Process Safety and Environmental Protection 153 (2021) 320–328

Contents lists available at ScienceDirect

Process Safety and Environmental Protection

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

Construction and analysis of a coal mine accident causation network based on text mining

Zunxiang Qiu[a], Quanlong Liu[a,*], Xinchun Li[a], Jinjia Zhang[b], Yueqian Zhang[a]

[a] School of Economics and Management, China University of Mining & Technology, Xuzhou, 221116, China
[b] Institute of Public Safety Research, Tsinghua University, Beijing 100084, China

A R T I C L E I N F O

Article history:
Received 17 April 2021
Received in revised form 28 June 2021
Accepted 21 July 2021
Available online 23 July 2021

Keywords:
Coal mine safety
Text mining technology
Complex network theory
Accident causation
Risk analysis

A B S T R A C T

It is important to systematically identify the contributing factors in coal mine accidents from a large-scale analysis of accident reports. However, previous scholars have mainly used human analysis methods to define accident-causing factors, leading to incomplete and biased cause checklists due to personal experience and knowledge. Furthermore, a data-driven method is needed to quantify the importance of each factor and clarify the mechanism of different types of accidents. Considering these, this study creatively combined text mining technology and a complex network to explore the coal mine accident-causing mechanism. Through the text mining of 307 accident reports, 52 main accident-causing factors were identified, and a coal mine accident causation network was constructed based on the strong association rules among factors. Second, eight core factors and their associated sets, as well as seven critical links for different accident types, were clarified through network centrality analysis and accident path analysis. This study shows that regulatory authority is the most influential level of accident causation, gas accidents are the most easily triggered accident type, a lack of effective mechanism for safety supervision→failure to arrange full-time safety inspectors to follow the shift→lack of serious and thorough on-site hidden danger investigations→inadequate anti-surge measures are the key links in gas accident causation. This study contributes new perspectives on identifying contributing factors and their complex interaction mechanisms from accident report data for practical applications in risk analysis and accident prevention.

1. Introduction

Underground coal mining is considered one of the most dangerous operations worldwide (Liu et al., 2019c; Nieto et al., 2014). In China, in particular, there is an appalling record of fatalities in the coal mining industry, ranking China as the third-worst coal producer regarding underperformance in coal mine safety (Liu et al., 2019b). Although the number of coal mine accidents has gradually decreased in recent years, the complexity of coal mine accidents has significantly increased (Qiao et al., 2019). Coal mine safety problems are a complex social-technical system problem (Sun and Liu, 2010). According to accident causation theory, the occurrence of accidents is the result of the interactions among multiple hazards, including people, equipment, environment and management (Liu et al., 2019d). The hidden dangers generated by any hazard may cause risk to spread in the system and eventually trigger accidents. Therefore, to effectively prevent coal mine accidents and improve risk management and control, it is essential to study coal mine accident causation mechanisms.

The purpose of studying the accident-causing mechanism is to explain the accident mechanism and provide a reference for risk assessment and accident analysis (Fu et al., 2018). According to the development of accident causation theory, research on the cause mechanisms of coal mine accidents has gone through three main stages. The first stage was the chain accident causation model. Zhang and Lowndes (2010) used the fault tree model to study the characteristics of 8 typical accident causes in coal and gas outburst accidents. Zhang et al. (2014) used fault tree analysis to investigate truck-related fatal accidents in surface mining, suggesting that inadequate or improper preoperational checks and poor maintenance of trucks were the two most common root causes of these accidents. The second stage was the epidemiological accident causation model; for example, Lenne et al. (2012) used human factor analysis and classification system models to find that organizational culture and inadequate implementation of safety supervision were the main factors to address to prevent coal mine accidents. Liu et al. (2019e) used the HFACS method combined with a struc-

* Corresponding author at: University RD. 1, Xuzhou, Jiangsu Province, 221116, PR China.
E-mail address: qll2016@cumt.edu.cn (Q. Liu).

https://doi.org/10.1016/j.psep.2021.07.032
0957-5820/

tural equation model to find that the external environment has the most significant direct and indirect impact on the unsafe behavior of miners, and unsafe leaders play an essential intermediary role. The third stage was the systemic accident causation model. For example, Zhu (2018) first applied the systems-theoretic accident model and processes (STAMP) to analyze a vast gas explosion accident from the perspective of social work, regulatory authorities, and corporate employees, finding that more than 12 organizations have control mistakes, and the lack of communication and feedback loops between all levels are the key factors leading to poor decision-making or inappropriate control. Fu et al. (2020) applied the 24 model to analyze gas accidents and proposed accident prevention measures for the risk control of gas and the behavioral safety of miners. Zhang et al. (2020) used the 24 model and the logical thought of why-because analysis (WBA) to identify the characteristics of safety culture deficiencies driving typical coal accidents.

While existing research has been carried out based on classic accident causation theory, which effectively deepens understanding of the cause mechanism of coal mine accidents, it has the following shortcomings. (1) When conducting large-scale analyses of coal mine accidents to determine their accident-causing factors, previous scholars mainly applied human-analysis methods such as brainstorming, the Delphi method and questionnaires, which inevitably consume extensive time and cost and may be influenced by incomplete and biased cause checklists due to personal experience and knowledge. (2) Although existing studies have analyzed the accident-causing mechanism from a systematic perspective, they still cannot fully reflect the hierarchy and correlation among accident-causing factors, nor the complex nonlinear interactions between causative factors. Specifically, a data-driven method is needed to systematically determine the key causal factors and mechanisms of different types of accidents and to formulate more accurate and efficient accident prevention strategies. Based on these results, this study creatively combines text mining technology and complex network theory to construct and analyze the causation network of coal mine accidents. First, text mining technology is used to extract major accident-causing factors from vast quantities of accident reports, and the risk management model in a sociotechnical system is applied to classify the hierarchies of the factors. After that, the strong association rules among factors are exposed by the association mining method, forming the basis to construct the coal mine accident causation network. Finally, the complex interaction within accident-causing factors is identified by the analysis method of complex network theory to determine the core factors and causation links in the different types of accidents, providing a systematic theoretical guideline for safe management in the coal mine industry. The workflow proposed in this study contributes to new perspectives of conducting large-scale analyses of accident reports to identify contributing factors and their complex interaction mechanisms.

