

Risk analysis of railway system based on the community detection

S.S. WANG & K.P. LI & F.B. SHAO

State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China

KEYWORD: railway accident; complex network; community detection; risk quantification

ABSTRACT: Railway safety is an important issue of transportation system. In this paper, a network model is proposed for the risk analysis of railway system combined with the Fast Newman community detection algorithm. In the network model, nodes represent accident causation factors and links are generated if the two factors occur in one accident. Employing Fast Newman algorithm, the network is divided into communities. By defining occurrence possibility and consequence severity, the risk of accident is quantified. The rail equipment accident data from 2010 to 2014 is chosen for analyzing. The analyzing results show that the risk caused by the factor of switch improperly lined is the highest in the system and human factors play an important role in railway accident causation. The constructed model can quantify the risk, and hence, it gives methodological support for railway accident prevention.

INTRODUCTION

Railway transportation is one important transportation mode in the whole transportation system. However, accidents have a big impact on the utility of railway system. With the development of railway transportation, the demand to improve railway safety is steady increasing. Now, more and more technologies are widely used in the railway system and amount of accident data can be easily obtained. Then, it is natural to analyze railway accidents by using these data. Through data mining, key causation factors can be identified and the passive safety management mode can be avoided by analyzing these key causation factor.

Different approaches were used to analyze the railway accident. Fan et al (2015) applied systems thinking approach to analyze the case of “7.23” Yong-Tai-Wen High-Speed train accident, and disclosed how an underlying systemic structure finally resulted in the major accident. Liu et al (2015) applied the method of fault tree combined with quantitative analysis to analyze high-speed railway accidents, and gave suggestions to decrease the occurrence possibilities of accidents. Hyun ju Kim et al (2016) analyzed railway casualty accidents and proposed that we could prevent the occurrence of accidents by establishing preventive response systems. Ouyang et al (2010) applied the STAMP-based method to analyze the case of Jiaoji railway accident. And the analysis is helpful to analyze accidents spreading in a broad sense. In addition, an effective emergent measures for accident response management was established.

Complex network emerged in the 1980s (Watts D J et al. 1998). Recently, complex network is applied into railway accident analysis and many important results have been achieved. Wang (2008) applied complex network to analyze the demographic characteristics of urban transit network, and then evaluated the connectivity reliability based on two different modes of attack. Ye (2009) evaluated the reliability of the national railway network based on two different modes of attack. Ma et al (2013) used complex network theory to analyze the causation of railway accident and gave a theoretical clue for railway accident prediction.

However, there are little work to analyze the relationship between accident causation factors and to quantify the accident risk based on the statistical accident data. And most of studies about community detection focus on the optimization of communities. There is little research focusing on analyzing actual railway accidents.

This paper aims at identifying important factors which cause high risk in railway system and quantifying the risk of the railway system. In this paper, a railway accident causation network model combined with the community detection is proposed for railway accident analysis. Based on this accident causation network, the risk of railway system is quantified. After data filtering, the causa-

tion factors are identified, and then the accident causation network is established. Employing Fast Newman algorithm, relationships between accident causation factors are obtained. At last, the risk of railway system is quantified.

THE PROPOSED MODEL

Model assumptions

Assumption 1: In complex network, nodes represent causation factors and edges represent relationships between nodes.

The edges are generated according to the following rule. If two causation factors co-occur in one accident, there is an edge between the two nodes and the weight is the number of co-occurrence accidents.

Assumption 2: Community structure exists in railway accident causation network.

The railway system is composed of several subsystems, overall, the closeness of factors in the same subsystem is higher than that of factors in different subsystems. However, the closeness degree between them are different, hence, the possibility of triggering failure between each subsystem is unequal. Then, the community cannot be divided according to the partition of subsystems. A community detection algorithm is needed to divide factors into communities.

Assumption 3: The railway accident causation network is an undirected network.

In reality, the relationships between causation factors include causality and incidence relationships. For community detection, the community structure is only decided by the closeness between nodes. Thus, the railway accident network can be regarded as an undirected graph.

