Study on characteristics and community structure of high-speed rail network centered on Shanghai

Jiayi Sun *

School of Sciences, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018

* Corresponding Author Email: jy_sun1205@163.com

Abstract. The high-speed railway is a typical representative of China 's high-speed transportation infrastructure, with typical complex network characteristics. Based on the complex network theory, this paper conducts an empirical study on a Space-P network model for high-speed train originating from Shanghai. Then the network is divided into communities by louvain algorithm. The average distance of the network is 1.76, the modularity of community division is 0.3683, and the clustering coefficient is 0.6859, which is typical of small-world network. Due to Space-P model, the network is highly connected, resulting in a relatively low modularity. To a certain extent, the network reflects the developed high-speed rail level and wide direct range of Shanghai, and also reflects the convenience of China 's high-speed rail.

Keywords: Complex networks, Community divisions, High-speed rail networks.

1. Introduction

As a typical representative of China's high-speed transportation infrastructure, the network and modernization of high-speed railway constantly reshape the transportation mode in China. The nodes in the high-speed rail network have become a hot topic of current academic research. Cities are widely distributed and widely connected

In recent years, many scholars at home and abroad have conducted in-depth analysis and research on the high-speed rail network from the perspective of complex network. Zhang Lanxia et al. used the complex network theory to analyze the characteristics of the two network models of high-speed railway geographic network and traffic flow network [1]. The study shows that the geographical network of high speed railway has tree structure and the distribution of power distribution. Li xin to urban agglomeration as the research object, respectively build the Beijing-Tianjin-Hebei, Yangtze river delta and the pearl river delta urban agglomeration of high-speed geographic network and service, for the three two network models of topology structure and centrality of analysis, and combined with the problems existing in the network structure of urban agglomeration high-speed infrastructure layout reasonable Suggestions for [2]. Fang Dachun et al. used network analysis methods such as network density, centrality index and condensed subgroup to conduct an in-depth study on the spatial structure evolution of the network of the high-speed railway network in the Yangtze River Delta [3]. Qi Qingyu uses the data of high-speed railway stations and high-speed railway lines in 2019 to reveal the topology structure of China's high-speed railway network from the non-scale, clustering coefficient and centrality[4]. Wang Wei et al., through the analysis of the constructed high-speed railway service network, the study showed that the high-speed railway service network has no scale characteristics and small world characteristics[5]. GuX et al. constructed China railway passenger network according to the railway ticket database, and studied and analyzed the correlation between the node degree index and other characteristic indexes in the passenger network[6]. Zhang J Using the complex network theory, they have analyzed the robustness of high-speed railway networks in China, the United States and Japan in the face of two different deliberate attacks, and obtained the of high-speed railway stations with weak robustness in the network[7]. Liu Guo et al. used the complex network theory to analyze the evolution characteristics of China's railway network from 2003 to 2015, and found that the density of China's high-speed railway network is constantly increasing, but there are still node cities in the east, central and western divided into [8].

At present, the research on high-speed railway network mainly analyzes the topological structure of high-speed railway network by using the complex network theory, while the related research on community structure is rarely applied to high-speed railway network. This paper studies the community structure of the society structure in the high-speed railway network. Through the study of the division of community, can better find high-speed rail network connectivity in better urban areas and poor connectivity between cities.At the same time from the perspective of community structure for high-speed rail line layout and optimization put forward reasonable Suggestions[9-12].

2. Network construction

2.1. Data source and processing

The data used in this paper are mainly obtained through the website of Railway 12306, obtained all high-speed train information from December 30-31,2022 through Python crawler, and extracted the trains with Shanghai as the starting point or destination as network research samples, a total of 173 high-speed trains.

However, since a city or county has multiple sites, this paper builds a network based on the following principles: if a city has multiple sites, these sites will be treated as one and named after the city. For example, Shanghai Hongqiao station, Shanghai station are all in Shanghai, so these stations are regarded as a station, and named Shanghai. A total of 259 nodes were obtained.

