Reliability Engineering and System Safety 207 (2021) 107352

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

A knowledge graph-based approach for exploring railway operational accidents

Jintao Liu [a,b,*], Felix Schmid [c], Keping Li [d], Wei Zheng [a,b]

[a] National Research Center of Railway Safety Assessment, Beijing Jiaotong University, China
[b] Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, China
[c] Birmingham Centre for Railway Research and Education, University of Birmingham, UK
[d] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China

ARTICLE INFO

Keywords:
Network analysis
Topological analysis
Knowledge graph
Accident analysis
Railway operational accident

ABSTRACT

Drawing lessons from past accidents is an essential way to improve the operational safety of railways. Various railway operational accidents and their related hazards constitute a causation network due to the interactions among the hazards. Some useful lessons can be captured from such a network. In this paper, a new knowledge graph-based approach to explore railway operational accidents is proposed, aiming to reveal the potential rules of accidents by depicting accidents and hazards in a heterogeneous network. This work serves as an extension and complement to classical homogeneous network-based accident analyses. Its originality is to apply the knowledge graph theory to railway operational accident analysis, by means of some topological indicators adapting to the heterogeneous structural features of knowledge graphs. To facilitate the construction of the accident knowledge graph, a modelling method is developed. The outcomes of the knowledge graph-based analysis provide railway operators with the decision-making basis for the investment of accident prevention efforts. An application on real railway operational accidents in the UK is presented. The results show the effectiveness of the proposed approach in terms of discovering the latent features of the corresponding railway operational accidents and assisting in formulating targeted preventive measures.

1. Introduction

Railways play essential roles in people's lives. The operation of railways depends on many elements, such as rolling stock, infrastructures, human operators, management and organisation [1]. Problems of these elements may lead to accidents during the operation of railways, i.e., railway operational accidents. Such accidents may result in fatalities and injuries. In the past decade, several severe railway operational accidents have happened. For example, the China-Yongwen railway operational accident occurred in July 2011, leading to 40 fatalities and 172 injuries [2]. In December 2015, an Amtrak passenger train derailed in Philadelphia, resulting in 8 fatalities and 185 injuries [3]. Less than one year later, the Santiago de Compostela in Spain train derailment happened and resulted in 78 fatalities and 145 injuries [4]. To prevent railway operational accidents, exploring and understanding past accidents is a fundamental tool [5, 6]. From the understanding of past railway operational accidents, the valuable knowledge contributing to accidents prevention can be captured.

A direct and widely used way to explore railway operational accidents is accident causation analysis, i.e., figuring out the causes of an accident and how they lead to the accident. The cornerstones of accident causation analyses are various accident models [7, 8]. The models reveal why and how accidents occur, such as the sequential accident model [9, 10], the epidemiological accident model [11, 12], and the systematic accident models [13, 14]. Underpinned by the accident models, a number of accident causation analysis methods have been developed and applied to railway accidents. Liu et al. [15], for example, analysed the China-Yongwen railway operational accident by use of the Fault Tree Analysis (FTA) method underpinned by the sequential accident model. Lin et al. [16] researched the fault tree analysis of adjacent track railway accidents. As for the limitations of FTA on static structure and uncertainty, they were overcome by mapping fault tree into Bayesian network (BN) [17, 18]. Baysari et al. [19] analysed 40 railway operational accidents in Australia using the Human Factors Analysis and Classification System (HFACS) derived from the epidemiological accident model. Zhan et al. [20] applied the HFACS to the China-Yongwen

* Corresponding author at: National Research Center of Railway Safety Assessment, Beijing Jiaotong University, China.
E-mail address: ljt@bjtu.edu.cn (J. Liu).

https://doi.org/10.1016/j.ress.2020.107352
Received 13 August 2020; Received in revised form 28 October 2020; Accepted 8 November 2020
Available online 26 November 2020

railway operational accident analysis. The Bow-tie (BT) model that is based on the epidemiological accident model was also able to be applied to accident modeling and analysis [21, 22]. Similar to the limitations of FTA, the static features of BT were also be overcome by the flexible structure of BN [23]. Besides, the System Hazard Identification Prediction and Prevention (SHIPP) method that combines the advantages of both sequential accident model and epidemiological accident model was also widely used in accident analysis [24, 25]. In terms of systematic accident models-based analyses, Belmonte et al. [26] utilised the Functional Resonance Accident Model (FRAM) to analyse accidents caused by railway switch locking failures. Salmon et al. [27] explored the causal factors underlying a railway crash accident with the Accimap method. Taking the China-Yongwen railway operational accident as a study case, Fan et al. [28] used system causal loop diagrams to find the contributory factors to the accident, and Li et al. [29] combined HFACS and Systems-Theoretical Accident Modelling and Processes (STAMP) together to reveal a number of human factors in the accident. Accident causation analysis is undoubtedly a significant way to explore and understand railway operational accidents.

Another widely used way to understand railway operational accidents is by statistically analysing past accident data. In this way, the hidden knowledge among past data, such as the knowledge related to accident frequency, accident severity and accident distribution, can be extracted to support safety strategies [5]. In the statistical analysis of accident frequency, some metrics are usually used for further insights into railway accident frequency, such as railway level crossing type, railway traffic density, railway track class and wayside device type. Evans [30] investigated the fatal accident frequencies by different railway level crossing types in Great Britain from 1946 to 2009. Naznin et al. [31] explored the effects of service frequency, route section length and signal priority on tram crash accident frequency. Liu et al. [32] quantitatively assessed the freight-train derailment frequency by the track class, the method of operation and the traffic density. Jonsson et al. [33] found the influence of the wayside device type and the traffic volume of trains on railway level crossing accident frequency. Evans et al. [34] calculated the fatal accident frequency at different types of railway level crossings. Zhou et al. [35] investigated the 407 accidents/incidents' frequency by their types and found the major human errors with largest percentage of occurrence. In accident severity analysis, the casualties and the number of damaged rolling stock are usually utilised to measure the accident severity. Furthermore, some factors, e.g., accident type, train speed, or cause type, can be used to explore the accident severity. Ahmad et al. [36] explored the relationships among human errors and the severities and types of accidents from railway accident records. Kim et al. [37] categorised the levels of severity of 80 railway operation accidents in the UK by the types of adverse events leading to accidents. Dabbour et al. [38] identified the relationships between the severity of injuries in railway level crossing collision accident and some factors, such as train speed, driver's gender and age. Zhang et al. [39] categorised the railway accident severity in the U.S. from 2000 to 2016 by the types of accidents. Lin et al. [40] counted the different severities of railway operational accidents caused by different causes in the U.S. from 1996 to 2017. In accident distribution analysis, the distribution of accidents by various factors, such as train type, train driver, location, and time, can be counted. Savage [41] investigated the distribution of fatal train-pedestrian collision accidents in Chicago from 2004 to 2012 by time of day, day of week, season, train line and train type. Naznin et al. [42] explored the distribution of tram-involved fatal crash accidents in Melbourne by various metrics, such as driver experience, tram length, speed and season. Haleem [43] counted the distribution of highway-railway level crossing crash accidents in the U.S. from 2009 to 2014 by a number of factors, e.g., train speed, time of crash, or crash location. Rudin-Brown et al. [44] investigated the distribution of Canada's freight rail accidents from 1995 to 2015 by fatigue, i.e., fatigue in materials, mental and physical fatigue in humans.

Whether in the accident causation analyses or the statistical analyses of railway operational accidents, there is a growing acceptance that a railway operational accident usually arises from multiple hazardous factors and their interactions [26, 27, 45]. The hazardous factors here can reside in any one or more of railway operation-related components, such as rolling stock, infrastructures, human operators, management and organisation. Due to complex interactions among the railway operation-related components [1], hazardous factors residing in the components are mutual interacted, e.g., a hazard can be the trigger of another one [46]. The interactions among hazards constitute an accident causation network, of which different railway accidents are manifestations [47].

To find additional knowledge derived from the complex interactions among accident causal factors, topological analyses of railway accident causation networks have received a great deal of attention. In railway accident causation networks, nodes represent accidents and/or their causal factors, e.g., hazards, and edges connecting nodes stand for the relationships between nodes. Various topological indicators in complex network theory [48], such as degree, diameter, clustering coefficient and betweenness centrality, are employed to explore the latent characteristics of accidents. By the directions of edges, existing railway accident causation networks can be categorised as undirected networks and directed networks. In the undirected railway accident causation networks, the relationships between nodes are abstracted as undirected edges. Klockner et al. [49] performed topological betweenness centrality analysis on an undirected network of driver misjudgment-related Signals Passes at Danger (SPAD) accidents and identified the main contributing factors. Shao et al. [50] constructed an undirected accident causation network based on the railway accident data in the U.S. from 2004 to 2013, and identified the key causal factors to accidents with degree and degree distribution analyses. Klockner et al. [51] employed the betweenness centrality of undirected networks to rank the contributing factors to a railway derailment accident. Li et al. [52] utilised the topological community analysis and node degree analysis to explore the hub node of an undirected railway operational accident causation network. In the directed railway accident causation networks, the relationships between nodes, e.g., cause-effect relationships, are abstracted as directed edges. Li et al. [46] established a directed metro operation hazard network, and applied seven topological indicators, such as network density, node degree and clustering coefficient, to reveal the structural properties of this network. Zhou et al.[53] constructed a railway causation network constituted by human factors and explored the importance of nodes. Liu et al. [54] employed three traditional topological indicators including input degree, output degree and betweenness centrality to rank the causes in the directed causation network of a train over speed hazard. Liu et al. [47] used several tailored topological indicators to reveal the latent characters of a directed railway operational accident causation network. Hou et al. [55] constructed a directed hazard network to reveal the topological features of subway construction accidents. Lam et al. [56] performed local, global and contextual topological analyses on a directed network of railway incidents in Japan. The above topological research on accident causation networks provides an overall and systematic view for understanding accidents. Through the topological analyses, some latent knowledge can be captured, such as the local influences and the intermediary roles of certain accident causal factors, the degree of closeness of relationships between accident causal factors, and the structural properties of causation networks. This is significant for deeply learning from past railway operational accidents.

It is worth noting that most of the above railway accident causation networks belong to a kind of single-dimensional complex networks [57], i.e., homogeneous networks consisting of the same type of nodes (e.g., hazard nodes) and the same type of edges (e.g., cause-effect relationship edges). Such an accident causation network can only depict one type of relationship between the same type of nodes. Comparatively speaking, multi-dimensional networks that include both various types of nodes (e.g., accident nodes and hazard nodes) and various types of edges (e.g.,

cause-effect edges and interdependent edges), i.e., a kind of heterogeneous networks, can provide more information for topological analysis [57, 58]. This is helpful in further exploring the nature of railway operational accidents. However, most of the widely used topological indicators are derived from single-dimensional complex networks, such as degree, clustering coefficient and betweenness centrality. It is difficult to use these topological indicators to reveal the inherent features of multi-dimensional railway accident networks. Although some improved topological indicators have been developed [47], the improved indicators can only be applied to railway accident causation networks that contain different types of nodes but no different types of edges. Therefore, to get further insights into railway operational accidents, multi-dimensional networks of railway operational accidents and the corresponding topological analysis approach still need to be further researched.

Knowledge graphs [59], as a kind of emerging networks, provide a new insight for multi-dimensional network modelling of railway operational accidents. By abstracting the interrelated domain knowledge entities as connected network nodes, a knowledge graph can depict a variety of relationships between various knowledge entities. Knowledge graphs have been applied to knowledge modelling in safety field [60-62]. This study aims to propose a knowledge graph-based modelling and topological analysis approach for exploring railway operational accidents. As one of the major aims of this work, a modelling method is developed to construct the railway operational accident knowledge graph (ROAKG). In the construction of the ROAKG, three kinds of structure matrices are utilised to extract a variety of relationships between knowledge entities. To perform topological analysis on the ROAKG, some new topological indicators adapting to the multi-dimensional features of the ROAKG are then proposed. Through the application to real railway operational accidents, the usefulness of the proposed approach is validated.

