**Analysis of Shenzhen Public Transportation Network**

21S051006 李峥岑

21S151082 常新乐

21S151120 尤志恒

1. **INTRODUCTION**

As the leading city in China, Shenzhen's high economic development is inseparable from its convenient public transportation. For most migrant workers in Shenzhen, the high coverage of public transportation network makes them easily to get around Shenzhen. In order to explore the influence of Shenzhen’s metro and bus on Shenzhen public transportation, we analyze the metro and bus route map of Shenzhen. We have collected the information about metro lines and bus lines in Shenzhen, as well as their correlation (i.e. the connection between metro stations and bus stations). Through computer simulation, we analyze the Shenzhen public transportation network properties and dynamical behaviors to make some conclusion about the network.

1. **Data acquisition and data preprocessing**
   1. **Data acquisition**

At first, we obtained the metro line information from the official website of Shenzhen Metro and bus line information from the official website of Shenzhen Transport Bureau. From these two types of public transport data, we can get two networks of very different sizes. In order to combine the two networks, we also use crawler to get the transfer relationship between metro station and bus station (i.e. the transfer bus station at the exit of each metro station) from the official website of Shenzhen Metro.

In the bus route information, most bus stations have latitude and longitude information, which can be used to map the network. Similarly, we also crawled the latitude and longitude information of each metro station from Baidu Map API for the subsequent mapping.

* 1. **Data preprocessing**

In most cases, collected data are complex and of different types, so data collection and preprocessing are necessary.

As shown in Figure 2-1, some data formats are inconsistent in the documents downloaded from the official website. Therefore, such problems need to be solved before data reading. In the simulation of Shenzhen's public transport network, we connected every two adjacent stations in the same metro or bus line. However, in the data of Shenzhen metro, Line 2 and Line 8 are two separate lines, while in the Shenzhen metro line map, the two lines are connected, so we uniformly stipulate they are the same line in the simulation.

示意图

描述已自动生成

Figure2-1: Inconsistent data formats

图表, 瀑布图

描述已自动生成

Figure2-2: Partial sketch of metro line

In total, Shenzhen has 10 metro lines (we regard Line 2 and Line 8 as the same line, because there is an edge between them that connects them end to end on the Shenzhen metro map) and 927 bus lines.

1. **Analysis of metro network**
   1. **Network and basic Information**

According to the data obtained and preprocessed by us, we draw the schematic diagram of Shenzhen metro network as shown in Figure 3-1, in which the location of each metro station corresponds to their longitude and latitude position on the actual map.

图表, 折线图

描述已自动生成

Figure3-1: Metro network graph

From Figure 3-1, we can know that the metro network graph consists of 237 nodes and 272 edges.

* 1. **Network properties**

From Figure 3-2 and Figure 3-3, we can know that in the metro network, most of the nodes have a degree of 2, because many stations on the metro line have only one line passing through them. And there are 11 nodes with degree 1 so we can infer that there are 11 stations in the whole metro network that are only starting stations (terminal stations) but not way stations. Similarly, other nodes with degree greater than 2 mean that they are the stations in the metro line which can transfer to different metro lines, such as Shenzhen North station. And the node with the maximum degree is the station with the most metro lines passing through.

图表, 直方图

描述已自动生成

Figure3-2: Degrees of nodes of metro network

图表, 直方图

描述已自动生成

Figure3-3: Degrees distribution of nodes of metro network

Since we simulate the metro line network as a connected component, we can directly find the shortest path between all nodes in the connected component. When calculating the shortest path between nodes, we write our own algorithm to calculate the shortest-path length. We choose Floyd algorithm between Dijkstra and Floyd algorithm because it adopts the idea of dynamic programming and can record the shortest path between two nodes, so as to facilitate the solution of the shortest-path length between the following nodes. As a result, we calculate that the average shortest-path length of the metro network is 13.583851820067224.

According to Figure 3-4, we can know that the clustering coefficient of every node is less than 0.2, even 0, and by calculating that, we find that the clustering coefficient of the network is 0.0036568213783403653. Obviously, it is a so small number that we can easily know the Shenzhen metro network has poor robustness and might be weak under attack. Generally speaking, if one station is attacked, there is a high probability that the entire metro line will be paralyzed (i.e. The network will be divided into two connected components, thus making many nodes unable to communicate with each other).