2.2. Network construction method

At present, the main methods to constructing topological model of transportation network include Space-P, Space-L and Space-R. Space-P method is mainly used to build the transportation network when node city i and node city j through the same high-speed train is connected between two nodes city space-L method is used to express the real network, when two node city is located in the same line and adjacent edge, this method to build the transportation network is closer to the actual line planning. The Space-R method is suitable for complex networks with a single structure. Because the constructed high rail network only represents the connectivity between node cities and only analyzes the network topology. Therefore, the Space-P method was selected for the analysis.

The complex network of high-speed railway constructed by the institute is composed of nodes and edges. The nodes are the city where the high-speed railway is located, and the side represents whether this mode of transportation is accessible between any two cities. Node set V (G) and the edge set E (G) constitute a complex network graph G (E, V) each edge in the edge set corresponds to a pair of point set (i, j) and (j, i), any edge $|e|=1$.

It can be concluded that the network has 259 nodes and 8058 edges. The size and connectivity of the network reflect the developed traffic of Shanghai to a certain extent.

3. Statistical parameter

3.1. Overall network index

3.1.1 Average degree

The degree of a node refers to the number of edges connected to a certain node, denoted as k_i. Network average degree is the mean degree of all nodes in the network, denoted as $<k>$.

$$< k \geq \frac{1}{N}\sum_{i=1}^{N} k_i \tag{1}$$

Where: N is the total number of nodes in the network.

3.1.2 Average path length

In the network, the shortest path between node *i* and node *j* is recorded as d_{ij}. The average path length is the average of the shortest path length between any two nodes in the network and is represented in *L*.

$$L = \frac{1}{N(N-1)} \sum_{i \geq j} d_{ij} \tag{2}$$

3.1.3 Clustering coefficient

The clustering coefficient describes the closeness of the connection between a node and its adjacent nodes. C_i is the clustering coefficient of the node *i*, E_i is the number of adjacency pairs that actually exist between node *i*, K_i is the number of edges in the network connected to the node *i*.

$$C_i = \frac{2E_i}{k_i(k_i-1)} \tag{3}$$

3.2. Network centrality index

3.2.1 degree of centrality

In a network, if a node is connected to many other nodes, then the node is in the central position, that is, the greater the degree of the node, the more important the node is.

$$DC_i = \frac{k_i}{N-1} \tag{4}$$

3.2.2 Mediation number centrality

Mediation centrality measures the transfer ability of a node and the ability to control other nodes. The higher the mediation centrality of a node, the stronger the transit ability and control ability over other nodes.

$$BC_i = \frac{1}{(N-1)(N-2)} \sum_{i=1, j \neq k}^{N} \sum_{j \neq k}^{N} \frac{\delta_{ij}^{k}}{\delta_{ij}} \tag{5}$$

In formula: δ_{ij} is the shortest number of paths, from node *i* to node *j*, and δ_{ij}^{k} is the number of times that all the shortest paths from node *i* to node *j* pass through node *k*.

4. Network feature analysis

4.1. Degree distribution

In this network, the degree values reflect the number of nodes directly connected by each node. Relevant software is used to calculate the degree value of each node in the whole network is shown in Figure 1.

The overall network node degree value is 259, of which there are seven stations with degrees greater than 150, indicating that these high-speed rail stations are directly connected to more than 150 high-speed rail stations in the network. These nodes are Shanghai, Zhengzhou, Xuzhou, Nanjing, Suzhou, Hangzhou and Changzhou. Since there are few nodes with high degree value in the network, while most nodes have low degree value, the network is dominated by a few node cities with high degree value and connected with node cities with low number of degree value, the degree value k and degree distribution probability *P* (*k*) is a power law pass, which conforms to the characteristics of scale-free network.