The remainder of this paper is organized as follows. Section 2 introduces the data sources, the basic concepts of each method and the research framework. In Section 3, manifestations of the accident-causing factors are extracted by text mining, and a coal mine accident causation network is constructed and analyzed. Finally, the conclusions of this study are presented in Section 4.

2. Material and methods

2.1. Data set

Accident investigation reports are an important data source for accident analysis because they contain various types of information, such as accident descriptions and direct/indirect causes of accidents. This study collected 307 coal mine accident investigation reports from 2007 to 2020, from the websites of the State Administration of Work Safety and the relevant administrative departments of various provinces and cities, as the corpus for text mining. The accident reports involve 24 provinces and 7 accident types, ensuring the objectiveness and universality of the subsequent application of text mining technology to identify the causes of coal mine accidents.

2.2. Text mining

In recent years, the application of data analysis in accident investigation reports has provided new paths to study accident-causing factors (Zhang et al., 2018; Zhu et al., 2019). Furthermore, text mining technology has been widely applied to study the causes of accidents. By providing an extensive analysis of accident reports or narratives, text mining can support a better understanding of the contributors to accidents, which significantly improves the predictive accuracy for accidents (Brown, 2016). For example, Raviv et al. (2017) found that in the tower crane domain, technical failures are the most hazardous risk factors using text mining and a k-means clustering analysis of 212 crane-related near-miss and accident reports. Hughes et al. (2018) introduced a semiautomated technique for classifying text-based close call reports from the GB railway industry, which enabled the categorization of a massive number of unstructured text documents. Singh et al. (2019) identified the nine most frequently recurring accident paths and corresponding preventive strategies by text mining both proactive (workplace observation and high-risk control program) and reactive data (incident records).

2.3. Complex network

Research on complex networks for accident prevention and risk assessment is also emerging. Complex networks can reveal the evolution of the interactions of complex factors and then explain the overall function of the complex system and the impact of individuals on the system. For instance, Li et al. (2017) established a metro operation hazard network to understand interrelations among metro system hazards. Liu et al. (2019a) proposed a new method for understanding railway operation accidents based on network theory, aiming to reveal potential hazard patterns from a high-level perspective. Guo et al. (2020) explored the behavioral risk chain of accidents based on complex network theory and case studies of the Chinese construction industry.

2.4. Research framework

As shown in Fig. 1, the research process of this paper includes the following steps. First, text mining technology is used to mine the text features of coal mine accident reports and reduce the initial text features' dimensionality based on the chi-square statistical method. On this basis, the leading causes of accidents are extracted, and the levels of the causes are classified according to the risk management model in the sociotechnical system. Second, the Apriori algorithm is used to extract the strong association rules between the accident causes, forming the basis to construct the network. Finally, a coal mine accident causation network is built, and thereby, the core accident causes and their closely associated-cause sets are obtained through network centrality analysis. Furthermore, through an accident path analysis, the degree of influence of different cause types on accidents is obtained, and the critical causation links in the different types of accidents are identified.

