Analysis of the Complex Characteristics of the New York City Subway Network

Xiaolong Ren [#, 1, *], Zhenyu Zhang [#, 2], Zixvan Yu [#, 3]

[1] Department of NUC University, College of Mathematics, Taiyuan, Shanxi, 030051, China

[2] School of GUT University, College of Mathematics and Applied Mathematics, Guangzhou, Guangdong 510006, China

[3] School of BNU, Zhuhai University, College of Future Education, Zhuhai, Guangdong, 519087, China

* Corresponding Author Email: yfbz0825@163.com

[#]These authors contributed equally

Abstract. Complex networks are important to the real world. Complex networks are instructive for the study of real networks such as metro networks, social networks and information networks. This paper study the New York City subway network and use complex network theory to research the static properties of subway complex network. Conclude that 79% of the nodes in the New York City subway network have degree 2, the network diameter is 59, more than 90% of the network nodes have a clustering coefficient of 0, the standard robustness of the network is 0.326, and the average shortest path length of the network is 18.38. This paper use deliberate attacks and random attacks on the network to study the dynamic Properties. The New York City subway network sites have good robustness against random attacks. In contrast, the network exhibits vulnerability when the maximum mesonumber nodes are targeted for deliberate attacks, and the entire network will be nearly paralyzed when more mesonumber nodes fail.

Keywords: Complex Network Theory, Robustness, Network Attacks.

1. Introduction

Complex networks are large-scale networks with complex topology and dynamic behavior. 1960s, the theory of complex networks emerged, and metro networks, as a typical realistic complex network, can naturally apply complex network theory to study its structural characteristics and provide theoretical basis and suggestions for government planning.

At present, people mainly construct Space-P or Space-L models of metro networks, calculate metrics such as degree and degree distribution, shortest path and its distribution and clustering coefficient of the networks, and analyze the metrics to study their characteristics. In the past decade, people have studied the topological properties and robustness and other characteristics of metro networks in many first-tier cities in China.

Tianzhi Gao et al. used 10 typical cities in China as research samples and analyzed their network topology characteristics and obtained that the nodes are in triangular layout [1]. Tong Bao et al. analyzed the characteristics of the Shanghai metro network and found that the whole metro network basically had no clustering characteristics [2]. Sujiang Zheng et al. analyzed the topological nature of Shanghai transportation network and found that the clustering coefficient of each station is mostly 0, and the whole metro network basically has no clustering characteristics [3]. Tiangang Qiang et al. analyzed the characteristics and robustness of the multi-modal transportation network in Harbin and obtained the conclusion that the clustering coefficients of the superimposed urban multi-modal transportation network increased compared to the single-modal sub-networks and the network had typical small-world characteristics [4]. Xin Chen et al. combined complex network theory and GIS technology to study the complexity characteristics and spatio-temporal evolution process of Guangzhou metro network, and obtained the conclusion that while the complexity of the Guangzhou metro network increased, most nodes in the network spread outward causing the network efficiency

to decrease [5]. Benkai Xie studied Zhengzhou city and obtained that the complex network is more robust to random attacks but presents vulnerability to deliberate attacks [6] Junjie Gan et al. tested the network robustness by simulating attacks on stations in the New York City subway network and found that only 15 nodes had to be destroyed to reduce the proportion of nodes in the maximum connected subgraph of the global network by half [7]. Chunchao Zhu et al. investigated the hazards of different deliberate attack strategies on network remoteness in dynamic attack mode and concluded that different deliberate attack algorithms have better attack effectiveness in dynamic attack mode [8]. J Zhang et al. studied the robustness of rail transportation networks by uncovering the characteristics of central networks [9].Ferber C V et al. compared the transport networks in London and Paris after suffering structural vulnerability after random and deliberate attacks, demonstrating that a sustained disruption of 0. 5% of the transport stations would lead to network collapse [10].

