Traffic & Transportation

Importance Analysis of Causative Nodes for Accident Chains of Railway Locomotive Operation Based on STPA-PageRank Method

Ping WAN[1], Wei-Lun YANG[2], Jie-Wen LUO[3], Xiao-Feng MA[4]

Original Scientific Paper
Submitted: 26 Mar 2024
Accepted: 21 June 2024

[1] Corresponding author, pingw04@ecjtu.edu.cn, East China Jiaotong University, School of Transportation Engineering
[2] 2318769056@qq.com, East China Jiaotong University, School of Transportation Engineering
[3] jiewenluo22@163.com, East China Jiaotong University, School of Transportation Engineering
[4] maxiaofeng@whut.edu.cn, Wuhan University of Technology, Intelligent Transport Systems Research Centre

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT

Nowadays, in terms of complex and random incidents for locomotive operation, the prevention and control for every tiny and possible influencing factor is not only costly, but also brings great psychological burden to locomotive drivers. Firstly, 68 sets of data of railway locomotive operation accidents happened in recent two years were collected and compiled. Secondly, the system theory process analysis (STPA) method was adopted to extract 68 accident chains based on those data. Then, the complex network theory and PageRank algorithm were utilised to calculate the importance of every node in directed-weighted network formed by those accident chains. The results showed that the importance of human factors is significantly higher than other layers including environment, facility and management. Especially, no effective control behaviour (H7) and false control behaviour (H10) are the top two important causative nodes among all human factors. Besides, being forced to stop (D39) and overrunning of signal (D42) are the top two important causative nodes among unsafe events. For those nodes with high value of PageRank, some targeted security measures should be adopted, so as to save risk management investment and improve the overall safety level of the locomotive operation system.

KEYWORDS

railway locomotive operation; risk management; complex network; STPA; PageRank.

1. INTRODUCTION

Since the 21st century, China railway has achieved leapfrog development when compared with the past. The total operational mileage of railways in China has reached 155,000 km in year 2022, ranking the second in the world, while the total operational mileage of high-speed railways has reached 42,000 km, ranking the first in the world [1]. Therefore, railway transportation has gradually become one of the most important transportation modes in China [2]. As the main artery of the national economy, railway bears the popular travel demands, and the transportation demands of bulk commodities. However, the severe consequences of railway accidents mean that the resulting property damage and social impact on people are much greater than those of other ground transportation accidents. Generally, there are four main causes for railway operation accidents which include human factors, facility factors, environmental factors and management factors [3]. The diversification of accident causes makes risk management of railway operation system relatively difficult. If the key factors (nodes) in the entire railway operation system are not effectively managed or prevented, even substantial resource investment in risk management will yield minimal results. Moreover, the locomotive operation is the most important link for risk management of railway operation system [4]. The locomotive driver position possesses special characteristics including wide distribution of personnel, high mobility and working day and night, which puts forward higher requirements for the risk management. Therefore, how to adopt a scientific method to improve risk management level of locomotive work, to dissect the evolution law of safety incidents and to propose targeted security measures have become urgent tasks. Risk management refers to the management process of minimising risk in a risky system, which usually involves identifying risk, analysing the causes and its frequency of occurrence, the potential consequences and severity of hazardous events, and ultimately taking a series of measures to prevent risk [5-8]. In previous studies on risk management of transportation operation system, qualitative analysis and quantitative analysis are two typical inductive methods [9-11].

Considering the qualitative analysis methods for risk analysis, Szkoda et al [12] adopted a failure mode and effects analysis (FMEA) to forecast risk in railway operations. Although FMEA could be used to dynamically describe problems, it was only suitable to qualitatively analyse simple systems with low interactivity. Leveson et al [13]. established a widely used risk analysis method named systems-theoretic process analysis (STPA) in which the occurrence of unsafe events was caused by incorrect control process, which could analyse the interaction between human, facility, environment, management and other factors causing the accident. The STPA-based method is conducive to find hidden dangers in the evolution process of an accident and the formulation of safety constraints, while it could not perform quantitative analysis. Considering quantitative analysis methods for risk analysis, Castillo et al [14] proposed a new Markovian Bayesian network model to predict the probability of accidents associated with the circulation of trains along a given high speed or conventional railway line with special consideration to human error. However, the prediction effectiveness based on the proposed model relies heavily on prior knowledge and the prediction process could become computationally complex and time-consuming when dealing with large datasets. Zhang et al [15] considered that fault tree analysis (FTA) would make safety risk assessment more accurately for high-speed railway station in China as it is a graphical representation method that presented the logical relationship of failure events in a tree format. Nevertheless, FTA usually regards the fault as a technical problem within the system, ignoring the influence of human factors which are usually one of the most important reasons for failure. Sun et al [16] applied the PageRank network centrality measure to effectively observe the changes in the importance of Dongdaegu Station with multiple datasets of railway passenger origin-destination flow.

In terms of the combination of qualitative and quantitative analysis for risk management in transportation system, Zhao et al [17] proposed STPA-Bayes model to analyse the safety of airborne head-up display system and used the quantitative analysis ability of Bayesian network theory [18-20] to reduce the influence of human factors on STPA method, and to remedy for the deficiency of STPA method without quantitative analysis ability. However, the Bayesian network construction needs the reference of the prior probability of accidents, which is still dependent on the prior knowledge of experts.

In the past, qualitative research lacked the presentation of quantitative data, while quantitative analysis lacked a systematic macro description. This study aimed to address these research gaps by adopting a method that combines qualitative and quantitative analysis. This study will utilise the STPA (systems theoretic process analysis) and PageRank algorithm to determine the importance of causative nodes in railway locomotive operation accidents.

2. METHODOLOGY

In this study, the STPA method was applied to systematically analyse the workflow of railway locomotive operations and identify all accident chains based on traffic accidents involving railway locomotives in the past two years. Then, in terms of those accident chains, a directed-weighted network and further an adjacent matrix were constructed according to complex network theory. Finally, PageRank algorithm was utilised to calculate the importance of every node for the accident chains. Consequently, some targeted risk management measures were proposed to curb the further evolution of accident chains as economical as possible.

2.1 STPA

STPA is an improved qualitative analysis method based on systems theoretic accident model and processes (STAMP) [21-23] which converts safety issues into system control process. According to the analysis of the system control process, safety constraint deficiencies during system operation are clearly identified. STAMP typically consists of three parts: safety constraints, layered control structure and process model [24]. Moreover, based on the STAMP model, STPA model [25] has been expanded and divided into four steps: defining the purpose of the analysis, modelling the control structure, identifying unsafe control actions and identifying loss scenarios. As a risk analysis method based on system-theoretic accident models, STPA has been widely applied to the analysis of various safety-critical systems, as well as the design of aerospace, defence, transportation, chemical, pharmaceutical and power generation industries [26].

