

Statistic Characteristics Analysis of Railway Hazard Causation based on Complex Network

Jintao Liu^{1,2,*}, Keping Li³, Wei Zheng^{1,2} and Chenling Li³

¹National Research Center of Railway Safety Assessment, Beijing Jiaotong University, China

²Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, China

³State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China

*Corresponding author

Abstract—Quantitative analysis of the potential causes of railway accidents is essential for ensuring the safety of railway. In this context, this paper proposes a directed railway hazard causation network (DRHCN) based on the complex network theory. According to the characteristic of the DRHCN, four statistical characteristic indicators of hazard causation are obtained. The statistic characteristic analysis of the train collision hazard causation is performed. The experimental results show that the proposed DRHCN model and its statistic characteristics can assist us to reveal the latent rules of the railway hazard causation and, hence, greatly prevent the railway accidents.

Keywords—statistic analysis; railway hazard; complex network; hazard causation

I. INTRODUCTION

The railway plays a significant role in daily life. Any accident of it may lead to casualties and property losses. For avoiding the railway accidents, it is essential to perform hazard analysis of the railway systems to identify the potential causes (called hazard causation) of accidents before accidents occur. Nowadays, the effective and widely used hazard analysis methods are based on systemic accident models, such as AcciMap method [1], STAMP approach [2]. However, these systemic methods are not quantitative but only qualitative. Only qualitative identification of the hazard causation is not sufficient to establish the effective control measures for the potential causes. Therefore, it is required to quantitatively analyze the hazard causation.

Railway systems are complex social-technical systems consisting of humans, technical equipments and environment. The railway hazard results from external disturbances, technical component failures, and dysfunctional interactions between humans and equipments. The component failures can be quantitatively evaluated by occurrence probabilities. Nevertheless, in reality, the characteristic of the dysfunctional interaction are not obvious. It is hard to obtain the occurrence probabilities of it. Hence, the traditional quantitative causation analysis approaches, such as FTA method [3], Bayesian network method [4], is not suitable for quantitative analysis of the railway hazard causation.

Since the railway systems are complex in terms of the large number of subsystems and their complex functions, hazard causation and relationships among them are always complex. As one of the most important part of the complexity theory, the complex network [5] is a powerful tool to statistically analyze

the complex objects and has been applied to the quantitative causation analysis by some researchers. For example, Ma et al. [6] and Zhou et al. [7] built an undirected railway accident causation network and utilized the statistic characteristics of the complex network to evaluate the causation. Klockner [8] utilized centrality analysis of the complex network to quantitatively investigate the contribution of causes to a railway accident. Lin et al. [9] revealed the hidden traffic accident causation through the cluster analysis of the causation network. However, these studies utilized the undirected complex network to model the causation, which has limitations in depicting the real relationships among causes. For example, design errors may lead to signal equipment faults but the signal equipment faults have no influence on the design errors, which cannot be depicted by an undirected network.

Differing from the aforementioned undirected causation network, in this paper, a directed railway hazard causation network (DRHCN) is modeled based on the complex network theory. Then based on the DRHCN, the statistic characteristics of the complex network are utilized to quantitatively analyze the hazard causation. As a case study, the hazard of train collision and its causation are analyzed. The experimental results show that the proposed DRHCN model and its statistic characteristics are effective for quantitatively analyzing the railway hazard causation.

II. MODELING DIRECTED RAILWAY HAZARD CAUSATION NETWORK

A. Definition of DRHCN

In the systemic accident models, the hazards occur when external disturbances, component failures, and dysfunctional interactions between humans and equipments are not adequately controlled. Hence, the causation identified by systemic hazard analysis methods can be divided into unsafe control actions (UCAs) and the potentially causal scenarios (CSs) that could cause the UCAs.

