Association analysis of accident factors in petrochemical storage tank farms

Liu, Zhihao; Zhou, Jianfeng; Reniers, Genserik

DOI
10.1016/j.jlp.2023.105124

Publication date
2023

Document Version
Final published version

Published in
Journal of Loss Prevention in the Process Industries

Journal of Loss Prevention in the Process Industries 84 (2023) 105124

Contents lists available at ScienceDirect

Journal of Loss Prevention in the Process Industries

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

Association analysis of accident factors in petrochemical storage tank farms

Zhihao Liu [a], Jianfeng Zhou [a,*], Genserik Reniers [b,c,d]

[a] School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou, 510006, China
[b] Faculty of Technology, Policy and Management, Safety and Security Science Group (S3G), TU Delft, 2628 BX, Delft, the Netherlands
[c] Faculty of Applied Economics, Antwerp Research Group on Safety and Security (ARGoSS), Universiteit Antwerpen, 2000, Antwerp, Belgium
[d] CEDON, KULeuven, 1000, Brussels, Belgium

ARTICLE INFO

Keywords:
Tank farm accidents
Association rule mining
Complex networks
Important accident factors

ABSTRACT

In order to identify and clarify the association between the factors leading to accidents in a petrochemical tank area, this study analyzes investigation reports of 212 petrochemical tank farm accidents and combines this with the "association rule" mining and science related to complex networks. The main risk factors are determined and a risk factor data set is constructed; 75 association rules are extracted from the factor data set based on the Apriori algorithm. Then the obtained association rules are used to construct an accident factors network of the petrochemical storage tank area, and the topology characteristics of the network are further analyzed to reveal the importance of factors. Factors with large node degree, betweenness, and clustering coefficients are obtained, such as "violation of operating regulations", "high concentration of flammable gas in the air", "lack of experience and professional skills", etc. These factors play an important role in the formation and development of accidents. The results also show that the accident cause network of the petrochemical storage tank area has a small average shortest path length and a large cluster coefficient, indicating a relatively close connection between the accident factors. The contributions of this study is not only extracting the hidden relationships among contributory factors to tank farm accidents using association analysis, but also revealing which factors are more important for the tank farm safety through the complex network.

Author contribution statement

Zhihao Liu: Investigation; Methodology; Writing - Original Draft. Jianfeng Zhou: Conceptualization, Supervision; Writing - Original Draft. Genserik Reniers: Validation, Writing - Review & Editing.

1. Introduction

In petrochemical enterprises, a large amount of hazardous chemicals (liquid or gas) are stored in storage tanks, and with the rapid development of the petrochemical industry, the capacity of a single storage tank and the scale of storage tank becomes even larger. Dangerous goods such as flammable and explosive substances are stored in the storage tank area, and accidents often lead to severe consequences. Accidents in petrochemical storage tank farms often involve a variety of factors, and there is an interaction and influence between these factors. Accident factors interact with each other, increasing the possibility of accidents or making the consequences of accidents more serious.

There have been many serious accidents in petrochemical storage tank areas in history, such as the explosions and fires at the Buncefield oil depot in northeast London, England on December 11, 2005. The fires lasted for 5 days, most of the storage tanks at the oil depot were destroyed, and 43 injuries were caused (Buncefield Major Incident Investigation Board, 2008); Another major accident, on October 23, 2009, severe explosions and fires occurred at the Puerto Rican Caribbean Oil Company (CAPECO), damaging 17 storage tanks (U.S. Chemical Safety and Hazard Investigation Board, 2015). In recent decades, some progress has been made in the risk assessment of hazardous chemical tank farms. Argyropoulos et al. (2012) proposed a systematic method for hazard identification of liquid hydrocarbon fuel storage tanks by applying the checklist technique to the causes of accidents and related protective measures. Wu and Chen (2016) developed a method for quantitative risk assessment of different accident forms caused by lightning in the tank area. Luo et al. (2018) developed a comprehensive risk assessment method for the safety assessment of natural gas spherical tank leakage by combining the improved fishbone diagram and risk matrix model. Guo et al. (2021) proposed an improved Similarity Aggregation Method (SAM) based FBN model to better handle various