2.2 Complex network

Complex network [27, 28] usually refers to a large-scale network that exhibits complex topological structures and dynamic behaviours. It is generally represented as a graph comprised of many network nodes interconnected by edges. For a specific network *G*, its edges and nodes are often denoted as *V* and *E*, respectively. In terms of a directed network, the edge between node α and node β is represented as (α, β), while (β, α) represents the edge from node β to node α. In a weighted network, the strength of the connection between nodes is quantified by the weight assigned to the edge. Consequently, when both concepts are combined, a directed weighted network emerges. The structure of a directed weighted network graph with N nodes can be represented using an adjacency matrix A, shown as follows. In this representation, if node *j* points to node k =1; otherwise, it is *0*. The weight of the connection between nodes is represented by $\mathrm{w_{jk}}$.

$$A = \begin{cases} ajk \times wjk \\ 0 \end{cases} else \tag{1}$$

2.3 PageRank algorithm

Page et al. firstly proposed the PageRank algorithm [29, 30] for calculating the importance of web pages and ranking the pages on Google's search engine. Given an arbitrary directed graph containing *n* nodes, a general random walk model is defined based on the directed graph, namely first-order Markov chain [31]. The transition matrix of the general random walk model is composed of a linear combination of two matrixes. One is the basic transition matrix (*M*) of the digraph, which means the transition probability from a node to all linked nodes is equal. The other one is a completely random transition matrix, which means that the transition probability from a node to any node is *1/n*. The linear combination coefficient is damping factor d $(0 \le d \le 1)$. This general random walk Markov chain usually takes on a stationary distribution denoted by vector R which represents the general PageRank of this digraph. R is determined by the following PageRank iterative algorithm [32], where *1* is a n-dimensional vector with all components of *1*. The specific expression of page PageRank value is illustrated as *Equation 2*. The calculation flow of PageRank algorithm is shown in *Table 1*. $\mathrm{P_i}$ represents the page to be evaluated; n represents the total number of web pages; s represents the jump probability, generally s=0.15; $\mathrm{E(P_i)}$ represents the set of links of the link to user page $\mathrm{P_i}$; $\mathrm{G(P_j)}$ represents a collection of links from page $\mathrm{P_j}$ to other pages.

$$\mathrm{PageRank}(P_i) = \frac{s}{n} + (1-s) \sum_{P_j \subset E(P_i)} \frac{\mathrm{PageRank}(P_j)}{G(P_j)} \tag{2}$$

Table 1 – The calculation flow of PageRank algorithm

Process	Description
Input	Directed graph with n nodes, transition matrix M, damping factor d, initial vector R_0
Output	PageRank vector R of the digraph.
Step 1	Let t = 0 ;
Step 2	Calculate ; $\mathrm{R = dMR = \frac{1-d}{n} 1}$ (3)
Step 3	If $\mathrm{R_{t+1}}$ is close enough to $\mathrm{R_t}$, Let R =$\mathrm{R_{t+1}}$, then stop iteration.
Step 4	Otherwise, t = t + 1, execute step 2.

3. RISK ANALYSIS RESULTS WITH STPA

In this part, the method of system-theoretic process analysis (STPA) was utilised to identify potential unsafe control actions for railway locomotive operation. The verification and scenario analysis were conducted using a rigorous formal language, resulting in the presentation of a scenario analysis framework containing 60 general factors extracted from those 68 accidental chains.

3.1 Defining purpose of the analysis

Firstly, 68 sets of data of railway locomotive operational accidents occurred in the past two years were collected from 16 railway bureaus in China. The data are raw and have not been processed in any way. Secondly, the raw data were processed step by step according to STPA's analysis process which was expressed as *Figure 1*. Through the analysis of accident cases, six main types of system-level losses can be identified: affecting shunting order; affecting driving order; the locomotive is small broken, and no one is injured; the locomotive is small or medium broken, and people are injured; the locomotive is large broken, causing casualties; the locomotive is damaged or scrapped, causing serious casualties.

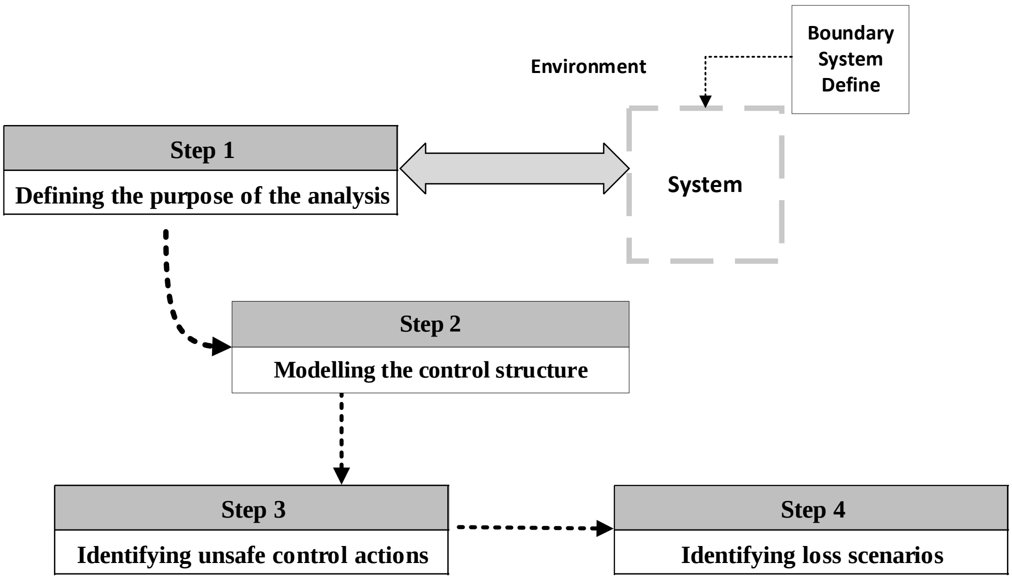


Figure 1 – Flow chart of STPA analysis

3.2 Modelling the control structure

Under the comprehensive action of various influencing factors including human, facility, environment and management, a potential unsafe situation for locomotive operation would be created. Furthermore, unreasonable control behaviour of locomotive drivers is often a key factor leading to the deterioration of the situation. Hence, the identification of unreasonable control behaviour can provide clues for searching the key factors causing accidents. Then, it is necessary to determine the control structure of railway locomotive operation security incidents.