To depict the hazard causation and their relations, the complex network is employed. The complex network is a topological description for the complex system. It is represented as a graph consisting of nodes and edges. The nodes are connected by the edges that can be undirected or directed. If two nodes are connected, it means that these two nodes have a certain relationship between each other, such as a causal relationship, a physical connectivity relationship. Based

on the complex network theory, directed railway hazard causation networks (DRHCNs) of railway systems can be constructed to depict hazard causation and their relationships. The definition of the DRHCN is shown as follows.

Definition 1 The DRHCN consists of the representing characters of $\langle V, E \rangle$, where,

- 1) $V = \{V_c, V_s, V_h\}$ refers to a finite set of nodes, where the V_c refers to the nodes representing the UCAs, the V_s refers to the nodes representing the CSs, and the V_h refers to the node representing the hazard;
- 2) E refers to a finite set of directed edges in the network. The nodes in the network are connected by these directed edges that represent the causal relationships among the hazard causation.

In the DRHCN, if a node v_i representing a cause may be caused by some nodes, the v_i is connected with these nodes by incoming edges, and these nodes are denoted as $N(v_i)$. If the node v_i may lead to the presence of some other nodes, the v_i is connected with the nodes by outgoing edges, and the nodes are denoted as $\tilde{N}(v_i)$. A simple DRHCN graph with five nodes is shown in Figure 1. In Figure 1, $N(v_2) = \{v_1, v_3\}$, and $\tilde{N}(v_2) = \{v_4\}$. It means that the presence of v_2 may be caused by the v_1 and the v_3 , and the v_2 may lead to the presence of the v_4 .

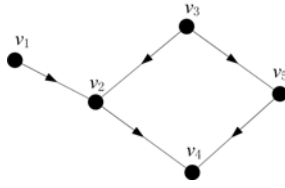


FIGURE I. A SIMPLE DRHCN GRAPH WITH FIVE NODES

B. Steps of Modeling DRHCN

According to the definition above, the DRHCN of railway systems can be established in following steps, which are shown in Figure 2.

Step 1: Construct causal networks of unsafe control actions. An UCA may be caused by the CSs, the first step of modeling the DRHCN is that each UCA is connected with some CSs by incoming edges. It is noteworthy that an incoming edge represents a possibly causal relationship instead of an inevitably causal relationship.

Step 2: Connect the causal networks. The high-level control actions will impact on the low-level control actions. It means that the UCAs at high level may also cause the low-level UCAs. Hence, the second step is to connect the causal networks of the UCAs according to the relationships between the UCAs at different levels.

Step 3: Connect the system-level hazard with the causal networks. The presence of the system-level hazard may be directly caused by some UCAs. Therefore, the hazard is connected with these UCAs by some incoming edges. As each UCA is modeled in a causal network, a causation network for the hazard is constructed, which is a DRHCN.

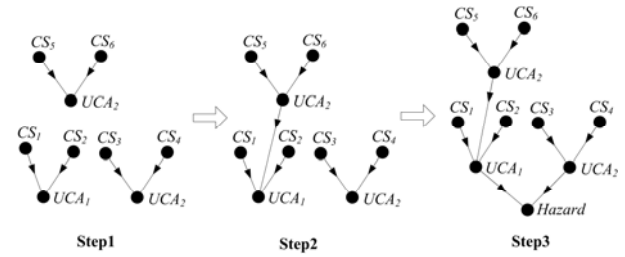


FIGURE II. STEPS OF MODELING A DRHCN

III. STATISTIC CHARACTERISTICS OF DRHCN

Complex networks are topological graphs with some parameters that reflect the statistic characteristics of networks. In this paper, four statistic characteristic parameters including output degree, degree distribution, shortest path length and betweenness are utilized to statistically analyze the hazard causation.

A. Output Degree and Degree Distribution

The node degree of a node is the number of edges connecting with this node [5]. If a node has larger node degree than others, it means that this node is a key point of the network. Considering the directions of edges, the node degree can be divided into input degree (the number of incoming edges) and output degree (the number of outgoing edges). As this paper focuses on the contribution of causes to the hazard, it is more significant to calculate the output degrees of nodes. The output degree of a node is calculated in terms of the adjacency matrix of the DRHCN:

$$k_i^{out} = \sum_{j \in N} a_{ij} \quad (1)$$

where N is the number of the causation nodes in DRHCN, a_{ij} is the value of the element in the i th row and the j th column of the adjacency matrix.

