The Evolution Analysis of Guangzhou Subway Network by Complex Network Theory

Li Zhu, Jinlong Luo

School of Transportation, Southeast University, 35 Jinxianghe, Nanjing, Jiangsu, China, 210096 and Jiangsu Transportation Institute, 223 Shuiximen Avenue, Nanjing, Jiangsu, 210017, China

Jiangsu Transportation Institute, 223 Shuiximen Avenue, Nanjing, Jiangsu, 210017, China

Abstract

As traffic congestion becoming increasingly prominent, more and more cities begin to focus on the development of subway. It is of great significance to study the subway network characteristics and evolution properties based on the complex network theory. In this paper, we take Guangzhou subway network (GSN) as a case and present its six characteristics. Furthermore, the properties of the future GSN are systematically analyzed and its evolution properties are also investigated. The conclusion may be useful for the development of GSN in the future.

Keywords: Subway network; Complex network; Stochastic network; Network evolution;

1. Introduction

With the continuous expansion of urban scale, traffic congestion has becoming more and more serious, especially in major cities, such as Shanghai, Beijing, and Guangzhou. Providing priority for the development of public transport is recognized as an effective policy to mitigate traffic congestion by many scholars. Furthermore, subway is the best choice in big cities with its advantages of large capacity, rapid, and time reliability. There are 16 cities in...
China have already built subway, and meanwhile, subway is under construction in 16 other cities. In addition, although many cities do not have subway, they have long-term planning to construct. The construction cycle of subway is long; and the investment is huge. However, its bringing consequences are far-reaching, and thus studying the subway network’s characteristics and making reasonable planning and operation for subway are of importance for the development of urban transport system.

Since Watts [1] introduced a small-world network model and Barabási [2,3] proposed a scale-free network model, the complex network theory developed quickly. This theory is widely used in the field of physics, computers, and transportation systems currently. Many scholars investigated traffic network properties with the complex network theory. Latora and Massimo [4] studied the preliminary properties of Boston subway network, and they presented that the Boston subway network had high global efficiency and low local efficiency. Sienkiewicz and Holyst [5] analyzed the public transport network topology properties of 21 cities in Poland and found that their degree distribution obeyed either power law distribution or exponential distribution. Gao [6] discussed scale-free characteristics and the degree distribution exponent of the urban public transport network. Lu and Shi [7] explored the complexity of the public transportation network in China and indicated that it had the scale-free characteristics and small world characteristics. Gautreau [8] investigated the evolution of the US aviation network from 1990 to 2000, and found that with many routes added or disappeared, the statistics values changed in a very low level. Luis [9] studied the evolution of the Brazil aviation network from 1995-2006, and found that the network was a dynamic network. With changes in the importance of airports or airlines, traffic flow on the network had doubled during a period in which the topology structure of network had shrunk. Liu and Song [10] established the topology of Guangzhou subway network by space L method, and analyzed the value and distribution of the network degree, clustering coefficient, and average shortest way length. They also discussed the influencing degree and reliability of the network under the situation that the transfer station with malfunction. Wu [11] examined the complexity of Beijing subway network with statistical analysis. Liu and Tan [12] established the complex network models based on the method of space L and space P and then analyzed the stability under the failure of Wuhan subway. Zou [13] adopted the space-of-stations to build the network topology of Foshan transit network and discussed its static and dynamic robustness.

The main contributions of this paper are as follows.

- The six typical complex network indices are presented in this study.
- We analyze the topological characteristics of GSN from complex network perspective.
- The evolution properties of GSN in the future are investigated through the six typical complex network indices.

Therefore, we can have a better understanding of the network structure. The results will provide a reference for the other subways that under construction or planning to construct in the future.

2. Networks evaluation index

Subway network is composed by stations and lines, where the stations are denoted by nodes and rail lines connection between the stations are denoted by lines. We would use actual distance to indicate the connection length between two adjacent stations. Due to the fact that it is difficulty to collect the data and for the convenience of investigation, we assume that the length of each line is 1. That will not have much impact on the analysis results. The subway network can be seen as a graph G, and assuming that it has N nodes and K edges and its adjacent matrix is 

\[ A = (a_{ij}) \]

\[ a_{ij} = \begin{cases} 
1 & \text{node } i \text{ connecting node } j, \\
0 & \text{otherwise}. 
\end{cases} \]

In order to analyse the properties of GSN, the following six indexes are selected to analyze the evolution of this complex network.