Based on railway technical management regulations and the railway operation rules, locomotive depot, locomotive drivers, station attendants, train dispatcher, locomotive, ground signal interlocking communication equipment, the steel rail, overhead contact system and operational environment are mutually affected in railway locomotive operation system. Furthermore, the mutual control relationships among those components were analysed to create a security incidents control structure diagram for railway locomotive operation, as shown in *Figure 2*.

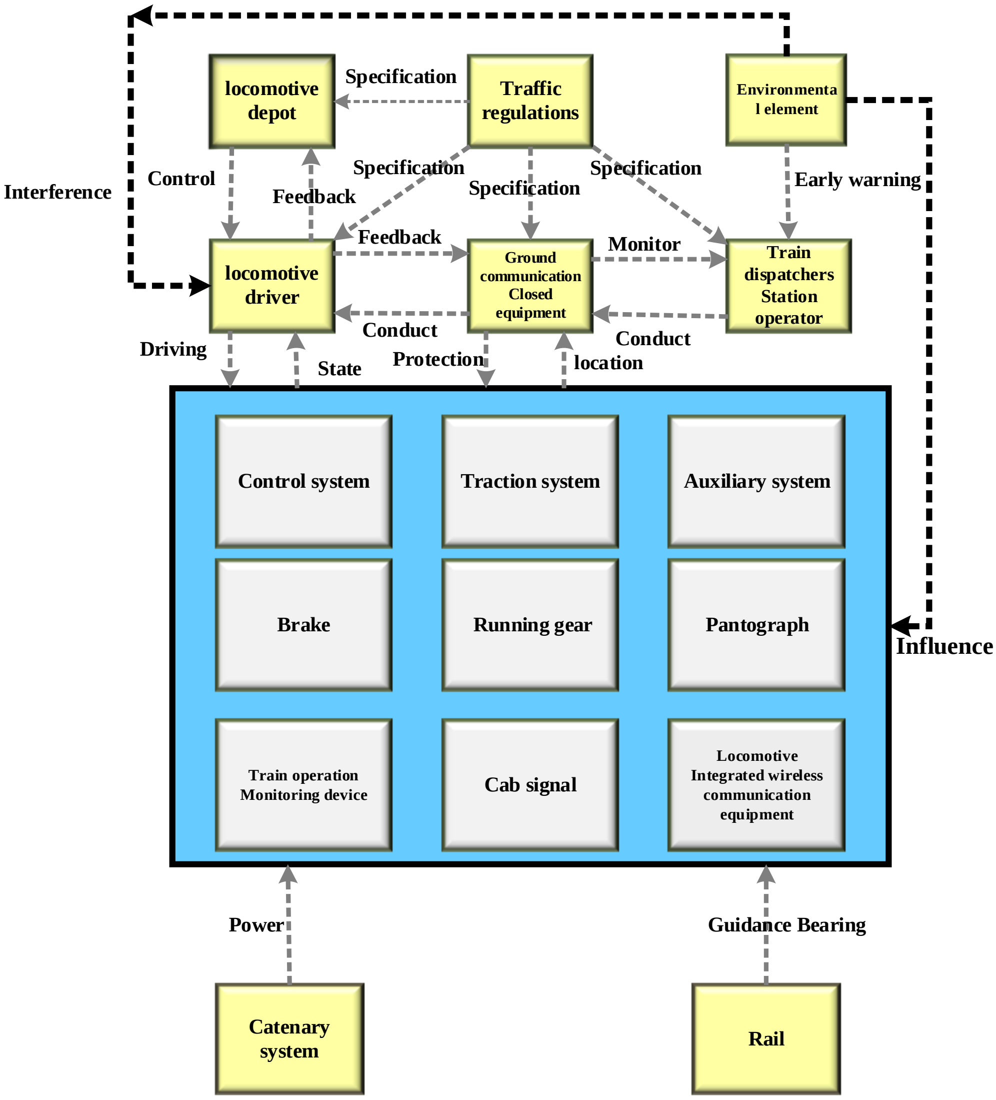


Figure 2 – Control structure diagram of railway locomotive operation security incidents

3.3 Identifying unsafe control actions

The traditional method, such as FTA, for classifying unsafe control behaviours may not be universally applicable to all types of accident analysis. This limitation can lead to ambiguity in behaviour classification, as the method overlooks the influence of human factors. To overcome the deficiency, STPA was utilised in this study to categorise the unsafe control behaviours into four types: no effective control behaviour, premature control behaviour, delayed control behaviour and erroneous control behaviour. Through STPA the unsafe control behaviours of locomotive drivers can be classified more accurately, which is helpful to identify causative factors of railway locomotive operation accidents.

3.4 Identifying loss scenarios

The cause of railway locomotive operation accidents generally includes principal factors and secondary factors. However, preventing the secondary events can be costly and bring little effect. In this regard, this study conducted analysis of the causes of 68 railway locomotive operation accidents based on the STPA method to identify principal factors and then propose targeted countermeasures.

According to *Figure 2*, the triggering events and evolution process of all railways locomotive accidents could be sorted out clearly, which was beneficial to draw accident chains. In this study every accident chain was consisted of accident causation (including management, human, environment, facility), unsafe events and system-level loss, as shown in *Table 2*.

Table 2 – Classification of accident chain nodes

Node categories	**Node numbers**	**Node ID**	**Node name**
Management	6	M1-M6	M1 Management training deficiencies for personnel professional skills; M2 Management deficiencies for train operation reveal; M3 Management deficiencies for one crew operation process; M4 Management process deficiencies for personnel backup on-duty and rest; M5 Management process deficiencies for locomotive overhaul servicing; M6 Safety management system deficiencies;
Human	10	H7-H16	H7 No effective control behaviour; H8 Control behaviour is premature; H9 Control behaviour is too late; H10 False control behaviour; H11 Control is not in place; H12 Tired; H13 Nervous; H14 Lack of concentration; H15 Mishear/misread the signal command, improper understanding; H16 Bad situational decisions;
Environment	9	E17-E25	E17 Long and steep grade; E18 Sharp curves; E19 Nighttime; E20 Fog, Haze, Sand-dust storm/Low visibility; E21 Torrential rain; H22 Snowstorm; E23 The rail surface is wet and icy; E24 Icing of catenary; E25 Oxygen rarefaction;
Facility	12	F26-F37	F26 Fault in the traction system; F27 Fault in the accessory system; F28 Fault in the control system; F29 Fault in the brake; F30 Fault in the running gear; F31 Fault in the pantograph; F32 Poor in roof insulation; F33 Fault in the train monitor and record device; F34 Fault in the cab signal; F35 Fault in the cab integrated radio communication equipment; F36 Fault in the digital train tail exhaust device, fault in the driver controller; F37 Car body structure is broken, locomotive couple is fractured;
Unsafe events	17	D38-D54	D38 Overspeed; D39 Being forced to stop; D40 Slope stop; D41 Start the locomotive blindly; D42 Overrunning of signal; D43 Forcing open of the point; D44 Derailment; D45 Rear-end conflict; D46 Lateral conflict; D47 Head-on conflict; D48 Hitting people; D49 Hitting livestock and cars; D50 Train separation; D51 Rail deficiencies and failures; D52 Passing neutral section with load; D53 Overhead catenary tripping and disconnection; D54 Pull cut the coupler, pull cut the air duct;
System-level loss	6	B55-B60	B55 Affecting shunting order; B56 Affecting driving order; B57 The locomotive is small broken, and no one is injured; B58 The locomotive is small or medium broken, and people are injured; B59 The locomotive is large broken, causing casualties; B60 The locomotive is damaged or scrapped, causing serious casualties.

