IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 1, 2020, accepted July 9, 2020, date of publication July 16, 2020, date of current version August 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009627

An Ineffective Transport-Focused, Causality-Based Approach to Station-to-Station Railway Freight Network Design

YUJIE YUAN, XIUSHAN JIANG, AND LI YUANXUN
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Xiushan Jiang (xshjiang@bjtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017 YFB 1201301, and in part by the Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong University, China.

ABSTRACT Rail ineffective transport refers to meaningless and unvalued transportation in the rail freight system, which including convective transportation, roundabout transportation, and unprocessed transportation. Ineffective transport does not only result in a strain on transportation capacity and increase costs but also imposes significant external pollution on low-carbon development. This paper attempts to study ineffective transport among rail stations. Base on defining the concept of ineffective transportation, it analyzes the ineffective transport volume and ineffective transport propagation on each station, which helps to seek more efficiently the core ineffective transport stations in the rail freight network. To better understand the mechanism of ineffective transport propagation effect at the rail freight transport system-level, an ineffective transport causality network (ITCN) was built based on the Granger causality test. Through the topology model of ineffective transport at 50 stations in China from 2013 to 2016, the results show that the ineffective transport of each station affects approximately 12 stations and also affected by 12 stations on average. Large-sized stations are affected by more stations than downstream. Small-sized and medium-sized stations are opposed. There are four core stations in this ITCN, and optimization of the ineffective transport at these four stations will save nearly fifty million kilogram of transport volume per year.

INDEX TERMS Rail freight network design, ineffective transport, ineffective propagation, ineffective transport network.

I. INTRODUCTION

Freight transportation is a space transfer of goods [1]. Rising freight demand led to a burgeoning volume of rail transportation. However, the rosy figures have the pressing issues behind. Due to the information asymmetry and the unbalanced distribution of freight capability, etc., rail systems are faced with increased convective transport, round-about transportation, and unprocessed transport, leading to an imbalance of the supply and demand, a waste of transportation cost, a developing obstacle of low-carbon economic and a shortage of transport capacity worse [2][4]. Thus reducing the ineffective transport volume appears extraordinarily essential. A study by Wen (1981) [3] reported that the total annual direct cost of coal haulage induced by ineffective transport was nearly 58 million RMB in China in 1979, and the ineffective coal turnover portion accounts for around 7% of total amount (nearly 7.4 billion ton-kilometer). The ineffective transport in rail freight network that has resulted from this meaningless and unvalued transportation.

The definition of ineffective transportation has not been a unified explanation, but the related concepts have been considered by many scholars. Initially, the inefficient transportation of rails mainly refers to the ineffective capacity in the utilization of rail transportation capacity [9]. With reverse thinking, this concept is proposed by analyzing the characteristics of spare freight volume in the rail transportation system and its capacity. Through the analysis of the ineffective capacity at the two levels of planning and execution, it is possible to determine the dialectical relationship between the ineffective capacity and the completion of the transportation production task, and to adjust and stabilize

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed.

the random disturbance effect of the transportation capacity, thereby completing the given transportation task [10]. The development of a low-carbon economy, which is continuously promoted by the world, has made the study of inefficient transportation more extensive. Scholars have begun to consider the inefficient transportation volume brought by the rail transportation freights' attributes and freight structures. Tongquan (1995) [11] showed that unwashed coal would result in 84.48 million tons of inefficient transportation, and unreasonable transportation phenomena such as high vacancy rate, repeated transportation, convective transportation, and inefficient transportation would also bring a tremendous environmental burden on the low-carbon economy [12]. At the same time, Initial ineffective transport can be attributed to several reasons, such as handling equipment, extreme weather, optimal version dispatching, etc. [5]. However, the different rail-freight volume is one of the most significant determinants that give rise to the propagation effect of ineffective transport. The most common phenomenon is that the same station generates multiple ineffective ODs (Origin-Destination); the ineffective of an earlier station can affect the subsequent station of the single-category freight [6]. If ineffective transport of stations is not free from a previous station, the next station will be restricted by the volume of rail freight. For these reasons, ineffective transport of small initial stations may lead to larger ineffective transport later, inducing much worse situations [7], [8].

Besides, taking ineffective transportation as a constraint on the optimization of the rail freight network is another research direction. Kinds of literature on rail transportation network optimization considering inefficient transportation, inefficient transportation cost calculation, and rail transport network analysis were published. Xue (2013) [14] improved the concept of rail transportation capacity and analyzed the characteristics of its concept system, pointing out the importance of how rail transportation capacity can improve the efficiency of rail network construction and operation. Zhonglin [13] improved accuracy in parameter optimization for irregular railcar traffic volumes is achieved by adjusting the duration of the base periods at the railway node which constitute the optimization period in the dynamic problem. In addition to traditional rail transportation, container-loaded rail freight transportation has also attracted the attention of scholars. Inefficient transportation caused by container transportation includes low container turnover rates, empty container transportation, and high allocating volumes [15]. Li *et al.* (2007) [16] proposed a strategy for transporting empty containers similar to the economic inventory model. This strategy can better transfer the right amount of empty containers to the appropriate position at the right time. Although advances have been made in the optimization of rail freight transportation and the analysis of the rail transport system under inefficient transportation, few studies consider the interdependence of the time series of inefficient transportation of ineffective transport at each station and look into the impact of inefficient transportation between each station. Therefore, the system framework for exploring causal relationships between stations and the link relationships in the system network is still elusive. Through the above introduction, several problems have proposed according to existing research:

1) There is no systematic, comprehensive definition of ineffective transport.
2) There is no specific method for calculating the amount of ineffective transport at each station.
3) There is no detailed analysis of the ineffective transport volume of each station. How to clearly and accurately determine the ineffective core stations in the rail transportation network?
4) Whether the ineffective transport of each station correlates with other multiple stations. If there is ineffective propagation between stations in the rail system network, what kind of ineffective propagation effect will exist between each station (region)?

In this study, by referring to a large number of documents and books on the ineffective rail transport, combined with the actual situation of rail freight transportation, the complete concept of ineffective transport on the rail is proposed, and the calculation method of its system is standardized. Meantime, we have used Granger causality [17] as a major method for causal problems [18]. Then, based on the Granger causality test, we establish an ineffective transport causality network (ITCNs) and study the topology and time characteristics of ITCNs in combination with the concept of graph algorithm, to understand the characteristics of ineffective transport of specific stations in the same category. According to the calculation and analysis of the ineffective traffic and ineffective propagation of each station, identify the core stations, "bridge" stations, and scope of influence of the ineffective transport system in the rail freight system. The results not only help us better understand the complex rail freight network but also support the decision of the carrier and the rail company.