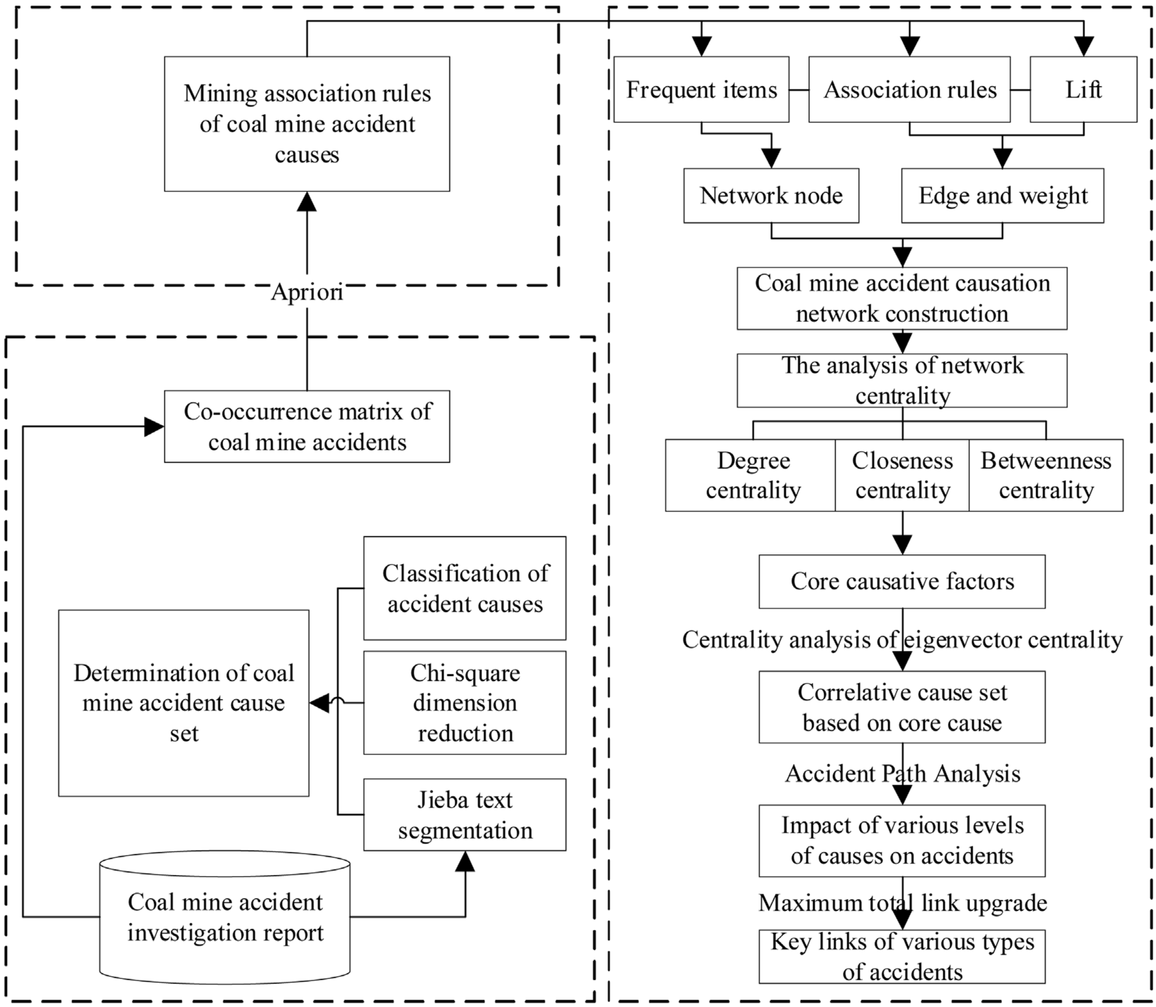

Fig. 1. Construction and analysis process of a coal mine accident causation network based on text mining.

3. Results and analysis

3.1. Identification of the causes in coal mine accidents

With the application of Python and Jieba, 307 investigation reports of coal mine accidents were initially subjected to text segmentation, and 860 initial text features were obtained. The initial feature dimension was too high, which would seriously hinder subsequent analysis; therefore, it was necessary to reduce the dimension and filter the initial text feature items. Qi (2008) found that the chi-square statistics method has apparent advantages over other dimensionality reduction methods in text recall and precision. Therefore, chi-square statistics were used to reduce the dimensionality of the initial feature items. The calculation formula is as follows:

$$\chi^2\left(t_k, c_j\right) = \frac{N(AD-BC)^2}{(A+C)(B+D)(B+D)(C+D)}$$

where t_k represents the feature item; c_j is the j-th category; N represents the total number of texts in the corpus; A represents the number of texts that contain t_k and belong to c_j ; B is the number of texts that contain t_k but do not belong to c_j ; C is the number of texts that belong to the c_j type but do not contain t_k ; D is the number of texts that do not contain t_k and do not belong to c_j (Qi, 2008).

The above chi-square statistical method is used to reduce the dimensionality of the initial features in Python. The first 100 text features are filtered according to the feature items' chi-square value for subsequent extraction and stratification of accident causes. Furthermore, the StyleCloud toolkit was used to visualize the top 100 text features, as shown in Fig. 2. The feature items are distributed from the inside to the outside according to the size of the chi-square value. Specifically, the larger the feature text is, the better it reflects the inherent characteristics of coal mine accidents.

The initial text features cannot fully reflect the causes of accidents, such as risk investigation, technical safety measures, safety monitoring, safety monitoring systems, etc. On the one hand, there is duplication of semantics, and on the other hand, they do not specifically reflect hidden dangers. To address this issue, this study standardizes the text characteristics based on the initial text characteristics and the content of the accident report. For example, risk investigation is regulated by a lack of serious and thorough investigation into on-site hidden dangers; technical safety measures are regulated by the inadequate formulation and implementation of technical safety measures; safety monitoring and safety monitoring systems are regulated by a lack of security monitoring systems. Ultimately, 52 standardized text features are selected for further stratification and analysis of the causes of coal mine accidents.

3.2. Classification of the causes of coal mine accidents

The coal mine safety problem is a sociotechnical system problem. The occurrence of accidents results from the coupling effect of multiple hazards in the links between people-equipment-

Fig. 2. Word cloud map of the top 100 text features.

environment-management, which is complex and systematic. However, existing research usually isolates various causes when analyzing coal mine accidents, ignoring the correlation between causes and their hierarchical nature. Therefore, based on the risk management model in sociotechnical systems (Rasmussen, 1997), this study divides the 52 coal mine accident causes into five levels, including the regulatory authority, the coal mining enterprise, on-site management, operation staff, and the environment and equipment, as shown in Table 1.

3.3. Network construction and analysis of coal mine accidents

3.3.1. Construction of coal mine accident causation network

In the coal mine accident causation system, consisting of 52 main accident-causing factors, the failure to control one or more of the causal factors will directly or indirectly affect the closely associated factors, causing risks to spread within the system and ultimately leading to accidents. Association rule analysis can mine the association relations between many items, and the Apriori algorithm is a classic algorithm for mining Boolean association rules (Li and Li, 2020). Therefore, the Apriori algorithm is applied to mine the strong association rules among the causes of coal mine accidents and use them to build a coal mine accident causation network. All accident types involved in the accident reports are numbered, as shown in Table 2. According to the requirement for Apriori algorithm analysis, the minimum support threshold is set to 0.06, the minimum confidence threshold is 0.1, the maximum number of antecedents is set as 2, and the minimum lift threshold is 1. A total of 1175 strong association rules were mined, reflecting the close connection between the causes of coal mine accidents. The top five association rules with regard to lift are shown in Table 3.