* Corresponding author.
E-mail address: jf.zhou@gdut.edu.cn (J. Zhou).

https://doi.org/10.1016/j.jlp.2023.105124
Received 28 January 2023; Received in revised form 29 April 2023; Accepted 13 June 2023
Available online 14 June 2023
0950-4230/

types of uncertainty in risk assessment of storage tanks. Ikwan et al. (2021) quantitatively analyzed the relevant risks by developing fault tree analysis and risk analysis methods to assist real-time risk prediction and safety assessment of tank leaks.

Many accident theories believe that the occurrence of accidents is the result of a combination of multiple factors (Leveson 2004; Rasmussen 1997). Usually, the main accident factors can be divided into four categories: human factors, equipment factors, environment factors (natural environment as well as operational environment and system environment) and management factors. Many risk assessment methods also consider the influence of multiple factors on risk, such as the fault tree analysis method, which uses logic gates ("AND gate", "OR" gate, etc.) to express the influence of multiple factors on the upper layer event, and finally on the top event. The fuzzy comprehensive assessment method is also a widely used risk assessment method, which usually uses weights to reflect the impact of multiple factors on risk. However, to use these methods, we need a prior-knowledge of the factors that influence the risk.

Analysis of historical accidents is also a way to know which factors are likely to cause accidents. For example, Nivolianitou et al. (2006) performed a statistical analysis of some of the characteristics of major accidents in the petrochemical industry included in the European major accident reporting system (MARS), and found that human factors and equipment failures are the main causes of major accidents, followed by natural phenomena and random events. Boyd (2015) analyzed 376 accidents involving twin-piston engine aircrafts from 2002 to 2012, and found that malfunctions, poor instrument approach procedures, and failure to maintain obstacle/terrain clearance at night were the most common causes of fatal accidents. Night operations, off-airport landings and post-impact fires all present additional risks of fatal flights. Ren et al. (2019) analyzed tunnel fire accidents (TFAs) which occurred in China from 2000 to 2016, and revealed that vehicle technical problems and vehicle traffic accidents are the main causes of tunnel fire accidents in China. In the study of Chu et al. (2020), they established the accident information statistics template (AIST) using key indicators extracted from accident investigation reports, four factors including season, time of day, weather and terrain were considered key external causes. Xia et al. (2021) conducted a statistical analysis of 120 fatal accidents of confined space operations in China between 2008 and 2018, and summarized that inadequate safety culture, inadequate supervision, violation of supervision, organizational process vulnerability, decision error, and violation of operation were the main causes of accidents.

Although these statistical analyses of historical accidents revealed factors that led to accidents, they didn't describe the associations between the factors. However, the "association rule" mining method can be used to discover the existing connections between factors from a large amount of data. Association rule mining can help us discover the correlation between accident factors and reveal the potential rules of accidents, thus providing a basis for safety management and accident prevention. Association rules reflect the interdependence between multiple objects. If there is a certain relationship between objects, one of them can be predicted by other objects (Savasere et al., 1995). Market Basket Analysis is the most typical application of association rules. It can grasp the purchasing habits of customers by analyzing the different commodities that customers put into their shopping baskets. At present, association rules have been applied in many fields, such as marketing, e-commerce, case analysis, risk management, etc. (Kavsek et al., 2008). There are also some studies using association rule mining to analyze accident data, for example, Mirabadi and Sharifian (2010) applied association rule mining techniques to analyze the data of past railway accidents in Iran to reveal unknown relationships and patterns among the data. Hou et al. (2020) employed the association rule mining approach to analyze tank farm accidents to discover rules of the most likely causal sequences inducing domino accidents. Wang et al. (2016) studied correlated industrial alarm sequences using an association rule mining approach combined with fuzzy sets. Jiang et al. (2020) proposed an association rule mining based framework to determine key factors associated with motorcycle injury severity. Yu et al. (2020) developed a Functional Resonance Analysis Model (FRAM) based hybrid simulator to aid hazard analysis in the process industries, and association rule mining was used for the interaction analysis of the simulated data. Ozaydin et al. (2022) used Bayesian network (BN) and association rule mining methods to analyze the data of unreported occupational accidents for fishing vessels in Turkey.