The remainder of this paper is organized as follows: Section 2 is the definition of ineffective transport. Section 3 introduces the calculation method of ineffective transport, meanwhile presents analytical methods of the ineffective transport in the rail freight network, including Granger causality testing, ITCN construction, and network analysis. Section 4 is an example analysis of the China Rail ineffective transport Network, and Section 5 is the conclusion and discussion.

II. INEFFECTIVE TRANSPORT

Ineffective transport is the meaningless and unvalued transportation volume that exists in rail freight transportation. The ineffective transport of the station is the count (or number) of ineffective transportation at each station considering convective transport, round about transport, and unprocessed transport.

Convective Transport (CT) is transported in the opposition direction on the same transport route (like a parallel line) in

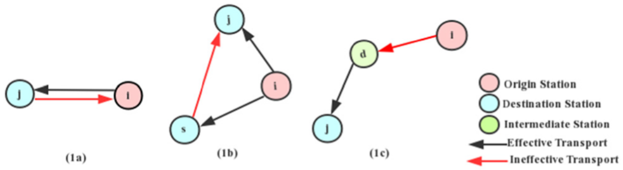

FIGURE 1. **Characteristic of ineffective transport.**

the same period of time, and overlaps with all (or part) of the other side of the transportation. It shows in Figure (1a). The transport volume of CT at a single station is as follows: If there is CT between station i and station j, $Q_{ij} \geq Q_{ji}$

$$\mathrm{CT}_{\mathrm{i}}^{\mathrm{m}} = Q_{ij}^{m} - Q_{ji}^{m} \tag{1}$$

where, $Q_{\mathrm{ij}}^{\mathrm{m}}$ is the freight volume of station i to station j at time m, CT_{i}^{m} is the ineffective freight volume of CT of station i at time m.

Round about transportation (RAT) is an unreasonable transport that bypasses the delivery of freights when there are more than two optional routes (without transportation capacity restrictions) that can be used. The main reasons for the RAT are as follows: First, due to insufficient capacity of individual sections of the line or other objective restrictions affecting the ability to pass; second, because the shipper does not start from the entire transportation system, the organization is poor, or the plan is not completed. RAT shows in the figure (1b). The transport of RAT of a single station is as follows:

If Station i is the originating station, Station j and Station s are the destinations of Station i, and $Q_{is} - Q_{ij} \geq Q_{sj}$,

$$\mathrm{RAT}_{\mathrm{i}}^{\mathrm{m}} = Q_{is}^{m} - Q_{ij}^{m} \tag{2}$$

Among them, $Q_{\mathrm{is}}^{\mathrm{m}}$ indicating the freight volume of the station i to the station s at the time m, $Q_{\mathrm{ij}}^{\mathrm{m}}$ indicating the freight volume of the station i to the station j at the time m, RT_{i}^{m} indicating the ineffective freight volume of the repeated transportation of the station i at the time m.

Unprocessed Transport (UT) refers to the volume of transportation that freights have not been screened and processed during transportation, resulting in excessively occupied transportation capacity. UT shows in figure (1c). From the origin station to the destination station, discardable freights (not caused by damage during transit) are processed at the intermediate station. The useless amount is ineffective transport volumes. The transport volume of unprocessed transportation at a single station is as follows: If station i is the originating station, station j is the terminating station of station i, station d, is intermediate stations, and transport the same batch of freights, while $Q_{id} - Q_{\mathrm{d}j} > 0$

$$\mathrm{UT}_{\mathrm{i}}^{\mathrm{m}} = Q_{id}^{m} - Q_{dj}^{m} \tag{3}$$

Among them, $Q_{\mathrm{id}}^{\mathrm{m}}$, indicating the total freight volume from station i to station d at time m, $Q_{\mathrm{dj}}^{\mathrm{m}}$ indicating the freight volume from station d to station j at time m, $\mathrm{UT}_{\mathrm{i}}^{\mathrm{m}}$ indicating that the inefficient freight volume of the UT of i. The total ineffective transport formula is expressed as follows:

$$\mathrm{InT}_{i}^{m} = CT_{i}^{m} + \mathrm{UT}_{i}^{m} + RAT_{i}^{m} \tag{4}$$

III. METHODOLOGY

As described in Section I, one of the main goal of this study is to determine the ineffective core stations and its ineffective propagation. In order to perform such goals, a detailed computational method framework will be present. The method consists of three parts.

Step 1 is to form an ineffective transportation time series of each station.

Step 2, the Granger causality (GC) test is used to determine the causal effects of the time series between the two stations, and the ITCN is generated among the multiple station-pairs.

Step 3, using the network science tools to explore the topological relationship impact of the ineffective transport propagation in the ITCN on the rail freight transport network, and finding the core stations with high-ineffective propagation.

IV. INEFFECTIVE TRANSPORT ON EACH STATION

First, the corresponding calculation method of ineffective transport is given to stations base on Section 2. Then, according to the calculation method, the total inefficient transport at each station will be calculated. The total inefficient transport at station i at different times is expressed as Formula (5). Among them, $\mathrm{InT}_{ij_{\mathrm{N}}}^{\mathrm{M}}$ indicating the ineffective transport volume from station i to station j at time M, N = 1, 2, 3, ..., n, M = 1, 2, 3, ..., m, InT_{i}^{k} indicating k freight category ineffective transport for station i.

$$\mathrm{In}T_{i}^{k} = \sum_{M=1}^{m} \sum_{N=1}^{n} \mathrm{InT}_{ij_{\mathrm{N}}}^{M} \tag{5}$$

We use an ineffective transport time series to indicate the station's inefficient transport capacity. For the station, we construct its ineffective transport time series y_i by using the monthly interval as the monthly interaction is the best time resolution in the station data-set. The value of each time interval represents the number of ineffective transports in the station $\mathrm{InT}(t)$. Through the statistics of the ineffective transport of each station, we can sort the stations by the ineffective transport volumes and select the major ineffective transport stations.

A. INEFFECTIVE TRANSPORT CAUSALITY NETWORK

The rail transportation system is also a typical large-scale complex system. Due to its complexity, the mechanisms of ineffective transport propagation are not fully understood, especially for the inter-dependencies of different stations. The same vehicle passes multiple stations; the ineffective transport of an earlier segment can affect the subsequent segments of the same vehicle. If freights are not free from a previous ineffective segment, the next segment will be affected by waiting for it. Thanks to theoretical innovation, the application of causality tests has grown to every field include detecting the interaction and propagation patterns in large-scale complex systems by time series analysis. All of the previous studies have shown that causality tests bring new insights into large-scale complex systems.