Based on strong association rules and the Pajek software, a coal mine accident causation network is constructed. The antecedent and subsequent items of all strong association rules are used as nodes in the network, the relationship between the antecedent and subsequent items is treated as the edge, and the lift is used as the edge's weight. The final coal mine accident causation network is shown in Fig. 3. The network contains 100 nodes and 1175 edges. The density of connections around nodes reflects the complexity of the area where the nodes are located. The closer the node is to the center of the network, the higher its influence on other nodes. Furthermore, among the causes of accidents, the causal factors at the level of the regulatory authority and the coal mining enterprise are closer to the network's center. For the types of accidents, gas accidents tend to be more toward the center of the network.

Table 1
Coal mine accident cause collection.

Cause level	Number of accident-cause
Regulatory Authority	G01 Unscientific formulation of safety supervision plan
	G02 Lack of effective mechanism for safety supervision
	G03 Insufficient investment in safety supervision
	G04 Ineffective control of illegal production
	G05 Inadequate management of safety supervision
	G06 Inadequate tracking of hidden-dangers rectification
	G07 Inappropriate arrangements for safe production
	G08 Production approval for unqualified companies
	G09 Lax daily supervision and inspection
Coal Mining Enterprise	E01 Chaotic security management
	E02 Inadequate formulation and implementation of technical safety measures
	E03 Insufficient staffing of safety inspections
	E04 Inadequate safety education and training
	E05 Incomplete safety management institutions
	E06 Disordered labor organization
	E07 Production organized in violation of laws and regulations
	E08 Focus on production and light on safety
	E09 Insufficient professional and technical personnel
On-Site Management	P01 Falsification of the security monitoring system
	P02 Poor quality of work arrangements in preshift meeting
	P03 Inadequate implementation of the roof management system
	P04 Insufficient mine safety supervision and inspection
	P05 Lack of attention to risk management and control in mining
	P06 Nonenforcement of the system wherein “Mine leaders lead workers down the shaft”
	P07 Conducting workers’ adventures against regulations
	P08 Failure to arrange full-time safety inspectors to follow the shift
	P09 Lack of serious and thorough investigation of on-site hidden dangers
Operation Staff	H01 Insufficient capacity for safety protection and emergency response
	H02 Nonstandard blasting operation
	H03 Inadequate anti-surge measures
	H04 Enter the mine without self-rescuer
	H05 On-site security officers’ failure to stop violations in time
	H06 Insufficient hazard identification ability
	H07 Illegal operations
	H08 Belief in luck;
	H09 Weak safety awareness among staff
	H10 Inadequate performance by the mine safety inspector
Environment and Equipment	M01 Lack of security monitoring system
	M03 Complex geological structure
	M07 Soft coal seam
	M10 Gas over limit
	M11 No personnel positioning system installed
	M12 Failure to install gas, methane and CO sensors as required
	M13 Local ventilators are not installed according to the regulations or used irregularly
	M15 Broken roof
	Equipment and locations prone to accidents:M02 Mined-out area; M04 Wind channel; M05 Return laneway; M06 Intake airflow roadway; M08 Coal mining machine; M09 Conveyer; M14 Roadheader

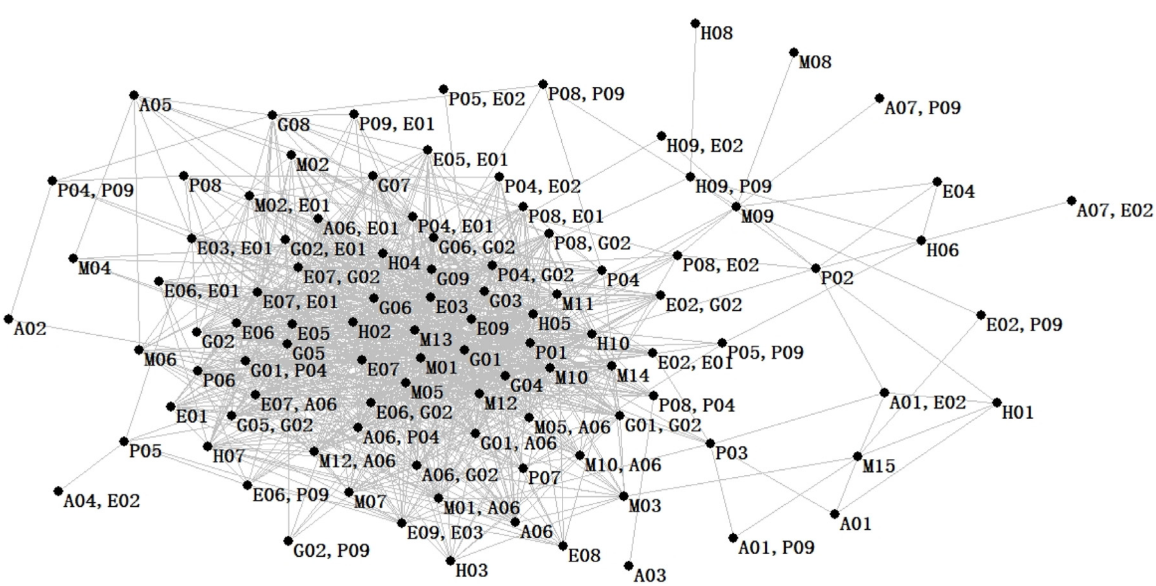

Fig. 3. Coal mine accident causation network.

Table 2
Accident type and number.

Accident type	Number
Roof accident	A01
Fire accident	A02
Mechanical or electrical accident	A03
Other accident	A04
Water disaster accident	A05
Gas accident	A06
Transport accident	A07

Table 3
Association rules of the top 5 on lift.