Based on the node degree, the degree distribution $p(k)$ is defined as the probability of a randomly selected node with degree k . It reflects the statistical properties of a network. Nevertheless, the degree distribution $p(k)$ is often noisy because of the finite network size. Therefore, the cumulative degree distribution $P(k)$, the probability of a randomly selected node with degree greater than or equal to k , is used for analysis instead. In this paper, we utilize the cumulative degree distribution to evaluate the causes in the DRHCN.

B. Shortest Path Length in DRHCN

The shortest path length is the number of steps along the shortest path connecting two nodes. Since this paper focuses on the influence of causes on the hazard, the shortest path length refers specifically to the shortest path length from each node to the node *Hazard*. For a node i , its shortest path length is denoted as d_{ih} . For example, in Figure 2, the shortest path lengths of the nodes CS_2 and CS_6 are 2 and 4 respectively.

According to the above definition, the smaller value of the shortest path length, the more direct influence of the cause on the hazard. Some safety measures can be implemented for

increasing the shortest path length so as to reduce the possibility of the hazard.

C. Betweenness in DRHCN

In the network theory, the betweenness of a node is defined as the proportion of the number of shortest paths going through this node to the number of all shortest paths in a network. It is used to measure what extent a node appears in the connection paths between the other nodes. Since the shortest paths refer specifically to the shortest paths leading to the node *Hazard* in this paper, the traditional betweenness should be tailored. The definition of betweenness b_i of a node i in the DRHCN is shown as follows.

$$b_i = \sum_{j,h \in N, j \neq h} \frac{n_{jh}(i)}{n_{jh}} \quad (2)$$

Where, both j and h are the nodes in the DRHCN, h refers specifically to the node *Hazard*, $n_{jh}(i)$ is the number of the shortest paths going through the node i and connecting the j and h , and n_{jh} is the number of the shortest paths connecting the j and h . According to the definition above, the betweenness centrality of a node reflects the role of intermediary the node plays in the emergence of the system hazard. In the DRHCN, the larger the betweenness centrality of a node, the larger numbers of shortest paths the node appears in.

IV. STATISTIC ANALYSIS OF TRAIN COLLISION CAUSATION

In this section, we take the hazard of train collision and its causation as a study case, and statistically analyze the causation.



FIGURE III. THE DRHCN MODEL OF TRAIN COLLISION

The hazard of train collision has been identified in [10]. And the causes of this hazard are related to five kinds of entities: dispatcher (D), centralized traffic control system (C), radio block center (R), vital computer (V) and train driver (T). According to the steps of modeling DRHCN, the directed causation network can be constructed and shown in Figure 3. The meanings of the nodes can be found in [10].

A. Output Degree and Degree Distribution

The average output degree of total 78 nodes in this DRHCN is 1.57. It indicates that each causal factor averagely leads to one or two causal factors. Only 25 nodes in the DRHCN have the output degrees greater than or equal to 2, as shown in

Figure 4. The values of other 53 nodes are equal to 1. As shown in Figure 4, the node “DA01” has the highest output degree with the value of 5. This indicates that the “DA01” may lead to five other causal factors and is a key point in the DRHCN. As shown in [10], the “DA01” means that the dispatcher violates working procedures because of misoperation. It is important to prevent such human failure by employee training and safety culture.

