The change of node importance and influencing factors in China's high-speed rail network

Xuedong liang1,a, Rui Liu2,b, Shifeng Lin3,c and Lufan Yang*4,d

1Business School, Sichuan University, China
2Business School, Sichuan University, China
3Business School, Sichuan University, China
4Business School, Sichuan University, China

liangxuedong@scu.edu.cn,b15680437690@163.com,clinshifeng@stu.scu.edu.cn,dyanlufan@stu.scu.edu.cn

Abstract. High-speed railway can change the geographical location of cities and enhance the importance of cities. This paper based on the complex network theory, combined with data of 2015-2019 year China's high-speed railway lines, constructs city as the node of the high-speed railway (HSR) network model, and probed the cities of PageRank in network, examined the high-speed rail network in the change of city node importance, analysed city node importance with time evolution and the regression model was used to determine the high-speed rail network in the city the affecting factor on the changes of node importance. The results show that: (1) Many cities in western China, such as Chengdu, Chongqing, have become significantly more important, while that of eastern coastal cities remains stable or even declines. (2) The level of economic development and population have the most obvious and positive influence on the importance of urban nodes. However, geographical location (latitude) and coastal area have negative effects on node importance.

1. Introduction
China is a vast country, and the distribution of natural resources and social and economic factors is extremely uneven. High-speed rail is a key platform for the spatial allocation of important economic factors such as population, capital, information and technology. High-speed trains are characterized by high capacity and speeds of more than 250 kilometers, improves the accessibility of cities, expands the scope of influence of cities, enhances spatial economic integration. The construction and use of railway can reduce traffic costs and spatial transaction costs, improve location conditions, and promote the reconfiguration of the agglomeration and diffusion space of economic elements [1]. High-speed railway affects the relationship between cities, promotes the redistribution of economic activities, and brings a large amount of capital, population and goods to the surrounding areas. China only completed 116.9 km of Beijing-Tianjin intercity high-speed railway in 2008. By the end of 2019, the railway has been reached 35,000 km, carrying 2.3 billion people[2].

In recent years, traffic infrastructure network has been widely used to determine the topological characteristics of the traffic network, such as: China railway network topological properties[3], the rise of complex network theory and urban network provide a new perspective for understanding the complexity of various transportation networks in the real world, the complex global transportation network[4]. Jiao[5] used the graph index, average path length, clustering coefficient and three centrality
indexes to investigate the impact of China's high-speed rail network on the overall connectivity of urban network, indicating that the high-speed rail network has improved the overall connectivity of the urban network based on passenger trains.

Node importance is an important feature of complex networks and provides more valuable solutions to urban development problems from the network perspective. Li [6] studied the importance of shipping network nodes and found that Europe had been at the center of the global shipping network to a large extent from 2001 to 2012. Sun's [7] studied the importance of provincial nodes in the national railway freight network and its influencing factors, and found that uneven distribution of natural resources would lead to spatial changes in the importance of nodes in the freight network.

In this paper, complex network theory is used to establish HSR network and PageRank is used as a measure of city node importance to study the evolution law of city, through the screening of related factors, establish regression model, to find important nodes in the network, and explore the affecting factor on the changes of the importance of the central city.

2. Methodology and Data

2.1. The Methodology and Data for Constructing an High-speed rail (HSR)

China's high-speed rail network is defined as a network has to have the right to $G = (V, E)$. Node set is defined as $V = \{V_i : i = 1, 2, 3...n\}$, $V$ is the number of nodes, each node is a city of China. The edge set is defined as $E = \{e_i : i = 1, 2, 3...m\}$, where $e$ is the number of edges and the edge is an interaction of transport between cities. The edge set is defined as $E = \{e_i : i = 1, 2, 3...m\}$, the edge of node between city traffic interaction relations, said the city has a high speed train services. The research data include cities of all high-speed railway stations in China from 2015 to 2019 is sourced from China Railway Ticketing website (www.12306.cn), China's HSR in 2019 is shown in Figure 1.