The contributions of this study are twofold. First, by mining the association rules, we obtain many interesting rules to investigate the hidden relationships among contributory factors to tank farm accidents. Second, with complex network analysis, we not only interpret these association rules, but also reveal which factors are more important for the tank farm safety. So far, in the accident factor analysis of petrochemical storage tank farms, the association rule mining method is rarely used to study the correlation of factors. This study utilizes association rule mining techniques to identify the set of accident contributing factors that frequently occur together in petrochemical storage tank accidents. On this basis, the obtained association rules are combined with the complex network analysis method to find out the important or key factors that influence the occurrence of accidents. Complex network is a method that has emerged in the last 20 years, and it has analysis methods to reveal the relationship between the internal factors of a complex system. In recent years, the complex network method has also been applied in the field of industrial safety, e.g., Li and Wang (2018) proposed a complex network-based railway system risk monitoring model to quantify the risk of accident causes, whereby the complex network model was used to identify accident factors and analyze how they affect each other. Ma et al. (2022) extracted the chain of events involved in 39 reports of ship grounding accidents based on event tree analysis, then developed a directed complex network, and analyzed the network through criticality evaluation and sensitivity analysis.

By identifying contributing factors and their associations, the approach presented in this study can provide useful insights into understanding the reasons behind the occurrence of tank accidents and developing effective safety policies and countermeasures.

The rest of this article is organized as follows: Section 2 presents the methodology used in tis study, including the introduction of the association rule mining method and the complex network method; the analysis results are presented in Section 3; finally, some conclusions of this study are drawn in Section 4.

2. Methodology

The occurrence of an accident often involves multiple factors. Analysis of the appearance of different factors leading to an accident can reveal their correlation. In this study, the association rule mining method is therefore used to obtain the relationship between accident factors. Association rules are an important subject of data mining, which are used to mine the correlation between valuable data items from a large amount of data. The task of association rules is to reduce a potentially large amount of disordered data into a small amount of static regular data that is easy to observe and understand.

The main process of this study includes 4 steps as shown in Fig. 1.

Step1 Data collection. Association rule mining is based on data analysis, thus a number of data are required for the analysis. This study is to analyze the correlation between factors for tank farm accidents, cases of tank accidents are to be collected.

Step2 Accident factors extraction. From the collected accident cases, contributing factors should be extracted for further analysis. The occurrence of an accident usually involves multiple factors, including human factors, equipment factors, management factors, etc. They work together to cause an accident, and it also shows that they are correlated in the occurrence of the accident.

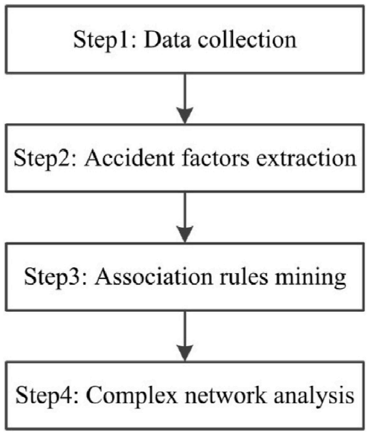

Fig. 1. Flowchart of the study.

Step3 Association rules mining. Use appropriate association rule mining algorithm to obtain association rules of the accident factors.

Step4 Complex network analysis. Based on the association relations of accident factors, using complex networks can not only directly display these association rules in a graphical way, but also analyze which factors in the accident factors play a more important role in the association network.