4. IDENTIFICATION OF KEY NODES FOR ACCIDENT NETWORK BASED ON COMPLEX NETWORK AND PAGERANK METHOD

The integrated approach was proposed based on complex network theory and PageRank algorithm to quantify the importance of all causative nodes involved in the accident chains.

4.1 Accident network diagram construction

According to the connection relationship among those nodes, 68 accident chains were integrated into an accident network diagram, as shown in *Figure 3a*. To make it convenient to directly observe connection relationship between different nodes, the space between different nodes was adjusted equally roughly. Then, after merging edges with the same starting and ending nodes, a directed weighted network was constructed, as shown in *Figure 3b*.

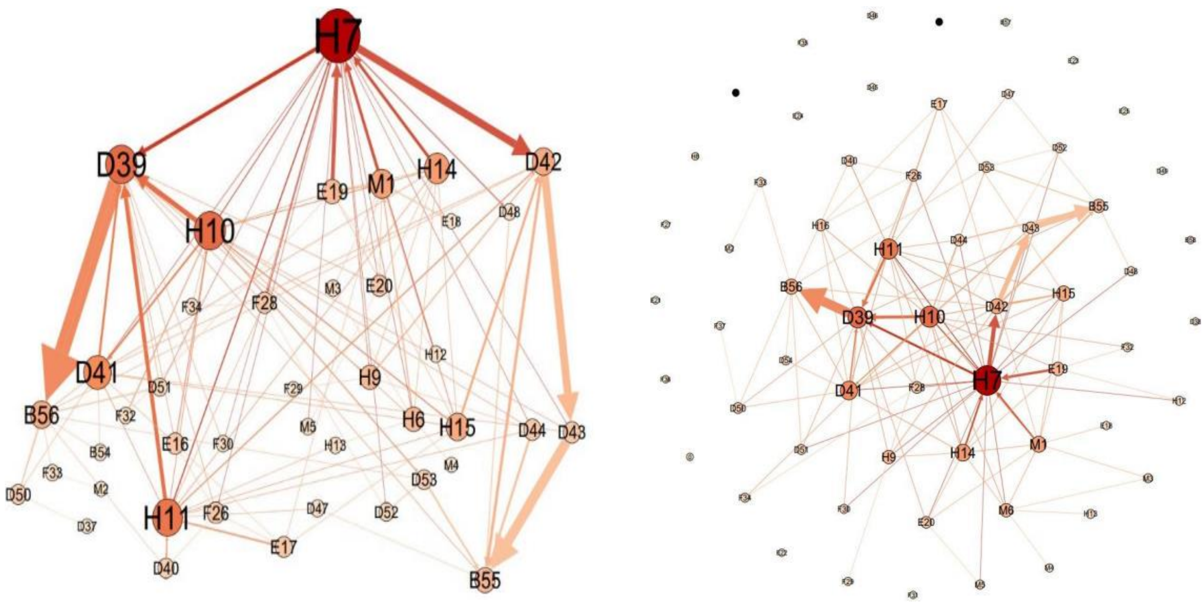


Figure 3 – Network diagram: a) accident network diagram; b) accident directed weighted network diagram

4.2 Constructing accident network adjacency matrix based on complex network theory

In terms of *Figure 3*, the accident network adjacency matrix was constructed according to calculation of node degree and node strength based on complex network theory.

Calculation of node degree

The degree of a node in the network refers to the number of edges connected to that node. As edges are directional in a directed weighted network, the degree of a node can be further divided into in-degree and out-degree. For instance, the in-degree of node m indicates the number of edges pointing towards node m from other nodes in the network. Whereas the out-degree of node m indicates the number of edges pointing towards other nodes from node m in the network. The total degree of node m is denoted by Km which can be expressed as follows:

$$k_m = \sum_{i \neq m} a_{im} + \sum_{j \neq m} a_{mj} \tag{4}$$

where the first term on the right side of the equation is in-degree of node m, while the second term is out-degree of node m. According to *Equation 4*, every node's in-degree and out-degree, as well as degree which is the sum of in-degree and out-degree, were calculated as shown in *Figure 4a*.

Calculation of node strength

The degree of a node can only represent the number of connecting edges with the other nodes, but it cannot indicate the strength of the connections between the node and the other nodes. Therefore, the concept for the strength of a node was introduced in complex network. Similar to degree of a node, the strength of a node can also be classified as out-strength and in-strength. Out-strength of a node refers to the sum of the edge weights from that node to other nodes, meaning it quantifies the cumulative strength of all outgoing connections from the node. On the other hand, in-strength refers to the sum of the edge weights directed towards that node from the other nodes, reflecting the combined strength of incoming connections. To assess the overall strength of a node, the total strength is generally adopted, which is the sum of its out-strength and in-strength. The calculation of the total strength of node m can be expressed as follows:

$$S_m = \sum_{m \neq i} S_{im} + \sum_{m \neq j} S_{mj} \tag{5}$$

where the first term on the right side of the equation represents the in-strength of node m, while the second term is the out-strength of node m. On the basis of *Equation 5*, every node's in-strength and out-strength, as well as strength which is the sum of in-strength and out-strength, were calculated with the results shown in *Figure 4b*.