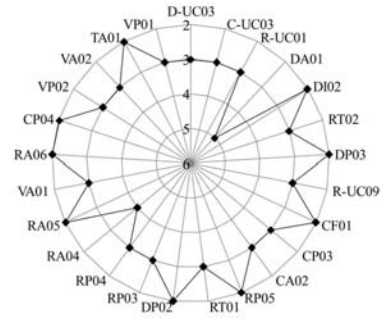


FIGURE IV. NODES WITH OUTPUT DEGREE GREATER THAN OR EQUAL TO 2

The cumulative degree distribution $P(k)$ of the DRHCN is illustrated in Figure 5. As shown in Figure 5, the $P(k)$ obeys power-law distribution with the approximate fit $P(k) \sim 2.534k^{-1.91}$. This indicates that the DRHCN can be treated as a kind of scale-free network. The average degree of nodes is 3.1, where the nodes with degree less or equal to 2 account for 51.3% of total 78 nodes. A small fraction of nodes have high degrees, such as the $V-UC01$ (degree is 15) and the $V-UC03$ (degree is 11) that are related to vital computer. This scale-free network characteristic means that the DRHCN is robust to random occurrence of causation, but is vulnerable to the occurrence of nodes with high degrees due to deliberate attacks. Eliminating or controlling the nodes with high degrees, such as the $V-UC01$ and the $V-UC03$, can cut off much of spread routes of causes. It has positive effects on the safety of the railways.

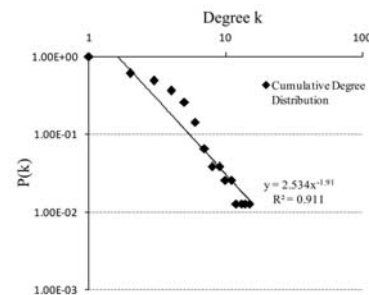


FIGURE V. CUMULATIVE DEGREE DISTRIBUTION

B. Shortest Path Length

The shortest path length from each node to the node *Hazard* is shown in TABLE I. All the shortest path lengths are less than 5. It means that any cause could lead to the presence of the hazard within five steps. Besides, it is shown in TABLE I that some causes related to both the vital computer and the train driver have the shortest path lengths equal to 1. This conforms to the actual situation that both the vital computer and the train driver control the train directly. The occurrence of these causes

may lead to the hazard directly. This indicates that controlling these causal factors can play a direct role in preventing the hazard.

TABLE I. SHORTEST PATH LENGTHS

Object	Shortest path length from each node to the node Hazard				
Dispatcher	D-UC01: 4 D-UC02: 4 D-UC03: 4	D-UC04: 2 D-UC05: 4 DP01: 3	DP02: 5 DP03: 3 DA01: 3	DI01: 5 DI02: 5	DF01: 3 DF02: 5
centralized traffic control system	C-UC01: 3 C-UC02: 3 C-UC03: 3	C-UC04: 3 C-UC05: 3 CP01: 4	CP02: 4 CP03: 4 CP04: 4	CP05: 4 CA01: 4 CA02: 4	CA03: 4 CF01: 4
Radio Block Center	R-UC01: 2 R-UC02: 2 R-UC03: 2 R-UC04: 2 R-UC05: 2 R-UC06: 2	R-UC07: 2 R-UC08: 2 R-UC09: 2 R-UC010: 2 RP01: 3 RP02: 3	RP03: 3 RP04: 3 RP05: 3 RA01: 3 RA02: 3 RA03: 3	RA04: 3 RA05: 3 RA06: 3 RI01: 3 RI02: 3 RI03: 3	RI04: 3 RI05: 3 RI06: 3 RT01: 3 RT02: 3
Vital Computer	V-UC01: 1 V-UC02: 1 V-UC03: 1	VP01: 2 VP02: 2 VP03: 2	VA01: 2 VA02: 2 VA03: 2	VI01: 2 VT01: 2 VT02: 2	VF01: 2
Train Driver	T-UC01: 1 T-UC02: 1	TP01: 2 TP02: 2	TA01: 2 TI01: 2	TI02: 2 TI03: 2	