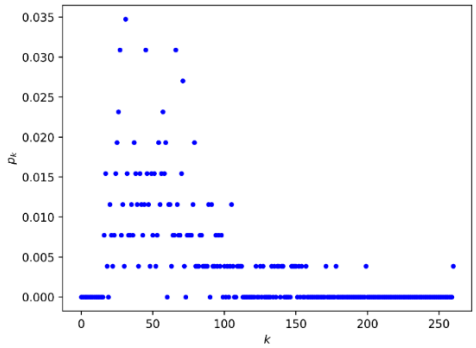


Figure 1. Degrees distribution

There are 111 sites in the network with nodes less than or equal to 30, and these nodes are at the edge of the network, most of them are small cities. The average degree of high-speed rail network <k> is 62.2239, indicating that each city can connect to 62 cities on average, and 57.14% of the cities can reach more than 50 cities, indicating that the overall connectivity of the network is good.

The cities with large degree are located in the eastern and central parts of China, indicating that the high-speed rail network in eastern and central cities is perfect and dense; due to their geographical location, population, economy and natural factors, northeast and western cities have fewer high-speed rail lines than the eastern cities, the network is mainly high-speed trains with Shanghai as the hub, and the train part in the western northeast is not recorded in the network, leading to western cities in Harbin and Dalian.

4.2. Average path length

The average path length and diameter of the network can reflect the transmission capacity and efficiency of the network. In the high-speed railway network, it reflects the number of trains that passengers need to take to reach the destination, namely direct or transfer, and reflects the convenience of travel.

The number of node pairs whose shortest path length is 1 accounts for 62% of the total number of edges in the network, and the average path length of high-speed rail network is 1.76. Therefore, the average path length obtained by the high-speed rail network based on Space-P is small, and the two cities can be reached by 1 or 2 high-speed rail lines. In this case, the complex high-speed rail network has high transfer and transportation efficiency, and the connectivity of each transport channel is also at a high level.

4.3. Clustering coefficient

The overall aggregation coefficient C in the constructed high-speed railway complex network C_g=0.6859 The global aggregation coefficient of the network reaches a higher level and the overall value shows high aggregation. A total of 83.78% of the urban clustering coefficient is higher than 0.5, indicating that the close connection between the node cities, and the network connectivity and aggregation are strong. The distribution of clustering coefficient is shown in Figure 2.

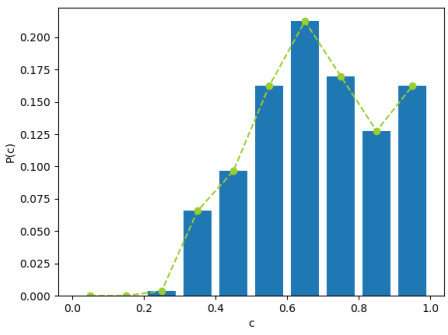


Figure 2. Distribution of the cluster coefficient

Based on the calculation results and analysis of the average path length in 4.2, it is clear that the high-speed rail network has strong aggregation coefficient and small average path length, and has the typical characteristics of a small world network.

4.4. Node centrality analysis

Betweeness centrality represents the ability of node cities as mediators. In general, in the comprehensive transportation and multimodal transport network, the cities with high Betweeness centrality have strong transit capacity, so such cities can be considered as transit hub, so as to achieve the purpose of optimizing the road network and hub layout.

The top ten cities with intermodal centrality are shown in Table.1.

Table 1. Ranking of the central cities

Id	City	Degree Centrality	Betweeness Centrality
4	Shanghai	1	4657.33
99	Zheng zhou	0.7713	1235.50
8	Xuzhou	0.6899	1062.37
3	Nankin	0.6628	977.54
7	Suzhou	0.6085	787.09
60	Hangzhou	0.5969	746.86
6	Changzhou	0.5814	726.49
9	Wuxi	0.5465	617.72
69	Changsha	0.5775	586.32
109	Wuhan	0.5891	575.69

In addition to Shanghai, Zhengzhou and Xuzhou rank second and third in their number centrality and clustering coefficient, even higher than more developed provincial capitals such as Hangzhou and Nanjing. Due to their unique geographical location, many railways, such as Longhai Railway, Beijing-Guangzhou Railway, Beijing-Hong Kong high-speed railway, Zhengzhou-West Railway and Zhengzhou-Xu high-speed railway, converge here.