The remainder of this paper is organised as follows. Section 2 introduces a knowledge graph modelling method for railway operational accidents, and how to use some tailored indicators to perform topological analysis on the railway operational accident knowledge graph. Section 3 then shows an application of the proposed approach to real railway operational accidents. Based on the results from Section 3, some safety strategies for railway operation are discussed in Section 4. Finally, some conclusions are drawn in Section 5.

2. Methodology

2.1. Railway operational accident knowledge graph (ROAGK) modelling

To explore railway operational accidents using a knowledge graph-based topological analysis, a ROAKG must be constructed. A knowledge graph consists of knowledge entities and their relationships [59]. To construct the ROAKG, identifying knowledge entities and their relationships, and mapping them into a network graph are three primary works.

2.1.1. Step1: Identification of knowledge entities

Knowledge entities of the ROAKG are basic knowledge units related to railway operational accidents. Identification of knowledge entities is the first step for learning from railway operational accidents with knowledge graphs. Railway operational accidents are manifestations of a latent network of interrelated hazards [46, 47], i.e., railway operational accidents are related by means of interacted hazards. Hence, railway operational accidents and their related hazards are treated as two kinds of primary knowledge entities of the ROAKG in this study. From the perspective of railway accident prevention, the identification of the role of a hazard in the hazard propagation (i.e., the role of a source hazard, a transition hazard or an accumulation hazard) can contribute formulating a general prevention strategy for the hazard, such as a strategy that prevents it from being caused and a strategy that prevents it from causing others. For turning the general strategy into specific prevention measures, the identification of the types of both immediate predecessors and successors of the hazard is helpful. Furthermore, the level of resource or efforts invested into each prevention measure can be provided by the risk associated with the hazard. The above information is closely relevant to the following factors, including accident consequences, hazard types as well frequencies. Hence, these factors are also considered as the basic knowledge entities in the paper. In addition to the above listed knowledge entities, contributing factors that may sustain the occurrence or exacerbate the consequences of railway accidents are also noteworthy in exploring accidents. However, eliminating or controlling contributing factors would not prevent railway accidents directly [63]. This is different from hazard control or elimination that would directly prevent the occurrence of railway accidents. Hence, contributing factors are not taken into account in this study.

The above knowledge entities can be obtained from well investigated accident data. It is worth noting that the coverage of the accident data determines the scope of the corresponding knowledge entities, i.e., the applicable objects of the knowledge entities-based analysis results. To illustrate how to identify knowledge entities, the publicly available accident investigation reports are taken as the data source. They are provided by many agencies, such the Rail Accident Investigation Branch (RAIB) of the UK, the National Transport Safety Board (NTSB) of the U. S., and the Japan Transport Safety Board (JTSB). The publicly available accident investigation reports published by the above agencies provide a kind of well investigated accident data that is about serious accidents and events that are significant for railway systems. The exploration of the ROAKG derived from these significant events is helpful for preventing such events. The knowledge entities of the ROAKG can be found or statistically obtained in well-designed investigation reports, e.g., the reports from the RAIB of the UK. Two examples of railway operational accidents recorded by the RAIB are shown in Table 1.

From a well-designed railway accident investigation report, a railway operational accident and its consequence can be found. Since the subsequent risk analysis based on ROAKG focuses on the safety risk, i.e., the estimation of harm to staff, passengers or public members, the consequence of the accident here is quantified by the fatalities and weighted injuries (FWI) shown in Table 2 [65]. For example, a derailment accident and its consequence of 0.008 FWI can be found in the report R182009 shown in Table 1. To identify hazards, the Checklist Method [66, 67], as a widely used hazard identification technique, is used in this study. Two railway hazard checklists [68, 69] provided by Rail Safety and Standards Board (RSSB) can be utilised to identify hazards leading to railway operational accidents. Meanwhile, the hazards are categorised by where they appeared in railway operation, such as human beings, equipment and infrastructure, the environment, and management and organisation [47]. It is worth noting that a hazard is a condition resulting in accidents potentially and could be caused by various types of specific causes. It is difficult to categorise a hazard by its causes' types. According to where hazards appear, the types of hazards include human (H)-type, equipment and infrastructure (EI)-type, the environment (E)-type and management and organisation (M)-type. For example, a hazard *misjudgment of the current hazardous situation* (denoted as H01) with H-type is identified from the report R072009 and shown in Table 1. Besides, the occurrence frequency of each hazard can be estimated by counting the number of times it's recorded in all the accident reports to be analysed. Here, the frequency of a hazard refers to the frequency with which the hazard led to the significant events to be analysed. For example, assuming the two accidents in 2008 shown in Table 1 are all the accident reports to be analysed, the frequency of the identified hazard *damage to track components* (denoted as H02) is estimated as 2 events/year because it's recorded a total of two times in 2008. Similarly, the frequency of the hazard *misjudgment of the current hazardous situation* is estimated as 1 event/year. Similarly, the average consequence of each type of accident can be statistically obtained. For example, in the case of the above assumption, the average consequence

Table 1
Two examples of accidents from the RAIB [64] and knowledge entities identified from them.

Report No.	Accident location	Time	Description of accident	Knowledge entities identified from each report
R072009	1.7 km south of Birmingham Snow Hill station on the Didcot and Chester line	25-03-2008	At 06:37 on 25 March 2008, due to the *misjudgment of track voids* by track inspectors using visual detection of static geometry, a *track twist* located 1.7 km south of Birmingham Snow Hill station on the route known as the Didcot and Chester line was *not detected* or corrected. As a result, four wagons of the freight train 6M15 *derailed* to the left and turned onto their sides. There were *no injuries*.	Misjudgment of the current hazardous situation (H01) with H-type; Count (H01) =1; Damage to track components (H02) with EI-type; Count(H02) =1; Derailment accident (A01); Consequence (A01) =0 FWI
R182009	Between Rhiw Goch and Tan-y-Bwlch and 5 miles (8 km) from Porthmadog Harbour station	03-05-2008	At 15:28 on 3 May 2008, due to the *insufficient maintenance management* by the Ffestiniog railway company, the *spreading of track gauge* located between Rhiw Goch and Tan-y-Bwlch and are 5 miles (8 km) from Porthmadog Harbour station *derailed* a passenger train. The rear two vehicles of the passenger train derailed. None of the passengers were injured, but one of the members of the train crew sustained a *minor injury* in the derailment.	Imperfect management of maintenance practices (H03) with M-type; Count (H03) =1; Damage to track components (H02) with EI-type; Count (H02) =1; Derailment accident (A01); Consequence (A01) =0.008 FWI

Table 2
Fatality and weighted injury [65].

Injury degree	Weighting
Class 1 Minor injury / multiple Class 2 injuries	0.008
Multiple Class 1 minor injuries / more severe injury	0.04
1-2 Major injuries	0.2
Multiple major / single fatality	1
Multiple fatalities	5

of derailment accident (denoted as A01) is 0.04 FWI/event that is the average of 0 FWI and 0.008 FWI shown in Table 1.

2.1.2. Step2: Identification of the links between knowledge entities

The identification of links is the step of determining the relationships between knowledge entities. This is also the primary step for constructing the ROAKG. The relationships here include the cause-effect relationships among hazards and accidents, the association relationships between hazards and hazard types, between hazards and occurrence frequencies, and between accidents and accident consequences.

In knowledge graphs, the knowledge entities and their links are usually depicted by knowledge triples [59, 70]. A knowledge triple is denoted as $<h, r, t>$, where h, t represent two entities, and r denotes the relationship between them, i.e., the link between them. In the present paper, the links between knowledge entities of the ROAKG are denoted by three kinds of keywords, i.e., `Cause-Effect', `TypeIs', and `hasValueof'.

- The keyword `Cause-Effect' denotes the cause-effect links between entities. For example, a pair of hazards (H01 and H02 in Table 1) form a knowledge triple <H01, Cause-Effect, H02> if and only if H01 is a direct causal factor leading to H02. Similarly, the hazard H02 and the derailment accident A01 in Table 1 can be denoted as <H02, Cause-Effect, A01>. It is worth noting that the existing publicly available railway accident data comes from different railway lines equipped with different infrastructures, vehicles, operational ways, etc. Any pair of causal hazards identified from such accident data may occur in different railway lines. Due to the different technical conditions and operational ways in different lines, the non-binary relationship between any pair of causal hazards may change with different railway line backgrounds. As a result, it is difficult to capture the meaningful non-binary relationships among the hazards by expert experience or by the statistics of the existing railway accident data. As an alternative, the binary system is utilized to describe the cause-effect relationships among factors in the paper.
- The keyword `TypeIs' is used to denote the association relationships between hazards and hazard types. For example, the type of the hazard H01 is H-type. As a result, this can be denoted as <H01, TypeIs, H-type>.
- The association relationships between hazards and occurrence frequencies are denoted by the keyword `hasValueof' that can also represent the specific frequency values. For example, the occurrence frequency of the hazard H01 is assumed to be 1 event/year. This can be denoted as <H01, hasValueof 1, Frequency>. Also, the keyword `hasValueof' is used to denote the association relationships between accidents and accident consequences. For example, the derailment accident and its average consequence in Table 1 can be denoted as <A01, hasValueof 0.004, Consequence>.

Taking the identified knowledge entities in Table 1 for example, the knowledge triples of them are shown in Table 3. The knowledge triples with the keywords `Cause-Effect' and `TypeIs' can be identified from each report, and the other triples are identified based on the statistics of all the accident reports, i.e. the triples related to hazard frequencies and accident consequences.

2.1.3. Step3: Construction of the ROAKG

After the links between knowledge entities has been identified from all the accidents to be analysed, the third and final step is to map the knowledge entities and their links into a network graph, i.e., the ROAKG. Due to the different types of links, we define three matrices to construct the ROAKG.

The first matrix is the Causality Adjacency Matrix (CAM) that is a square matrix with the entry CAM_{ij} defined by Eq. (1).

Table 3
Knowledge triples from the two accidents in Table 1.

Report No.	Knowledge entities identified from each report	Knowledge triples from each report	Knowledge triples based on statistics of the two reports
R072009	Misjudgment of the current hazardous situation (H01) with H-type; Count (H01) =1; Damage to track components (H02) with EI-type; Count(H02) =1; Derailment accident (A01);Consequence (A01) =0 FWI	<H01, Cause-Effect, H02>, <H02, Cause-Effect, A01>, <H01, TypeIs, H-type>, <H02, TypeIs, EI-type>	<H01, hasValueof 1, Frequency[a]>, <H02, hasValueof 2, Frequency>, <H03, hasValueof 1, Frequency>, <A01, hasValueof 0.004, Consequence[b]>
R182009	Imperfect management of maintenance practices (H03) with M-type; Count (H03) =1; Damage to track components (H02) with EI-type; Count (H02) =1; Derailment accident (A01); Consequence (A01) =0.008 FWI	<H03, Cause-Effect, H02>, <H02, Cause-Effect, A01>, <H03, TypeIs, H-type>, <H02, TypeIs, EI-type>	

[a] The unit of frequency here is event/year.
[b] The unit of accident consequence here is FWI/event.

$$CAM_{ij} = \begin{cases} 1 & < i,\ Cause-Effect, j > \in KTs \\ 0 & < i,\ Cause-Effect, j > \notin KTs \end{cases} \quad (1)$$

where i represents a hazard, j represents a hazard or an accident, and KTs represents all knowledge triples identified in Step 2. This means that the CAM is determined by the knowledge triples with the keyword `Cause-Effect'. By the CAM, the knowledge triples depicting cause-effect relationships can be mapped into a graph, i.e., there is a cause-effect relationship edge from the knowledge entity i to the knowledge entity j if the entry CAM_{ij}=1. Here, knowledge entities are abstracted as nodes of a network. It is worth noting that the knowledge entities connected by cause-effect links contains hazards and accidents. Due to the unidirectional cause-effect relationships between them, it is difficult to use the undirected edges to depict the cause-effect relationships. Hence, the cause-effect relationship edges are depicted by the directed edges in this work. Besides, in some research [52,55], the weight value of an edge is allocated by the occurrence number of the edge to represent the strength of cause-effect relationship. It is difficult to use such values to evaluate the different strengths of the cause-effect relationships. This is because the values are derived from the railway accident data coming from different railway line backgrounds. Therefore, as an alternative without loss of generality, the unweighted edges are employed in this paper. To distinguish the different relationships, the cause-effect edge is labeled with the keyword `Cause-Effect'. Taking the knowledge triples in Table 3 for example, Fig. 1 shows the CAM and its corresponding graph.