Antecedent	Consequent	Support (%)	Confidence (%)	Lift
M01	P01	7.82	62.50	12.79
M01, A06	P01	6.84	57.14	11.70
M10	H03	7.17	22.73	11.63
M10, A06	H03	6.19	31.58	9.67
E07, A06	H07	6.19	26.32	8.98

3.3.2. Network centrality analysis

To clarify the key causal factors for coal mine accidents and thus propose a targeted prevention strategy, it is necessary to conduct quantitative research on the importance of the causal factors. By applying the Pajek software to analyze the centrality of the network, the degree centrality, closeness centrality, and betweenness centrality of the nodes is obtained. Newman (2008) proposed these three types of centrality to measure the importance of nodes based on different assumptions. Degree centrality assumes that the more other nodes a node can directly affect, the greater its influence. Closeness centrality assumes that the smaller the average shortest path of a node to all other nodes in the network, the more critical that node is. Betweenness centrality assumes that the intermediary ability to establish the shortest path for different nodes is essential. Therefore, to fully reflect the importance of the accident-causing nodes, this study normalizes and summarizes the index values of degree centrality, closeness centrality, and betweenness centrality. Accordingly, the aggregated score is applied to measure the node importance, and the top eight causes of accidents were extracted as the core causes to address accident prevention in coal mines, as shown in Table 4. Specifically, among the regulatory authorities, the unscientific formulation of safety supervision plans and lax daily supervision and inspection are the main factors; among coal mining enterprises, insufficient professional and technical personnel and insufficient staffing of safety inspections are the main factors, reflecting that coal mines urgently need to improve their construction of and investment in safety agencies. For on-site management, falsification of the security monitoring system and inadequate performance by the mine safety inspector are the main factors, reflecting the lack of safety education and training in coal mines. Among the environment and equipment, conveyers are the most accident-prone equipment, and the lack of security monitoring systems requires a high degree of vigilance.

3.3.3. Analysis of the associated-cause set based on core causes

The core cause node plays a vital role in the network of accidents. When the core cause is not controlled, the risk will gradually spread to those nodes closely related to it. Specifically, when security monitoring systems are lacking, this is usually accompanied by potential hidden dangers, such as gas over the limit, which then triggers a gas accident. Therefore, when preventing the core cause, it is necessary to perform a joint defense against the closely associated-cause set. If the core cause is not controlled, the related cause set should be controlled to block the spread of risks in the network and avoid accidents.

Table 4
The core causes of coal mine accident prevention.

Centrality analysis	The top eight causes of coal mine accidents
(Aggregate score) *100	Insufficient professional and technical personnel: 11.37 Insufficient staffing of safety inspections: 10.99 Unscientific formulation of safety supervision plan: 10.90 Falsification of the security monitoring system: 8.32 Lax daily supervision and inspection: 7.59 Conveyer: 7.24 Lack of security monitoring system: 6.90 Inadequate performance by the mine safety inspector: 6.86
Degree centrality	Unscientific formulation of safety supervision plan: 77 Lack of security monitoring system: 71 Insufficient staffing of safety inspections: 70 Organize production in violation of laws and regulations: 67 Inadequate tracking of hidden-dangers rectification: 63 Insufficient professional and technical personnel: 62 Failure to install gas, methane and CO sensors as required: 57 Return laneway: 55
Close centrality	Unscientific formulation of safety supervision plan: 0.66 Insufficient staffing of safety inspections: 0.64 Falsification of the security monitoring system: 0.63 Insufficient professional and technical personnel: 0.63 Lack of security monitoring system: 0.62 Organize production in violation of laws and regulations: 0.61 Inadequate tracking of hidden-dangers rectification: 0.60 Inadequate performance by the mine safety inspector: 0.59
Betweenness centrality	Insufficient professional and technical personnel: 0.09 Insufficient staffing of safety inspections: 0.083 Unscientific formulation of safety supervision plan: 0.07 Conveyer: 0.06 Falsification of the security monitoring system: 0.05 Lax daily supervision and inspection: 0.04 Inadequate implementation of the roof management system: 0.04 Inadequate performance by the mine safety inspector: 0.03

To identify the associated-cause set of different core causes, a network analysis focused on each core cause was conducted using Pajek. Eigenvector centrality can quantify the degree of connection between nodes and core nodes; therefore, this study determines the associated-cause set of core causes based on the eigenvector centrality value of each node in an individual center network. Using the inadequate performance by the mine safety inspector (H10) as an example, the antecedent and subsequent items of all strong association rules that include H10 were used as nodes in the network, the relations between the antecedent and subsequent items were treated as edges, and the associated lift was used as each edge's weight. The individual center network is shown in Fig. 4. The size of any node in the figure is proportional to the value of eigenvector centrality for the node.

For the purpose of excluding the influence of other core causes, eigenvector centrality values of nodes representing these causes were not considered when exploring the associated-cause set of H10. Consequently, G03, G07, M11, M12, G04, and G06 were the

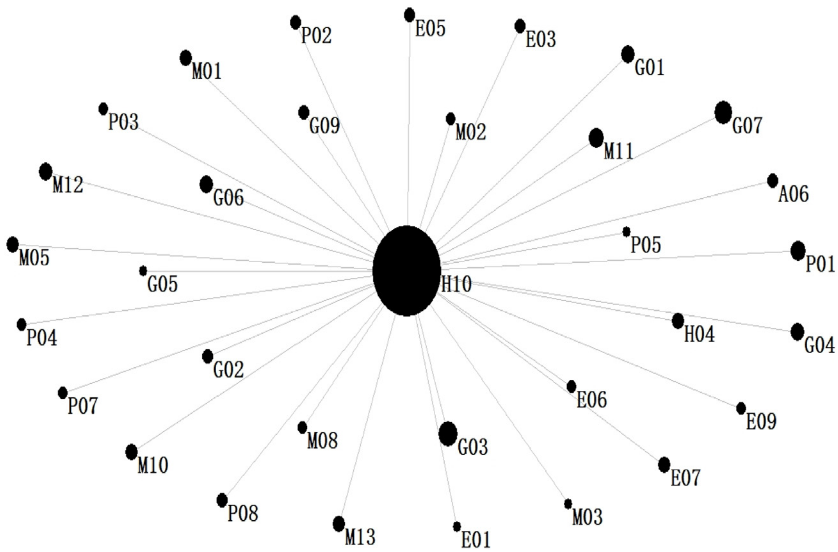

Fig. 4. Egocenter network based on inadequate performance by the mine safety inspector (H10).