Nanjing and Hangzhou are the capital cities of important provinces in the southeast coastal region and are close to Shanghai, thus having a high centrality in this network. Changsha and Wuhan is the provincial capital city developed in the central region, and it is also an important hub in the central region. Therefore, the medium number centrality and degree centrality are relatively high. However, due to the high-speed railway network with Shanghai as the hub, the more developed cities in the Yangtze River Delta region of Suzhou, Changzhou and Wuxi in the Yangtze River Delta have a higher centrality than the central cities.

5. Discovery of the community structure

5.1. Community detection algorithm

Louvain Algorithm is a heuristic algorithm based on modular degree optimization. The algorithm is two layers of iteration, the outer iteration idea is the bottom-up condensation method, and the memory iteration is the condensation algorithm plus the exchange strategy, so as to avoid the defect problem that once two nodes in the simple condensation method are merged, they can not be separated again[13].

5.2. Precision evaluation of community division

Since the connections between some nodes are stronger than those between others in real networks, this often leads to the generation of associations. Modularity is a measure of the density of links within and between communities. The modular formula of the network is shown in formula (6).

$$Q = \frac{1}{2m}\sum_{ij}\left[w_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i, c_j) \tag{6}$$

m is the total number of edges of the network, and w_{ij} represents the adjacency matrix, δ (c_i, c_j) is the 0 - 1 function, and the c_i, c_j is the community in which the nodes i, j are located, if $c_i = c_j$, That is, node i and node j are in the same community, then (c_i, c_j) Is given a value of 1 and 0 otherwise. The larger the module Q, the higher the accuracy of the community division results, which in general is [-0.5,1). In the actual network, the module degree is generally 0.3-0.7, and the network with high module degree is very rare.

5.3. Analysis of club results

At different resolutions, the community classification results obtained by the algorithm division are shown in Table.2.

Table 2. Association classification results

	Resolution parameter γ	The number of community	modularity
1	0.8	4	0.3683
2	1	5	0.3154
3	1.2	7	0.2643

Where the resolution parameter γ> 1, detect smaller module; 0 <= γ <1, detect larger module γ = 1, classical modularity (default).

The module degree is above 0.3, indicating that the high-speed rail network with Shanghai as the hub has a significant community structure.Due to the high connectivity of this network, the module degree is higher when the resolution is lower.According to the principle that the larger the module degree Q, the better the community division effect, this paper selects the first community division results for development analysis. The specific community division is shown in Table.3.

Table 3. Cities and association sizes included in the associations

	Includes the city	Society size (Including the number of cities.)
Society 1	Wuhan, Huanggang, Hefei, Anqing, Xianyang,.....	71
Society 2	Beijing, Tianjin, Nanjing, Jinan, Suzhou, Shenyang, Changchun, Harbin, Dalian, Qingdao...	52
Society 3	Jiaxing, Hangzhou, Nanchang, Changsha, Guiyang, Chengdu, Guangzhou, Shenzhen...	50
Society 4	Shanghai, Xuzhou, Zhengzhou, Taiyuan, Chongqing, Lianyungang, Yinchuan...	86

Association 1 is mainly located in central China, most of which are Hubei, Anhui, Shaanxi, and a few are from Hunan.

Association 2 cities are mainly located in northern China, such as Jilin, Liaoning, Beijing, Tianjin, Hebei and Shandong provinces, as well as some cities in Jiangsu province.

Association 3 cities are mainly located in the south of China, mainly in Zhejiang, Jiangxi, Hunan, Guangdong, Guizhou and other provinces.

Association 4 takes Shanghai and Zhengzhou as the community centers, and the cities are mainly in northwest China and the marginal cities with low connectivity in the network.

