2 introduced the basic methods of the simulation system. In the Section 3, the simulation system implementation and analysis are displayed. The Section 4 is given to conclude this paper and suggest some works in the future.

2. Basic Methods

2.1. Network Construction

The network is composed of a set of nodes \( \{v_i\} \) and a set of edges \( \{e_{ij}\} \). This paper uses the adjacency matrix[4] to further express the national railway network as the adjacency matrix of graph \( G=(V, A) \) is defined as follows,

\[
M = (m_{ij})_{n \times n} \in \{0,1\}^{n \times n} \quad \begin{cases} 1(i, j) \in A \\ 0(i, j) \notin A \end{cases}
\] (1)

Average shortest path length, cluster coefficient, degree distribution and betweenness centrality are the four characteristic parameters [5-8], which are often used to quantitatively describe the complex network are.

2.1.1. Average shortest path length. The average path length of node i is defined as,

\[
L_i = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} L_{ij}
\] (2)

Where \( L_{ij} \) represents the number of edges that exist between the two nodes i and j, and the average path length L of the network is defined as follows,

\[
L = \frac{2 \sum L_{ij}}{n(n-1)}
\] (3)

2.1.2. Cluster coefficient. For a node in the network, its cluster coefficient represents the probability that any two nodes connected to this node are connected to each other. The specific definition is as follows,

\[
M_i = \frac{E_i}{k_i(k_i-1)/2}
\] (4)

Among them, \( M_i \) is the cluster coefficient, \( k_i \) is the number of nodes connected to the node, and \( E_i \) is the number of edges connected to the node.

The cluster coefficient of the network is defined as follows,

\[
M = \frac{\sum M_i}{N}
\] (5)
The value range of $M$ is $[0,1]$. When $M=1$, any two nodes in the network are connected, and the network is called a fully connected network. When $M=0$, the network may be a fully isolated node, or it may be a connected network where the cluster coefficient of each node is 0.

2.1.3. Degree distribution. The average degree of the network is defined as follows,

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i$$

(6)

Among them, $\langle k \rangle$ represents the average degree of the network, and $n$ represents the number of nodes.

The probability that the degree value of a node in the network is a certain fixed value is usually used as the degree distribution function of the node in the network.

2.1.4. Betweenness centrality. The definition of betweenness centrality is as follows,

$$B(v) = \sum_{i \neq v \neq j} \frac{\sigma_{ij}(v)}{\sigma_{ij}}$$

(7)

$B(v)$ represents the importance of node $v$ in the entire network, and $\sigma_{ij}(v)$ represents the number of shortest paths from node $i$ to node $j$ and through node $v$.

2.1.5. Distance distribution and characteristic length. Referring to the definition process of the distribution function of degree distribution, we can define a distribution function $P(l)$ to describe the distribution of length $l$ in the network. This distribution function represents the ratio of the edge with length $l$ in the network to the total number of edges in the network. This distribution function $P(l)$ is the distance distribution function of the network. In complex networks, nodes are more inclined to connect nearby nodes than remote nodes. In order to characterize this characteristic, the concept of characteristic length is introduced. After research, it is found that the characteristic length and the distance distribution have an exponential distribution relationship.

2.2. The Fault Propagation Process of the Network

Complex networks typically face two types of attacks, random attack and selective attack. One for the network's own reasons, and one for deliberate destruction [9]. In complex network theory, fault propagation is defined as the process of system component faults causing a series of other faults and even spreading to the entire system is called fault propagation. That is, a small number of node faults may cause the entire network to collapse and cause catastrophic consequences.

In complex networks, the evaluation indicators are network efficiency and the relative size of the maximal connected subgraph. This paper describes fault tolerance by calculating the relative size of the maximal connected subgraph of a rail transit network [10].

2.3. Recovery Process after Network Fault

Regarding the network recovery process, because different system recovery resources are different, there are generally two situations. One situation is that there are fewer recovery resources, and only a few nodes or even one node and its adjacent nodes can be recovered at a time. One situation is that there are many recovery resources, and multiple nodes can be recovered at a time. If only a few nodes or one node can be recovered at a time, the recovery sequence is particularly important, which can directly determine the recovery ability of the recovery strategy. If considering the situation of more recovery resources, the recovery strategy is more flexible and the recovery speed is greatly accelerated. Under normal circumstances, when only a few nodes or one node can be recovered at a time, there are three common methods, random recovery, degree priority recovery, and betweenness
priority recovery. When multiple nodes can be recovered at one time, the most common recovery method is edge recovery.

3. Implementation and Analysis
This paper takes Beijing subway network data as an example.

3.1. Simulation of the Fault Spreading Process of the Network
3.1.1. Fault spread without recovery process. Through the Monte Carlo simulation of the urban rail transit network, the relationship between its giant component, its characteristic length and node survival rate after the occurrence of fault is calculated. Fig. 1 is the Monte Carlo fault spread simulation results of the urban rail transit network, \( p^\infty \) represents the size of the giant component, \( \zeta \) represents the length of the characteristic path, and \( P \) represents the survival rate of the node after the fault. It can be seen that the curve is similar to the first-order phase transition curve, and the critical threshold of network phase transition is about 78%. That is to say, when the network is attacked and the survival rate is higher than 78%, the fault spreading speed is relatively slow. When the survival rate is lower than 78%, the network collapses sharply.