The second matrix is the Type Matrix (TM) that is a matrix with the entry TM_{ij} defined by Eq. (2).

$$TM_{ij} = \begin{cases} 1 & < i,\ TypeIs, j > \in KTs \\ 0 & < i,\ TypeIs, j > \notin KTs \end{cases} \quad (2)$$

where i represents a hazard, j represents the hazard types, and KTs represents all knowledge triples identified in Step 2. The TM is determined by the knowledge triples with the keyword `TypeIs'. By the TM, knowledge triples depicting hazards and hazard types can be mapped into a graph, i.e., there is an edge labeled with `TypeIs' from the entity i to the entity j if the entry TM_{ij}=1. Similarly, taking the knowledge triples in Table 3 for example, Fig. 2 gives the TM and adds its information into the knowledge graph shown in Fig. 1.

The third matrix is the Value Matrix (VM) that is a matrix with the entry VM_{ij} defined by Eq. (3).

$$VM_{ij} = \begin{cases} a & < i,\ hasValueof a, j > \in KTs \\ 0 & < i,\ hasValueof a, j > \notin KTs \end{cases} \quad (3)$$

where i represents a hazard or an accident, j represents frequency or consequence, a refers to the specific value of frequency or consequence shown in the corresponding knowledge triple, and KTs represents all knowledge triples identified in Step 2. This means that the VM is determined by the knowledge triples with the keyword `hasValueof'. By the VM, knowledge triples depicting the value of frequency or consequence can be mapped into a graph, i.e., there is an edge labeled with `hasValueof' from the entity i to the entity j if the value of the entry VM_{ij} is greater than 0. Taking the knowledge triples in Table 3 for example, Fig. 3 shows the VM and adds its information into the knowledge graph shown in Fig. 2 to form the final ROAKG of the examples.

With the aid of the above three matrices, the ROAKG can be constructed to provide a multi-dimensional network perspective for understanding railway operational accidents. To get further insights into the structural features of the ROAKG, two more matrices are introduced, which can be used in the subsequent analysis. One is the Causality Shortest Path Matrix (CSPM) that is a square matrix with the entry $CSPM_{ij}$ defined by Eq. (4).

$$CSPM_{ij} = \sum_{p,q \in N} CAM_{pq} \quad (4)$$

where $CSPM_{ij}$ represents the length of the shortest path from a hazard i to a hazard or an accident j, p and q represent two entities, N represents all entities on the shortest path, CAM_{pq} is obtained by Eq. (1). The CSPM can capture the length features of causal paths of the ROAKG. An example of a CSPM derived from the CAM shown in Fig. 1 is shown in Fig. 4(a). Here, the $CSPM_{ij}$ can be obtained by using of the Ahn and Ramakrishna's shortest path algorithm [71]. The other matrix is the Causality Reachability Matrix (CRM) that is a square matrix with the entry CRM_{ij} defined by Eq. (5).

$$CRM_{ij} = \begin{cases} 1 & CSPM_{ij} > 0 \\ 0 & CSPM_{ij} = 0 \end{cases} \quad (5)$$

a

	H01	H02	H03	A01
H01	0	1	0	0
H02	0	0	0	1
H03	0	1	0	0
A01	0	0	0	0

b

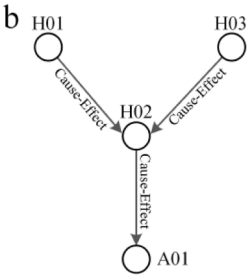

Fig. 1. Illustration of the CAM and its corresponding graph, (a) the CAM of the knowledge triples in Table 3, (b) the knowledge graph derived from the CAM.

a

	H-type	EI-type
H01	1	0
H02	0	1
H03	1	0

b

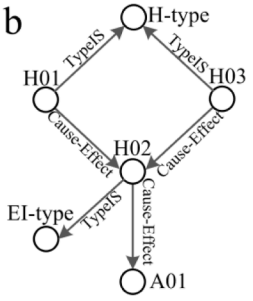

Fig. 2. Illustration of the TM, (a) the TM of the knowledge triples in Table 3, (b) the extended knowledge graph with the TM.

a

	Frequency	Consequence
H01	1	0
H02	2	0
H03	1	0
A01	0	0.004

b

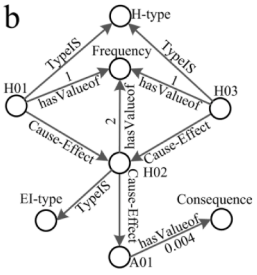

Fig. 3. Illustration of the VM, (a) the VM of the knowledge triples in Table 3, (b) the final ROAKG of the two examples shown in Table 1.

a

	H01	H02	H03	A01
H01	0	1	0	2
H02	0	0	0	1
H03	0	1	0	2
A01	0	0	0	0

b

	H01	H02	H03	A01
H01	0	1	0	1
H02	0	0	0	1
H03	0	1	0	1
A01	0	0	0	0

Fig. 4. Illustration of the CSPM and the CRM, (a) CSPM and (b) CRM of the knowledge triples in Table 3.

where CRM_{ij} indicates whether there is causal path from a hazard i to a hazard or an accident j, and $CSPM_{ij}$ is defined by Eq. (4). The CRM can capture the causal reachability features of the ROAKG. Similarly, an example of a CRM derived from the CAM shown in Fig. 1 is shown in Fig. 4(b).

2.2. Topological analysis of the ROAKG

The ROAKG is a multi-dimensional network of railway operational accidents, in which a variety of relationships between various knowledge entities are depicted. To explore the latent features of past accidents, network topological analysis can be performed on the ROAKG by the use of topological indicators.

However, most of commonly used topological indicators [48] are derived from single-dimensional complex networks that consist of the same type of nodes and the same type of relationships. It is difficult to use the traditional topological indicators to reveal the inherent features of the multi-dimensional ROAKG. Although some traditional topological indicators have been applied to the topological analyses of knowledge graphs [72-75], they are only available for the partial single-dimensional structures in knowledge graphs rather than the whole multi-dimensional structures. Therefore, it is important to tailor some topological indicators that adapt to the structural characteristics of the multi-dimensional ROAKG to perform network topological analysis. The tailored topological indicators can help discovering the latent rules of accidents from the perspective of multi-dimensional network structure, thereby assisting in formulating more targeted safety strategies.

For the definition of new topological indicators, let N represent the nodes of the ROAKG. N={HN, AN, TN, VN}, where HN represents the hazard nodes, AN represents the accident nodes, TN represents the hazard type nodes, and VN represents the nodes related to values, i.e., the frequency nodes and the consequence nodes.

The causal closeness of hazards provides an indication of the causal connection degree between hazards. Two indicators related to causal closeness are studied, namely active causal closeness and passive causal closeness. The active causal closeness indicates how difficult it is for a given hazard to cause other hazards and can be calculated by Eq. (6). The active causal closeness of a given hazard h is the reciprocal of the average shortest path length between h and the hazards that it can cause directly and indirectly. This means that the higher the value of the active causal closeness of a hazard is, the more easily the hazard can cause other hazards. Similarly, the passive causal closeness indicates how difficult it is for a given hazard to be caused by other hazards and can be calculated by Eq. (7). It is the reciprocal of the average shortest path length between a given hazard and the hazards that can cause the given hazard directly and indirectly. This means that the higher the value of the passive causal closeness of a hazard is, the more easily the hazard can be caused by other hazards.

$$C_h^A = 1 \Big/ \left(\sum_{i \in HN} CSPM_{hi} \Big/ \sum_{i \in HN} CRM_{hi} \right) \tag{6}$$

$$C_h^P = 1 \Big/ \left(\sum_{i \in HN} CSPM_{ih} \Big/ \sum_{i \in HN} CRM_{ih} \right) \tag{7}$$

In order to get further insights into the relationships between a given hazard and other hazard types, a direct reachability indicator and a direct source indicator are studied. The direct reachability indicator P_{hT}^R denotes the percentage of the hazards with the type T in all hazards that a given hazard h can cause directly. It can be calculated by Eq. (8) and used for revealing the proportion of each type of hazards that a given hazard can cause or impact directly. Similarly, the direct source indicator P_{Th}^S denotes the percentage of the hazards with the type T in all hazards that can cause a given hazard h directly. It can be calculated by Eq. (9) and used to explore the proportion of each type of hazards which a given hazard can directly originate from.

$$P_{hT}^R = \sum_{j \in HN} \left(CAM_{hj} \cdot TM_{jT} \right) \Big/ \sum_{i \in HN} CAM_{hi} \tag{8}$$

$$P_{Th}^S = \sum_{j \in HN} \left(CAM_{jh} \cdot TM_{jT} \right) \Big/ \sum_{i \in HN} CAM_{ih} \tag{9}$$

As mentioned, hazards can be categorised into different types. The exploration of the connectivity between different hazard types can help revealing the causality between hazard types. The number of direct edges between two types of hazards is the local connectivity of the two hazard types and can be calculated by Eq. (10). As shown by Eq. (10), it indicates the local direct connectivity between the type E and the type F. The hazard type E or F here represents any one of the four hazard types, namely human (H)-type, equipment and infrastructure (EI)-type, the environment (E)-type and management and organisation (M)-type. Furthermore, in the sense of global connections, the number of paths connecting two types of hazards reflects the global connectivity of the two hazard types and can be calculated by Eq. (11). The indicators of connectivity between hazard types, i.e., the local connectivity indicator and the global connectivity indicator, can help us reveal the causal relationships between hazard types from both local and global perspectives.

$$L_{EF}^C = \sum_{i,j \in HN} \left(CAM_{ij} \cdot TM_{iE} \cdot TM_{jF} \right) \tag{10}$$

$$G_{EF}^C = \sum_{i,j \in HN} \left(CRM_{ij} \cdot TM_{iE} \cdot TM_{jF} \right) \tag{11}$$

For the purpose of formulating prevention strategies, the risk of harm related to each hazard is also included into the topological analysis. It can provide decision-making basis for the investment of accident prevention efforts. The risk of harm here refers to the safety risk, i.e., the estimation of harm to staff, passengers or public members involved in railway operational accidents. The safety risk arising from a hazard h can be calculated by Eq. (12).

$$Risk_h = VM_{hFrequency} \cdot \left(\sum_{Ai \in AN} CRM_{hAi} \cdot P_{Ai \mid h} \cdot VM_{AiConsequence} \right) \tag{12}$$

where $VM_{hFrequency}$ is an entity of the Value Matrix and represents the occurrence frequency of h; CRM_{hAi} is an entity of the Causality Reachability Matrix and has a value of 1 if and only if the h can cause the accident Ai; $P_{Ai|h}$ represents the probability that the hazard h results in the accident Ai and can be estimated by the ratio of the number of h resulting in the accident Ai to the number of h in all accident reports; $VM_{AiConsequence}$ is also an entity of the Value Matrix and represents the average consequence of the accident Ai.