Table 5
Core causes and their associated-cause sets.

Core cause	Associated-cause set based on core cause
Insufficient professional and technical personnel	Incomplete safety management institutions; Gas over limit; Organize production in violation of laws and regulations; Conducting workers' adventures against regulation; Roadheader; Chaotic security management
Insufficient staffing of safety inspections	Incomplete safety management institutions; Production organized in violation of laws and regulations; Inappropriate arrangements for safe production; Roadheader; Ineffective control of illegal production; Gas over limit
Unscientific formulation of safety supervision plan	Incomplete safety management institutions; Organize production in violation of laws and regulations; Inappropriate arrangements for safe production; Roadheader; Ineffective control of illegal production; Gas over limit
Falsification of the security monitoring system	Gas accident; Lack of effective supervision mechanism for safety supervision; Insufficient mine safety supervision and inspection; Failure to arrange full-time safety inspectors to follow the shift; Failure to install gas, methane and CO sensors as required; Insufficient capacity for safety protection and emergency response
Lax daily supervision and inspection	Inappropriate arrangements for safe production; Inadequate tracking of hidden-dangers rectification; Ineffective control of illegal production; Wind channel; Local ventilators are not installed according to the regulations or are used irregularly; Incomplete safety management institutions
Conveyer	Inadequate safety education and training; Failure to arrange full-time safety inspectors to follow the shift; Coal mining machine; Transporting accidents; On-site security officers' failure to stop violations in time; Failure to install gas, methane and CO sensors as required
Inadequate performance by the mine safety inspector	Insufficient investment in safety supervision; Inappropriate arrangements for safe production; No personnel positioning system installed; Failure to install gas, methane and CO sensors as required; Ineffective control of illegal production; Inadequate tracking of hidden-dangers rectification

top six nodes in terms of their eigenvector centrality values, which were 0.083, 0.069, 0.045, 0.041, 0.040, and 0.037, respectively. Therefore, this group of nodes is regarded as the associated-cause set for inadequate performance by the mine safety inspector. Furthermore, the analysis of the accident investigation reports shows that coal mining enterprises that have inadequate mine safety inspector performance are the most likely to pay insufficient attention to safety supervision by the local regulatory authority and to conduct insufficient tracking of hidden dangers of the enterprise. These problems further lead coal mining enterprises to underestimate the construction of safety institutions and ignore operators' safety equipment, which significantly increases the hidden dangers of accident risks. Repeating the above analysis for other core causes, the associated-cause set of all core causes, shown in Table 5, is obtained.

3.3.4. Analysis of the accident path

(1) Analysis of the average accident path at different causal levels

From a system perspective, studying the average path length at different causal levels that trigger accidents can unearth the underlying causes of accidents. Specifically, the average path length from the cause node to various accidents is calculated at the five levels of regulatory authority, coal mine enterprise, on-site management, operation staff, and environment and equipment, as shown in Fig. 5. Accordingly, the most likely trigger of an accident is the level of regulatory authority, in which the middle path to various accidents is 2.54, indicating that the regulatory authority's causal node only needs 2.54 steps on average to induce accidents. Therefore, the regulatory authority is the first entity responsible for resolving various potential risks, and it is also the most influential level of accident causation. Among the various accident types, the most easily triggered is a gas accident, in which the middle path is only 1.9, indicating that when two accident causes exist simultaneously, a gas accident may occur.

(2) Analysis of critical links causing different types of accidents

To further clarify the internal mechanism of different types of accidents and formulate precise prevention and control strategies for them, it is necessary to identify the critical cause links in the accidents. Lift can reflect the strength of the correlation between different factors (Xu et al., 2020). For example, the lift between A and B is 2, which means that under condition A, the probability that B will occur doubles. Therefore, this paper constructs each type of accident causation network separately and takes the path with the

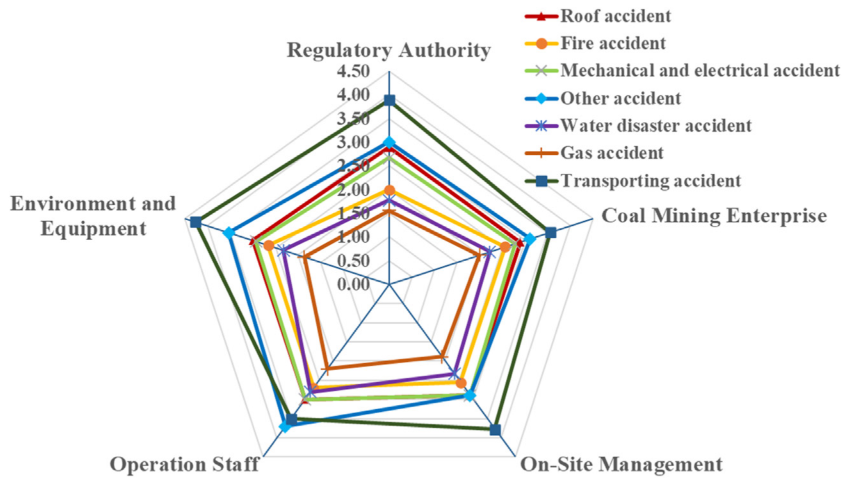

Fig. 5. Average path length between different causal levels and accident types.