Assumption 4: The severity of consequence caused by factor is positively correlated with its occurrence frequency.

In actual railway operation, the occurrence probability of accident is small, but once the accident happens, the loss is enormous. Thus, in the premise of quantifying accident risk, it is assumed that the relationship between factor occurrence frequency and consequence-severity is positive correlation. And it will provide support for further analysis.

Model construction

Step 1 data filtering

The railway accident data is obtained from the website of Federal Railroad Administration (<http://www.fra.dot.gov>). There are four accident databases including highway rail accidents, operational data, rail equipment accidents and railroad casualties.

In this paper, the rail equipment accident data set, which includes reported cases of collisions, derailments, explosions, or other events involving the operation of railroad on-track equipment and involving damages exceeding the reporting threshold for the year reported, from 2010 to 2014 is chosen for analyzing. The data includes primary cause, contributing cause and total reportable damage et al.

Step 2 edge weight definition

The edge between two nodes is generated if the two nodes co-occur in the same accident. And the weight of the edge is the number of accidents caused by the two factors. Considering that factors appear as primary cause and contributing cause, the occurrence number is counted, respectively. And the weight is sum of the occurrence number.

Step 3 complex network construction

Complex network is usually represented by the adjacency matrix. If the adjacency matrix of accident causation network is A , a_{ij} represents the connection condition between causation factor i and factor j . If factor i and factor j never appear in one accident, $a_{ij}=0$, if factor i and factor j appear in one accident, $a_{ij}>0$, and the number equal to the edge weight. Thus, the railway accident causation network is established.

Step 4 community detection

In community detection algorithms, the modularity Q are mainstream for community detection. And the Fast Neman algorithm is a representative method which can deal with plenty of nodes.

The general formula of Q (Wang X et al. 2006) is:

$$Q = \frac{1}{2m} \sum_{ij} (W_{ij} - P_{ij}) \delta(C_i, C_j) \quad (1)$$

where, W_{ij} is the number of edges between node i and node j ; m is the number of total edges of network; P_{ij} is the number of expected edges between node i and node j ; C_i is the community that node i belonged; C_j is the community that node j belonged. When $C_i = C_j$, $\delta(C_i = C_j) = 1$, otherwise $\delta(C_i = C_j) = 0$

The steps of Fast Newman (Xin H. 2016) algorithm are as follows. First, every node is assumed as a community, and then the communities will merge from the direction of fastest growth or slowest reduction of modularity Q . Algorithm stops when all nodes combine into one community. It is the optimal community structure when Q is largest.

Step 5 risk quantification

Risk is quantified based on the definition of risk (Mitchell V W. 1995), the risk R is defined as

$$R = P \times L \quad (2)$$

where R represents risk; P is risk occurrence possibility; L is the severity of risk consequence. After getting the risk of every node, community risk can be got by accumulate the risk of every node contained in the community.

NUMERICAL COMPUTATION

After data filtering, 1585 valid data are obtained from the accident database from 2010 to 2014. Table 1 shows 254 causation factors.

Table 1 Federal Railroad Administration accident causation factors.

node	causation factor
1	Air hose uncoupled or burst (LOCO-MOTIVE)
2	Broken brake pipe or connections
...	...
253	Catenary system defect
254	Other way and structure defect

Analyzing the community structure

The accident causation network is divided by FN algorithm. Table 2 shows the community obtained by FN algorithm is with modularity $Q=0.5891$ and 18 communities. The range of Q is $[0, 1]$. If Q is closer to 1, the community structure is more obvious. The range of Q in the actual network is $[0.3, 0.7]$, thus, the result shows that the constructed railway accident causation network has obvious community structure which is consistent with the fact. For generality, the community which contains more than 9 nodes is regarded as an effective community.