The $Risk_h$ in Eq. (12) can be used to estimate the harm arising from the occurrence of a hazard. It is an estimation of harm from the perspective of each individual hazard. However, from the perspective of the whole knowledge graph, if a hazard h is controlled or eliminated, the eventually reduced risk might be larger than $Risk_h$. This is because some causal paths from other hazards to accidents will be cut off if the hazard h plays the role of intermediary in these paths, i.e., the harm derived from other hazards is also avoided by the elimination of the hazard h. In order to capture the risk related to the intermediary role of a hazard h, an intermediary risk is studied and can be calculated by Eq. (13).

$$Risk_h^B = \sum_{\substack{i \in HN, \\ Ai \in AN}} (CRM_{ih} \cdot CRM_{hAi} \cdot Risk_i) \tag{13}$$

where CRM_{ih} and CRM_{hAi} are entities of the Causality Reachability Matrix, and the product of them is equal to 1 if and only if the hazard h plays the role of intermediary in the path from the hazard i to the accident Ai; $Risk_i$ is the risk arising from the hazard i and can be calculated by Eq. (12).

In order to get further insights into the intermediary risk, the cumulative intermediary risk distribution $P(r)$ of the ROAKG can be used. As defined in Eq. (14), it is the probability of a randomly chosen hazard with the intermediary risk value greater than or equal to r. The cumulative intermediary risk distribution can help revealing the inherent intermediary risk features of the ROAKG. This contributes to formulating accident prevention strategies.

$$P(r) = P\left(Risk_h^B \geq r\right) = \sum_{r}^{\infty} (N(r) / N_{HN}) \tag{14}$$

3. Case study

3.1. Data collection

The railway operational accident data in the UK from 2005 to 2015 is chosen for this study. The data can be obtained from the accident investigation reports published on the website of the Rail Accident Investigation Branch (RAIB) of the UK (https://www.gov.uk/raib-reports). In addition to railway operational accidents, both metro accidents and tram accidents are described and reviewed in the investigation reports, but not considered in our study. As a result, 214 railway operational accidents in the period of 2005-2015 are collected from the accident investigation reports numbered R022006 to R182016. The collected accidents can be divided into seven types, i.e., derailment, near miss, collision between two trains, struck-by-object, train runaway, overrun the stopping point, and electric shock. Here, the near miss events published by the RAIB are some events that are regarded as being significant for the entire railway system by the RAIB. Hence, the published near miss events are also collected into our study. Besides, in the data collection, if two or more types of accidents appear in one reported event, e.g., a collision accident is caused by a train runaway accident, the accident that directly lead to the loss, i.e., injures, fatalities or financial damage, is chosen as the final accident in the reported event. The other accidents are taken as the causes of the final accident. The type of the final accident is taken as the accident type reported by the event. Appendix A shows the raw data with the assignment of individual events to the adopted categories. As shown by Appendix A, two types of accidents that appear in one report are marked with a causal arrow. For example, in the Report102011, a runaway accident (A04) caused a collision accident (A03) that directly led to the loss. They are denoted as `A04->A03′, in which A03 is the final accident and its cause is A04. Fig. 5 shows the statistics of the 214 railway operational accidents. It can be found that the derailment accidents take up the highest proportion of total, which occurred 73 times and account for 34%, followed by near miss (26%), collision (17%), struck-by-object (17%), and the rest of the accidents (6%).

It should be noted that the publicly available railway operational accident data from the RAIB are about serious accidents and some events that are regarded as being significant for the entire railway system by the RAIB. As for the most events that are analysed within the framework of

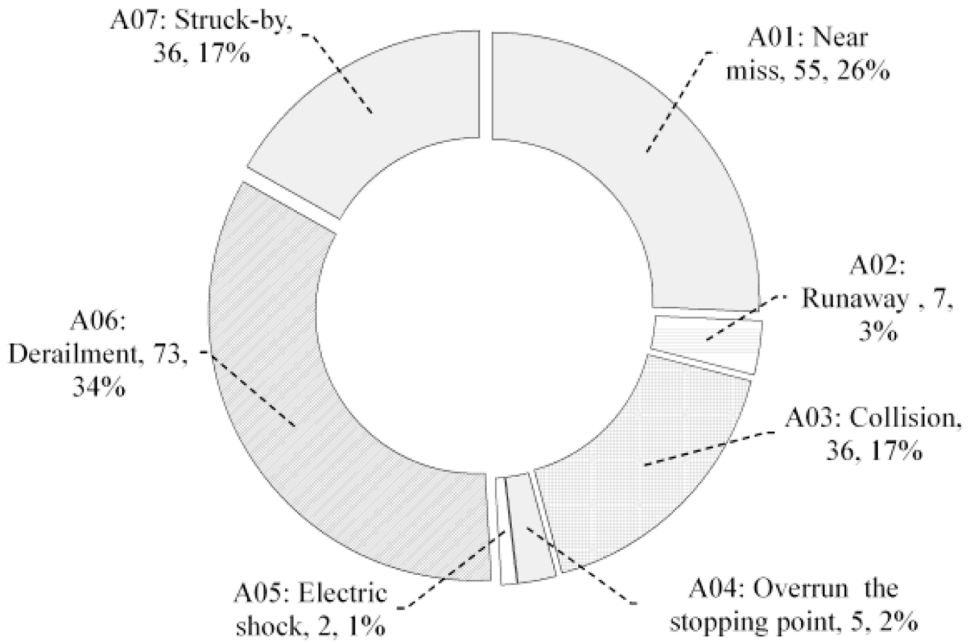

Fig. 5. Statistics of the collected railway operational accidents [64].

the safety management systems of infrastructure managers and railway undertakings, the investigation results of them are not publicly available. In this context, the publicly available railway operational accident data are taken as study case to show how to perform the proposed approach. On this basis, the latent knowledge analysed from such study case is just the knowledge of the publicly available events, not the knowledge of all operational accidents or the entire railway system. Although the publicly available events are rare and unusual, they are significant for the entire railway system. The exploration of the knowledge of them is helpful for preventing these events.

3.2. ROAKG modelling

The knowledge entities are firstly identified from the collected 214 railway operational accidents by Step 1 introduced in Section 2.1. As shown in Section 2.1, the knowledge entities include accident-related knowledge entities (i.e., railway operational accidents and their consequences) and hazard-related knowledge entities (i.e., hazards, hazards' types and hazards' frequencies). Through the identification of knowledge entities, the average consequences of the seven types of accidents are obtained and shown in Table 4. Moreover, 31 hazards, the types of the hazards and the occurrence frequencies of the hazards are obtained and listed in Table 5. As shown by Step 1, the consequences of accidents and the frequencies of hazards are statistically obtained on the basis of the 214 railway operational accidents in the period of 2005-2015. Here, the frequency of a hazard refers to the frequency with which the hazard led to the significant events to be analysed.

The links between the knowledge entities are then identified by Step 2 in Section 2.1. The identified links are depicted in the form of knowledge triples. Due to limited space, partial knowledge triples are shown in Table 6. As shown in Table 6, the knowledge triples are identified by different ways. The knowledge triples with the keywords `Cause-Effect' and `TypeIs' are identified according to the content of each accident report, and the other triples originate from the statistics of all the accident reports.

Based on the identified knowledge entities and their links, the railway operational accident knowledge graph (ROAKG) is constructed by Step 3 in Section 2.1, as shown in Fig. 6. The ROAKG consists of 44 nodes and 203 edges, including 134 edges denoting cause-effect relationships, 38 edges denoting hazard types, and 31 edges denoting the values of hazard frequencies or accident consequences. The nodes of the ROAKG include 7 accident nodes numbered from A01 to A07, 31 hazard nodes numbered from H01 to H31, and four type nodes numbered H-Type (human-type), EI-Type (equipment and infrastructure-type), E-Type (the environment-type) and M-Type (management and organization-type), respectively.

3.3. Topological analysis results

The active causal closeness and passive causal closeness of hazards are calculated by Eqs. (6) and (7). They are displayed in the form of matrix diagram in Fig. 7. As shown in Fig. 7, some hazards with passive causal closeness of zero, but that have high values of active causal closeness may play the roles of source hazards, such as H20 (Snow lying on railway tracks or braking equipment) and H28 (Inadequate management of railway staff competence). In other words, it is easier for them to cause other hazards than to be caused. In contrast, some hazards with active causal closeness of zero, but that have high values of passive causal closeness, are treated as accumulation hazards, such as H16 (Poor wheel and rail adhesion force) and H24 (Fallen trees or branches lying on tracks or hitting the rolling stock directly). The hazards in the middle of Fig. 7 play the roles of transition hazards. Some of them are closer to the accumulation hazards due to their higher values of passive causal closeness, such as H25 (Fallen rocks lying on tracks or hitting the rolling

Table 4
Accident-related knowledge entities.

Accident and its number	Average consequence of each type of accident (FWI/event)
Near miss (A01)	0
Runaway (A02)	0.007
Collision (A03)	1.947
Overrun the stopping point (A04)	0.096
Electric shock (A05)	0.600
Derailment (A06)	0.283
Struck-by-object (A07)	0.189

Table 5
Hazard-related knowledge entities.

Hazard and its number	Hazard types	Frequency of each hazard (event/year)
Inadvertent operational behavior of a railway staff (H01)	H-type	13
Missing or belated deceleration applied by a train driver (H02)	H-type	26
Distracted or fatigue driving (H03)	H-type	10
Missing or belated detection of problems (H04)	H-type	35
Exposure to hazardous working environment (H05)	H-type	19
Inconsistency with operation standards or rules (H06)	H-type	21
Misjudgment of the current hazardous situation (H07)	H-type	17
Degradation of or damage to rolling stock including bogies and wheels (H08)	EI-type	36
Damage to track components including rail, sleepers and points (H09)	EI-type	45
Unbalanced train load or overload condition (H10)	EI-type	6
Settlement of or damage to track subgrade (H11)	EI-type	13
Subsiding of or damage to a bridge over or carrying railway tracks (H12)	EI-type	14
Excessive or interrupted power supply (H13)	EI-type	8
Failure of a signal system or equipment (H14)	EI-type	23
Insufficient or interrupted brake force of a braking system (H15)	EI-type	10
Poor wheel and rail adhesion force (H16)	EI-type	15
Objects (e.g., containers) carried by a train falling down (H17)	EI-type	4
Train movement without authority or in hazardous conditions (H18)	EI-type	7
Heavy rain that could bring damage to railway infrastructures (H19)	E-type	10
Snow lying on railway tracks or braking equipment (H20)	E-type	12
Strong wind affecting the stability of train components or trackside objects (H21)	E-type	10
Flood or ponding water destroying the stability of track subgrade (H22)	E-type	9
Freezing weather causing the ice on tracks, braking systems or other equipment (H23)	E-type	9
Fallen trees or branches lying on tracks or hitting the rolling stock directly (H24)	E-type	5
Fallen rocks lying on tracks or hitting the rolling stock directly (H25)	E-type	7
Concrete debris falling from artificial structures onto railway tracks (H26)	E-type	6
Insufficient safety training of railway staff (H27)	M-type	11
Inadequate management of railway staff competence (H28)	M-type	16
Imperfect emergency management or organisation (H29)	M-type	10
Insufficient or ineffective supervision on operational practices (H30)	M-type	6
Imperfect management of maintenance practices (H31)	M-type	6

Table 6
Partial knowledge triples identified from the accident reports.

Report No.	Knowledge triples from each report	Knowledge triples based on statistics of all the reports
R022006	<H06, Cause-Effect, A06>, <H06, TypeIs, H-type>	<H01, hasValueof 13, Frequency>, <H02, hasValueof 26, Frequency>, <H03, hasValueof 10, Frequency>,
R072006	<H03, Cause-Effect, H02>, <H02, Cause-Effect, A03>, <H03, TypeIs, H-type>, <H02, TypeIs, H-type>	<H04, hasValueof 35, Frequency>, <H05, hasValueof 19, Frequency>, <H06, hasValueof 21, Frequency>, <H07, hasValueof 17, Frequency>, <H08, hasValueof 36, Frequency>, ...
R082006	<H04, Cause-Effect, H08>, <H08, Cause-Effect, A06>, <H04, TypeIs, H-type>, <H08, TypeIs, EI-type>	<A07, hasValueof 0.189, Consequence>
...	...	
R182016	<H30, Cause-Effect, H03>, <H03, Cause-Effect, A04>, <H30, TypeIs, M-type>, <H03, TypeIs, H-type>	

stock directly) and H22 (Flood or ponding water destroying the stability of track subgrade); some others are closer to the source hazards, for example, H29 (Imperfect emergency management or organisation) and H07 (Misjudgment of the current hazardous situation); other hazards near the middle line, like H14 (Failure of a signal system or equipment) and H18 (Train movement without authority or in hazardous conditions), have approximately equal values of active causal closeness and passive causal closeness respectively. This categorisation of hazards on the basis of the closeness to other hazards is useful for formulating targeted hazard control strategies.