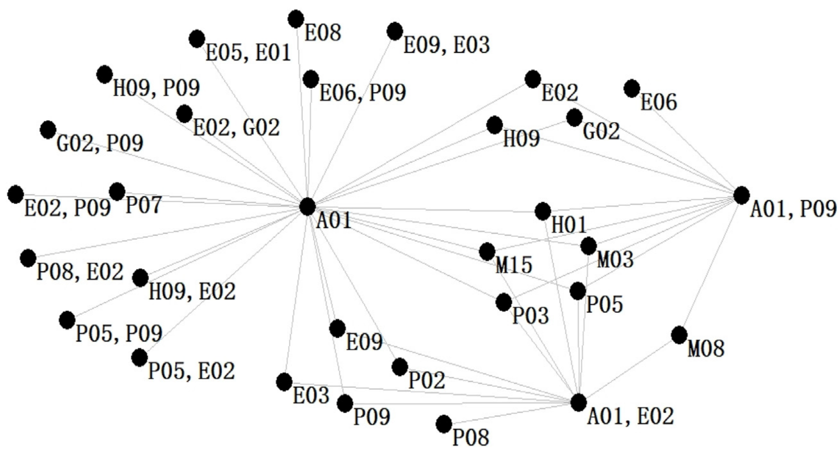

Fig. 6. Roof accident causation network.

largest total lift as the critical link in the accident. Taking the roof accident (A01) causation network as an example, as shown in Fig. 6, the network nodes are all the antecedents or consequences of roof accidents. The lift of the association rule is used as the weight of the edge.

Using Pajek to find the longest path based on the edge's weight, the critical links in a roof accident are obtained as follows: E06→P02→P09→M15→A01, shown in Fig. 7. The total lift, 5.46, equals the sum of all edge weights on the critical link. Specifically, when chaotic security management, poor-quality work arrangements in the preshift meeting, a lack of serious and thorough on-site hidden danger investigation, and a broken roof coexist, roof accidents are more likely to occur than other accident paths in the network. Similarly, the critical links in the causes of other accident types are shown in Table 6.

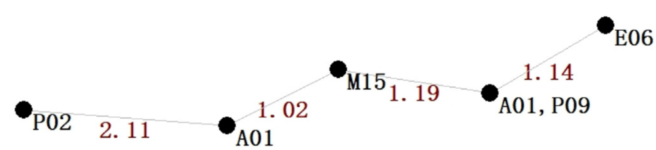

Fig. 7. Critical link for a roof accident.

4. Conclusions

(1) Based on the text mining of 307 coal mine accident investigation reports, combined with chi-square statistical dimensionality reduction and word cloud analysis methods, 52 accident causes, such as the falsification of the security monitoring system and insufficient staffing of safety inspections, were identified as the leading causes of coal mine accidents. Furthermore, based on the risk management model in the sociotechnical system, the 52 identified causes of accidents are divided into five levels: regulatory authority, coal mining enterprise, on-site management, operation staff, environment and equipment.

(2) Through the construction and analysis of the coal mine accident causation network, the eight core causes and their associated-cause sets are identified: unscientific formulation of the safety supervision plan, lax daily supervision and inspection, insufficient professional and technical personnel, insufficient staffing of safety inspections, artificial falsification of the security monitoring system, inadequate performance by the mine safety inspector, conveyer, and lack of security monitoring systems. Furthermore, by accident path analysis, regulatory authority is recognized as the most influential level of cause. Indeed, an incomplete supervision system and the government's failure to implement primary responsibility are the deep-level reasons for coal mine accidents. Finally, chaotic security management→poor quality of work arrangements in preshift meetings→lack of serious and thorough on-site hidden danger investigation→broken roof→roof accident and other critical links for 7 types of accidents were identified, providing a sci-

Table 6
Critical links that cause different types of accidents.

Critical link of accident cause	Total Lift
Chaotic security management→Poor quality of work arrangements in preshift meeting→Lack of serious and thorough on-site hidden danger investigation→Broken roof→Roof accident	5.46
Unscientific formulation of safety supervision plan→Insufficient mine safety supervision and inspection→Lack of serious and thorough on-site hidden danger investigation→Fire accident	7.94
Insufficient mine safety supervision and inspection→Failure to arrange full-time safety inspectors to follow the shift→Conveyer→Mechanical or electrical accident	3.61
Insufficient investment in safety supervision→Inadequate formulation and implementation of technical safety measures→Chaotic security management→Weak safety awareness among staff→Other accident	4.47
Inappropriate arrangements for safe production→ Insufficient staffing of safety inspections → Insufficient professional and technical personnel →Water disaster accident	5.46
Lack of effective supervision mechanism for safety supervision→Failure to arrange full-time safety inspectors to follow the shift→Lack of serious and thorough on-site hidden danger investigation→Inadequate anti-surge measures→Gas accident	9.46
Inadequate formulation and implementation of technical safety measures→Insufficient staffing of safety inspections→Lack of serious and thorough on-site hidden danger investigation→Staff's weak safety awareness→Transport accidents	5.68

entific basis for improving the efficiency of risk prevention and more precise control of risk.

(3) This study focuses on data mining of coal mine accidents and the identification of core causes. However, the construction of an accident prediction system using the characteristics of potential causes is only preliminarily discussed, for which in-depth research is needed in the future.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors gratefully acknowledge the Major Program of Philosophy and Social Science Research in Jiangsu University (Grant no. 2020SJZDA085), the Jiangsu Province Postgraduate Research and Practice Innovation Project (Grant no. KYCX21_2094), the Project Funded by China Postdoctoral Science Foundation (Grant nos. 2019M662003 and 2020M680593), the Jiangsu Planned Projects for Postdoctoral Research Funds (Grant no. 2019K171), and the Key Construction Bases of Philosophy and Social Sciences in Jiangsu Universities (Grant no. 2018ZDJD-B009).