![Figure 1. HSR in 2019](image)

2.2. Measuring the Node Importance of the High-speed rail (HSR)

In order to comprehensively consider the interaction between cities, pagerank can consider to measure the relative importance of each city by taking into account all direct and indirect links (incomings). Represents a network of node cities for all other nodes, thus providing a reasonable and measurable sense of city importance. In this paper, 90 cities with high-speed rail frequency over 1000 in 2019 are selected as objects to measure their PageRank value from 2015 to 2019 to study the interaction between cities.

2.3. Methodology for Detecting the Influencing Factors on the Node Importance

According to the new economic geography theory [8], transportation hubs are generally cities with good geographical locations. The advantages of geographical locations can enhance regional economic vitality and promote regional economic development. Transportation hubs can highlight the
geographical conditions of cities, expand regional advantages, and have an important impact on urban development. Based on the above discussion, the regression model is as follows:

\[ Y_{it} = \beta_0 + \beta_1 GDP_{pcit} + \beta_2 AvgPOP_{it} + \beta_3 AppT_{it} + \beta_4 Latitude_i + \beta_5 Longitude_i + \beta_6 Air_i + \beta_7 Coastal_i + \mu_i + \epsilon_{it} \]  

(1)

Where: \( Y_{it} \) is the PageRank value of city i in year t, \( \beta_0 \) is a constant common to all cities and years, while the other \( \beta \) parameters refer to the model coefficients. \( GDP_{pcit} \) is the Gross Domestic Product per capita in city i in year t to capture the impact of the average income on each city. \( AvgPOP_{it} \) is the average population of city i in year t. \( AppT_{it} \) is the increase value of tertiary Industry of city i in year t. \( Latitude_i \) denotes the latitude of the city, \( Longitude_i \) denotes the longitude of the city. \( Air_i \) indicates whether city i has an airport and is a dummy variable. If city i has an airport counted as 1, otherwise, it is 0. \( Coastal_i \) indicates whether city i is in a coastal region and is a dummy variable. If city i is in a coastal region counted as 1, otherwise, it is 0.

In (1), \( \epsilon_{it} \) is an error term, and we assume the error term to consist of a time-invariant province-specific unobservable \( \mu_i \). In the model, we use panel-specific first-order autocorrelated AR (1) residuals; that is, \( \epsilon_{it} = \epsilon_{it-1} + \nu_{it} \). Because the Breusch-Pagan [9]/Cook-Weisberg [10] test rejects the null hypothesis of homoscedasticity (or constant error variance), the sample data have heteroscedastic properties. Psarian’s test of cross-sectional correlation also rejects the null hypothesis that the residuals are uncorrelated, and therefore, the model is specific to the panel-specific AR (1) residuals[11]. Since the number of cross sections for a given dataset is greater than the time series. Therefore, “We chose a panel corrected Standard Error (PCSE) estimator for the linear cross-sectional time series model[12]. In this study, continuous variables of explanatory variables in the model take logarithms, which eliminates nonlinearity and limits variance variation. The one-year lag value of all time-varying variables used was estimated to avoid potential endogeneity problems[11].

3. Results and Analysis

3.1. PageRank

From the change of PageRank calculation results, the high-node importance cities in the high-speed rail network have been transferred from the core cities (such as Beijing and Shanghai) to the new first-tier cities (such as Zhengzhou, Changsha, Chongqing and other cities). From 2015 to 2019, southwestern cities with relatively low levels of economic development will gain in importance. Chongqing moved up to ninth place, Chengdu rose from 13th to 11th, while Beijing, which has a higher level of economic development, dropped from fourth to 20th place, while Shanghai and Jinan dropped from 2nd and 7th to 5th and 10th respectively. The results show that the development of high-speed rail transportation in cities in central and western China shows a growing trend, the importance of nodes increases, the potential is huge, and the level of economic development improves rapidly. The importance of nodes in eastern cities remained stable or even decreased, and economic development slowed down.