The direct reachability indicator and the direct source indicator are calculated by Eqs. (8) and (9), and shown in Fig. 8 and Fi.9 respectively. Fig. 8 shows the proportions of different types of hazards that can be directly caused by a given hazard, i.e., the different types of the immediate successors of a given hazard. From Fig. 8, it can be seen that some hazards can only cause a single type of hazards, such as H02 (Missing or belated deceleration applied by a train driver) and H31 (Imperfect management of maintenance practices), and some others have no successors, such as H05 and H16, which is consistent with the active causal closeness of zero in Fig. 7. Fig. 9 shows the proportions of different types of hazards that can directly cause a given hazard, i.e., the different types of the immediate predecessors of a given hazard. In Fig. 9, similarly, some hazards have only one type of predecessors and, some others have no predecessors, such as H20 and H28 which act as source hazards in Fig. 7. The exploration of the proportion of immediate successor/predecessor types of a hazard can provide further insights into the relationships between the hazard and other hazards. This can assist in developing specific safety strategies for each hazard, by being combined with the categorisation of hazards based on closeness in Fig. 7.

Fig. 10 shows the number of both direct and indirect connections between different types of hazards, introduced in Eqs. (10) and (11). Furthermore, the numbers of connections between hazards and accidents are also calculated by Eqs. (10) and (11), and shown in Fig. 10. In Fig. 10(a), it can be seen that most of direct connections between hazards are related to hazards with H-type and EI-type. Meanwhile, most of direct connections between hazards and accidents are related to the two types of hazards, indicated by the values of 29 and 20 in Fig. 10(a). This means that most of hazards evolve into accidents by triggering the two types of hazards. Controlling of the two types of hazards can block the causal paths to accidents. Besides, although hazards with E-type and M-type have low direct correlations with accidents, respectively given by the values of 7 and 1 in Fig. 10(a), there is a significant increase in indirect connections, indicated by the values of 109 and 71 in Fig. 10(b). This indicates that the two types of hazards play the role of underlying

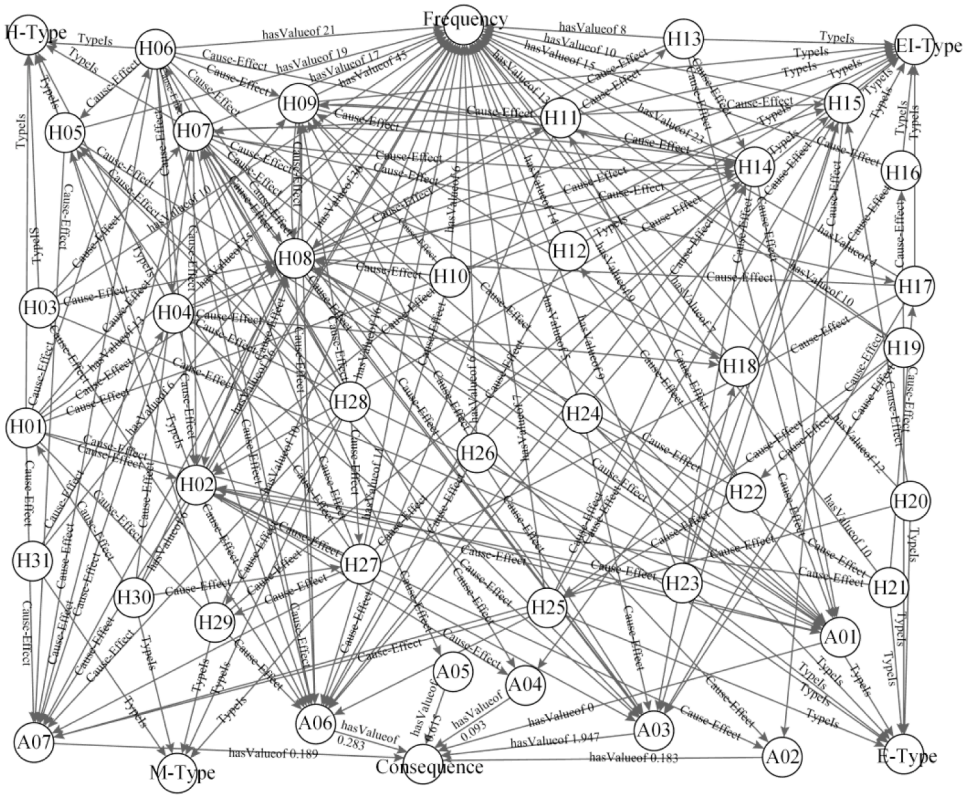

Fig. 6. The ROAKG derived from the 214 accidents in our study.

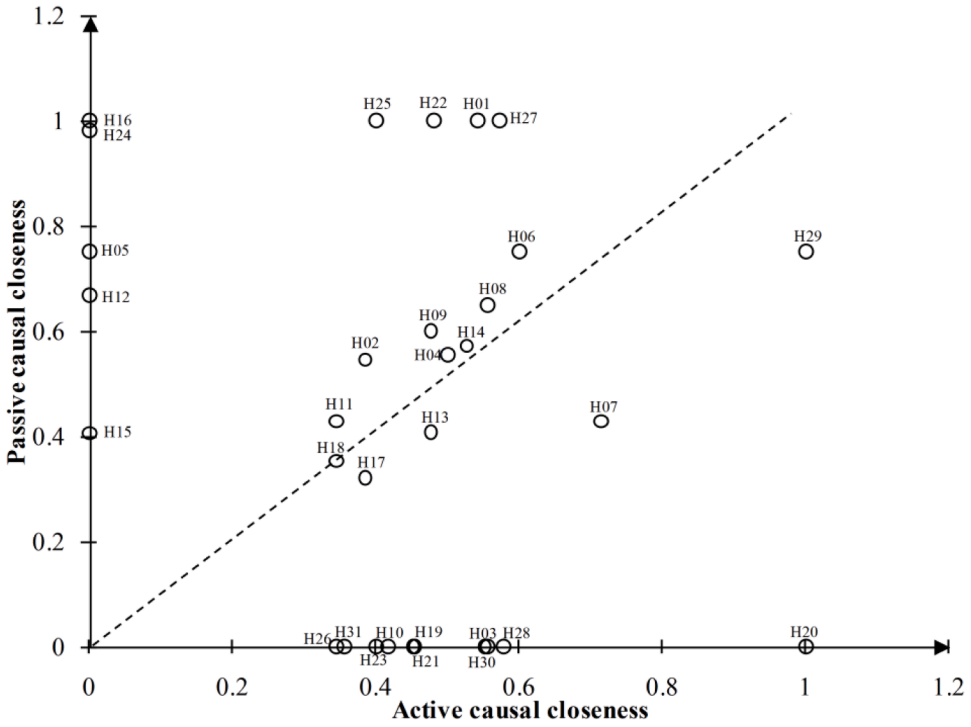

Fig. 7. Causal closeness of hazards.

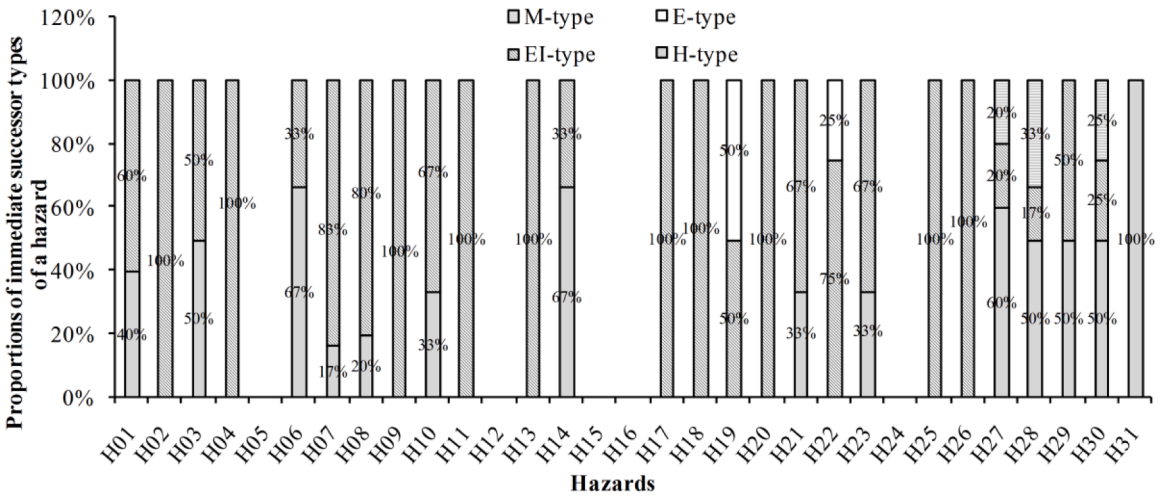

Fig. 8. Proportions of immediate successor types of a hazard.

factors resulting in accidents, and hence deserve more attention. The exploration of connectivity between different types of hazards contributes revealing the nature of railway operational accidents from an overall perspective.

Fig. 11 displays both the safety risk and the intermediary risk related to hazards, which are calculated by Eqs. (12) and (13). From Fig. 11, some hazards with high safety risk values can be found, such as H04 (Missing or belated detection of problems), H07 (Misjudgment of the current hazardous situation), H09 (Damage to track components including rail, sleepers and points) and H16 (Poor wheel and rail adhesion force). From the perspective of each individual hazard, each of these hazards can bring high harm to people involved in railway operational accidents. In Fig. 11, there are eleven hazards with intermediary risk value equal to zero. This is because these hazards do not play the intermediary role in the causal paths to accidents. From Fig. 11, the hazards with high intermediary risk values can be identified, such as H02 (Missing or belated deceleration applied by a train driver), H07, H08 (Degradation of or damage to rolling stock including bogies and wheels) and H14 (Failure of a signal system or equipment). If these hazards are eliminated or controlled, much risk will be brought under control due to the hazards' roles of intermediary in the causal paths to accidents, indicated by the values of 21.86 FWI/year, 20.63 FWI/year, and 20.55 FWI/year, respectively.

3.4. Comparison of the network topology-based analyses

In the above application of the proposed approach, different railway accident factors, i.e., the above different knowledge entities, are depicted in a network model and analysed, by means of both the heterogeneous knowledge graph network and the proposed indicators adapting to the heterogeneous network. It is difficult for the existing railway accident topological analysis approaches [49-56] to model and

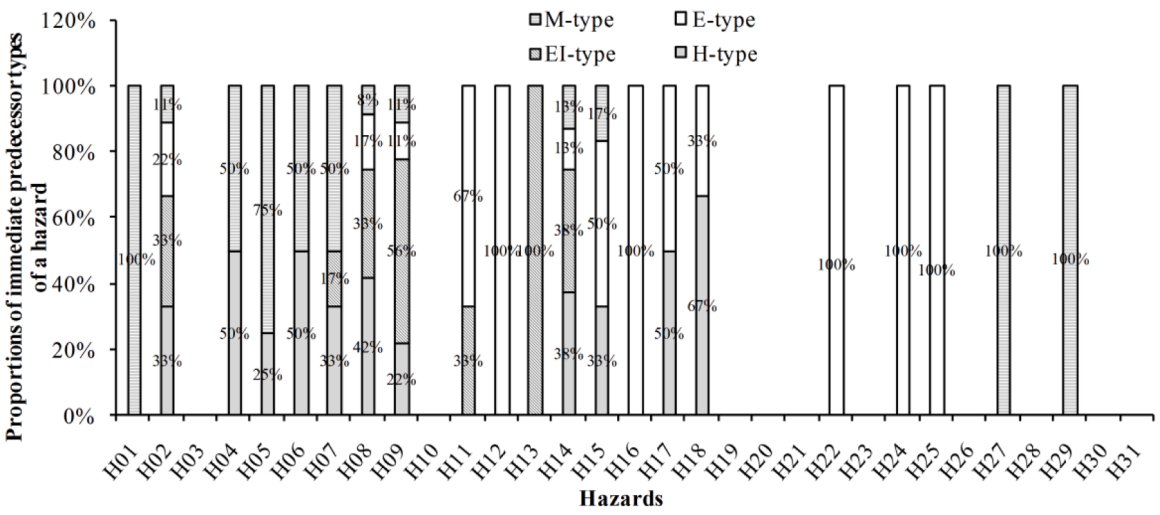

Fig. 9. Proportions of immediate predecessor types of a hazard.

a

	H-type	EI-type	E-type	M-type	Accidents
H-type	8	15	0	0	29
EI-type	4	14	0	0	20
E-type	2	13	3	0	7
M-type	10	4	0	4	1

b

	H-type	EI-type	E-type	M-type	Accidents
H-type	17	48	0	0	67
EI-type	16	64	0	0	131
E-type	12	52	3	0	109
M-type	18	33	0	5	71

Fig. 10. Connections between different hazard types, (a) direct connections, (b) indirect connections.