In ITCN, causality revealed the effects of stations and reflected the interaction of ineffective transport. If the ineffective transports counted at the starting station can explain the amount of ineffective transport counted at the end station at the same time, there is a causal relationship. Here, GC will help to understand the existence and direction of impact between the two stations based on the ineffective transport time series. If the conditional distribution of y_i determined by the y_i and y_j hysteresis values is the same as the conditional distribution determined only by the y_i lag value, such as the formula (6).

$$f\left(y_i^t \middle| y_i^{t-1},\ldots,y_j^{t-1},\ldots\right)=f\left(y_i^t \middle| y_i^{t-1},\ldots\right) \qquad (6)$$

It is said that there is no Granger causal relationship between y_j^{t-1} and y_i^t. Another expression is that if it can be shown that the value in y_i can provide statistically significant information about future values in y_j, it can be said that the time series y_i results in a time series.

First, the GC test uses an unrestricted regression equation to obtain the squared residual sum:

$$y_i^{\mathrm{t}}=\sum\nolimits_{m=1}^{L_{ij}}\alpha_m y_i^{t-m}+\sum\nolimits_{m=1}^{L_{ij}}\beta_m x_j^{t-m}+u^t \qquad (7)$$

y_i^{t} is the current value of the ineffective transport time series y_i, y_i^{t-m} is the past value of the time series y_i, y_j^{t-m} is the past value of the time series y_j m, u^t is the error term, α_{m} and β_{m} are the coefficients. In addition, L_{ij} represents a hysteresis, indicating that the current value should be regressed with the value in the past L_{ij} hours.

Then, test the null hypothesis that y_i has no GC for y_j;

$$\mathrm{H}_0:\ \beta_1=\beta_2=\ldots=\beta_m=0 \qquad (8)$$

Finally, the F-statistics and F-values are used to test the null hypothesis:

$$\mathrm{F}=\frac{(SSE_r-SSE_u)/L_{ij}}{SSE_u/\left(W-L_{ij}\right)} \qquad (9)$$

Among them, SSE_r represents the sum of squared residuals of the model after the constraint is imposed (the null hypothesis is established). SSE_u represents the sum of squared residuals of the model without applying constraints. $\mathrm{L_{ij}}$ indicates the maximum lag period. W indicates the sample size. Under the condition that the null hypothesis is established, the F statistic gradually obeys $\mathrm{F}_{(L_{ij},W-L_{ij})}$ distribution. If the F value calculated by the sample falls within the critical value, accept the null hypothesis that y_j has no GC for y_i.

Through the above causality testing process, we can assess the impact of ineffective transportation at each station pair. However, before applying the method, the ineffective transport that needs to meet the current station should not be independent, but should be closely related to the previous station or the subsequent station, that is, the conditions for rail connection between the two stations. However, multiple stations are involved in the rail freight transport system. Therefore, determining the relationship between different stations is very complicated. Pairwise analysis cannot handle the complexity of system-level inefficient transport propagation.

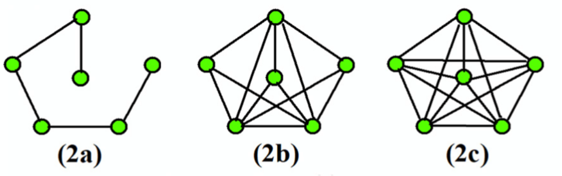

FIGURE 2. Density of ITCN. The density in figure (2a) is Sparse (D = 0.3), figure (2b) is Dense (D = 0.8), and the density of figure (2c) is complete/Clique (D = 1).

As mentioned earlier, due to the large number of rail stations and the complex interactions, only the information at the individual station level cannot be used to invalidate the characteristics of transport propagation. Complex network theory and its associated metrics and tools provide a way to study inefficient transportation systems beyond traditional techniques [19]. Therefore, network-level analysis is used to capture the global structure of functional interactions. A system-level ITCN can be built with the Individual test and analyzed by a network analysis tool. In ITCN, each station is a node, and each directed edge represents the Granger causal relationship between the stations. The statistical method for the number of relationships between stations is:

$$r_{ij}=\begin{cases}1 & if s_i \leftrightarrow s_j \in \ edge_{ij} \ \cup edge_{ji}\\ 0 & otherwise\end{cases} \qquad (10)$$

Individual testing of the station can form a matrix of ITCN:

$$\mathrm{ITCN}=\left\{\begin{matrix} r_{i_1j_1} & \cdots & r_{i_1j_n}\\ \vdots & \ddots & \vdots\\ r_{i_nj_1} & \cdots & r_{i_nj_n}\end{matrix}\right\} \qquad (11)$$

The maximum density is used to evaluate the sparsity or extreme density of the network. The sparsity of a graph is based on the comparison of its relationship number with the maximum possible number of relationships (if there is a relationship between each pair of nodes). A graph in which each node has a relationship with other nodes is called a complete graph or a clique of component, as shown in Figure 2. Although the sparsity or extreme density of the network does not have a strict dividing line, any network with an actual density close to the maximum density (D = 1) is considered to be dense. The formula for maximum density and full network is expressed as

$$\mathrm{D_{ITCN}}=2\sum\nolimits_{i\neq j}^{n}\mathrm{R}_{ij}\Big/n\,(n-1) \qquad (12)$$

where

$$\mathrm{Max\,D}=\mathrm{n\,(n\text{-}1)}\big/2 \qquad (13)$$

where n is the total number of nodes in the ITCN and $\sum_{i\neq j}^{n} R_{ij}$ is the total number of causal relationships r_{ij} between all stations in the ITCN.

When the algorithm is executed on a very sparse or very dense network, it returns a meaningless result, as shown in Figure (2). Because if the graph is too sparse ($D < 0.3$), there may not be enough relationships for the algorithm to calculate useful results, or because the connections between the nodes are very tight ($D > 0.8$), they don't give too much additional information. High density can also distort certain results or increase the computational complexity. In this case, you need to filter out the business-related sub-graphs and then analyze and calculate them.