3.2. Regression analysis

| Variables  | N  | Mean   | Sd   | Min   | Max   | Unit       |
|------------|----|--------|------|-------|-------|------------|
| PageRanks  | 450| 0.00589| 0.00346| 0.000527| 0.021| Unit       |
| GDPpc      | 450| 70,638 | 34,929| 19,004| 189,568| RMB        |
| AvgPOP     | 450| 618.3  | 404.6| 105.8| 3,397| Ten thousand people |
| AppT       | 450| 274.6  | 380.2| -2,407| 2,572| hundred million RMB |
| Latitude   | 450| 30.25  | 5.068| 22.33| 45.44| Unit       |
| Longitude  | 450| 115.5  | 4.261| 104| 126.4| Unit       |
Before reporting the regression results, we first discuss the correlation between descriptive statistics and explanatory variables. Table 1 reports the basic descriptive statistics of the main variables of the model, including sample observations, mean and standard deviation, maximum and minimum values, and variable types. Table 2 shows that the correlation coefficient between explanatory variables is generally low, with a high positive correlation between AppT and GDPpc, and a high positive correlation between AvgPOP variables, indicating that there is a multicollinearity problem between these variables. However, the multicollinearity caused by the high correlation between variables only affects the estimation efficiency and does not affect the consistency, indicating that we can still obtain unbiased estimators. Therefore, we still include these variables in the estimation model to avoid potential endogeneity problems caused by missing variables. See Table 3 for the result of calculating the determinant of PageRank by using PCSE evaluators.

Table 3. Results of the determinants of PageRank using the PCSE and SCCSE estimators

| Explanatory variable | PageRank PCSE |
|----------------------|---------------|
| lnGDPpc              | 0.0019***     |
| lnAvgPOP             | 0.0025***     |
| lnAppT               | 0.0003**      |
| latitude             | -0.0001***    |
| longitude            | 0.0002***     |
| coastal              | -0.0011***    |
| airport              | 0.0010**      |
| Constant             | -0.0507***    |

R-square 0.343
Observations 450
Number of code 90

The results showed that: GDPpc was used to reflect the influence of average level of economic activity in each region, and the regression coefficient was 0.0019. The higher the per capita GDP is, the higher the local economic level and node importance are, the more enterprises and talents it attracts. Second, AvgPOP is used to represent the influence of agglomeration economy. Annual average
population is an important positive factor proves that the higher the urban development potential and the larger the annual average population are, the higher the level of development, the more opportunities, the higher the degree of specialization and diversification of services are. AppT is the added value of the tertiary industry, The greater the added value of the tertiary industry, the greater the impact of the benefits from high-speed railway. Next, geographical location has an impact on the PageRank regression results of HSR nodes. In longitude, the symbol of regression coefficient is positive, indicating that the greater the longitude, the greater the importance of nodes. Because eastern cities have better economic foundation, more complete supporting facilities. The regression coefficient of latitude is negative, indicating that the higher the latitude, the less important the node is. That is to say, the relative importance of high-speed rail in China is gradually increasing from north to south. This may be because the south has a more dynamic economy, better terrain and more development potential than the north. Result shows that the geographical location of the coastal areas has a negative effect on the importance of the nodes in the HSR network. Relatively speaking, this is because coastal provinces have good natural seaports and focus on developing coastal advantages. Coastal cities did well before the construction of high-speed railways. However, panel data in recent years show that there is little growth in coastal cities, so the coastal areas show negative effects.