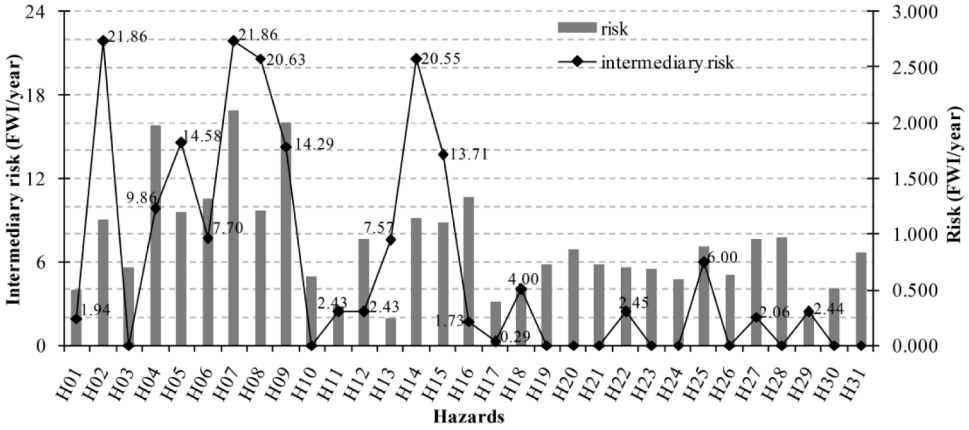

Fig. 11. Safety risk and intermediary risk.

analyse the above various accident factors because of both the used homogeneous networks and their corresponding topological indicators. Although both are network topology-based analyses on railway accidents, there are different emphases in this work and the existing railway accident topological analysis approaches. Table 7 shows the different features of them. As shown in Table 7, the existing analyses focus on exploring latent rules from accident causation factors and their causal relationships by use of homogeneous networks and their topological indicators. In this paper, the knowledge graph-based analysis focuses on revealing the hidden information from various accident factors, by means of heterogeneous knowledge graph network and the tailored indicators. The work of this paper serves as a complement to the existing railway accident topological analysis approaches.

It is worth noting that the reference [47] proposed a heterogeneous railway accident causation network and several topological indicators adapting its features. In that work, the traditional homogeneous network was extended into a heterogeneous one. However, its heterogeneous network is a kind of heterogeneous nodes network that consists of different types of nodes (i.e., hazard nodes and accident nodes) and the same types of edges (i.e., cause-effect relationship edges). In other words, such a network is still a kind of accident causation network adopted in traditional research [49-56]. The topological analysis on it focuses on revealing the latent information in the cause-effect links among accident causation factors and accidents, such as network densities, path length information and betweenness information. Differing from its work, in this paper, both the nodes and the edges in the

Table 7
Features of the network topology-based analyses on railway accidents.

Topological analyses of railway accidents	Features of network model	Topological analysis indicator	Indicator category	Topological characteristic reflected by the indicator
Existing topological analyses of railway accidents [49-56]	Homogeneous structure consisting of the same type of nodes (i.e., causal factor nodes) and the same type of edges (i.e., causal relationship)	Node degree (input degree, output degree)	Local	The direct causal connection of an accident causal factor to other causal factors
		Degree distribution	Global	The distribution of causal factors' direct connections over the whole accident causation network
		Clustering coefficient	Local	The degree to which the neighboring factors of an accident causal factor are connected to each other
		Path length (shortest, average or longest path length)	Global	The number of edges connecting a pair of accident causal factors, i.e., the length of the causal path between them
		Network density	Global	The direct connectivity among all accident causal factors of the whole accident causation network
		Betweenness	Global	The degree to which a causal factor acts as an intermediary in the causal paths among others
Knowledge graph-based topological analysis in this paper	Heterogeneous structure consisting of various types of nodes and various types of edges	Causal closeness	Global	The role of a causal factor plays in the causation propagation, i.e., the role of a source factor, a transition factor or an accumulation factor
		Direct reachability /source indicator	Local	The proportions of different types of immediate predecessors/successors of a causal factor
		Local and global connectivity between causal factor types	Local and Global	The causal relationships between causal factor types from both local and global perspectives
		Intermediary risk	Global	The risk associated with the intermediary role of an accident causal factor

heterogeneous knowledge graph network are heterogeneous. The accident causation network is only one part of the heterogeneous knowledge graph network. The point of the topological analysis on such a heterogeneous network is the exploration of the hidden knowledge among various knowledge entities, by means of the topological characteristics shown in Table 7. It is an extension to the work in the reference [47].

4. Discussion

The results obtained in the case study show the usefulness of the knowledge graph-based approach for exploring railway operational accidents. The proposed topological indicators provide useful information for revealing the potential rules and hidden knowledge of the collected railway operational accidents, by adapting to the multi-dimensional structural characteristics of the ROAKG.

4.1. Comparison of the risks associated with hazards

Some hazards with high safety risk values are identified, the top four of which are the hazards H07, H09, H04 and H16, as shown in Fig. 11. They should be eliminated or controlled with more prevent efforts to reduce the harm arising from the occurrence of them. However, from the perspective of whole knowledge graph, the intermediary risk indicator (shown in Eq. (13)) provides additional decision-making information for the investment of accident prevention efforts, i.e., the intermediary risk values of hazards in Fig. 11. In this indicator, the harm related to the intermediary role of a hazard is considered. To explore the effects of the two risk results, we measure how the sum of safety risk values of hazards changes with the elimination or control of each hazard. Here, the elimination or control of each hazard is represented by removing the corresponding hazard nodes from the ROAKG. The lower the sum of safety risk values is after the removal of a hazard, the lower the potential harm to the people involved in railway operation is. Two removal strategies of ten hazard nodes are implemented, including the safety risk sequence and the intermediary risk sequence. The effects of removal strategies are depicted in Fig. 12. It can be seen that the removal strategy of intermediary risk sequence has better effects in reducing the potential harm. Therefore, it is important to give priority to investing more prevent efforts in eliminating or controlling the hazards with high intermediary risk values.

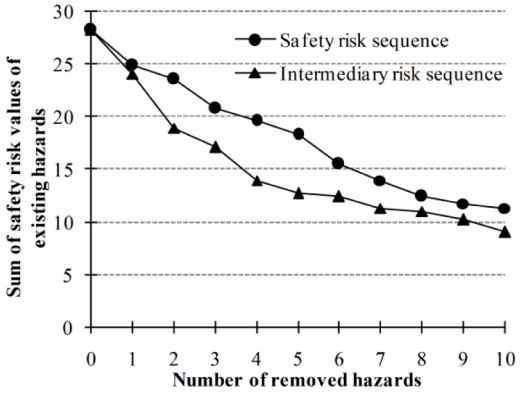

Fig. 12. Two strategies of removing hazard nodes.

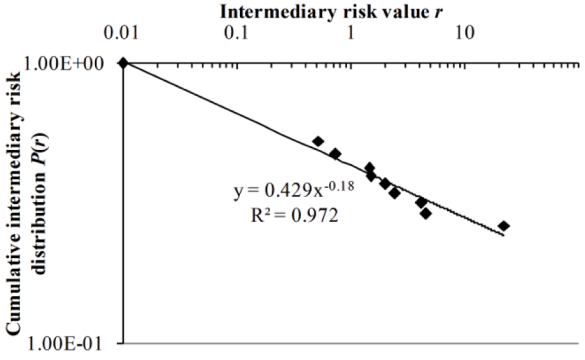

Fig. 13. Cumulative intermediary risk distribution.

The cumulative intermediary risk distribution $P(r)$ is obtained by Eq. (14) and depicted in Fig. 13. As shown in the double logarithmic coordinate system in Fig. 13, it can be seen that the cumulative intermediary risk distribution obeys a power-law distribution with the approximate fit $P(r) \sim 0.429*r^{-0.18}$ (R^2=0.972). This means that a few of hazards have high intermediary risk values while most of hazards have low values. For

instance, the top four of the hazards with high intermediary risk, i.e., the hazards H02, H07, H08 and H14 (12.9% of all 31 hazard nodes) in Fig. 11, account for 47.6 % of the sum of intermediary risk values. Furthermore, by Eq. (12), the safety risk reduced by removing the four hazard nodes is calculated to be 14.23 FWI/year, accounting for 50.4% of the initial sum of safety risk values, as shown in Fig. 12. This is consistent with the rule of power-law distribution in Fig. 13. If the four hazards are eliminated, most of the potential harm to the people involved in railway operation will be effectively reduced.

4.2. Use of the topological analysis results

The obtained results reveal different topological characteristics of hazards, as shown in Table 7. By use of the topological analysis results, the targeted prevention measures for key hazards can be formulated. A step-by-step procedure of using them is shown in Fig. 14. The first step is to identify the key hazards by the results of intermediary risk of hazards, as shown in Fig. 11. Then, according to the roles of key hazards in the hazard propagation, as shown in Fig. 7, general prevention strategies can be formulated. In this step, the role of a hazard refers to the role of a source hazard, a transition hazard or an accumulation hazard. For a source hazard, a strategy that prevents it from causing others can be formulated. Similarly, a strategy that prevents a hazard from being caused can be indentified for an accumulation hazard. The third and last step is to turn the general strategies into specific prevention measures by both the major immediate predecessors and the major immediate successors of hazards, as shown in Figs. 8 and 9.