B. ANALYTICAL METHOD OF ITCN

We mainly study the centrality and community detection of ITCN. The centrality mainly helps us understand the speed of the ineffective transportation of the station, the importance of specific stations in the network and its impact on the network, and help us understand the ineffective transportation dynamics of each region and identify the special routes that act as bridges between the regions. With the establishment of ITCN, centrality analysis and community detection, we can solve several problems in the ineffective transport network and its propagation characteristics:

How many connections of ineffective transport do each station have, how many stations will each station affect or affect, which stations are the most important in ITCN, and which nodes have the most control overflow between nodes and groups, which station is a "bridge"?

Is the ineffective transport propagation link between the station pairs bidirectional, what is the aggregation trend of the station, can ineffective transport propagation between stations be divided into several sub-areas, and how serious is the transmission of ineffective transport?

Figure 3 is an example of ITCN network consisting of 30 nodes and 41 ineffective transport causation (edges). The calculations combined with centrality and community detection show the results of the analysis of the ITCN network presentation and answer the above questions. Network $D_{\text{ITCN}} = 0.09 < 0.3$, network sparse and ineffective transport propagation has little effect, the whole network can be divided into 4 sub-communities. Among them, station A is the highest degree of the sub-community. Station B is the easiest to affect other stations in the sub-community. Station D is the most important station. Station C is the station that exists as a "bridge" throughout the network. The symmetry between the stations E, F, and G is the largest, forming a ring effect. The methods were elaborated as followed:

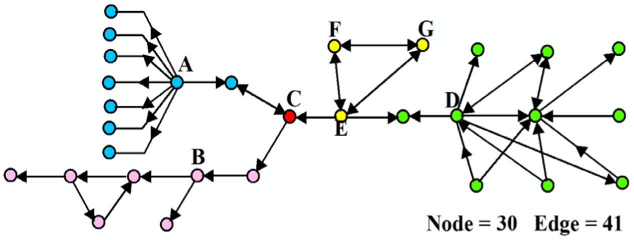

FIGURE 3. Topology of ITCN.

1) CENTRALITY DETECTION

Centrality detection mainly analyze the degree of centrality, close centrality, intermediary centrality, page ranking Page Rank and Reciprocity Parameter.

Degree Centrality is used as a baseline measure of connectivity, reflecting the number of stations with ineffective causal propagation links. In a directed network, station i has in-degree and out-Degree links, indicating the number of stations affected by station i and the number of stations affecting station i, $d_i^{in} = \sum_{j=1}^{N} r_{ji}$ and $d_i^{out} = \sum_{j=1}^{N} r_{ij}$ respectively. The total degree of station i is $d_i = d_{L}^{in} + d_i^{out}$. The average degree of the network in the middle is $d_i = d_i / n_i$, to calculate the number and average number of other stations that will affect other stations. In Fig. 3, station D is affected by other 7 stations. Thus, in-Degree is 4, whereas Out-Degree is 4. The average degree of the network of station D is 1.15, meaning that station D affects 1.15 others on average.

Closeness Centrality is used to calculate the meaning of a station to the network. Granger [20] proposed a way to detect the effectiveness of nodes propagating information over a network. The method metric is the extent to which the node is close to all other nodes. The formula is expressed as:

$$C_{WF}(i) = \frac{n-1}{N-1}\left(\frac{n-1}{\sum_{j=1}^{n-1} d(i,j)}\right) \tag{14}$$

where i is a station, N is the number of all stations, n is the number of nodes on the component where i is located, and d (i, j) is the shortest number of paths between other stations i and j. The maximum close centrality value is 1, and the larger the value, the greater the influence of the station directly affecting other stations in the network. In Fig. 3, Station A (0.022) and station E (0.018) have higher closeness centrality in the network.

Betweenness Centrality is used to find the main control stations in the ITCN and identify the stations that are most affected. Because sometimes the most important gear in the system is not the most obvious gear with the highest impact. Sometimes, some intermediate stations link groups or stations that have the greatest control over resources or information flow, so intermediaries can be used to find bridge stations from one community to another in the network. The formula is as follows:

$$B(i) = \sum_{s\neq i\neq j} \frac{p(i)}{p} \tag{15}$$

where i is a node, p is the number of shortest paths between nodes s and j, and p(i) is the number of shortest paths through s between s and j. Station C ($B_{(C)} = 0.046$) is the station that exists as a "bridge" throughout the network.

Page Rank is used to understand the impact of the overall impact and to analyze the ranking of the station's impact on the overall network. Calculate the formula (16). Among them, is the collection of all the stations linked to the station i, the station j is a connection station belonging to the collection B_i, and L (j) is the number of external links of the station j (the degree of outbound).

$$\mathrm{PR}\left(i\right)=\sum\nolimits_{j\in B_i}\frac{PR\left(j\right)}{L\left(j\right)} \tag{16}$$

The Reciprocity Parameter indicates that the ineffective transport between the pairs of stations affects the bidirectional nature of the propagation link. Reciprocity means that station i affects station j, and station j also affects station i ($r_{ij}=r_{ji}=1$). The parameter R is used to measure the overall symmetry of the directional network. It is defined as

$$\mathrm{R}=\sum\nolimits_{i\neq j}^{\mathrm{N}}\left(r_{ij}-\bar{r}\right)\left(r_{ji}-\bar{r}\right)\Big/\sum\nolimits_{i\neq j}^{N}\left(r_{ij}-\bar{r}\right)^2 \tag{17}$$

where,

$$\bar{r}=\sum\nolimits_{i\neq j}^{N}r_{ij}\Big/N\left(N-1\right) \tag{18}$$

The maximum value of R is 1, which means that the inefficient transport propagation between all pairs of stations is bidirectional. The larger the R value, the more symmetric the network. The R in Fig (3) is 0.15.

2) COMMUNITY DETECTION

Largest Connected Cluster represents the extent of ineffective transport propagation (means disaster area). The largest connected cluster [21] is a group in which stations are connected by propagation links. To make it represent the disaster area of ineffective propagation, we set an effective baseline for the members of a cluster. A station is considered when it affects several other stations (out-degree exceeds a threshold). If we postulate that the threshold is in-degree = out-degree = w, the size of the largest connected cluster (Md) is counts of the station which degree is higher than w.