- **Degree**
  
  Degree is used to describe the node characteristics in complex network. The degree \( k_i \) of a node \( i \) is the number of edges connecting with the node.

- **Degree correlation**
The degree correlation is defined as the average degree of the nodes adjacent to node i, as shown in Eq.2.

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in v(i)} k_j,$$

(2)

Where $v(i)$ is set of points adjacent to node i. Thus, average adjacent degrees of all nodes degree, their degree being $k$, is $k_{nn}(k)$.

$$k_{nn}(k) = \frac{1}{N_k} \sum_{i \in N_k} k_{nn,i}.$$  

(3)

Where $N_k$ is the number of node whose degree is k. The degree correlation represents the preference of choice among the nodes. If $k_{nn}(k)$ increases with k, namely the node with large degree prefers to connecting with other large degree nodes, the network is positive correlation. Conversely, if $k_{nn}(k)$ diminishes with k, i.e. the node with large degree prefers to connecting with the nodes owning small degree, the network is negatively correlated[13]. The degree correlation can be conductive to understanding that one fixed station will connect stations degree similar to it or not.

Newman proposed a convenient quantitative method to judge the correlation of network, namely calculate Pearson correlation coefficient of network node degree by the following formulation [14],

$$r = \frac{M^{-1} \sum_{ij} j_i k_i - (M^{-1} \sum_{i} \frac{1}{2} (j_i + k_i))^2}{M^{-1} \sum_{ij} \frac{1}{2} (j_i + k_i)^2 - (M^{-1} \sum_{i} \frac{1}{2} (j_i + k_i))^2}$$

(4)

where $j_i$ and $k_i$ are the two end side degrees of the i th edge, respectively. $i=1\cdots M$, where $M$ is the edges number. If $r>0$, the network is the positive correlation; If $r<0$, the network is negative correlation; and if $r=0$, the network has no correlation.

- **Average shortest path (L)**

The definition of $d_{ij}$ is the distance between station i and station j in the network, and it is namely the number of the shortest path edges connecting the two nodes. The network average shortest path length L is defined as the average distance between two nodes and has the form

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

(5)

where $N$ is the number of nodes. The network average shortest path length is also called network characteristic path length. L is smaller, the distance of any node in the network topology is smaller, and the accessibility of the overall network is better.

- **Average betweenness (B)**

The average betweenness can be used to characterize the number of shortest paths between two nodes that run through node i. Therefore, the node with larger average betweenness will has more traffic flow to pass. The average betweenness reflects the role and influence of the corresponding node in the entire network. The more the shortest paths pass the node, the bigger average betweenness will be, and the connectivity role of the node in the network is greater.

- **Clustering coefficient (C)**

In a given network, assuming that there are $K_i$ nodes adjacent to node i. Clustering coefficient is used to describe the aggregation of the nodes in the network, i.e. the tightness of the network. It has the form as follows.

$$C_i = \frac{E_i}{\frac{1}{2} K_i(K_i - 1)}.$$  

(6)
The entire network’s clustering coefficient is the mean of the clustering coefficient of each node, as shown in Eq.7.

$$C = \frac{1}{n} \sum_{i=1}^{n} C_i. \quad (7)$$

- Network efficiency (E(G))
  The network efficiency describes the overall connectivity of the urban mass transit network, as presented in Eq.8.

$$E(G) = \frac{1}{N(N-1)\sum_{i \neq j \in G} \frac{1}{d_{ij}}} \quad (8)$$

where N is the number of nodes, dij is the shortest path between node i and node j[15].

- Network standard entropy (E)
  Entropy is the measurement of the degree of "order". The network can be divided into the "disorder" and "order". The "disorder" network refers to the network that is randomly connected, and each node is roughly the same. However, the "order" network refers to the network that is scale-free, and this means that there are a few "core nodes" (known as the Hub node) with a large number of "peripheral nodes" in the network and the importance of nodes are different. Thus, we can use the network structure entropy to quantitatively measure the "order".