Table 2 Community structure

Number of community	nodes contained
1	151
2	177,195
3	21,23,24
4	69,75,100,102,134,145,148
5	18,22
6	52,53
7	95,194
8	197,198
9	3,4,6,9,10,13,47,48,78
10	68,76,80,81,133,141,144
11	14,15,162,163,164,165,166,167,168,169,170,178,254
12	71,72,73,77,82,83,84,85,86,87,89,90,91,99,101,103,104,105,115,130,135,137,138,139,140,142,143,153
13	7,8,12,17,59,60,61,63,64,66,74,92,93,96,97,98,107,119,128,129,146,146,156,181,190,191
14	11,16,19,20,45,62,67,4,106,108,109,110,111,112,113,116,117,118,120,121,122,123,124,125,126,127,131,192
15	29,34,38,39,41,42,50,79,114,147,161,176,185,189,222,233,237,238,239,240,241,242,243,244,245,248,249,252
16	2,25,26,28,30,31,33,43,49,70,132,136,157,158,159,160,173,199,200,201,202,203,204,205,206,207,208,209,210,211,212,213,214,215,216,217,218,219,221,223,224,225,226,227,228,229,230,231,232,234,235,236,246,247,250,251
17	1,5,27,36,37,40,44,46,51,54,55,56,57,58,65,88,149,150,152,154,155,172,174,175,179,180,182,183,184,186,187,188,193,196,220,253
18	32,35

By analyzing the contained factors of 7 effective communities, the following conclusions can be obtained.

(1) For every community, the number of types of factors is more than one.

The factors of every effective community are belonging to different types. Such as community 13, it contains mechanical and electrical failures factors, train control-human factors and track, roadbed and structures factors. Community 15 contains factors of all five types. This confirms that the correlation between subsystems exists. The result of community detection is not identical with the partition of subsystems. It shows that the system view of supervision and management should be adopted in railway operation.

(2) Human factors appear in all communities.

Human factors appear in all effective communities, and human improper operation always accompany with the failure of corresponding system. Such as community 10, improper operation and mechanical-electrical failure of brake system exit at the same time. Human proper operation can reduce the possibility of triggering other factors and improve the railway safety. Thus, the supervision of human is one important work in railway accident prevention.

(3) External factors or confounders are non-essential factors in network.

External factors are the interference from people or environment outside the railway system, such as hurricane rain and other inclement weather, and other unconventional operations. These factors don't appear in all effective communities and its proportion is small. It shows that external factors will cause the accident, but its impact is small. Thus, more attention should be paid on internal system factors.

Risk quantitative analysis

According to Ref. 11, P is defined as the ratio of appearance frequency to total frequency, and L is defined as the product of occurrence possibility and total damage of 1585 accidents. After getting the risk of every node, community risk can be obtained by accumulating the risk of every node contained in the community. Table 3 shows the frequency, occurrence possibility and risk of every node. Table 4 shows the risk of 7 effective communities.

Table 3 the frequency, occurrence possibility and risk of every node

node	Frequency	Possibility	Risk
1	1	0.000315	27.51662
2	1	0.000315	27.51662
3	4	0.001262	440.2659
...
252	7	0.002208	1348.314
253	5	0.001577	687.9154559
254	8	0.002524	1761.064

Table 4 the risk of 7 effective communities.

community	risk	community	risk
12	5084741	16	495106.5
11	335455.1	4	294895.6
15	136042.2	17	86484.73
14	73689.5		

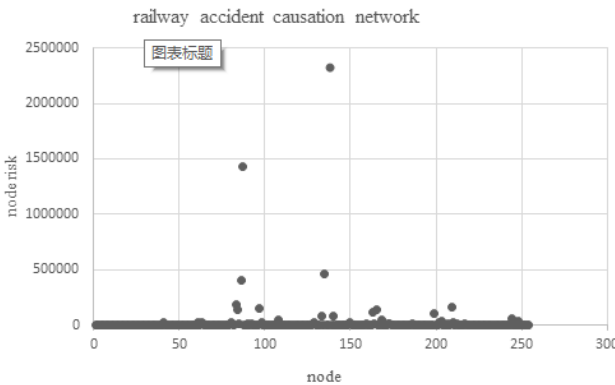


Figure 1. The risk of nodes contained in network.