Triangle counts and clustering coefficients are used to define the inherent clustering trends of the station. The aggregation coefficient of a station is the score of a direct ineffective transport propagation link (the number of triangles in the network) in its neighbors (stations with ineffective transport propagation links to the station). In the Figure (4), the network nodes are the same, but because of the relationship between the points, the strength of the formed community is also different. In Fig.(4a), CCT(i) = 0, CC = 0, although there is a correlation between the points, but there is no agglomeration. The local clustering coefficient coefficients of Fig.(4b), Fig.(4c) and Fig.(4d) are CCT (i)=0.2, CCT (i)=0.6, CCT (i)=1, respectively, indicating that there is a ring association in the network, and Form a community gathering. The largest connected cluster (SCC) to indicate the extent to which ineffective transport impacts (the disaster area). The Largest Connected Cluster [22] can be used to divide the groups that connect the stations through the propagation link, identify the disaster areas that represent the ineffective transport propagation, and set a valid baseline for the cluster.

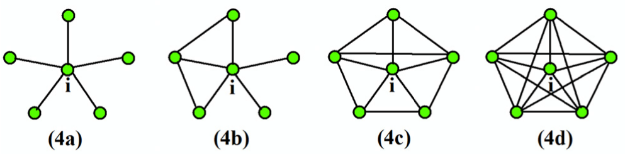

FIGURE 4. Clustering coefficients of ITCN.

For the network, the local clustering coefficient is calculated as

$$\mathrm{CC}^T\left(i\right)=2T_i\big/r_i\left(r_i-1\right) \tag{19}$$

And the total clustering coefficient is calculated as

$$\mathrm{CC}=\frac{1}{n}\sum_{i=1}^{n}\frac{\left(\frac{1}{2}\right)\sum\limits_{j}\sum\limits_{s}\left(r_{ij}+r_{ji}\right)\left(r_{is}+r_{si}\right)\left(r_{js}+r_{sj}\right)}{\left[\left(\sum\limits_{j\neq i}r_{ji}+\sum\limits_{j\neq i}r_{ij}\right)\left(\sum\limits_{j\neq i}r_{ji}+\sum\limits_{j\neq i}r_{ij}-1\right)-2\sum\limits_{j\neq i}r_{ij}r_{ji}\right]} \tag{20}$$

The community is used to assess whether ineffective transport propagation between stations can be divided into several sub-areas, where each sub-area of the station has dense inefficient transport propagation links, while the links to the rest of the system are sparse. Modularity is designed to measure the strength of dividing the network into communities. The formula for Mod in ITCN is defined as:

$$\mathrm{Q_{Mod}}=\sum_{ij}\left[\frac{R_{ij}}{2m}-\frac{r_i^{out}r_j^{in}}{\left(2m\right)\left(2m\right)}\right]\delta\left(c_i,c_j\right)=\sum_{i=1}^{e}\left(e_{ii}-r_i^2\right) \tag{21}$$

$$\mathrm{e}_{ii}=\sum_{j}\frac{R_{ij}}{2m}\delta\left(c_i,c_j\right)\quad r_i=\frac{r_i^{out}}{2m}=\sum_{j}r_{ij} \tag{22}$$

where R_{ij} is the adjacency matrix corresponding to ITCN, if there is an edge from i to j, then $\mathrm{R}_{ij}=1$, otherwise $\mathrm{R}_{ij}=0$. m is the total number of connections, and 2m is the total degree, $\mathrm{R}_{ij}\big/2m$ is the actual probability of the connection between the two nodes. r_i^{out} and r_j^{in} are the degrees of i and j, respectively. If we maintain the degree distribution of the network but randomly shuffle its sides, the probability of any pair of nodes being connected after shuffling is $\mathrm{r}_i^{ouy}r_j^{in}\big/(2m)^2$. $\delta\left(\mathrm{c}_i,c_j\right)$ indicates 1 if node ij belongs to the same community, otherwise the result is 0. According to Du (2018), a simplified formula can be proposed, which is expressed as follows:

$$\mathrm{Q_{Mod}}=\frac{1}{\mathrm{M}}\sum_{ij}\left(R_{ij}-\frac{r_i^{out}r_j^{in}}{M}\right)\delta\left(c_i,c_j\right) \tag{23}$$

At the same time, we can compare with ITCN through network randomization. The distribution of values in the ITCN is determined by constructing a random network of identical nodes and relationships.

V. CASE STUDY

A. DATA DESCRIPTION AND PROCESSING

The data set analyzed in this paper is provided by the railhead office and includes information on all freight lines in China from January 2013 to December 2016. The database contains 408,188 transportation lines connecting 3147 stations across the country. Through the statistics of all rail high-value-added scattered freight transportation, we calculated the ineffective transport volume of each category in Table 1 and selected the metal products with the largest amount of ineffective transport in white freights as the research object. Usually, convective transport (CT) is one of the most direct and most likely to cause inefficient transport between stations, so we consider the ineffectiveness and propagation impact of CT in the inefficient transport networks.

TABLE 1. The ineffective transport volume of each category.

Freight Category	Ineffective Transport/Kg	Stations (counts)	ODs (counts)
Metal Products	37897578	581	2331
Paper and Stationery	35021574	681	3841
Chemicals	32732427	476	2525
Textile	29694582	677	4510
Foodstuff and Tobacco Products	29320828	731	6381
Agricultural Products	28582571	639	4291
Electronics Machinery	28334754	668	3625
Industrial Machinery	15024932	680	7762
Medicine	12931776	321	1087
Fresh and Live Goods	3499408	302	961
Agricultural Machinery	893430	165	572

Through the statistics of the ineffective transport of metal products, we rank the ineffective transport of 50 stations, and we can get the total amount of ineffective stations such as Table 2. Among them, Station 1 has the most massive inefficient traffic (ITV/kg), which is 2187 times the minimum ineffective transportation (station 42), followed by Station 15, Station 3, Station 5, Station 2, Station 10, Station 19, Station 7., Station 22, Station 4. Just considering a single order of ineffective transport does not fully reflect the importance of ineffective transportation at the station. Therefore, by establishing an Ineffective Transportation Causal Network (ITCN), stations with high inefficiencies propagation are screened based on the ineffective transport volume of each station.

TABLE 2. The ineffective transport volume of each station.