As shown in Fig. 7, four key hazards with high intermediary risk values have been identified. Some targeted prevention measures for the four hazards, i.e., H02, H07, H08 and H14, can be formulated according to the procedure shown in Fig. 14. For example:

- In Fig. 7, it can be found that H02 (Missing or belated deceleration applied by a train driver) plays the role of a transition hazard but is closer to the accumulation hazards. This indicates that more resources should be allocated to blocking the causal paths to H02 or preventing the occurrence of the predecessors of H02. In Fig. 9, H02's immediate predecessors that account for high proportion include both H-type and EI-type hazards, both of which are 33%. As shown in the ROAKG in Fig. 6, these H-type and EI-type hazards includes H01, H06, H07 and H08. Hence, some targeted measures can be formulated, such as `further enhancing the competence of train drivers', and `introducing real-time detection and alarm systems for locomotive health statuses'. Besides, the immediate successor type of H02 only includes EI-type, indicated by 100% in Fig. 8. Some measures can be applied to blocking the causal paths from H02 to its immediate successor, i.e., the H18 that is the unique EI-type successor of H02 in Fig. 6, for example, `deploying the advanced train control systems'.
- It is shown in Fig. 7 that H07 (Misjudgment of the current hazardous situation) plays the role of a transition hazard but is closer to the source hazards. This means that more prevent efforts should be put in blocking the causal paths that begin at H07. From Fi.8, it can be seen that H07's immediate successors accounting for high proportion are EI-type hazards, indicated by 83%. Specifically, they are H08, H15 and H17, as shown in Fig. 6. To blocking the causal paths from H07 to them, some targeted measures can be implemented, such as `enhancing the supervision on operational practices', `introducing automatic strength and fatigue detection systems for bogies and wheels', and `equipping braking systems with performance testing devices'. As shown in Fig. 9 and Fig. 6, the main immediate predecessors of H07 are H-type hazards with the proportion of 50%, including H27 and H28. Some efforts should be also made to preventing H07 from being triggered, such as `strengthening the safety training of staff' and `improving the staff competence management'.
- As shown in Fig. 7, H08 (Degradation of or damage to rolling stock including bogies and wheels) plays the role of a transition hazard with approximately equal values of both active and passive causal closeness. This indicates that both blocking the causal paths to H08 and blocking the causal paths beginning at H08 should be paid more attention. From Fig. 8 and Fig. 6, it can be found that the main immediate successors of H08 are EI-type hazards with the proportion of 80%, including H09 and H11. Some targeted measures can be implemented to prevent H08 from causing these successors, such as `deploying rail track health status detection and alarm systems' and `introducing track subgrade settlement monitoring systems'. In Fig. 9, H08′s immediate predecessors accounting for high proportion include both H-type and EI-type hazards (42% and 38%, respectively). As shown in Fig. 6, they are H01, H04, H07 and H10. Some control measures can be formulated to block the causal paths to H08, such as `enhancing the competence of rolling stock maintainers', `strengthening the supervision on rolling stock inspection' and `deploying rolling stock load condition detection systems'.
- Similarly, in Fig. 7, H14 (Failure of a signal system or equipment) plays the role of a transition hazard with approximately equal values of both active and passive causal closeness. Prevention resources should be allocated to blocking both the causal paths to H14 and the causal paths beginning at H14. As shown in Fig. 8 and Fig. 6, the main immediate successors of H14 are EI-type hazards with the proportion of 67%, including H02 and H07. Some targeted measures can be implemented to prevent H14 from causing the successors, such as `improving the level of daily inspection and maintenance' and `enhancing the competence of dispatchers'. From Fig. 9 and Fig. 6, it can be seen that H14's immediate predecessors accounting for high proportion include H-type and EI-type hazards, both of which are 38%, including H01, H06 and H13. Some measures can be applied to blocking the causal paths to H14, such as `equipping signal systems with automatic checking devices', `enhancing the competence of signal workers' and `strengthening the supervision on signal equipments maintenance'.

As shown above, by the use of the topological analysis results, some key lessons and their corresponding recommended prevention measures have been identified and formulated, which are summarised in Table 8.

5. Conclusion

This paper presents a novel knowledge graph-based approach for exploring and understanding railway operational accidents. As the basis for the knowledge graph-based analysis, a modelling method is developed to construct the railway operational accident knowledge graph (ROAKG). To discover the hidden knowledge that contributes to accident prevention, some new topological indicators are proposed, adapting to the multi-dimensional structural features of the ROAKG. An

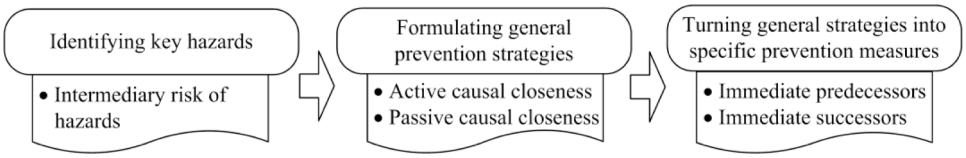

Fig. 14. Procedure of using the topological analysis results.

Table 8
Key lessons from the topological analysis results.

Key hazard	Role of the hazard	Major immediate predecessors (P) and successors (S)	Recommended prevention measures
H02 (Missing or belated deceleration applied by a train driver)	Transition hazard closer to accumulation hazards	P: H01, H06, H07 and H08 S: H18	Further enhancing the competence of train drivers. Introducing real-time detection and alarm systems for locomotive health statuses. Deploying the advanced train control systems.
H07 (Misjudgment of the current hazardous situation)	Transition hazard closer to source hazards	P: H27 and H28 S: H08, H15 and H17	Enhancing the supervision on operational practices. Introducing automatic strength and fatigue detection systems for bogies and wheels. Equipping braking systems with performance testing devices. Strengthening the safety training of staff.
H08 (Degradation of or damage to rolling stock including bogies and wheels)	Transition hazard	P: H01, H04, H07 and H10 S: H09 and H11	Deploying rail track health status detection and alarm systems. Introducing track subgrade settlement monitoring systems. Enhancing the competence of rolling stock maintainers. Strengthening the supervision on rolling stock inspection. Deploying rolling stock load condition detection systems.
H14 (Failure of a signal system or equipment)	Transition hazard	P: H01, H06 and H13 S: H02 and H07	Improving the level of daily inspection and maintenance. Enhancing the competence of dispatchers. Equipping signal systems with automatic checking devices. Enhancing the competence of signal workers. Strengthening the supervision on signal equipments maintenance.

application of the proposed knowledge graph-based approach on real railway operational accidents in the UK has been performed. The results show that the proposed approach is useful in discovering the latent features of the collected railway operational accidents, by means of the knowledge graph-based topological analysis. By further analysing the revealed features, the effectiveness of the proposed approach in providing decision-making basis for the investment of accident prevention efforts is shown. The knowledge graph-based approach is expected to be applied to exploring railway accidents of other types, as well as providing additional decision information for railway accident prevention.

In this paper, the cause-effect links between hazard nodes are simplified and qualitative for the sake of brevity and clarity. Actually, the cause-effect relationship between hazards are formed through various media, such as mechanical medium between wheels and rail track, informational medium between signal equipment and train driver, policy medium between management and supervision, energy medium between strong wind and falling objects. The different media will result in different strength and spreading speeds of cause-effect relationships. Both the depiction of the various cause-effect media and the quantification of cause-effect relationships will be further researched in future work. Furthermore, if the sufficient and well-described accident data about one certain railway line is publicly available in the future, the meaningful non-binary relationships among the hazards will be able to be captured by use of the probabilistic approach. On the basis of that, the sensitivity analysis based on a non-binary system can be utilized to identify the key causation factors.

The database used in this paper contains only some of railway operational accident events. If the well-formed and consistent data about the other events is generated and publicly available in the future, the proposed approach will be appropriate for exploring a wider range of more in-depth knowledge about all railway operational events. And the knowledge graph model in this paper can be further extended with the new data for the knowledge exploration. Furthermore, if there is systematical and consistent accident data about other complex transport systems, it is expected to play a role in exploring the latent rules. Besides, in addition to the existing railway hazard checklists, how to organise and utilise the railway expert interviews to improve the accuracy and objectivity of hazard identification is also urgent in further research.

CRediT authorship contribution statement

Jintao Liu: Conceptualization, Methodology, Writing - original draft. **Felix Schmid:** Validation, Writing - review & editing. **Keping Li:** Visualization, Investigation. **Wei Zheng:** Supervision.

Declaration of Competing Interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

Acknowledgments

This work is co-supported by the National Natural Science Foundation of China (No. 61803019, No. 61803020), the China Railway Corporation Foundation (No. K2019X010) and the Beijing Natural Science Foundation (No. 8202039). The authors also gratefully acknowledge the insightful comments of the editor and the reviewers that substantially improved the paper.

Appendix A

Report No.	Accident	Report No.	Accident	Report No.	Accident	Report No.	Accident	Report No.	Accident
R152006	A01	R102014	A01	R042014	A03	R052008	A07->A06	R162015	A06
R252006	A01	R112014	A01	R062014	A03	R072008	A06	R012016	A06
R272006	A01	R122014	A01	R092014	A04-> A03	R092008	A06	R052016	A06
R062007	A01	R132014	A01	R152014	A02->A03	R112008	A07->A06	R102016	A06
R102007	A01	R142014	A01	R252014	A03	R162008	A06	R112016	A06
R162007	A01	R172014	A01	R262014	A03	R182008	A07->A06	R162016	A06
R312007	A01	R232014	A01	R042015	A02->A03	R202008	A06	R162006	A07
R432007	A01	R082015	A01	R152015	A03	R022009	A06	R222007	A07
R062008	A01	R132015	A01	R022016	A03	R032009	A06	R232007	A07
R132008	A01	R192015	A01	R072016	A03	R042009	A06	R262007	A07
R222008	A01	R202015	A01	R092016	A02->A03	R072009	A06	R042008	A07
R262008	A01	R142016	A01	R132016	A04-> A03	R102009	A06	R192008	A07
R122009	A01	R202006	A02	R262006	A04	R182009	A06	R212008	A07
R142009	A01	R132007	A02	R182011	A04	R282009	A06	R272008	A07
R202009	A01	R082008	A02	R292014	A04	R012010	A06	R162009	A07
R042010	A01	R122008	A02	R152016	A04	R022010	A06	R172009	A07
R122010	A01	R052009	A02	R182016	A04	R032010	A06	R192009	A07
R172010	A01	R092011	A02	R142007	A05	R062010	A06	R212009	A07
R182010	A01	R152011	A02	R072015	A05	R072010	A06	R222009	A07
R022011	A01	R072006	A04-> A03	R022006	A06	R112010	A06	R232009	A07
R202011	A01	R122006	A03	R082006	A06	R142010	A06	R292009	A07
R032012	A01	R292007	A03	R142006	A06	R192010	A06	R302009	A07
R042012	A01	R302007	A03	R172006	A06	R032011	A04-> A06	R082010	A07
R052012	A01	R332007	A03	R192006	A06	R052011	A06	R152010	A07
R072012	A01	R352007	A03	R212006	A06	R072011	A02->A06	R012011	A07
R092012	A01	R362007	A03	R222006	A06	R162011	A07->A06	R062011	A07
R112012	A01	R452007	A03	R012007	A04->A06	R012012	A06	R112011	A07
R122012	A01	R012008	A03	R022007	A06	R182012	A06	R172011	A07
R132012	A01	R102008	A03	R032007	A06	R192012	A06	R192011	A07
R142012	A01	R142008	A03	R052007	A06	R242012	A06	R102012	A07
R262012	A01	R152009	A03	R072007	A06	R022013	A06	R152012	A07
R282012	A01	R242009	A03	R082007	A06	R042013	A06	R162012	A07
R052013	A01	R312009	A03	R202007	A06	R082013	A06	R172012	A07
R062013	A01	R332009	A03	R212007	A06	R222013	A06	R212012	A07
R072013	A01	R102010	A04-> A03	R242007	A06	R022014	A06	R232012	A07
R112013	A01	R042011	A03	R252007	A06	R072014	A06	R202013	A07
R132013	A01	R082011	A03	R322007	A06	R202014	A06	R212013	A07
R142013	A01	R102011	A02->A03	R342007	A06	R212014	A06	R012015	A07
R152013	A01	R142011	A03	R392007	A06	R272014	A06	R052015	A07
R182013	A01	R062012	A03	R422007	A06	R022015	A06	R062015	A07
R032014	A01	R202012	A03	R442007	A06	R032015	A06	R122015	A07
R052014	A01	R122013	A04-> A03	R022008	A06	R102015	A06	R172015	A07
R082014	A01	R172013	A04-> A03	R032008	A07->A06	R112015	A06		

References

[1] Kyriakidis M, Majumdar A, Ochieng WY. Data based framework to identify the most significant performance shaping factors in railway operations. Saf Sci 2015;78:60–76.

[2] SAWS. Investigation report of the 7.23 Yongwen Railway accident. http://www.gov.cn/gzdt/2011-12/29/ content_2032986.htm;2011 [accessed 13 August 2020].

[3] NTSB. Derailment of Amtrak Passenger Train 188. Washington, D.C., 2015.

[4] Shultz JM, Garcia-Vera MP, Santos CG, et al. Disaster complexity and the Santiago de Compostela train derailment. Disaster Health 2016;3(1):11–31.