The division of community structure in high-speed railway geographic network is based on the connection relationship of high-speed railway lines between cities. From the spatial distribution of community structure, it can be seen that the community structure in high-speed railway geographic network is not only affected by the strength of inter-city railway connection, but also affected by administrative division, geographical location and other factors, but the spatial pattern of community structure is roughly consistent with the distribution of railway trunk lines. The high-speed rail lines within the community are relatively dense, while the high-speed rail lines between the communities are relatively sparse.

Because the network studied in this paper is modeled by the Space-P method, the overall network has higher connectivity, and the module degree of different non-overlapping community detection algorithms is lower than that of the high-speed rail network with Space-L [10]. At the same time, it also reflects that the various urban agglomerations in China's high-speed rail network are closely connected and accessible.

6. Conclusions

Based on the complex network theory and the Space-P method, this paper conducts an empirical study on the high-speed rail network starting from Shanghai and analyzes the properties of the relevant complex networks. Meanwhile, it divides the communities and studies the community structure through louvain algorithm, and draws the following conclusions.

1. The network has 259 nodes and 8058 edges, with an average degree of 62.2239, indicating that Shanghai has a wide range of direct cities and high network connectivity.

2. The average distance of the high-speed rail network is 1.76, the aggregation coefficient C_g =0.6859, with strong aggregation coefficient and small average path length, which is typical of small-world networks.

3. In this network, apart from Shanghai, Zhengzhou, Xuzhou and other cities have high intermediate centrality, which indicates that these nodes have strong transfer capacity in this network.

4. The module degree of community division is 0.3683. Because the network studied in this paper is modeled by the Space-P method, the overall network has high connectivity, resulting in a relatively low module degree.

References

[1] Zhang Lanxia, Qin Yong, Wang Li, characteristic analysis of high-speed railway [J]. Journal of Railway Science and Engineering, 2016,13 (02): 201-209.

[2] The comparative analysis of the characteristics of high-speed railway network in the three urban agglomerations of Li Xin, Guo Jinli and Zhang Yu is based on the perspective of complex network [J]. Resource Development and Market, 2016,32 (06): 703-707.

[3] Spatial structure reconstruction of Yangtze River Delta urban agglomeration in the era of Sun Mingyue high-speed railway —— Based on social network analysis [J]. Economic Geography, 2015,35 (10): 50-56.

[4] Qi Qingyu's analysis of Chinese high-speed railway network based on complex network [J]. Logistics Engineering and Management, 2020,42 (09): 22-24 + 18.

[5] Wang Wei, Liu Jun, Jiang Xi, and other topological characteristics of China's railway network [J]. Journal of Beijing Jiaotong University, 2010,34 (03): 148-152.

[6] GU X, Li D. Study on the correlations of Chinese railway passenger network based on complex network theory[C]//International Conference on Information Management. IEEE, 2012, 2:376-379.

[7] ZHANG J, HU F, WANG S, et al. Structural vulnerability and intervention of high speed railway networks[J]. Physica A: Statistical Mechanics and its Applications, 2016, 462:743-751.

[8] Liu Guo, Ye Kunhui Research on the evolutionary characteristics of high-speed railway in China based on complex network [J]. Transportation Research, 2017,3 (03): 6-13 + 19.

[9] Zhang Lingtao's research on the community structure of High-speed railway network in China based on the complex network theory [D]. North China University of Science and Technology, 2021.

[10] Xiong Ruixin. High-speed network topology characteristics research and anti-destroying ability analysis [D]. Lanzhou jiaotong university, 2022. The DOI: 10.27205 /, dc nki. Gltec. 2022.000849.

[11] Lai Fengbo, Xu Bing, Xu Ying, etc. Topological characteristics and node centrality of complex high-speed railway network [J]. Journal of Shandong University (Engineering Science),2022,52(06):14-22.

[12] Xie Zhi-qiang, GUO Jin-li, Zhang Jing. Reliability analysis of high-speed rail network in Yangtze River Delta based on Complex Network [J]. Software Guide, 201,20(07):156-162.

[13] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of statistical mechanics: theory and experiment, 2008, 2008(10): P10008.