With the development of social economy, the scale of cities is getting bigger and bigger, and the traffic problem of cities becomes one of the important reasons to hinder the development of cities. Previously, people mainly study the topological characteristics of static complex networks, and less research is involved in the robustness of networks under attack. At the same time, domestic scholars are less likely to study complex networks in large foreign cities. In recent years, the United States has the closest contact with China. As New York is the largest city in the United States, the study of the topology and robustness of New York transportation network is of great significance both in theory and in reality.

In this paper, we first investigate the static topological properties of the subway network in New York, such as degree distribution, average degree, diameter, clustering coefficient, average path length, robustness, and local robustness. The network robustness is then studied using random and deliberate attacks on the network, while calculating the maximum connected subgraph size, connectivity, number of circles, efficiency, and average path length.

2. Network construction

A network can be abstracted as a graph G=(*V*,*E*) consisting of a point set *V* and an edge set *E*. Where *V denotes the set of urban rail transit network* stations, including *N*=|*V*| nodes. *E* denotes the set of urban rail network routes, including *M* = |*E*| edges. Each edge in *E* has a pair of points in V corresponding to it.

In this paper, we study the topological nature of the network, so that a subway network can be abstracted as an unweighted undirected graph. Nodes can be defined as subway stations, but there are different methods of describing the definition of edges between nodes. The description of a subway network can be generally summarized into two methods based on station connectivity: one is known as the Space-L method, where subway stations are considered as nodes, and if two stations are adjacent to each other on a particular subway line, then they are connected by edges. The other is the Space-P approach, where subway stations are considered as nodes, and if two stations have direct access to a subway line, then they are connected by an edge.

The network constructed by Space-L method reflects the geographical connection between subway stations and retains the basic topological characteristics of the subway network, while the network constructed by Space-P method can well reflect the interchange situation of the subway network. Since the subway is a fast transportation mode, both the geographical connection and the interchange situation are valuable to study. Therefore, in this paper, we study the Space-L method for the New York City subway network.

Two points are said to be connected if they are reachable between two transfer stations, i.e., there are edges between the two points. For the edge (v_i, vj), the adjacency matrix $A=(a_{ij})_{n\times n}$ can be constructed, where

$$a_{ij} = \begin{cases} 1, (v_i, v_j) \in E \\ 0, \text{other} \end{cases} \quad (1)$$

Figure 1 shows the New York City subway network. We number each station in turn and build the adjacency matrix. In this paper, we study the topological properties of the subway network. If a terminal station is also a transfer point, it is considered as a transfer point only. If there are duplicate paths between two stations, and two or more paths can reach each other, they are considered as one path, and finally the adjacency matrix is obtained.



Figure 1. New York City Subway Route Map

3. Network parameters

3.1. Static parameters

3.1.1 Degree distribution

The degree describes the importance of the nodes in the network. The degree k_i of a node i is defined as the number of edges connected to this node, and the average of the degrees k_i of all nodes i in the network is called the average degree of the network, defined as $\langle k \rangle$. The degree distribution of nodes in the network is represented by the distribution function $P(k)$, which describes the probability that the degree of an arbitrarily chosen node is exactly k.

3.1.2 Standardized average degree

The average degree of the network $\langle k \rangle$ divided by the maximum possible degree N-1 to obtain the normalized mean degree $\langle k \rangle b$ which is.

$$\langle k \rangle_b = \frac{\langle k \rangle}{N-1} = \frac{1}{N(N-1)} \sum_{i=1}^{N} k_i \tag{2}$$

3.1.3 Diameter

The distance d between two nodes i and j *in the network*$_{ij}$ is defined as the number of edges on the shortest path connecting these two nodes. The maximum value of the distance between any two nodes in the network is called the diameter D of the network.

3.1.4 Average path length

Average path length is a physical quantity that characterizes the small-world properties of complex networks. It can represent the average number of stations that need to pass through to complete a trip from the origin to the destination, indicating the average degree of separation between stations. The average path length *L of* the network is the average of the distances between all pairs of nodes.

$$L = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij} \tag{3}$$

3.1.5 Clustering coefficients

Suppose a node *i* in the network has k_i edges connecting it to other nodes, these k_i nodes are called neighbors of node *i*. The clustering coefficient c_i is the ratio of the actual number of edges e_i between all neighbors of node *i* and the total number of possible edges $k_i (k_i - 1)/2$, i.e.

$$c_i = \frac{2e_i}{k_i(k_i-1)} = \frac{\sum_{j,m} a_{ij} a_{im} a_{mi}}{k_i(k_i-1)} \tag{4}$$

a_{ij} is the adjacency matrix of the network.