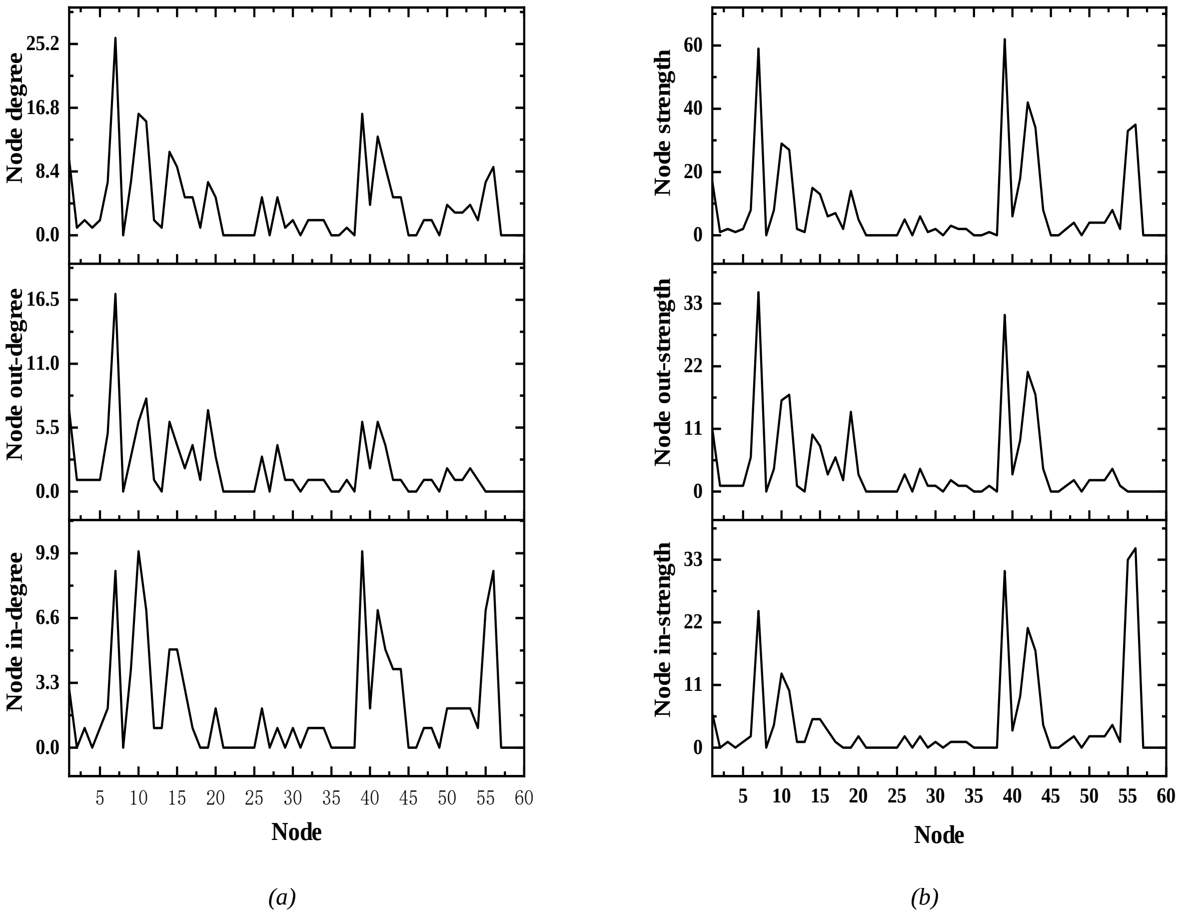


Figure 4 – Area diagram of all accident network nodes: a) degree value; b) strength value

Establishment of adjacent matrix of accident network

Through the calculation of node degree and node strength, accidents can be presented in the form of adjacency matrix of accident network, which can clearly show the relationship between different nodes and edges. After 3,600 calculations were conducted for those 68 locomotive operation railway accidents, the adjacency matrix was formally established. Due to the excessive volume of data, part of the data in the adjacency matrix has been extracted as shown in *Table 3*.

Table 3 – Adjacency matrix of the accident network

Node	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**.....**	**55**	**56**	**57**	**58**	**59**	**60**
1	0	0	1	0	0	0	5	.....	0	0	0	0	0	0
2	0	0	0	0	0	0	0	.....	0	0	0	0	0	0
3	0	0	0	0	0	1	0	.....	0	0	0	0	0	0
4	0	0	0	0	0	1	0	.....	0	0	0	0	0	0
5	0	0	0	0	0	0	1	.....	0	0	0	0	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
40	0	0	0	0	0	0	0	.....	0	1	0	0	0	0
41	0	0	0	0	0	0	0	.....	0	1	0	0	0	0
42	0	0	0	0	0	0	0	.....	5	1	0	0	0	0
43	0	0	0	0	0	0	0	.....	17	0	0	0	0	0
44	0	0	0	0	0	0	0	.....	4	0	0	0	0	0
45	0	0	0	0	0	0	0	.....	0	0	0	0	0	0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
57	0	0	0	0	0	0	0	.....	0	0	0	0	0	0
58	0	0	0	0	0	0	0	.....	0	0	0	0	0	0
59	0	0	0	0	0	0	0	.....	0	0	0	0	0	0
60	0	0	0	0	0	0	0	.....	0	0	0	0	0	0

Establishment of three-dimensional matrix graph of accident network

According to the preceding analysis, this study identified 60 causative nodes through the STPA process. We have collected 68 sets of data of railway locomotive operation accidents from 16 railway bureaus in China over the past two years. The whole evolution process for every accident is represented by an accident chain which is formed using a certain number of nodes. For example, the 59th accident chain is represented with Node6-Node7-Node10-Node41-Node43-Node55. In addition, the node degree and node strength of every accident chain was calculated to determine the connection relationship between any two nodes as indicated above. The connective relationship is represented by edge weight in this study as depicted in *Figure 5*.

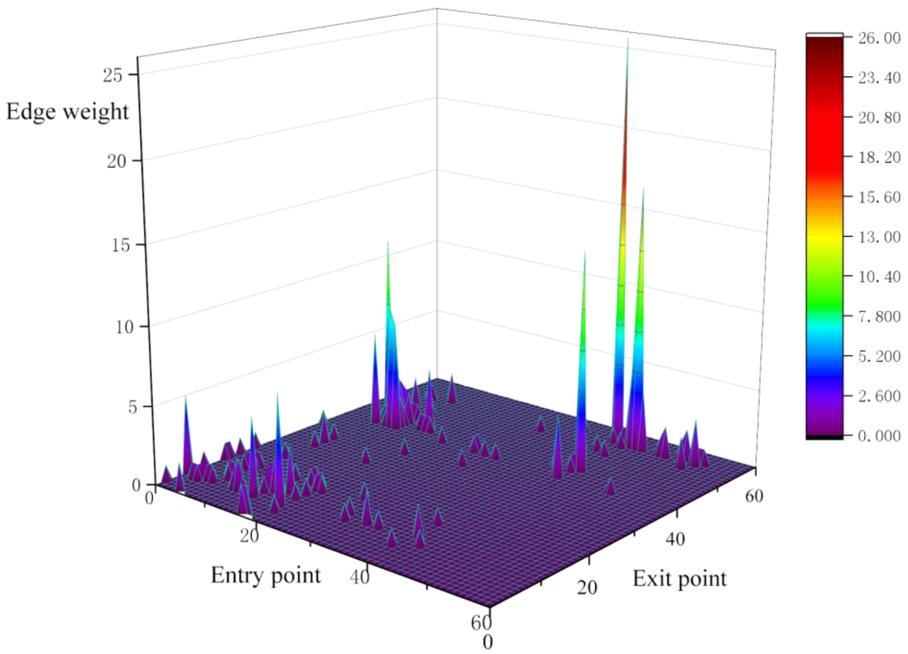


Figure 5 – A three-dimensional matrix graph for describing the connection relationship between different accident nodes