C. Betweenness

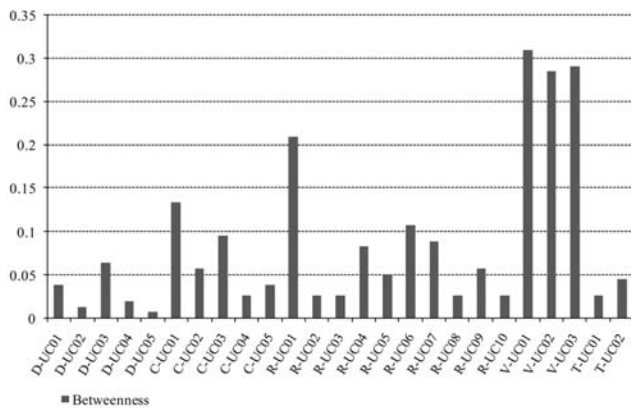


FIGURE VI. BETWEENNESS OF THE DRHCN

The betweenness of nodes in the DRHCN are calculated and illustrated in Figure 6. There are only 25 nodes having betweenness of more than zero. The values of the others are zero and invisible, because they do not play roles of intermediaries in the spread of the causes. There are 6 nodes having betweenness greater than 0.1. They are the *C-UC01*, the *R-UC01*, the *R-UC06*, the *V-UC01*, the *V-UC02* and the *V-UC03* respectively. The total values of their betweenness are equal to 1.34 accounting for 62.4% of the sum of all betweenness. This means that 62.4% of all shortest paths go through these 6 nodes. Controlling these causes can increase the lengths of the shortest paths between the most nodes, which will slow down the spread efficiency of the causation.

V. CONCLUSION

In this paper, the complex network theory is employed to model the directed railway hazard causation network (DRHCN). Based on the DRHCN, four statistic characteristics of hazard causation are obtained. The statistic characteristic analysis of the hazard causation is performed to quantitatively

evaluate the potential causes of the train collision. In the real case study, the key hazard causation, such as dispatcher errors, vital computer-related causes (*V-UC01*, *V-UC03*, et al.) and radio block center-related causes (*R-UC01* and *R-UC06*) are identified. These causes should be preferentially prevented by control measures.

ACKNOWLEDGMENT

This work is co-supported by National Natural Science Foundation of China (No.U1434209) and the Fundamental Research Funds for the Central Universities (No. 2015JBZ006).

REFERENCES

- [1] B. Rasmussen. Functional modelling as basis for studying individual and organisational factors—application to risk analysis of salmonella in pork. *Food Control*, 12(3), pp.157-164, 2001.
- [2] N. Leveson. *Engineering a Safer World: Systems Thinking Applied to Safety Engineering Systems*. Massachusetts: MIT Press, 2012.
- [3] P. Liu, L.X.Yang. Fault tree analysis combined with quantitative analysis for high-speed railway accidents. *Safety Science*. 79, pp.344-357, 2015.
- [4] D. Smith, B. Veitch. Understanding industrial safety: Comparing Fault tree, Bayesian network, and FRAM approaches. *Journal of Loss Prevention In The Process Industries*. 45, pp.88-101, 2016.
- [5] O.P. Kuznetsove. Complex networks and activity spreading. *Automation and remote control*. 76(12), pp. 2091-2109, 2015.
- [6] X. Ma, K.P. Li. Analyzing the causation of a railway accident based on a complex network. *Chin Phys*. 23(2), 028904, 2014.
- [7] J. Zhou, K.P. Li. “Railway Accident Analysis Using Information Theory and Complex Network”, *ICIEMS*, Shanghai, pp. 865-870, Sep 2013.
- [8] K. Karen, S. Lecturer. “Railway accident modelling in a connected world”, *Conference on Railway Excellence*, Adelaide, pp.1-10, May, 2014.
- [9] Lin, Lei, Q. Wang, and A. W. Sadek. *Data Mining and Complex Network Algorithms for Traffic Accident Analysis*. *Journal of the Transportation Research Board*. pp. 128-136, 2014.
- [10] Dong A R Application of CAST and STPA to railroad safety in China (MS Thesis). Boston: Massachusetts Institute of Technology (MIT), 2012.