图表, 直方图

描述已自动生成

Figure3-4: Clustering coefficient of nodes of metro network

图表

描述已自动生成 图表

描述已自动生成

1. 1-core (b) 2-core

图形用户界面, 文本, 应用程序, Word

描述已自动生成

1. 3-core

Figure3-5: Coreness calculating (metro network)

The calculating of the coreness of the network step by step as is shown in the figure 3-5. Obviously, metro network’s coreness is 2. Considering a higher-coreness node is more important than a lower-coreness, there are 170 most important nodes in 237 nodes, thus its rubustness is extremely low.

* 1. **Dynamical behaviors**

We attack nodes of metro network in both intentional and random ways, and take the average shortest-path length and size of the largest connected subgraph as the standard. The results are as Figure 3-6 and Figure 3-7. In the intentional attack, we attack the nodes with the maximum degree of the metro network. We can find that the intentional attack is clearer when the number of attacked nodes is small, and influence of random attack seems more average. We can also find that when the nodes with the maximum degree are attacked, the average shortest-path length and size of the largest connected subgraph drops dramatically. It shows that the metro network has poor robustness again.

图表, 折线图

描述已自动生成

Figure3-6: Average shortest-path length under attack (metro network)

图表, 折线图

描述已自动生成

Figure3-7: Size of the largest connected subgraph under attack (metro network)

1. **Analysis of fusion network**

In Shenzhen, the metro enables fast commutes between two distant locations. However, through the study of metro network, we know that it is difficult to guarantee the convenience and unimpeded of Shenzhen public transportation by metro alone. On the one hand, two adjacent metro stations may be far away from each other, and people in between still have trouble when they want to take the metro. On the other hand, because of the poor robustness of the metro network, the destruction of one node may lead to the breakdown of the whole network.

For the solution to the first problem, we thought of Shenzhen's bus network. As we know, the distance between bus stations is short, which is also a good means of transport. It can solve the "last mile" problem and make up for the deficiencies of the metro. As for the second problem, almost every metro station has transfer stations at its exits. We regard each metro station as connected with the bus stations that can be transferred at its exits, thus connecting Shenzhen metro network and bus network to get a fusion network.

* 1. **Network and basic Information**

When drawing the bus network, we also reserved only the bus station nodes with latitude and longitude information in order to draw the network better.

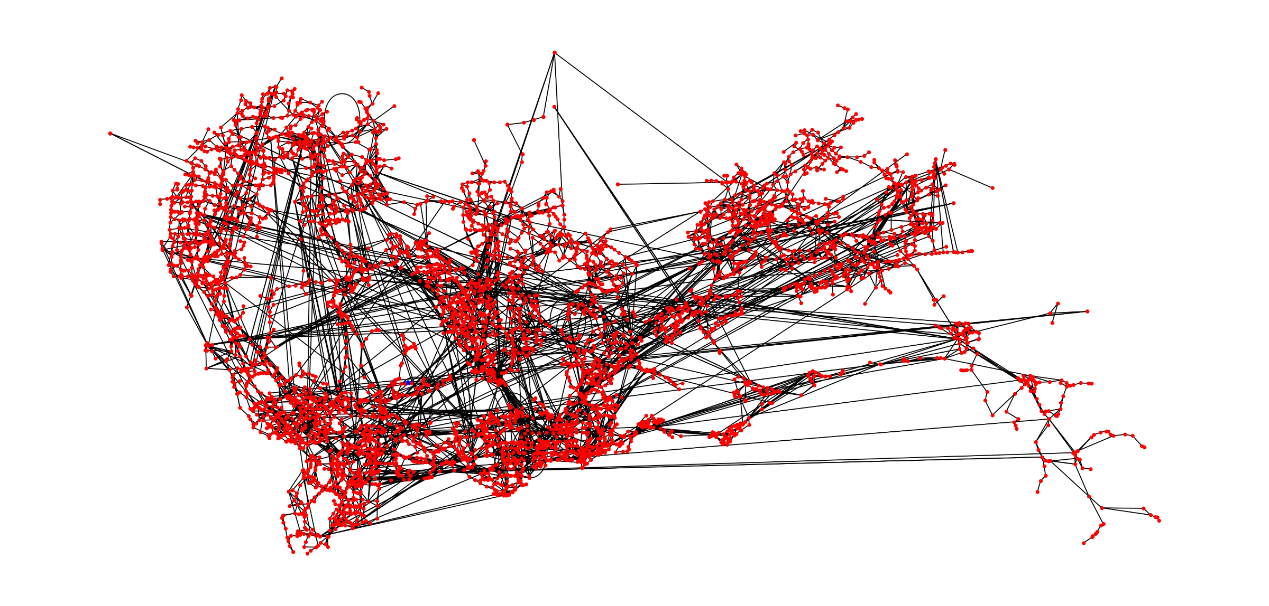


Figure4-1: Fusion network graph

From Figure 4-1, we can know that the fusion network graph consists of 5723 nodes and 11021 edges. It means there are more and denser public transport nodes spread over the same size area (Shenzhen), making Shenzhen public transport realize the real sense of convenience.