As illustrated in *Figure 5*, several peaks of edge weights were found in three different areas including human factors (H7-H16, namely node7-16), unsafe events (D38-54, namely node 38-54) and system-level loss (B55-B60, namely node 55-60). It was noticed that the maximum number of peaks occurred in the human factors. Therefore, it can be concluded that the nodes from the human factors have played a more dominant role than other nodes, making human factors more susceptible to trigger critical accident risk and leading to system-level losses. Moreover, it was found that the human factors had a closer relationship with environment factors like E23 (the rail surface is wet and icy). Therefore, greater emphasis should be placed on environmental factors contributing to human errors when proposing prevention measures for accidents related to human factors.

4.3 PageRank of accident network nodes

Neither node degree nor node strength could provide overall evaluation of importance of the nodes involved in the accident chains. Therefore, PageRank algorithm was introduced to comprehensively evaluate each accident node's importance. Assuming equal initial influence for all nodes, namely:

$$w_0 = \left[\frac{1}{60}\ \frac{1}{60}\ \cdots\cdots \frac{1}{60}\ \frac{1}{60}\right]. \tag{6}$$

Then, the transition matrix was calculated based on the adjacent matrix data of the accident network including node degree and node strength. Subsequently the PageRank iteration algorithm based on *Equation 3* was applied, and four consecutive iterations were performed with the results shown in *Figure 6*.

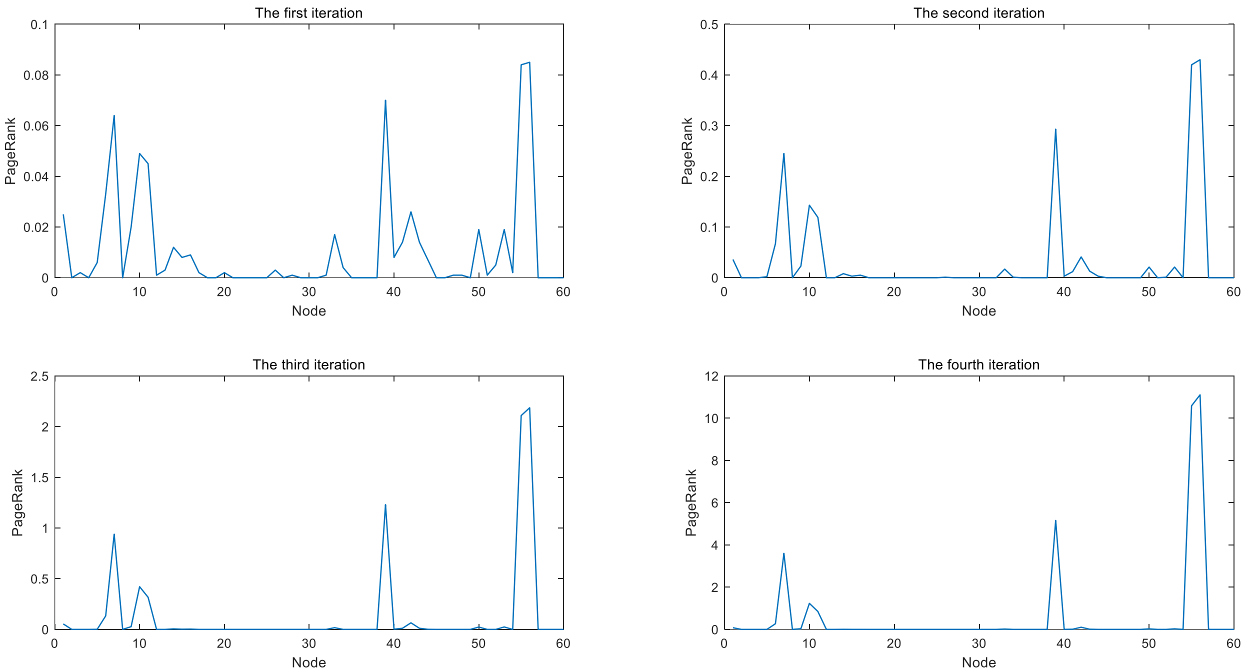


Figure 6 – The PageRank value after four iterative calculations for all accident network nodes

As illustrated in *Figure 6*, after the third iteration, the PageRank value of most nodes remained stable and reached convergence, meaning that the influence of most nodes achieved equilibrium state. As indicated in *Figure 6*, it is evident that human factors still played the most important role in the development and evolution of the accident chains. Among the top seven nodes, there were three nodes belonging to human factors, among which H7 (no effective control behaviour) ranked the first, followed by H10 (false control behaviour) and then H11 (control is not in place). Coincidently, the rank of the three causative nodes from human factors matched exactly with that according to node degree and node strength. Hence, it can be concluded that the three nodes including H7, H10 and H11 are the three most important human factors leading to railway locomotive accidents. Similarly, D39 (being forced to stop) was still the most easily triggered unsafe event, which was in line with the result calculated based on node degree and node strength. Besides, a node named with M6 (safety management system deficiencies), from management layer, appeared first on the list ranked by PageRank value.

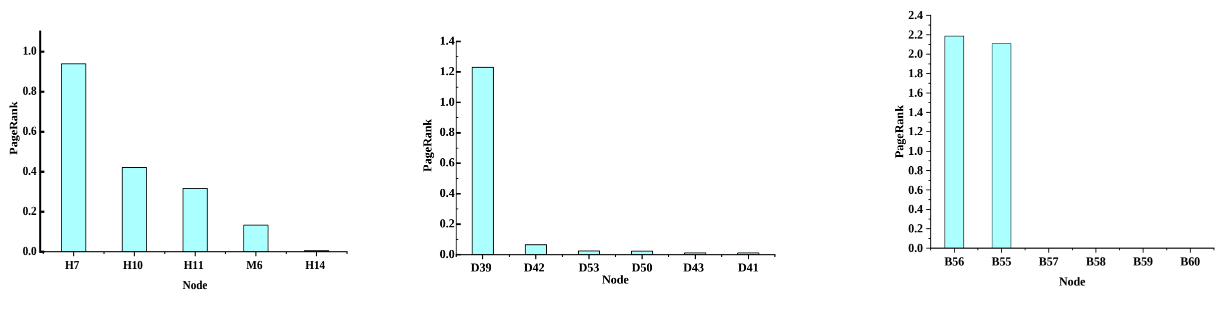
5. DISCUSSION

Firstly, this study applied STPA based method to identify all unsafe control behaviours and unsafe events leading to system-level loss for railway locomotive operation accidents. Through the four-step STPA analysis, 68 locomotive operation accident chains were formed by a total of 60 nodes which were divided into six categories including management (M1-M6), human (H7-H16), environment (E17-E25), facility (F26-F37), unsafe events (D38-D54) and system-level loss (B55-B60).