Network structure entropy E is defined as follows.

$$E = -\sum_{i=1}^{N} I_i \ln I_i \quad (9)$$

where Ii denotes the importance degree of node i, and it has the form,

$$I_i = \frac{k_i}{\sum_{i=1}^{N} k_i} \quad (10)$$

When the network is completely uniform, namely when Ii=1/N, E gets the maximum value, as shown in Eq.11.

$$E_{\text{max}} = -\sum_{i=1}^{N} \frac{1}{N} \ln \frac{1}{N} = \ln N. \quad (11)$$

When all nodes in the network are connected to a central node (we may assume that all connect with the first node), i.e. when k1=N-1 and kj=1 (j≠1), the network is the most uneven, and it has the minimum entropy (as shown in Eq.12).

$$E_{\text{min}} = -\sum_{j=2}^{N} \frac{1}{2(N-1)} \ln \frac{1}{2(N-1)} - \frac{1}{2} \ln \frac{1}{2} = \frac{\ln 4(N-1)}{2}. \quad (12)$$

To exclude the impact of the number of nodes, we make the network structure entropy normalized by the following formulation,

$$\bar{E} = \frac{E - E_{\text{min}}}{E_{\text{max}} - E_{\text{min}}} = \frac{-2 \sum_{i=1}^{N} I_i \ln I_i - \ln 4(N-1)}{\ln N^2 - \ln 4(N-1)}. \quad (13)$$

We call $\bar{E}$ as the network standard entropy, and obviously, $0 \leq \bar{E} \leq 1$. 
As the network becomes more even, the network standard entropy will be greater and it will be more "disorder" [16].

3. Network description

As one of the representatives of China’s subway network, GSN is China’s third largest city subway system, and its first line opened on June 28, 1997. There are a total of 9 operation lines and 160 stations in the network by 2013. The route length of GSN is 260.5 km. Subway has become the main means of transportation in Guangzhou City, and the average daily passenger flow is about 4.8 million.

At present, GSN will be planned to add Line 3 eastern extension segment, north-south extension segment of Line 4, Line 5 eastern extension segment, Line 6 eastern extension segment, Line 7, Line 8 north extension segment, Line 10, Line 11 (ring line), Line 12, Line 13, Line 14 and its branch, Line 16, and Line 20 in 2020. At that time, the total mileage of this network will arrive at 677 km, and there will be 305 stations, including 72 transfer hubs, as shown in Fig 1. The list of the various parameters of the planning for GSN by 2020 is illustrated in Table 1.

Table 1 show that there are a total of 305 stations and 370 connected edges for GSN by 2020. Each station is directly connected to an average of 2.43 stations, and it indicates that the cross-cutting between the GSN is relatively small, and the network is comparatively uniform. The Average shortest path is 14.39, and the network diameter is 50. The maximum travel distance of GSN is no more than 50 stops while the average distance is 14.39 stops. If the average station spacing is 1 km, passengers will take an average distance of 14 km to destination. Clustering coefficient of GSN is relatively smaller than that of random networks [17], and meanwhile GSN has a smaller average shortest path. Based on the complex network theory, the future GSN has the characteristics of random network.

4. Results analysis

4.1. Shortest Path

Statistics results of the path length for GSN are presented in Fig 2. It can be seen from Fig 2(a), the average shortest path length between two certain stations of GSN is about 11 stops, and the longest distance is 50 stops. Fig 2 (b) reveals that 60% of the stations can be reached by less than 14 stops, and more than 80% of the stations can be reached by less than 21 stops. That is, the average travel distance of passengers is small and subway network connectivity is relatively high. That shows that the expansion of GSN has already been quite complete.

<table>
<thead>
<tr>
<th>Characteristic Index</th>
<th>Plan Year (2020)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node</td>
<td>305</td>
</tr>
<tr>
<td>Edge</td>
<td>370</td>
</tr>
<tr>
<td>Betweenness</td>
<td>15.3947</td>
</tr>
<tr>
<td>Average clustering coefficient</td>
<td>0.0072</td>
</tr>
<tr>
<td>Average shortest path</td>
<td>14.3947</td>
</tr>
<tr>
<td>Network diameter</td>
<td>50</td>
</tr>
<tr>
<td>Average degree</td>
<td>2.4262</td>
</tr>
<tr>
<td>Path length equivalent of the random network</td>
<td>6.45</td>
</tr>
<tr>
<td>Clustering coefficient of random networks</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Fig. 1. Map of GSN in 2020.

Fig. 2. (a) the statistics probability of shortest path between stations; (b) the cumulative statistics probability of shortest path between stations.
4.2. Degree and Clustering Coefficient

Statistics results of degree distribution for GSN are showed in Fig 3. According to the results, we note that the GSN is relatively uniform, and line connectivity is higher. However, the degree of the subway network station is lower. By calculating the cumulative probability distribution of node degrees and fitting the splash, it shows that the degree of cumulative distribution probability $p(k)$ exhibits exponential distribution. Fitting results are $p(k) = 4.119e^{-0.936}$, $R^2 = 0.962$, and this fitting are good. Based on the complex network theory, notice that the GSN has characteristics of the random network.