Order	ID	ITV(kg)	Order	ID	ITV(kg)
1	1	12265007	26	18	613573
2	15	9509403	27	12	553209
3	3	8642503	28	34	515939
4	5	8028679	29	40	462114
5	2	6778292	30	24	425480
6	10	5163030	31	21	417687
7	19	4701482	32	39	368390
8	7	4302316	33	28	340821
9	22	4138330	34	35	291985
10	4	3812468	35	32	290239
11	14	2873540	36	27	265688
12	36	1997746	37	45	215036
13	23	1986389	38	30	174674
14	31	1951733	39	46	156401
15	8	1941065	40	38	138054
16	33	1817800	41	41	132754
17	9	1773888	42	47	101388
18	17	1643102	43	43	96297
19	11	1492154	44	29	60754
20	26	1376816	45	37	59143
21	6	1058861	46	44	39869
22	13	916930	47	49	25933
23	16	803907	48	50	23565
24	20	796992	49	48	17886
25	25	768257	50	42	5608

For the metal products, due to seasonal changes in market demand, considering the supply and demand requirements, we divide the length of the time series of ineffective transport by quarter, and screen and calculate of freight routes for CT in each quarter. From January 2013 to December 2016, there were a total of 6241 CT, involving 739 stations. The ineffective transportation volume of the metal products at each station is formed, and the ineffective transport time series of 739∗739 is formed. According to the delineation of the freight volume, We screened 50 stations with the largest ineffective transport and formed the ineffective transport time series of each station, as shown in Figure 5. The stations in the ITCN include 7 large-sized stations (Stations that account for more than 1% of the total freight volume, 29 for total), 17 medium-sized stations (account for 0.1% to 1% of the total freight volume,126 for total) and 26 small-sized stations (account for less than 0.1% of the total freight volume, 983 for total). Although small stations are a major component of the overall system, they are rarely involved in the spread of ineffective transportation. Large and medium-sized stations can easily fall into the inefficient transport. In terms of ineffective transport, the ineffective transport of the three-types stations tends to be between [0, 250000]. So, the time analysis of large-size, medium-size, and small-size stations have generated considerable ineffective transport, indicating that the main culprit for ineffective transport is not only caused by large-size stations.

B. ITCN ANALYSIS

To perform a system-level analysis of ineffective transport propagation, we built ITCN using paired GC tests based on

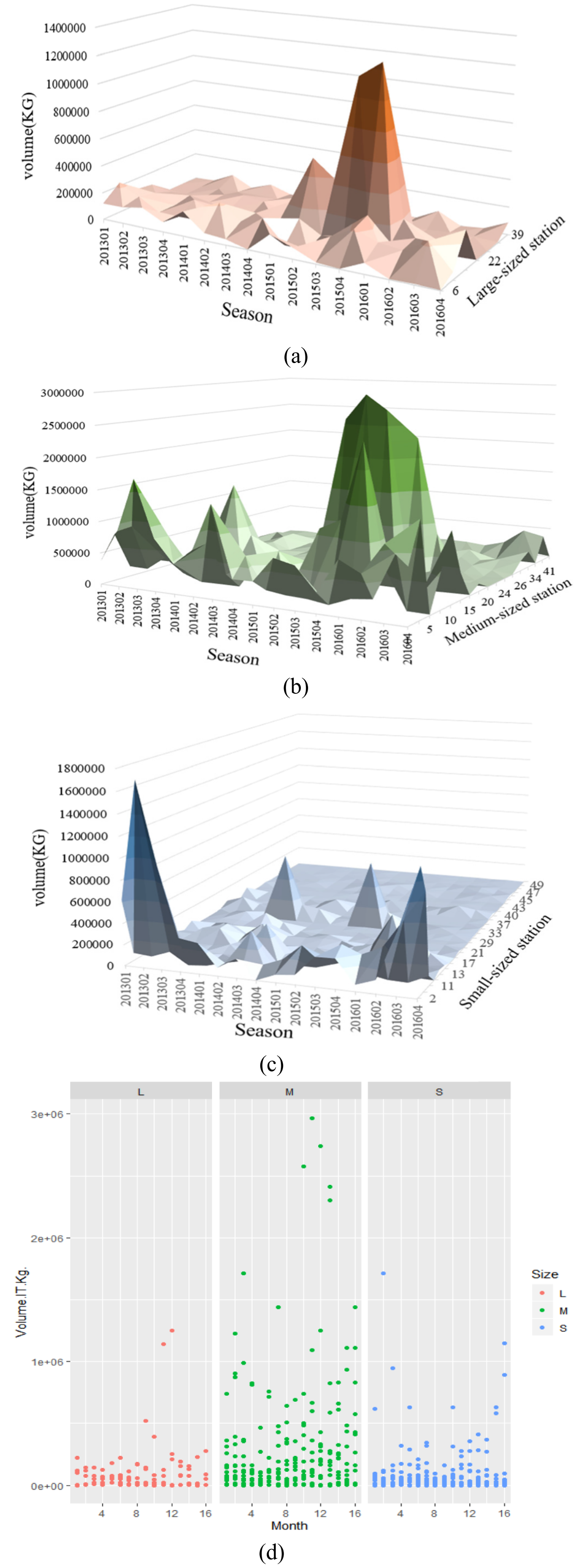

FIGURE 5. The ineffective transport time series of each station.

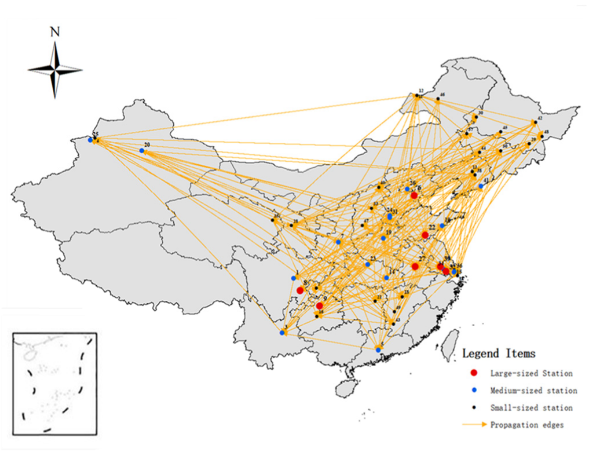

FIGURE 6. GC test of Station ineffective transport (results of first 100 ODs).

the data depicted in Figure 6. To calculate the GC test between each site pair, we can get the ineffective transport propagation matrix, shows in formula (24).

$$
ITCN = \begin{array}{c} ST \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 19 \\ \vdots \\ 50 \end{array}
\begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & \cdots & 50 \end{array} \\
\left\{ \begin{array}{cccccccccccc}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & \cdots & 0 \\
0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & \ldots & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & \ldots & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & \ldots & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & \ldots & 0
\end{array} \right\} \end{array} \tag{24}
$$

ITCN has 50 stations. A GC test was performed between each station pair, and 2500 ($50 * 50$) GC tests were performed to go out to the GC test of its own station. The effective test was 2450. After deleting the unconnected stations, we found that ITCN only contains 50 nodes and 606 edges (Fig.7), which means that only about a quarter of the stations have ineffective transport propagation links with other stations. Figure (6) is a network diagram of the effects of ineffective transport propagation. Large-sized stations (24% are affected) and medium-sized stations (13%) are vulnerable to ineffective transport propagation relative to the spread of inefficient transportation on small-sized stations (2%).