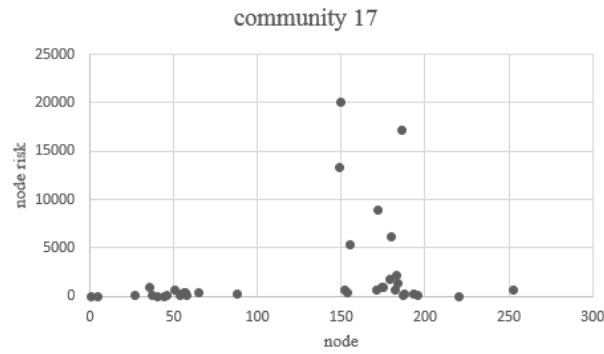


Figure 2. The risk of nodes contained in community 17.

Figure 1 shows the node risk before community detection. It shows that the risk of every node is different and only a small number of nodes have higher risk. The node with higher risk illustrates that its appearance frequency is great or it is closely related to most other factors. Then, it is a key factor in railway accident causation network.

Figure 2 shows the node risk of effective community 17 after community detection. The node risk of community 17 is consistent with non-uniformly distribution, and only a small number of nodes are with high risk. The node with higher risk is the key causation factor, and more attention should be paid to it. There is node clustering phenomenon in the Figure 2. It shows that relationships between the same type causation factors are closer which is in accordance with actual status.

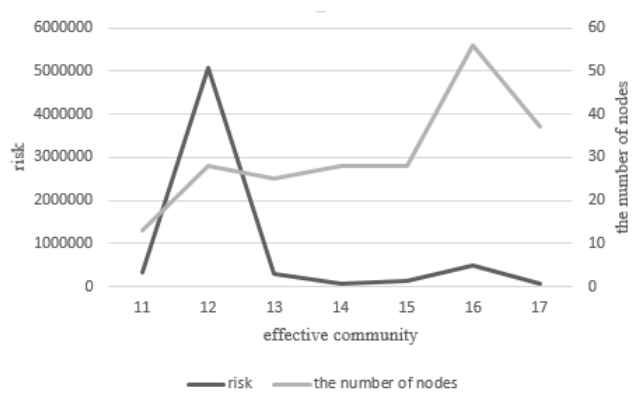


Figure 3 The risk and the number of nodes contained in the 7 effective communities.

Figure 3 shows that the risk of community 12 is the highest but the number of nodes contained in it is not the most. However, the number of nodes in community 16 is the most, and the risk is low.

The factors in community 12 are almost human factors. Most of accidents start from the unsafe behavior of human, and human is also injured in the accident. Thus, human factor is an important issue in the research of railway accident analysis. The supervision and control of train operation is based on the communication between human and environment. Human improper operation or react not in time is easy to cause serious railway accident. Thus, community 12 is with the highest risk.

The factors in community 16 are mainly track, roadbed and structure factors. These factors are rail hardware, and the occurrence possibility is low.