4. Conclusions
This paper uses the complex network theory to establish HSR and PageRank as the measurement of the importance of urban nodes to study the evolution rule of urban importance. By screening relevant factors, a regression model is established to identify the size and direction of factors affecting the importance of HSR nodes. The research results provide a theoretical basis for the planning of high-speed railway in Chinese cities. There are two main results: First, the PageRank value of many cities in the Midwest increased significantly. From 2015 to 2019, central regions such as Changsha continued to increase, while Wuhan and Zhengzhou tended to be stable. Many cities in the west, such as Chengdu, Chongqing and Xi’an, have grown in importance. High-speed railway transportation facilities have been constantly improved, geographical location has been improved, and the importance of cities has increased. With the gradual close contact with other cities, population, capital flow more frequent. Second, urban economic development level and urban geographical location are two important factors that affect the importance of a city in the high-speed rail network. Cities with high levels of economic development tend to have larger populations, higher per capita disposable income and more efficient public transport. At the same time, cities with a high level of economic development generally get more fiscal revenue, which can be used to develop high-speed rail stations. Finally, in recent years, the importance of cities in the high-speed rail network has increased year by year from north to south. The importance of the location of the nodes of the high-speed rail network in coastal areas has a negative impact, which is due to the construction of high-speed rail in coastal cities, which have a good development and the growth of the panel data after the completion of high-speed rail in recent years reflects that the space of coastal cities is not large. Therefore, we should strengthen the development of high-speed rail in inland cities in the future. In this paper, by studying the changes of important nodes and the temporal evolution trend of node importance in the high-speed rail network from 2015 to 2019, the important core cities of each year are found out and the influencing factors of importance are found out. The research results have important practical significance. The future development of high-speed railway can be planned by referring to the city's own conditions, geographical conditions and core city location. It can rapidly promote the linkage between cities, and then directly expand the urban radiation space, so that the radiation scope of cities to other cities in the high-speed rail network will be enhanced, and the economic development of its surrounding areas will be promoted, and the connection with other cities will be increasingly close.

References
[1] Zhang, W., P. Nian, and G. Lyu, A multimodal approach to assessing accessibility of a high-speed railway station. Journal of Transport Geography, 2016. 54: p. 91-101.
[2] Perl, A.D. and A.R. Goetz, Corridors, hybrids and networks: three global development strategies for high speed rail. Journal of Transport Geography, 2015. 42: p. 134-144.

[3] Li, W. and X. Cai, Empirical analysis of a scale-free railway network in China. Physica A: Statistical Mechanics and its Applications, 2007. 382(2): p. 693-703.

[4] Kaluza, P., et al., The complex network of global cargo ship movements. J R Soc Interface, 2010. 7(48): p. 1093-103.

[5] Jiao, J., J. Wang, and F. Jin, Impacts of high-speed rail lines on the city network in China. Journal of Transport Geography, 2017. 60: p. 257-266.

[6] Li, Z., M. Xu, and Y. Shi, Centrality in global shipping network basing on worldwide shipping areas. GeoJournal, 2014. 80(1): p. 47-60.

[7] Sun, Q., et al., Exploring the Node Importance and Its Influencing Factors in the Railway Freight Transportation Network in China. Journal of Advanced Transportation, 2019. 2019: p. 1-16.

[8] Fujita, M. and T. Mori, Frontiers of the New Economic Geography*. Papers in Regional Science, 2005. 84(3): p. 377-405.

[9] Breusch, T.S. and A.R.J.E.J.o.t.E.S. Pagan, A simple test for heteroscedasticity and random coefficient variation. 1979: p. 1287-1294.

[10] Cook, R.D. and S. Weisberg, Diagnostics for heteroscedasticity in regression. Biometrika, 1983. 70(1): p. 1-10.

[11] Tsekeris, T., Interregional trade network analysis for road freight transport in Greece. Transportation Research Part E: Logistics and Transportation Review, 2016. 85: p. 132-148.

[12] Beck, N. and J.N. Katz, What To Do (and Not to Do) with Time-Series Cross-Section Data. American Political Science Review, 2012. 89(3): p. 634-647.