The average of the clustering coefficients of all nodes in the network is called the clustering coefficient of the network and is defined as *C*, i.e.

$$C = \langle c \rangle = \frac{1}{N} \sum_{i \in N} c_i \tag{5}$$

3.1.6 Local Robustness

Typically, the local carrying capacity of a site is expressed in terms of node degree k_i, and the higher the value of node degree the higher the connectivity and the stronger the local carrying capacity and robustness of the site. After calculating the standard degree distribution, the transit site provides hubs for non-transit sites, and the point weight k_i of node v_i, the reciprocal *X* of the variance Y_i of the weight distribution is used to measure the local robustness, that is.

$$X = \frac{N}{\sum_{i=1}^{N} Y_i} = \frac{N}{\sum_{i=1}^{N} \sum_{j \in N_i} \left(\frac{1}{k_i}\right)^2} = \frac{N}{\sum_{i=1}^{N} \frac{1}{k_i}} \tag{6}$$

The variability of the weight distribution of node v_i Y_i reflects the degree of dispersion of the edge weight distribution. At this point, the non-transit sites are mainly concentrated on this one edge, and a failure at any point on the edge will cause the line to be disrupted, and the number of sites represents the probability of failure, i.e., the more sites need to be maintained, the weaker the robustness. In other words, the larger *X is the more robust the* network is.

3.1.7 Robustness

Robustness *r is a* physical measure of the robustness of a rail transportation network. It quantifies the robustness as the number of interchangeable paths in the network divided by the total number of stations, i.e.

$$r = \frac{ln(M-N+2)}{N} \tag{7}$$

r divided by the maximum possible network robustness to obtain r_b in order to compare the robustness of rail transit networks in different cities.

$$r_b = \frac{ln\left(\frac{N(N-1)}{2} - N + 2\right)}{N} \tag{8}$$

3.2. Dynamic parameters

3.2.1 Connectivity

The connectivity remains proportional to the resilience when external factors attack or internal faults disturb the network. Therefore, the network connectivity is chosen as one of the evaluation indexes for the robustness of metro line network in this paper.

The connectivity is the ratio of the actual number of edges of the network to the theoretical maximum number of edges. The connectivity is calculated as follows.

$$R = \frac{M}{3N-6} \tag{9}$$

3.2.2 Relative size of the maximum connected subgraph

The relative size of the maximum connectivity subgraph is a physical quantity that reflects the change in the topology of the subway network before and after damage or internal failure, and focuses on the degree of damage to the subway network.

A maximally connected subgraph is a subgraph that relies on a minimum number of edges to connect all nodes. The relative size of the maximum connected subgraph is the ratio of the total number of nodes in the subgraph to the number of all nodes in the network. If there is a change in the internal structure of the network or an external attack, the network topology will be split into multiple subgraphs and the relative size of the maximum connected subgraph will be reduced.

When a node in the original network is destroyed, the original network splits into several small collectives and a maximum connected subgraph. The relative size S of the maximum connected subgraph of the metro line network is.

$$S = \frac{N'}{N} \tag{10}$$

N' is the maximum number of nodes in the connected subgraph after an external attack or internal failure of the metro network topology.