![Figure 1. Simulation results of fault propagation](image)

3.1.2. The law of fault spread with edge recovery. Through Monte Carlo simulation, the relationship between the recovery probability of its the giant component and its edge recovery and the node survival rate after the occurrence of fault propagation is calculated, and the relationship between iteration times of fault propagation, recovery probability of edge recovery and node survival rate is discussed.

In the Fig. 2, \( \gamma \) is the edge recovery probability, and NOI represents the number of iterations of fault propagation. This indicator shows that using edge recovery can reduce the critical threshold of network phase change and improve the ability of the network to resist faults. It can be seen from Fig. 2 that when the edge recovery probability \( \gamma \) changes from 0 to 1, the number of iterations of fault propagation at the phase transition threshold is reduced from 7 to 6, which means that the time for fault propagation is reduced. Therefore, for the network, adjusting the edge recovery probability can play a role in adjusting the range of fault spread after a fault. Through the number of iterations, it can be found that the propagation time of the fault with edge recovery increases, which has an inhibitory effect on the fault propagation.
3.2. Simulation of Recovery Process after Network Fault
When discussing recovery, the model after fault propagation is the same.

3.2.1. Random recovery. As can be seen in the Fig. 3, random recovery nodes sometimes recovery multiple, sometimes one, sometimes not one, but relatively speaking, when the survival rate of nodes in the network is low, the number of nodes recovered each time is relatively large, in the network node survival rate is low, basically only one node is recovered or not recovered at a time. Reflected in the functional curve, the slope of the recovery initial time curve is significantly higher than that of the end of the recovery, i.e., the recovery speed is gradually slowing down.

3.2.2. Degree priority recovery. It is assumed that each iteration step takes 1 MS, and the time in the functional time variation curve of the following degree priority recovery and betweenness priority recovery is the iteration time.
As can be seen in Fig. 4, it has fewer iterative steps than random recovery, that is, the recovery time will decrease, because the number of nodes recovered during its recovery is higher than random recovery, resulting in a relatively fast pace of recovery, the recovery effect is better than random recovery. According to the network function curve of degree priority recovery, the recovery speed of degree priority recovery and random recovery are decreasing gradually.
3.2.3. Betweenness priority recovery. Where p is the survival rate of nodes in the failed network. As can be seen in Fig. 5, it has the fewest iterative steps and the shortest recovery time compared to random recovery and degree priority recovery, because the initial stages of its recovery are to recover many nodes, and the number of nodes that are fixed later will be smaller, and the recovery effect is best compared to random recovery and degree priority recovery. It can be concluded that the importance of using betweenness to describe nodes in a network is better than using degree values to describe the importance of nodes in a network. However, it is worth noting that although the betweenness priority recovery is relatively good, but not so obvious, compared with the degree priority recovery, the gap can be almost ignored. According to the network function curve of betweenness priority recovery, betweenness priority recovery and random recovery and degree priority recovery have one thing in common, and the rate of recovery is gradually decreased from high to low.

4. Conclusion

Based on the theory of complex network, this paper focuses on the fault propagation of urban rail transit networks and the recovery problems after faults. However, because the author's ability and knowledge are limited, the established model only stays at a shallow level, so the follow-up research can be carried out from several aspects,
(1) The perfect model of urban rail transit network. In this paper, only the model of the data level of the subway network is established, and if the practical application is considered, its power supply system and subway control network should be considered comprehensively, and the recovery strategy developed by this model is more instructive.

(2) This paper uses the existing random recovery, degree priority recovery, betweenness priority recovery and edge recovery and another recovery means to recovery network. I hope that in the future, we can study parameters that can better reflect the nature of the network, not limited to cluster coefficient, degree, betweenness and other parameters, through these parameters to propose more efficient and practical recovery methods.

Acknowledgments
This work was financially supported in part by Jilin Provincial Education Department’s “Thirteenth Five Year” scientific and technological research planning project (JJKH20190500KJ), supported in part by Jilin Provincial Special Foundation for Industrial Innovation (2019C047-6) and supported in part by Natural Science Foundation of Changchun Normal University (2018-015).

References
[1] D.J. Watts, S.H. Strogatz, Collective dynamics of ‘small-world’ networks, Nature. 393(1998) (6684) 440-442.
[2] L.M. Marchiori, Is the Boston subway a small-world network? Physica A: Statistical Mechanics and its Applications. 2002.
[3] P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, & S.S. Manna, Small-world properties of the Indian Railway network, Physical Review E Statistical Nonlinear & Soft Matter Physics. 67 (2002) (3) 036106.
[4] Bin, H., L. Fang, Z. Hou-Shun, Robustness of Complex Networks under Attack and repair, Chinese Physics Letters. 26 (2009) (12) 128901.
[5] Y.L. Chen, An algorithm for finding the quickest paths in a network, Computers and Operations Research. 20 (1993) 59-65.
[6] G. Concas, S. Marchesi, Pinna, N. Serra, Power-Laws in a Large Object-Oriented Software System, IEEE Transactions on Software Engineering. 33 (2007) (10) 687-708.
[7] J. Sienkiewicz, J.A. Holyst, Public transport systems in Poland: from Bialystok to Zielona Gora by bus and tram using universal statistics of complex networks, Acta Physica Polonica. 36 (2005) (5) 310-317.
[8] S.H. Strogatz, Exploring Complex Network, Nature. 410 (2001) (3) 268-276.
[9] R. Albert, H. Jeong, A.L. Barabasi, Diameter of the worldwide web, Nature. 401 (1999) 130-131.
[10] R. Albert, H. Jeong, A.L. Barabasi, Error and attack tolerance of complex networks, Nature. 406 (2000) 378-382.