* 1. **Network properties**

Similarly, bus lines and metro lines are divided by route, so the number of nodes with degree 2 is still large. From Figure 4-2 and Figure 4-3, we can know that in the fusion network, the nodes with degree 1 also represent the starting and ending stations of the line. The degree of many nodes greater than 2 indicates that the complexity of the fusion network has been improved. Generally speaking, people have more transfer options at these public transit stops, making public transport more flexible.

By our calculations, in the fusion network, the average degree of nodes is 3.5728186831008757.

图形用户界面, 图表, 直方图

描述已自动生成

Figure4-2: Degrees of nodes of fusion network

图表

描述已自动生成

Figure4-3: Degrees distribution of nodes of fusion network

We use Floyd algorithm which is the same as metro network to evaluate the average shortest-path length of the fusion network and calculate that the average shortest-path length of the fusion network is 10.688882573122035.

According to Figure 4-4, we can know that the clustering coefficient of some nodes is 1, which means that there are complete subgraphs in the network. At the same time, the proportion of nodes with clustering coefficient of 0 also decreased a lot. In addition, the clustering coefficient of the whole network calculated by us is 0.17157763714305918.

Therefore, the fusion network have better rubustness than metro network.

图表, 条形图, 直方图

描述已自动生成

Figure4-4: Clustering coefficient of nodes of fusion network

图表

描述已自动生成

1. 1-core

图表, 散点图

描述已自动生成

1. 2-core

图表

描述已自动生成

1. 3-core

图表

描述已自动生成

1. 4-core

图表

描述已自动生成

1. 5-core

图形用户界面, 文本, 应用程序, Word

描述已自动生成

1. 6-core

Figure4-5: Coreness calculating (fusion network)

The calculating of the coreness of the fusion network step by step as is shown in the figure 4-5. Obviously, the fusion network’s coreness is 5, thus its rubustness is better than the metro network.

* 1. **Dynamical behaviors**

We also attack nodes of the fusion network in both intentional and random ways, and take the average shortest-path length and size of the largest connected subgraph as the standard. The results are as Figure 4-6 and Figure 4-7. In the intentional attack, we attack the nodes with the maximum degree of the metro network.

We can find that when the number of the intentional attack is low, the average shortest-path length increases. This is because when some nodes of high degree are deleted, the fusion network does not become disconnected. And if the direct shortest path of the original two nodes cannot be reached, they might be connected by a longer path. However, the surge in average shortest-paths length when the nodes with the highest degree are deleted also shows that some important traffic stops are influential as transportation hubs. In contrast, random attack has little influence on the average shortest-path length of fusion network.

As for the size of the largest connected subgraph, its changing trend in the composite network is roughly the same as that in the metro network. Intentional attack makes the network easier to divide the whole connected network into many smaller connected subgraphs than random attack.

图表, 折线图

描述已自动生成

Figure4-6: Average shortest-path length under attack (fusion network)

图表, 折线图

描述已自动生成

Figure4-7: Size of the largest connected subgraph under attack (fusion network)

1. **Conclusion**

In the process of analyzing Shenzhen public transport network, we first studied Shenzhen metro network and drew some conclusions. Then we combine Shenzhen metro network and bus network to get a fusion network of Shenzhen public transport. The comparison of the network properties of the two networks is shown in Table 5-1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Network | Nodes | Edges | Average Degree | Average S.P. Length | Clustering Coefficient | Coreness |
| Metro | 237 | 272 | 2.29 | 13.58 | 0.0036 | 2 |
| Fusion | 5723 | 11021 | 3.57 | 10.68 | 0.1715 | 5 |

Table5-1: Difference of the metro network and fusion network

From the scale of network, the fusion network is obviously larger than the metro network, including the number of nodes and edges. Meanwhile, the fusion network has higher average degree and shorter average shortest-path length. It means the fusion network has better overall coverage and more flexible path selection between two nodes. What’s more, the higher clustering coefficient and coreness of the fusion network also indicates that it has better robustness than the metro network.

All in all, the convenient transportation in Shenzhen is inseparable from the combined action of metro network and bus network. The former ensures fast commuting over long distances, while the latter ensures wide area coverage and improves the flexibility of the overall transport network.

1. **GUI of the analysis**