Secondly, the human factors were found to have higher impact on railway locomotive accidents when compared with other categories of causative factors as human factors accounted for four of the top five causative factors, as shown in *Figure 7a*. It can be seen that H7 (no effective control behaviour) ranked the first, meaning that H7 represents the most important causative factors, followed by H10, H11 and H14. Therefore, some targeted training measures should be taken for training drivers to reduce their negative impact on the locomotive operation. Except for those human factors, it is worthy to note that M6 (safety management system deficiencies) ranked the fourth in the list of all causative factors, implying that safety management system deficiencies should be seriously dealt with over time for railway locomotive operation.

Thirdly, for the sake of grasping the appearance opportunity of different unsafe events in railway locomotive operation, the comprehensive importance of the nodes representing unsafe events between each other was compared based on the comprehensive ranking rule, as illustrated in *Figure 7b*. According to *Figure 7b*, it was found that D39 (being forced to stop) ranked the first among all nodes of unsafe events and followed by D42 (overrunning of signal) and D50 (train separation). Thereby, being forced to stop is the most easily triggered form of railway locomotive operation incidents.

Finally, the comprehensive importance of the nodes representing system-level loss consisting of B55-B60 was also compared according to the comprehensive ranking rule, with the result shown in *Figure 7c*. We can see that the top two nodes are B56 (affecting driving order) and B55 (affecting shunting order), meaning that the system-loss forms represented by the two nodes are the two most common ones in the railway locomotive operation accident chains. Therefore, some targeted contingency plan for remedying safety constraint deficiencies should be prepared to deal with the two system-level loss forms.



(a) *(b)* *(c)*

Figure 7 – Ranking of accident network nodes in terms of PageRank: a) causative nodes; b) unsafe event nodes; c) system-level loss nodes

6. CONCLUSION

The results show that H7 (no effective control behaviour), H10 (false control behaviour), H11 (control is not in place) and M6 (safety management system deficiencies) are the most prominent nodes in the accident causation network which should be paid more attention to for safety risk management for railway locomotive operation.

For H7, H10 and H11, the high rankings may stem from varied reasons. On the one hand, the driver's response time may be insufficient, which may be due to lack of driving experience. On the other hand, the three high ranking factors cannot be directly attributed to individual locomotive drivers, but rather to the performance of the whole human-facility-environment system; for example, shunting personnel failing to inspect the track in advance, station attendants failing to notify field workers of incoming trains according to the standards, car number checkers failing to verify car numbers and maintenance personnel conducting insufficiently thorough inspections, resulting in facility hazards not being promptly addressed. The high cause rate of M6 may be due to the irresponsibility of managers and the lack of discipline among cadres. The railway industry has strong professional staff and slow replacement. Usually, employees are slow to adapt to key positions and their emergency response skills cannot quickly match with the locomotive operation requirements.

In view of the above analysis, all operational regulations must be strictly enforced to improve the safety of the railway locomotive operation. Furthermore, emergency training and exercises should be strengthened to improve the emergency response ability of the locomotive operators. All railway workers need to understand the main points of emergency response so that they can quickly deal with abnormal situations to minimise the harm of accidents and even prevent accidents.

However, due to the limitations in this study, some deep research can be conducted in future. (1) As the limited sample size of the railway locomotive operation accidents, the results in this study are inevitably subject to some bias. By extending the time span and increasing the accident sample size, more accurate conclusions can be drawn by using the method proposed in this study. (2) With the rapid development of machine learning methods, the improved PageRank algorithm based on machine learning methods should be researched in future to calculate the importance of causative nodes of railway locomotive operation accidents. (3) As the subject of this study is about analysis of locomotive operation accidents, the problem of deficiency factors on wagons has not been considered yet. Some factors contributing to wagon operation incidents will be considered to comprehensively enhance railway transportation safety in the future.

ACKNOWLEDGEMENTS

The present study is one of the supported projects funded by National Nature Science Foundation of China (52062014), Social Science Foundation of Jiangxi Province (23GL15).

REFERENCES

[1] Zhou JP, Yang LC. The implications of high-speed rail for Chinese cities: Connectivity and accessibility. *Transportation Research Part A*. 2018;116:308-326. DOI: 10.1016/j.tra.2018.06.023.

[2] Mahdi R, Richard FF. Rail transport delay and its effects on the perceived importance of a real time information. *Frontiers in Psychology*. 2021;12:619308. DOI: 10.3389/fpsyg.2021.619308.

[3] Nagy E, Csiszár C. Analysis of delay causes in railway passenger transportation. *Periodica Polytechnica: Transportation Engineering*.2015;43(2):73-80. http://real.mtak.hu/id/eprint/23409.

[4] Finochenko TA, Dergacheva LV, Yaitskov IA. Risk management in transportation safety system. *IOP conference series: earth and environmental science*. 2021;666(2):022050. DOI:10.1088/1755-1315/666/2/022050.

[5] Chen F, et al. Exploring the association between quantified road safety target attributes and their success: An empirical analysis from OECD countries using panel data. *Journal of Safety Research*. 2023;85:296-307. DOI: 10.1016/j.jsr.2023.03.003.

[6] Wei C. Research on current situation and countermeasures of high-speed railway safety supervision in China. *China Safety Science Journal*. 2022;32(S1):34-38.DOI:10.16265/j.cnki.issn1003-3033.2022.S1.0204.

[7] Bekisz A, et al. Risk Management Using Network Thinking Methodology on the Example of Rail Transport. *Energies*. 2022;15(14):5100. DOI: 10.3390/en15145100.

[8] Araz OM, et al. Role of analytics for operational risk management in the era of big data. *Decision Sciences*. 2020;51(6):1320-1346. DOI: 10.1111/deci.12451.

[9] Darroch N, Beecroft M. A qualitative analysis of the interfaces between urban underground metro infrastructure and its environment in London. *Tunnelling and Underground Space Technology*. 2021;114:103930. DOI: 10.1016/j.tust.2021.103930.