By calculating $D_{ITCN} = 0.49$ ($0.3 < D_{ITCN} < 0.8$), the density of ITCN satisfies the topological analysis density. Based on the number of nodes and edges, we built a random network of 50 nodes with 606 edges. Through the establishment of the network, we can answer the above questions:

How many stations will each station affect or be affected? For a station i in ITCN, in-degree is the number of stations that each station will partially lead to ineffective transportation of station i, and out-degree is the number of stations where partial ineffective transportation is caused by station i. Here, the average in-degree($\bar{d}_i^{in}$) = the average out-degree ($\bar{d}_i^{out}$) = 12.12, indicating that each station affects approximately 12 stations and is affected by approximately 12 stations. Which stations is the most important in ITCN? Through the degree centrality calculation results, Station 28 ($d_{28} = 39$) is the core station of ineffective transport propagation in ITCN, indicating that station 28 will influence adjacent 39 stations (including large-sized stations, medium-sized stations, and small-sized stations) seriously. In addition, the Closeness of station 5 ($C_{WF}(5) = 0.01408$), station 13 ($C_{WF}(13) = 0.01449$), and station 15 ($C_{WF}(15) = 0.01428$) is nearly to station 28 ($C_{WF}(28) = 0.01450$) in ITCN, indicating that station 5, station 13, and station 15 are also crucial stations in ITCN.

Which nodes have the mostly control over flow between nodes and groups, which station is a "bridge"? The largest Betweenness Centrality B is 0.03121, which belongs to station 1, indicating that stations 1 is the bridge station in ITCN. Station 1 is a crucial station for OD optimization among other stations in the rail freight system. Because the most frequently traversed station in the OD of ineffective transport propagation is station 1 (as an intermediate station).

Is the ineffective transport propagation link between the station pairs bidirectional? ITCN's reciprocity parameter is R = 0.232. For a random network with the same number of nodes and edges (compared by 1000 networks generated by network randomization techniques), the average R of the random network is only 0.1, much smaller than R = 0.232. Therefore, ITCN is more symmetrical. One possible reason is that CT between pairs of stations results in ineffective transport effects of two-way propagation.

What is the aggregation trend of the station? The overall clustering factor CC is 0.08907, which is greater than twice the random network (CC = 0.04305), indicating that ITCN stations tend to cluster.

Can ineffective transport propagation between stations be divided into several sub-areas? The ITCN was analyzed using a community detection algorithm. Modularity is used to assess the strength of dividing the network into communities, with greater modularity and a more visible community structure. ITCN's modular value is 0.5052, and the average modular value of 1000 random networks is 0.4291. Therefore, there is a piece of evidence that ITCN's ineffective transport propagation can be clearly divided into several sub-areas, indicating that ineffective transport propagation is more regional.

How serious is the transmission of ineffective transport? We use the largest connection cluster to indicate the severity of ineffective transport propagation. The members of the connected cluster are selected by the out-degree threshold so that the cluster contains a collection of highly inefficient transport stations that affect many other stations. We define the out-degree threshold to be greater than the average network (12.12) representing stations that affect more than twelve stations. ITCN's Md is 16, indicating that 16 is included in the disaster area where ineffective transport is transmitted.

By considering the typical ineffective transport category (mental products) and station grade (by considering the volume of ineffective transport), combined with seasonal changes, select quarterly 50 typical station ineffective transport data to demonstrate ineffective transport network analysis results. We found that, on average, each station affects approximately 12 stations and also affects approximately 12 stations. However, stations of different sizes are different. Large stations (high-transport stations) are affected by more stations than downstream. Small and medium-sized stations are opposed. The relationship between the in-degree and out-degree of the station proves that some of the largest stations increase the route of ineffective transport. The reciprocity parameter shows the two-way nature of the inefficient transport propagation path between ITCN station pairs. Community and modularity indicate that ineffective transport propagation cannot be divided into sub-regions. Despite this, we found clusters consisting of high stations in our analysis. We also found that inefficient transportation of all stations is highly correlated with the largest contact cluster. In 2013-2016, only about a quarter of stations in China were under the influence of CT alone, faced up with the propagation of ineffective transportation on a single category. The results also show that stations in metropolis affected by many upstream ones, actually affecting fewer downstream ones. In addition, the stations belonging to the connected cluster are not fixed stations, which indicate that the culprit of ineffective transport propagation is not a fixed set of stations.

On the other hand, by Combining with the major station's statistic and high-ineffective propagation analysis, four stations should be considered its ineffective transport volume seriously in the rail freight system, which is station 1, 5,13, and 15. Among these four stations, the ineffective transport volume of station 13 is not large, but it is the "bridge" station in the influence of ineffective transportation propagation, and its ineffective transportation will directly affect 42 stations. Another interesting finding is that these four stations are not large-sized stations, which means the traffic volume of some small-sized and medium-sized stations also have a significant impact on the rail freight system when designing a transportation plan. To sum up, through the analysis of total ineffective transport and station inefficiency propagation at each station, we screened out station 1, 5, 13, and 15 as the core stations in the rail system. Traffic optimization needs to be prioritized for these four stations in this ITCNs. Form convective transportation, station 1 contains 34 ODs, station 5 contains 16 ODs, station 13 contains 14 ODs, and station 15 contains 8 ODs. Optimization of the ineffective transport at these four stations will save nearly fifty million kilograms (50,875,305.25 kg) of transport volume per year.

VI. CONCLUSION

In this study, we study the ineffective transport and its ineffective propagation mechanism between stations from a new perspective, that is, identify the concept of ineffective transport, rejig the method of counting ineffective transport, and establish an ineffective transport causal network (ITCN) based on the relationship between the ineffective transport time series of each station, and apply the network. Analytical tools to reveal the macro impact of the station and the inefficient transport of transport and study the ineffective impact propagation of complex rail ineffective transport systems. By considering the invalidity propagation problem from the perspective of the inter-dependency of ineffective transport time series between stations, the method can capture the interaction pattern of the ineffective transport effects between different stations. In our research, the edge of ITCN is the result of daily time interactions, representing functional connectivity and potential running conditions. To demonstrate this approach, we constructed the ITCN using the ineffective transport Data-set from January 2013 to December 2016 at the China Freight Rail Station.