![Fig. 3. degree distribution statistics.](image)

According to the topology of the planning for GSN in 2020, the clustering coefficient is 0.0072, and it is rather small. Fig.3 also indicates the poor station intensity of GSN, which can be seen from the difference between the GSN and other transport network. The characteristic statistical values of GSN are illustrated in Table 2.

<table>
<thead>
<tr>
<th>Year</th>
<th>Deg._Ave.</th>
<th>Deg.Cor.</th>
<th>L_Ave.</th>
<th>B_Ave.</th>
<th>C_Ave.</th>
<th>E_Ave.</th>
<th>E(G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>1.600</td>
<td>-0.33</td>
<td>2.000</td>
<td>3.000</td>
<td>0</td>
<td>0.7766</td>
<td>0.2000</td>
</tr>
<tr>
<td>1999</td>
<td>1.875</td>
<td>-0.07</td>
<td>5.667</td>
<td>6.667</td>
<td>0</td>
<td>0.9747</td>
<td>0.0625</td>
</tr>
<tr>
<td>2002</td>
<td>1.917</td>
<td>-0.01</td>
<td>5.840</td>
<td>6.8406</td>
<td>0</td>
<td>0.9536</td>
<td>0.0417</td>
</tr>
<tr>
<td>2003</td>
<td>1.936</td>
<td>-0.01</td>
<td>7.440</td>
<td>8.4409</td>
<td>0</td>
<td>0.9685</td>
<td>0.0323</td>
</tr>
<tr>
<td>2005</td>
<td>2.051</td>
<td>-0.09</td>
<td>8.041</td>
<td>9.0418</td>
<td>0</td>
<td>0.9712</td>
<td>0.027</td>
</tr>
<tr>
<td>2006</td>
<td>2.035</td>
<td>-0.07</td>
<td>10.083</td>
<td>11.0827</td>
<td>0</td>
<td>0.9701</td>
<td>0.0182</td>
</tr>
<tr>
<td>2007</td>
<td>2.034</td>
<td>-0.07</td>
<td>10.6651</td>
<td>11.6651</td>
<td>0</td>
<td>0.9714</td>
<td>0.0175</td>
</tr>
<tr>
<td>2009</td>
<td>2.100</td>
<td>0.038</td>
<td>10.538</td>
<td>11.5308</td>
<td>0</td>
<td>0.969</td>
<td>0.0133</td>
</tr>
<tr>
<td>2010</td>
<td>2.092</td>
<td>0.141</td>
<td>13.541</td>
<td>14.5414</td>
<td>0</td>
<td>0.9771</td>
<td>0.0081</td>
</tr>
<tr>
<td>2012</td>
<td>2.166</td>
<td>0.084</td>
<td>12.893</td>
<td>13.8927</td>
<td>0</td>
<td>0.9729</td>
<td>0.0075</td>
</tr>
<tr>
<td>2013</td>
<td>2.188</td>
<td>0.062</td>
<td>13.383</td>
<td>14.383</td>
<td>0</td>
<td>0.9753</td>
<td>0.0065</td>
</tr>
<tr>
<td>2014</td>
<td>2.192</td>
<td>0.022</td>
<td>14.778</td>
<td>15.7775</td>
<td>0</td>
<td>0.9758</td>
<td>0.0053</td>
</tr>
<tr>
<td>2015</td>
<td>2.304</td>
<td>0.069</td>
<td>13.979</td>
<td>14.9786</td>
<td>0.0022</td>
<td>0.9704</td>
<td>0.005</td>
</tr>
<tr>
<td>2020</td>
<td>2.426</td>
<td>0.035</td>
<td>14.395</td>
<td>15.3947</td>
<td>0.0072</td>
<td>0.9649</td>
<td>0.004</td>
</tr>
</tbody>
</table>

5. Evolution of GNS

According to the statistical data in Table 2, we discuss the evolution properties of GSN in this section. Firstly, the evolution of the average degree and degree correlation from 1997 to 2020 is analyzed, and the results are shown in Fig.4 (a) and Fig.4 (b).
From Fig 4(a), it can be seen that the average degree increase with the year evolution. The average degree almost remains the same during 2005-2007, because most newly opened subway lines are the extension of the old lines. In 2005, the first segment of the Line 3 (from Guang-Zhou-Zhan to Ke-Cun) and Line 4 (from Wan-Sheng-Wei to Xin-Zao) were went into operation. In 2006, first-stage of Line 3 (from Guang-Zhou-Dong-Zhan to Fan-Yu-Guang-Chang, from Tian-He-Ke-Yun-Zhan to Shi-Pai-Qiao) and Line 4 (from Xin-Zao Station to Huang-Ge) were went into operation. In 2007, the first-stage of whole Line 4 (from Wan-Sheng-Wei to Jin-Zhou) was went into operation. Most of the new lines lie on the edge of network topology, and conclusions conform to the reality. In a word, the average degree of the network is relatively small, and only a few transfer stations may appear crowded.