3.2.3 Efficiency

Compared with the average path length, the global efficiency of the network is extremely accurate in reflecting the changes in connectivity reliability of complex networks before and after internal failures and external attacks. The global efficiency directly represents the connectivity reliability, and the most significant feature is that there is a positive relationship between global efficiency and connectivity reliability. The higher the alternative path provisioning power of the network, the higher the robustness. In case of internal failures or external attacks, the connectivity reliability decreases and the global efficiency decreases simultaneously, resulting in low alternative path provisioning and low robustness of the network. The global efficiency is calculated as

$$E(V) = \frac{1}{N(N-1)} \sum_{i \neq j \in V} \frac{1}{d_{ij}} \tag{11}$$

Obviously, $0 \leq E(V) \leq 1$. When $E(V) = 1$, it means that any two nodes in the topology of the subway network are directly connected to each other, and the network connectivity reliability enters the maximum peak, and the network alternative path provisioning capability also rises to the maximum, when the robustness reaches the maximum. When $E(V) = 0$, it means that all nodes in the topology of the metro network are not connected, i.e., each node is independent compared to the other nodes. At this point the connectivity reliability reaches its minimum value, the network has no ability to provide alternative paths, and the network robustness value is zero.

3.2.4 Number of laps

The robustness metric of the urban metro system focuses on the alternative path provisioning capability possessed by the network. The number of turns μ=M-N+1 in Berge's definition is used to represent the alternative paths that can be provided.

The number of circles also increases when the size of urban rail network gradually increases, although it does not mean that the larger the network size will have stronger robustness. The reason is that the number of nodes in a metro network increases as its size increases, and the probability of internal failures or external attacks increases. Therefore, the robustness of the subway network should be reflected by the circle rate.

$$\mu^T = \frac{\mu}{N} \tag{12}$$

3.2.5 Average path length

The most common evaluation metric is the average path length of the network: if the length is smaller, the connectivity reliability is higher; if the length is larger, the connectivity reliability will be lower. If the system is disrupted beyond a critical value, the network will have isolated nodes. The initial network average path length is then replaced by the maximum connected subgraph average path length.

4. Calculation results

4.1. Calculation results of static parameters

As seen in Figure 2, the maximum value of the node degree of the New York City subway network is 7, the minimum value is 1, and the average value of the network node degree is 2.178; the stations with node degree of 2 account for 79%, which indicates that most of the stations in the New York City subway network have connections with only two neighboring stations in the same line.

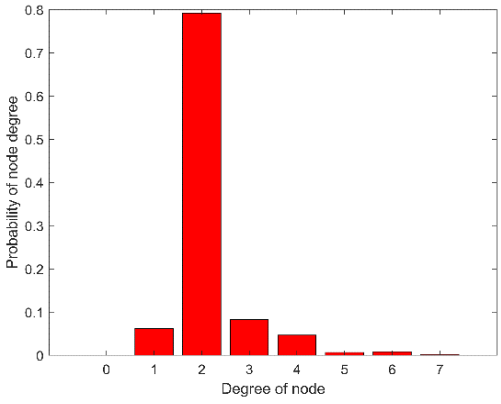


Figure 2. Degree distribution

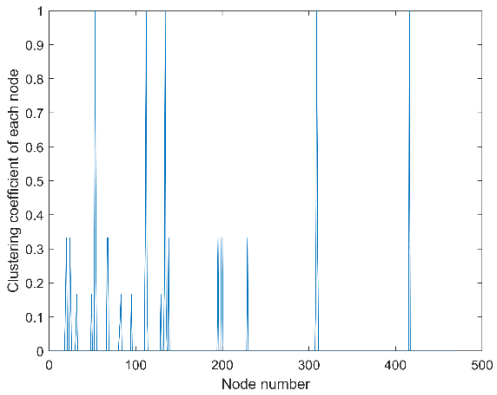


Figure 3. Clustering coefficients

The calculation yields the clustering coefficients of the nodes of the network in Figure 3

The results of the above analysis show that the Space-L network topology model shows very weak clustering among nodes, with 93% of the network nodes having a clustering coefficient of 0.

Table.1. Calculate the results of network static parameters

Network Characteristics	Standardized average degree	Average path length	Diameter	Local Robustness	Standardized Robustness
Value	0.0046	18.3834	59	2.0007	0.3259

4.2. Calculated results of deliberate and random attacks

The New York subway line network is used as the research object for random attack and deliberate attack simulation.

In order to observe more specifically the change process of the robustness indicators when Figure 1 is attacked, only one node is damaged in each attack, and the simulation is carried out using MATLAB, and the results are shown in Figure 4-8.