According to the survey results, the rail company can develop effective countermeasures to prevent the transmission of ineffective transport in a specific link, thereby reducing the transportation inefficiency of the entire network. Decision-makers can potentially use the proposed method to analyze station interactions to identify critical stations. The key stations and ODs will help them make decisions about resource allocation to improve the station's ability to transport. For rail companies, identifying ineffective transport routes can help them weigh the impact of transport management initiatives, such as rail transport control and timetable optimization, and choose the best combination to increase the efficiency of the freight system. For rail freight planners, by applying the proposed method, they can identify core stations in the network based on ineffective transport propagation. This information will help them make decisions about network transport expansion and resource allocation. The counter-intuitive results of downstream stations with less impact on large stations will cause rail freight planners to focus on controlling inefficient transportation for small and medium-sized stations, which may have been overlooked before.

Research using ITCN to study the transport of inefficient transport of freight can be further extended. For example, our ITCN is an unweighted network. Edge weights can be considered because they reflect the degree of causality. It is also interesting to compare ITCNs in different countries /regions' rail freight systems and investigate countries. We may find better insights from this international comparison to reduce ineffective transportation. From another perspective, the results of the calculations are taken as targets or constraints, taking into account the analysis of rail network optimization.

ACKNOWLEDGMENT

The authors would like to thank the insightful and constructive comments from anonymous reviewers.

REFERENCES

[1] G. Feng, "Transformation of rail traditional freight yard under modern logistics environment," Tech. Rep., 2016.

[2] J. Yuli, "Research on the development trend of rail transportation of iron and steel enterprises from the perspective of low carbon economy," *Enterprise Reform Manage.*, no. 14, p. 204, 2017.

[3] W. Suzhen, "Pay attention to improving the economic effect of rail coal transportation," *Rail Transp. Economy*, no. 1, pp. 38–42, 1981.

[4] C. Yijing *et al.*, "Research on the influence of rail capacity constraints on coal price in China," *Price Theory Pract.*, vol. 414, no. 12, pp. 81–84, 2018.

[5] T. Yuxun, "What is the invalid transportation?" no. 17, p. 7, 2004.

[6] W. Wei *et al.*, "Topological characteristics of China rail network," *J. Beijing Jiaotong Univ.*, vol. 34, no. 3, pp. 148–152, 2010.

[7] Z. Yaqin, "Research on the analysis of rail network performance and propagation impact under emergencies," Beijing Jiaotong Univ., Beijing, China, Tech. Rep., 2010.

[8] Z. Yubin, "Research on the complexity of rail freight transport network," Lanzhou Jiaotong Univ., Lanzhou, China, Tech. Rep., 2015.

[9] Y. Hao and H. Guoqi, "Quantitative study on delay and recovery of rail transportation capacity loss," *J. Rail*, no. 4, pp. 50–56, 1994.

[10] X. Ruihua, "Analysis of ineffective capability in capacity utilization of rail transportation system," *J. Shanghai Rail Univ.*, no. 4, pp. 69–75, 1996.

[11] Z. Tongquan, "Discussion on the countermeasures to be taken in the face of transportation restricting coal to the market," *Coal Economy Res.*, no. 9, pp. 27–28, 1995.

[12] H. Huang, "Discussion on the development of low carbon logistics under low carbon economy," *Commercial Economy*, no. 1, pp. 32–132, 2011.

[13] L. Zhonglin, "Research on theory and calculation method of transportation capacity of rail road network system," Beijing Jiaotong Univ., Beijing, China, Tech. Rep., 2006.

[14] D. Xue, "Research on transportation capacity and reliability of rail network system," Lanzhou Jiaotong Univ., Lanzhou, China, Tech. Rep., 2013.

[15] Z. Xin, "Research on optimization model and algorithm of container transportation path," Wuhan Univ. Technol., Wuhan, China, Tech. Rep., 2012.

[16] J.-A. Li, S. C. H. Leung, Y. Wu, and K. Liu, "Allocation of empty containers between multi-ports," *Eur. J. Oper. Res.*, vol. 182, no. 1, pp. 400–412, Oct. 2007.

[17] L. Yong, "Research on topology and routing of wireless sensor network for rail facilities security monitoring," M.S. thesis, Central South Univ., Changsha, China, Apr. 2011.

[18] W. Lei and M. Liang, "Method for network topology discovery of rail signal supervision system based on SNMP," *Rail Comput. Appl.*, vol. 27, no. 11, pp. 70–78, 2018.

[19] M. Grandjean, "Connected World: Untangling the air traffic network," Tech. Rep. hal-01521378, 2016.

[20] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[21] A. W. Wolfe, "Social network analysis: Methods and applications," *Amer. Ethnologist*, vol. 24, Nov. 1997.

[22] W.-B. Du, M.-Y. Zhang, Y. Zhang, X.-B. Cao, and J. Zhang, "Delay causality network in air transport systems," *Transp. Res. E, Logistics Transp. Rev.*, vol. 118, pp. 466–476, Oct. 2018.

YUJIE YUAN received the B.Sc. degree in management from the Keuka College, Keuka Park, NY, USA, in 2015, the B.Sc. degree in project management from the Yunnan University of Finance and Economics, Yunnan, China, in 2015, and the M.Sc. degree in science (management) from the Keuka College, in 2016. She is currently pursuing the Ph.D. degree in control system and engineering with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China.

Her research interests include system engineering, transport engineering, high-speed rail and civil aviation competition, and system control.

XIUSHAN JIANG received the B.S. degrees in management science and engineering from Beijing Jiaotong University, the M.S. degree in management, in 1999, and the Ph.D. degree in traffic and transportation planning and management, in 2005.

She is currently a Professor with the School of Traffic and Transportation, Institute of System Engineering, Beijing Jiaotong University. Research projects are funded by the Ministry of Science and Technology, the Chinese Academy of Engineering, the National Railway Corporation, the National Development and Reform Commission (NDRC), SF Express Company Ltd., and so on. Her research interests include traffic and transportation planning and management, traffic and transportation systems engineering, transportation planning, and transportation economics and policy.

LI YUANXUN received the B.Sc. degree in industrial engineering from the Xi'an University of Technology, Xi'an, China, in 2017. He is currently pursuing the M.Sc. degree in control system and engineering with the School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China.

His research interests include system engineering, transport engineering, high-speed rail and civil aviation competition, high-speed rail and civil aviation cooperation, comprehensive transportation planning, transportation economics, and system control.

• • •