As shown in the Fig. 4 (b), note that 2007 was a turning point. Before it (including 2007), the topology of GSN is a negative correlation, and those years can been seen as a period of rapid space expansion of GSN. We can clearly see that the network tends to connect the small degree nodes. This indicates that the new lines are rather more and new lines have a low connectivity with the old lines, which are also consistent with the evolution of degree and consistent with the reality. The network topology structure is positive correlation after 2009. This means that after a period of spatial expansion of GSN, it begins to increase its own density. In reality, most of the new stations are in the internal of network, and new stations make the network closer. To improve the transport environment, the policy-makers should also focus on improving the density of the subway network.

Secondly, we investigate the evolution of average shortest path from 1997 to 2020, as shown in Fig.5 (a). It can be seen that the average shortest path increases. In some years, the average shortest path is lower than that of the previous years, due to the fact that the opened new lines optimize the GSN and increase the number of transfer stations. With the expansion of the network, the average shortest path will increase to a certain extent, and then it will stabilize and decline slightly. The reason for this is that with further optimization of the network, the transfer will be more convenient, and the average shortest path will decline.

Thirdly, the evolution of network average betweenness from 1997 to 2020 is analyzed, and the results are illustrated in Fig.5 (b). One can notice that the average betweenness increases with the expansion of this network. With the increase of the subway lines, the number of the shortest path also increase. The average betweenness in
some years is less than that of the previous years, and the reason for this is same as that for the average shortest path. Generally speaking, passengers will choose the shortest path to reach their destination. The subway stations that have greater average betweenness will attract more passengers, and hence the transfer stations owning large average betweenness will be very congested, such as the Guang-Zhou-Zhan, Ti-Yu-Xi-Lu Station. The managers should pay more attention to the connection of transfer modes, station space layout, and passenger evacuate etc, when this type of transfer stations are planned and constructed. It can be found the variation of the average betweenness and the average shortest distance are almost the same. The reason for this is that the opening of new subway lines has the same effect on the change trend of these two parameters.

Then we discuss the evolution of efficiency from 1997 to 2020 and the results are presented in Fig.6 (a). The network efficiency gradually declines from 1997, and the conclusion is consistent with the average shortest path. Along with optimization of network, network efficiency will be reduced to a certain extent and then keep stable with a slight increase. Overall, the passengers will choose the shortest path to their destination.

At last, the evolution of network standard entropy from 1999 to 2020 (the first subway is opened in 1997, and thus there is no network. Consequently, we have not statistical value of that year) are investigated and shown in Fig.6 (b). According to the concept of by the standard entropy, the greater it becomes, the more uniform and "disorder" the network will be. As presented in Fig.6 (b), the GSN is rather "disorder". In the network, the majority of nodes degree is 2 and only a few nodes’ degree is 1or 3 or other values. Therefore, we conclude that the network is still relatively uniform. The maximum degree in 2020 is 8, and the node is Ti-Yu-Xi-Lu Station. The networks have a small number of "core nodes" and a large number of "peripheral nodes”. Once the main degrees are different, the network standard entropy will decline. When there are more hubs formed in the network relatively, the network structure entropy will be smaller. Notice that the topology of GSN is relatively uniform.

6. Conclusions

Based on the concept of the complex networks, firstly, this paper discussed the complexity of the medium-term planning of GSN. By the end of 2020, average degree, average clustering coefficient, and average shortest path of GSN were very small, and GSN showed the characteristics of random networks.

Secondly, this article investigated the evolution of GSN. By introducing six typical statistics parameters of complex network, we systematically analyzed the evolution of the network parameters, and the results indicated that the radial network development might lower passenger transport efficiency and bring enormous pressure to the transfer stations. This provided a theoretical reference for the policy-makers and was conductive to the development of GRN in the future. Furthermore, it also could provide a theoretical basis for the development of subway, and contribute to the further expansion of this idea.

References