[5] Kaeeni S, Khalilian M, Mohammadzadeh J. Derailment accident risk assessment based on ensemble classification method. Saf Sci 2018;110:3–10.

[6] Goh YM. Ubeynarayana CU. Construction accident narrative classification: an evaluation of text mining techniques. Accid Anal Prev 2017;108:122–30.

[7] Grant E, Salmon PM, Stevens NJ, et al. Back to the future: what do accident causation models tell us about accident prediction? Saf Sci 2018;104:99–109.

[8] Hulme A, Stanton NA, Walker GH, et al. What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. Saf Sci 2019;117:164–83.

[9] Alileche N, Olivier D, Estel L, Cozzani V. Analysis of domino effect in the process industry using the event tree method. Saf Sci 2017;97:10–9.

[10] Hemmatian B, Abdolhamidzadeh B, Darbra RM, Casal J. The significance of domino effect in chemical accidents. J Loss Prevent Proc 2014;29:30–8.

[11] Reason J, Hollnagel E, Paries J. Revisiting the ``Swiss Cheese'' model of accidents. https://www.eurocontrol.int/publication/revisiting-swiss-cheese-model-accidents; 2006 [accessed 13 August 2020].

[12] Duffield S, Whitty SJ. Developing a systemic lessons learned knowledge model for organisational learning through projects. Int J Proj Manage 2015;33(2):311–24.

[13] Rasmussen J. Risk management in a dynamic society: a modelling problem. Saf Sci 1997;27(2-3):183–213.

[14] Leveson N. A systems approach to risk management through leading safety indicators. Reliab Eng Syst Saf 2015;136:17–34.

[15] Liu P, Yang L, Gao Z, et al. Fault tree analysis combined with quantitative analysis for high-speed railway accidents. Saf Sci 2015;79:344–57.

[16] Lin C, Saat MR, Barkan CPL. Fault tree analysis of adjacent track accidents on shared-use rail corridors. Transp Res Record 2016;2546(1):129–36.

[17] Khakzad N, Khan F, Amyotte P. Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches. Reliab Eng Syst Saf 2011;96:925–32.

[18] Khakzad N, Khan F, Amyotte P. Risk-based design of process systems using discrete-time Bayesian networks. Reliab Eng Syst Saf 2013;109:5–17.

[19] Baysari MT, McIntosh AS, Wilson JR. Understanding the human factors contribution to railway accidents and incidents in Australia. Accid Anal Prev 2008;40(5):1750–7.

[20] Zhan Q, Zheng W, Zhao B. A hybrid human and organizational analysis method for railway accidents based on HFACS-railway accidents (HFACS-RAs). Saf Sci 2017;91:232–50.

[21] Rostamabadi A, Jahangiri M, Zarei E, et al. A novel fuzzy Bayesian network approach for safety analysis of process systems; an application of HFACS and SHIPP methodology. J Clean Prod 2020;244:118761.

[22] Khakzad N, Khan F, Amyotte P. Dynamic risk analysis using bow-tie approach. Reliab Eng Syst Safe 2012;104:36–44.

[23] Khakzad N, Khan F, Amyotte P. Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. Process Saf Environ 2013;91(1-2):46–53.

[24] Rathnayaka S, Khan F, Amyotte P. SHIPP methodology: predictive accident modeling approach. Part I: methodology and model description. Process Saf Environ 2011;89(3):151–64.

[25] Baksh A, Khan F, Gadag V, Ferdous R. Network based approach for predictive accident modelling. Saf Sci 2015;80:274–87.

[26] Belmonte F, Schön W, Heurley L, Capel R. Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: an application to railway trafficsupervision. Reliab Eng Syst Saf 2011;96(2):237–49.

[27] Salmon PM, Read GJM, Stanton NA, Lenné MG. The crash at Kerang: investigating systemic and psychological factors leading to unintentional non-compliance at rail level crossings. Accid Anal Prev 2013;50:1278–88.

[28] Fan Y, Li Z, Pei J, et al. Applying systems thinking approach to accident analysis in China: case study of ``7.23'' Yong-Tai-Wen high-speed train accident. Saf Sci 2015;76:190–201.

[29] Li C, Tang T, Chatzimichailidou MM, et al. A hybrid human and organisational analysis method for railway accidents based on STAMP-HFACS and human information processing. Appl Ergon 2019;79:122–42.
[30] Evans AW. Fatal accidents at railway level crossings in Great Britain 1946–2009. Accid Anal Prev 2011;43(5):1837–45.
[31] Naznin F, Currie G, Logan D, Sarvi M. Application of a random effects negative binomial model to examine tram-involved crash frequency on route sections in Melbourne, Australia. Accid Anal Prev 2016;92:15–21.
[32] Liu X, Rapik Saat M, Barkan CPL. Freight-train derailment rates for railroad safety and risk analysis. Accid Anal Prev 2017;98:1–9.
[33] Jonsson L, Björklund G, Isacsson G. Marginal costs for railway level crossing accidents in Sweden. Transp Policy 2019;83:68–79.
[34] Evans AW, Hughes P. Traverses, delays and fatalities at railway level crossings in Great Britain. Accid Anal Prev 2019;129:66–75.
[35] Zhou JL, Lei Y. Paths between latent and active errors: analysis of 407 railway accidents/incidents' causes in China. Saf Sci 2018;110:47–58.
[36] Ahmad M, Shabnam S. Application of association rules in Iranian Railways (RAI) accident data analysis. Saf Sci 2010;48:1427–35.
[37] Kim DS, Yoon WC. An accident causation model for the railway industry: application of the model to 80 rail accident investigation reports from the UK. Saf Sci 2013;60:57–68.
[38] Dabbour E, Easa S, Haider M. Using fixed-parameter and random-parameter ordered regression models to identify significant factors that affect the severity of drivers' injuries in vehicle-train collisions. Accid Anal Prev 2017;107:20–30.
[39] Zhang Z, Turla T, Liu X. Analysis of human-factor-caused freight train accidents in the United States. J Transp Saf Secur 2019:1-29.
[40] Lin C, Rapik Saat M, Barkan CP. Quantitative causal analysis of mainline passenger train accidents in the United States. P I Mech Eng F-J Rai 2020;234(8):869–84.
[41] Savage I. Analysis of fatal train-pedestrian collisions in metropolitan Chicago 2004–2012. Accid Anal Prev 2016;86:217–28.
[42] Naznin F, Currie G, Logan D. Exploring the impacts of factors contributing to tram-involved serious injury crashes on Melbourne tram routes. Accid Anal Prev 2016;94:238–44.
[43] Haleem K. Investigating risk factors of traffic casualties at private highway-railroad grade crossings in the United States. Accid Anal Prev 2016;95:274–83.
[44] Rudin-Brown CM, Harris S, Rosberg A. How shift scheduling practices contribute to fatigue amongst freight rail operating employees: findings from Canadian accident investigations. Accid Anal Prev 2019;126:64–9.
[45] Chen D, Xu C, Ni S. Data mining on Chinese train accidents to derive associated rules. P I Mech Eng F-J Rai 2017;231(2):239–52.
[46] Li Q, Song L, List GF, et al. A new approach to understand metro operation safety by exploring metro operation hazard network (MOHN). Saf Sci 2017;93:50–61.
[47] Liu J, Schmid F, Zheng W, Zhu J. Understanding railway operational accidents using network theory. Reliab Eng Syst Saf 2019;189:218–31.
[48] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: structure and dynamics. Phys Rep 2006;424(4-5):175–308.
[49] Klockner K, Toft Y. Accident modelling of railway safety occurrences: the safety and failure event network (SAFE-Net) method. In: International conference on applied human factors and ergonomics (AHFE); 2015. p. 1734–41.
[50] Shao F, Li K. A complex network model for analyzing railway accidents based on the maximal Information coefficient. Commun Theor Phys 2016;66(4):459–66.
[51] Klockner K, Toft Y. Railway accidents and incidents: complex socio-technical system accident modelling comes of age. Saf Sci 2018;110:59–66.
[52] Li K, Wang S. A network accident causation model for monitoring railway safety. Saf Sci 2018;109:398–402.
[53] Zhou JL, Lei Y. A slim integrated with empirical study and network analysis for human error assessment in the railway driving process. Reliab Eng Syst Saf 2020;204:107148.
[54] Liu J, Li K, Zheng W, Zhu J. An importance order analysis method for causes of railway signaling system hazards based on complex networks. P I Mech Eng O-J Ris 2018;233(4):567–79.
[55] Hou G, Jin C, Xu Z, et al. Exploring evolutionary features of directed weighted hazard network in the subway construction. Chinese Phys B 2019;28(3):38901.
[56] Lam CY, Tai K. Network topological approach to modeling accident causations and characteristics: analysis of railway incidents in Japan. Reliab Eng Syst Saf 2020;193:106626.
[57] Barrett L, Henzi SP, Lusseau D. Taking sociality seriously: the structure of multi-dimensional social networks as a source of information for individuals. Philos T R Soc B 2012;367(1599):2108–18.
[58] Boccaletti S, Bianconi G, Criado R, et al. The structure and dynamics of multilayer networks. Phys Rep 2014;544(1):1–122.
[59] Sheth A, Padhee S, Gyrard A, Sheth A. Knowledge graphs and knowledge networks: the story in brief. IEEE Internet Comput 2019;23(4):67–75.
[60] Hughes P, Robinson R, Figueres-Esteban M, van Gulijk C. Extracting safety information from multi-lingual accident reports using an ontology-based approach. Saf Sci 2019;118:288–97.
[61] Rashidy RAHE, Hughes P, Figueres-Esteban M, et al. A big data modeling approach with graph databases for SPAD risk. Saf Sci 2018;110:75–9.
[62] Liu H, Chen H, Hong R, et al. Mapping knowledge structure and research trends of emergency evacuation studies. Saf Sci 2020;121:348–61.
[63] Dindar S, Kaewunruen S. Assessment of turnout-related derailments by various causes. In: International congress and exhibition on sustainable civil infrastructures; 2018. p. 27–39.
[64] RAIB. Rail accident investigation branch reports. https://www.gov.uk/raib-reports; 2018 [accessed 13 August 2020].
[65] RSSB. Guidance on hazard identification and classification (GE/GN8642). London: Rail Safety And Standards Board; 2014.
[66] Gerbec M. Safety change management-A new method for integrated management of organizational and technical changes. Saf Sci 2017;100:225–34.
[67] Clark JR, Stanton NA, Revell KMA. Identified handover tools and techniques in high-risk domains: using distributed situation awareness theory to inform current practices. Saf Sci 2019;118:915–24.
[68] RSSB.. Engineering safety management (The Yellow Book). London: Rail Safety and Standards Board; 2007. Volumes 1 and 2.
[69] RSSB. Hazard analysis and risk assessment for rail projects (T955) reports. https://www.rssb.co.uk/ Standards-and-Safety/Tools–Resources/; 2012 [accessed 13 August 2020].
[70] Chen X, Jia S, Xiang Y. A review: knowledge reasoning over knowledge graph. Expert Syst Appl 2020;141:112948.
[71] Ahn CW, Ramakrishna RS. A genetic algorithm for shortest path routing problem and the sizing of populations. IEEE T Evolut Comput 2002;6(6):566–79.
[72] Castillo E, Cervantes O, Vilarino D. Authorship verification using a graph knowledge discovery approach. J Intell Fuzzy Syst 2019;36(6):6075–87.
[73] Huang X, Zheng Q, Zhang C. Extracting learning features of knowledge unit in knowledge map. In: International conference on intelligent information hiding and multimedia signal processing; 2014. p. 345–8.
[74] Liu J, Wang J, Zheng Q, et al. Topological analysis of knowledge maps. Knowl-Based Syst 2012;36:260–7.
[75] Ding YH, Yu HT, Huang RY, Gu YJ. Complex network based knowledge graph ontology structure analysis. In: IEEE International conference on hot information-centric networking. chongqing; 2018. p. 193–9.