[10] Huang Y, Zhang Z, Tao Y. Quantitative risk assessment of railway intrusions with text mining and fuzzy Rule-Based Bow-Tie model. *Advanced Engineering Informatics*.2022;54:101726. DOI: 10.1016/j.aei.2022.101726.

[11] Yin J, Ren X, Liu R. Quantitative analysis for resilience-based urban rail systems: A hybrid knowledge-based and data-driven approach. *Reliability Engineering & System Safety*. 2022;219:108183. DOI: 10.1016/j.ress.2021.108183.

[12] Szkoda M, Satora M. The application of failure mode and effects analysis (FMEA) for the risk assessment of changes in the maintenance system of railway vehicles. *Technical Transactions*. 2019;116(8):159-171. DOI: 10.4467/2353737XCT.19.086.10865.

[13] Leveson N. A new accident model for engineering safer systems. *Safety science*. 2004;42(4):237-270. DOI: 10.1016/S0925-7535(03)00047-X.

[14] Castillo E, et al. A Markovian bayesian network for risk analysis of high speed and conventional railway lines integrating human errors. *Computer-Aided Civil and Infrastructure Engineering*. 2016;31(3):193-218. DOI: 10.1111/mice.12153.

[15] Zhang Q, et al. Railway safety risk assessment and control optimization method based on FTA-FPN: A case study of Chinese high-speed railway station. *Journal of advanced transportation*. 2020;2020(1):3158468. DOI: 10.1155/2020/3158468.

[16] Kyounga P, Lee JH. A before-and-after study in changes of the railway network centrality on Dongdaegu multi-modal transit hub: Focused on PageRank centrality. *Journal of Korean Society of Transportation*. 2022;40(5):669-682. DOI:10.7470/jkst.2022.40.5.669.

[17] Zhao CX, Li H. Safety analysis and evaluation of airborne HDS based on STPA-Bayes model. *Systems Engineering and Electronics*. 2020;1083-1092. DOI: 10.3969/j.issn.1001-506X.2020.05.15.

[18] Duttweiler L, Thurston SW , Almudevar A. Spectral Bayesian network theory. *Linear Algebra and its Applications*.2022;2023:674. DOI: 10.1016/j.laa.2023.06.003.

[19] Yazdi M, Kabir S. Fuzzy evidence theory and Bayesian networks for process systems risk analysis. *Human and Ecological Risk Assessment: An International Journal*. 2020;26(1):57-86.DOI: 10.1080/10807039.2018.1493679.

[20] Kembłowski WM, Grzyl B, Kristowski A. Risk modelling with Bayesian networks-case study:Construction of tunnel under the Dead Vistula River in Gdansk. *Procedia Engineering*. 2017;196:585-591. DOI: 10.1016/j.proeng.2017.08.046.

[21] Patriarca R, et al.The past and present of System-Theoretic Accident Model and Processes (STAMP) and its associated techniques: A scoping review. *Safety science*. 2022;146:105566. DOI: 10.1016/j.ssci.2021.105566.

[22] Sadeghi R, Goerlandt F. A proposed validation framework for the system theoretic process analysis (STPA) technique. *Safety science*, 2023;162:106080. DOI:10.1016/j.ssci.2023.106080.

[23] Yang P, et al. Automated inspection method for an STAMP/STPA-Fallen barrier trap at railroad crossing. *Procedia Computer Science*. 2019;159:1165-1174. DOI: 10.1016/j.procs.2019.09.285.

[24] Liu HT, et al. High-speed railway emergency dispatching safety analysis based on STAMP/STPA. *China Safety Science Journal*. 2021;31(6):113. DOI: 10.16265/j.cnki.issn 1003-3033.2021.06.015.

[25] Benhamlaoui W, Rouainia M , Liu Y. Comparative study of STPA and Bowtie methods: Case of hazard identification for pipeline transportation. *Journal of Failure Analysis and Prevention*. 2020;20(6):2003-2016. DOI: 10.1007/s11668-020-01010-9.

[26] Howard G, et al. Formal analysis of safety and security requirements of critical systems supported by an extended STPA methodology. *IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Paris, France*.2017;174-180. DOI: 10.1109/EuroSPW.2017.68.

[27] Cheng X, Scherpen J. Model reduction methods for complex network systems. *Annual Review of Control, Robotics, and Autonomous Systems*. 2021;4:425-453. DOI: 10.1146/annurev-control-061820-083817.

[28] De Bona, et al. A reduced model for complex network analysis of public transportation systems. *Physical A: Statistical Mechanics and its Applications*. 2021;567:125715. DOI:10.1016/j.physa.2020.125715.

[29] Bojchevski A, et al. Scaling graph neural networks with approximate PageRank. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020;2464-2473. DOI: https://doi.org/10.48550/arXiv.2007.01570.

[30] Zhou Y, et al. Metro station risk classification based on smart card data: A case study in Beijing. *Physical A: Statistical Mechanics and its Applications*. 2022;594,127019. DOI: 10.1016/j.physa.2022.127019.

[31] Spanninger T, Büchel B, Corman F. Train delay predictions using Markov chains based on process time deviations and elastic state boundaries. *Mathematics*. 2023;11(4):839. DOI: 10.3390/math11040839.

[32] Park S, Lee W, Choe B. A Survey on personalized PageRank computation algorithms. *IEEE Access*. 2019;7:163049-163062. DOI:10.1109/ACCESS.2019.2952653.

万平, 杨伟伦, 罗杰文, 马晓凤

基于 STPA-PageRank 方法的铁路机务行车事故链致因节点重要度分析

摘要

如今,针对铁路机务行车中的复杂随机事件,若是预防和控制每个微小潜在的影响因素,不仅费用高昂,还给机车驾驶员带来了巨大的心理压力。本文首先收集整理了近两年发生的 68 组铁路机务行车事故数据。接着,采用系统理论过程分析(STPA)方法提取了 68 条事故链。然后,利用复杂网络理论和 PageRank 算法计算由事故链构成的有向加权网络中各节点的重要度。结果表明:人为因素的重要度显著高于环境、设施和管理等其它因素。其中,未进行有效的控制行为(H7)和进行了错误的控制行为(H10)是所有人为因素中最重要的两个事故致因节点。此外,被迫停车(D39)和冒进信号(D42)是不安全事件中最重要的两个致因节点。因此,对于 PageRank 值较高的致因节点,应采取针对性的安全措施,进而节省风险管理投入,并提高机务行车系统的整体安全水平。

关键词

铁路机务行车; 风险管理; 复杂网络; STPA; PageRank