After random attacks, the relative size of the connected subgraphs of the New York City subway network decreases slowly, while after deliberate attacks, the network shows three cliff-like decreases,

indicating that nodes with higher mesonumbers are more important in the network and are responsible for connecting smaller subgraphs in parts of the network.

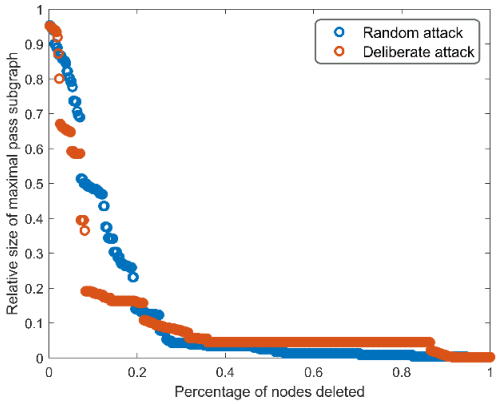


Figure 4. Relative size of maximum connectivity subgraph

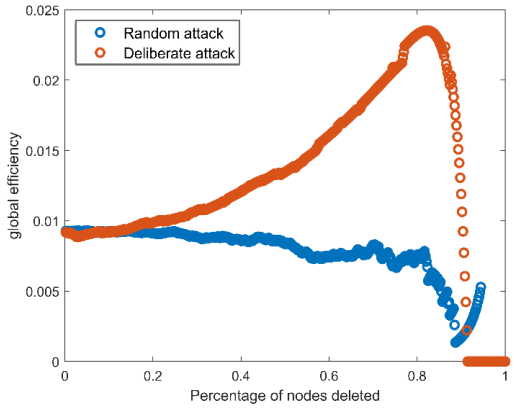


Figure 5. Global efficiency

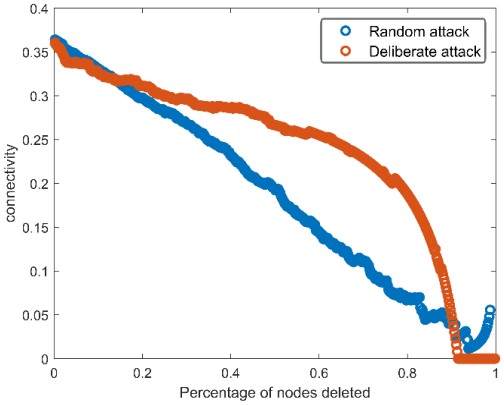


Figure 6. Network connectivity changes

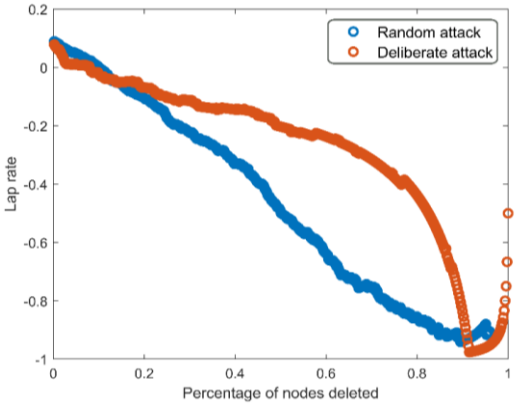


Figure 7. Network circle rate change

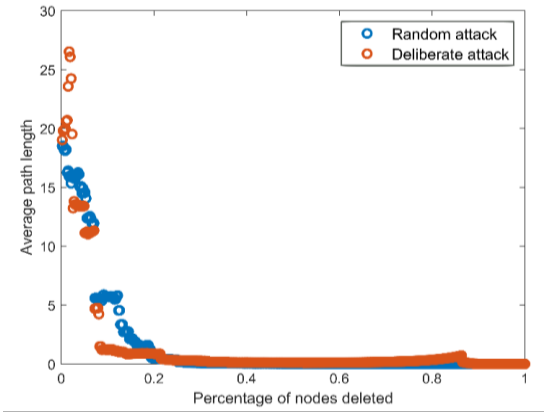


Figure 8. Network average path length variation

5. Conclusion

In this paper, an empirical study of the New York City subway network is conducted. Firstly, a network topology model of the New York City subway network is constructed using complex network theory, and its complex network characteristics are analyzed. Then the dynamic properties of the New York City subway network are studied based on the Space-L network topology model, and two strategies are formulated for deliberately attacking the nodes with high network mesonumber and randomly attacking the nodes of the New York City subway network, and simulation experiments are conducted to test the relative size of the maximum connectivity subgraph, network efficiency,