References

Brown, D.E., 2016. Text mining the contributors to rail accidents. IEEE Trans. Intell. Transp. Syst. 17, 346–355.
Fu, G., Suo, x., Jia, Q., Fu, m., 2018. Comparative study on ten accident causation models. J. Saf. Sci. Technol. 14, 58–63.
Fu, G., Xie, X.C., Jia, Q.S., Tong, W.Q., Ge, Y., 2020. Accidents analysis and prevention of coal and gas outburst: understanding human errors in accidents. Process. Saf. Environ. Prot. 134, 1–23.
Guo, S.Y., Zhou, X.Y., Tang, B., Gong, P.S., 2020. Exploring the behavioral risk chains of accidents using complex network theory in the construction industry. Phys. a-Stat. Mech. Appl. 560, 12.
Hughes, P., Shipp, D., Figueres-Esteban, M., van Gulijk, C., 2018. From free-text to structured safety management: introduction of a semi-automated classification method of railway hazard reports to elements on a bow-tie diagram. Saf. Sci. 110, 11–19.
Lenne, M.G., Salmon, P.M., Liu, C.C., Trotter, M., 2012. A systems approach to accident causation in mining: an application of the HFACS method. Accid. Anal. Prev. 48, 111–117.
Li, J., Li, S., 2020. Analysis of causes and association rules for falling accidents based on text mining method. J. Changsha Univ. Sci. Technol. (Natural Science) 17, 61–67.
Li, Q.M., Song, L.L., List, G.F., Deng, Y.L., Zhou, Z.P., Liu, P., 2017. A new approach to understand metro operation safety by exploring metro operation hazard network (MOHN). Saf. Sci. 93, 50–61.
Liu, J.T., Schmid, F., Wei, Z., Zhu, J.B., 2019a. Understanding railway operational accidents using network theory. Reliab. Eng. Syst. Saf. 189, 218–231.
Liu, Q., Li, X., Hassall, M., 2019b. Regulatory regime on coal mine safety in China and Australia: comparative analysis and overall findings. Resour. Policy, 101454.
Liu, Q., Li, X., Meng, X., 2019c. Effectiveness research on the multi-player evolutionary game of coal-mine safety regulation in China based on system dynamics. Saf. Sci. 111, 224–233.
Liu, Q., Meng, X., Li, X., Luo, X., 2019d. Risk precontrol continuum and risk gradient control in underground coal mining. Process. Saf. Environ. Prot. 129, 210–219.
Liu, R.L., Cheng, W.M., Yu, Y.B., Xu, Q.F., Jiang, A.W., Lv, T., 2019e. An impacting factors analysis of miners' unsafe acts based on HFACS-CM and SEM. Process. Saf. Environ. Prot. 122, 221–231.
Newman, M.E., 2008. The mathematics of networks. The New Palgrave Encyclopedia of Economics, 2., pp. 1–12.
Nieto, A., Gao, Y., Grayson, L., Fu, G., 2014. A comparative study of coal mine safety performance indicators in China and the USA. Int. J. Min. Miner. Eng. 5, 299–314.
Qi, S., 2008. Network Text Mining and Its Application in Incident Correlation Intelligence Analysis. National University of Defense Technology.
Qiao, W.G., Li, X.C., Liu, Q.L., 2019. Systemic approaches to incident analysis in coal mines: comparison of the STAMP, FRAM and "2-4" models. Resour. Policy 63, 10.
Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27, 183–213.
Raviv, G., Fishbain, B., Shapira, A., 2017. Analyzing risk factors in crane-related near-miss and accident reports. Saf. Sci. 91, 192–205.
Singh, K., Maiti, J., Dhalmahapatra, K., 2019. Chain of events model for safety management: data analytics approach. Saf. Sci. 118, 568–582.
Sun, A., Liu, M., 2010. Cause analysis model for major and serious accidents in a social-technical perspective. J. China Coal Soc. 35, 876–880.
Xu, W., He, S., Liu, C., Wang, X., 2020. Construction and analysis of railway accident causation network based on association rules. Railway Transp. Econ. 42, 72–79.
Zhang, R.L., Lowndes, I.S., 2010. The application of a coupled artificial neural network and fault tree analysis model to predict coal and gas outbursts. Int. J. Coal Geol. 84, 141–152.
Zhang, M., Kecojevic, V., Komljenovic, D., 2014. Investigation of haul truck-related fatal accidents in surface mining using fault tree analysis. Saf. Sci. 65, 106–117.
Zhang, J.J., Cliff, D., Xu, K.L., You, G., 2018. Focusing on the patterns and characteristics of extraordinarily severe gas explosion accidents in Chinese coal mines. Process. Saf. Environ. Prot. 117, 390–398.
Zhang, J.S., Fu, J., Hao, H.Y., Fu, G., Nie, F.C., Zhang, W.Y., 2020. Root causes of coal mine accidents: characteristics of safety culture deficiencies based on accident statistics. Process. Saf. Environ. Prot. 136, 78–91.
Zhu, K., 2018. Application of STAMP model in coal mine accident analysis. Xitong Gongcheng Lilun yu Shijian/Syst. Eng. Theory Pract. 38, 1069–1082.
Zhu, Y.F., Wang, D.M., Shao, Z.L., Xu, C.H., Zhu, X.L., Qi, X.Y., Liu, F.M., 2019. A statistical analysis of coalmine fires and explosions in China. Process. Saf. Environ. Prot. 121, 357–366.