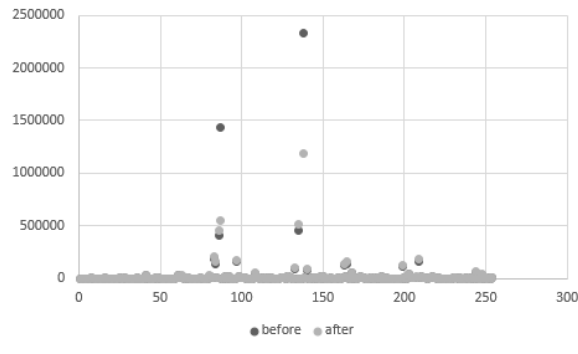


Figure 4 The risk before and after removing the hub node

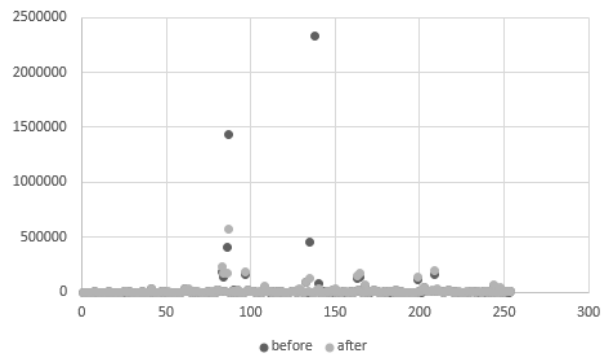


Figure 5 The risk before and after removing the edge

Node 138 is the hub node of the network based on the degree analysis of complex network. Figure 4 shows the results of the risk after removing the hub node. Figure 4 shows the risk before and after removing the hub node. There are some nodes with high risk before the hub node is removed, but the risks reduced after removing. It shows that protecting the key factor in network can effectively reduce the overall risk.

The edge between node 87 and 138 is with the greatest weight based on the analysis of complex network. Figure 5 shows the risk after removing this edge. There are some nodes with high risk before the edge is removed, but the risks reduced after removing. It shows that protecting the key factors that easily trigger others in network can effectively improve the overall safety.

CONCLUSIONS

A network model for railway accident causation and risk quantification is proposed combined with community detection algorithm. Based on the rail equipment accident data from 2010 to 2014, the causation factors are identified. In the network model, nodes represent accident causation factors and links are generated if the two factors occur in one accident. Employing Fast Newman algorithm, the network is divided into communities. By defining occurrence possibility and consequence severity, the risk of accident is quantified. The results show that the community which contains more human factors will be with higher risk, such as community 12. Thus, more attention should be paid to human factors in railway operation. At the same time, strengthening the check of other causation factors can further reduce the occurrence possibility of accident. And the risk caused by the factor of switch improperly lined is the highest in the system, hence, keeping the check of this factor is an important work in railway operation.

It is worth mentioning that the analysis is based on statistical data. However, the occurrence of accidents is a dynamic process. The emergency of accidents has not yet been considered in this paper and it will be our future research.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. U1434209), Research Foundation of State Key Laboratory of Railway Traffic Control and Safety, Beijing Jiaotong University (Grant No.RCS2016ZJ001) and Fundamental Research Funds for the Central Universities (Grant No.2015JBZ006).

REFERENCE

- 1) Fan, Y., Li, Z., Pei, J., Li, H., & Sun, J. 2015. Applying systems thinking approach to accident analysis in China: Case study of “7.23” Yong-Tai-Wen High-Speed train accident. *Safety science*, 76, 190-201.
- 2) Ju Kim, H., young Jeong, J., Wook Kim, J., & Kyung Oh, J. 2016. A Factor Analysis of Urban Railway Casualty Accidents and Establishment of Preventive Response Systems. *Procedia-Social and Behavioral Sciences*, 218, 131-140.
- 3) Liu, P., Yang, L., Gao, Z., Li, S., & Gao, Y. 2015. Fault tree analysis combined with quantitative analysis for high-speed railway accidents. *Safety science*, 79, 344-357.
- 4) Mitchell, V. W. 1995. Organizational risk perception and reduction: a literature review. *British Journal of Management*, 6(2), 115-133.
- 5) Ouyang, M., Hong, L., Yu, M. H., & Fei, Q. 2010. STAMP-based analysis on the railway accident and accident spreading: Taking the China–Jiaoji railway accident for example. *Safety Science*, 48(5), 544-555.
- 6) Watts, D. J., & Strogatz, S. H. 1998. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684), 440-442.
- 7) Wang X, Li X, Chen G. 2006. The theory and application about complex network [M]. Tsinghua University Press.
- 8) Wang, Y. 2008. Research on connectivity reliability of urban transit network based on theory of complex network. Beijing: Beijing Jiaotong University.
- 9) Xin H. 2016. Research on Railway Accident Causation Modeling Analysis. Beijing, Beijing Jiaotong University.
- 10) Xin, M., Ke-Ping, L., Zi-Yan, L., & Jin, Z. 2013. Analyzing the causation of a railway accident based on a complex network. *Chinese Physics B*, 23(2), 028904.
- 11) Ye, T. 2009. Research on Reliability of Countrywide Railway Network Based on theory of Complex Network. Beijing, Beijing Jiaotong University.