connectivity, number of circles, and average path length after the network suffers from random failure and deliberate attack. dynamic properties of the New York City subway network. The main conclusions obtained are as follows:

(1) By analyzing the complex network characteristics of the Space-L network topology model of the New York City subway network, the results show that 79% of the nodes in the network have a degree of 2, and the majority of stations are connected to only two stations before and after; the network diameter is 59, and the farthest distance between two stations in the New York subway network is 59 stations; 93% of the network nodes have a clustering coefficient of 0; the standardized robustness of the network is 0.326; the average shortest path length of the network is 18.38. This indicates that the average number of accessible stations between subway stations in the network nodes is 18; the standardized average degree of the network is 0.004635; 59% of the path lengths between nodes are between 10 and 30, which indicates that the average number of stations passed by traveling on the New York City subway is relatively high.

(2) The robustness of the New York City subway network was tested in a simulation experiment by deliberately attacking nodes with high meshes and randomly attacking them to disable the stations in the New York City subway network. The results show that the NYC subway network is robust to random attacks, while the network is vulnerable to deliberate attacks with nodes of the largest meshes, and the entire network will be nearly paralyzed when more nodes of large meshes fail.

In summary, the New York City subway network has a long average path length and individual nodes have a high clustering coefficient. The subway network has a typical network cluster structure and locally has the characteristics of a small-world network. With a limited number of attacks, the nodes with large meshes need to be maintained in a focused manner. In addition, metro managers should strengthen the daily maintenance and security checks on stations with high node degree values to guarantee the normal operation of important nodes and prevent causing large area paralysis of the metro network. In order to improve the global efficiency of the network, metro stations can be added between different network clusters to enhance the robustness of the network.

References

[1] Gao Tian-Zhi, Chen Kuan-Min, Li Feng-Lan. Topology analysis of urban rail transit network [J]. Journal of Chang'an University (Natural Science Edition),2018,38(03):97-106.

[2] Bao Tong, Li Jun-Jing, Guo Shu-Ting. Analysis of topological characteristics and passenger flow prediction of Shanghai rail transit network [J]. Information and Computer(Theory Edition),2021,33(20):62-67.

[3] Zheng Su-Jiang. Analysis of the nature of Shanghai metro network topology [J]. Intelligent Computers and Applications,2019,9(04):205-208

[4] Qiang Tian-Gang, Zhao Ming-Ming, Pei Yu-Long. Research on complex network characteristics and robustness of urban multi-modal transportation networks [J]. Transportation Information and Safety,2019,37(01):65-71.

[5] Chen Xin, Han Yin. Characterization of Guangzhou metro network topology and reliability evaluation [J]. China Water Transport(Second half month),2023,23(02):127-129.

[6] Xie Ben-Kai, Tu Xin-Yu, Li Qin. Robustness analysis of urban rail transportation network [J]. Science, Technology and Engineering,2020,20(28):11693-11697.

[7] Gan Jun-Jie, Nie Gui-Hua, Xu Di. Research on complex characteristics and robustness of Wuhan subway network [J]. Safety and Environmental Engineering,018,25(06):120-126.

[8] Zhu Chun-Chao. Research on the robustness of complex networks based on different deliberate attack strategies [D]. East China Jiaotong University, 2022.

[9] J Zhang, et al. Characteristics on hub networks of urban rail transit networks [J]. Physica A Statistical Mechanics & Its Applications,2016,447 . 502 - 507.

[10] C V Ferber, et al. A tale of two cities. vulnerabilities of the London and Paris transit networks [J]. Computer Science, 2012, 4(1): 74 - 89.