2.1. Data collection

An accident investigation report usually provides detailed causes of the corresponding accident, which facilitates the analysis of the accident factors. Thus, this study is based on investigation reports of petrochemical storage tank farm accidents. Most investigation reports are collected from the official websites of the Ministry of Emergency Management and some local work safety administrations of China, and a small amount of petrochemical storage tank farm accident reports are collected from other safety management websites. In the end 212 accidents are identified as the research data in this work.

Although there is no writing standard or normalized approach for accident investigation reports, they contain some basic parts, such as the time and location of an accident, on-site investigation records, the accident process, the direct causes, the indirect causes and the consequence caused by the accident. Table 1 demonstrates the extracted parts from an accident report.

2.2. Accident factors extraction

There are no standards for the writing of accident investigation reports, which also leads to different descriptions of the same or similar factors in different accident investigation reports. The accident factors are extracted based on the two main steps: first, extracting direct and indirect causes of an accident from its investigation report, and second, extracting factors from the direct and indirect causes. The first step can be easily implemented from the report. The second step is somewhat more complicated. In different reports, the causes of the accidents are described differently, so it is necessary to standardize the accident factors, and combine similar factors into one factor. For example, in different reports, operation error may be mistakenly opening a valve, opening a wrong valve, failure to start a device at the required time, etc. In the study, they are all represented as “operation error”. Therefore, this study analyzes the accident factors mentioned in a report from four categories: human factors, equipment factors, environmental factors, and management factors, where, human factors and equipment factors are usually the direct causes of accidents, and environment factors and management factors are the indirect causes of accidents.

Accident factors involved in the 212 accident reports are summarized into the categories. Table 2 lists the extracted factors from the report shown in Table 1.

Human factors mainly include psychological factors, physiological factors, professional quality of employees and so on. The human factors involved in the 212 accident reports include: violation of operating procedures, operators are not certified, illegal command, hazards are not identified, improper emergency response, weak safety awareness, operation error, lack of experience and professional skills, and so on. Among the human factors, “violation of operating procedures” occurs most frequently, accounting for 39% of the total number of accidents.

Equipment factors mainly include faults and damages during the operation of equipment, equipment defects, etc. The equipment factors involved in the 212 accident reports include: rupture damage of tank or pipeline, equipment failure, a safety device is missing or damaged, a runaway reaction, defects in equipment design or quality, and so on. Among the equipment factors, “equipment failure” and “defects in equipment design or quality” appear most frequently in the studied accidents, whereby they each account for 18% of the total number of accidents.

Environment factors can be divided into three aspects: natural environment factors, operation environment factors and system environment factors. Natural environment factors mainly include lightning, hurricane, etc.; operation environment factors mainly include humidity, static electricity, etc.; system environment factors mainly include the flammable gas concentration of the environment, the distance between storage tanks, etc. Among the environmental factors, “high concentration of flammable gas in the air” occurs most frequently, and this factor exists in 61% of the accidents. If there is a high concentration of flammable gas in the air, generally it is caused by a leak of a vessel. This factor is usually by itself caused by other factors. Another environmental factor that occurs frequently is “spark of static electricity”, which exist in 29% of the accident cases.

Management factors mainly include omissions and mistakes in the management process, involving personnel training, setup of safety management departments, equipment management, etc. Among the 212 accident reports, management factors mainly include: work plan is not good enough, inadequate safety management of outsourcing, inadequate special operation management, lack of safety education and training regulations, lack of safety supervision and inspection, safety management is not implemented, potential hazards are not eliminated in time, equipment is not regularly maintained and repaired, poor operating procedure, and so on. Among management factors, “lack of safety education and training regulations” is the most frequent factor, existing in the 26% of the accidents.

2.3. Association rules mining

Compared with traditional statistical analysis methods, association

Table 1
Extracted contents from an accident report.

Report number	Rpt_1
Accident time	November 2, 2020 11:45
Accident location	A LNG limited liability company
Consequence	Seven people died and two people were seriously injured, with a direct economic loss of 20.29 million Yuan
Direct causes	During the cutting of the outlet pipe of a low-pressure pump, the LNG in the low-pressure outgoing transmission pipe leaked; the mixture of the vaporized gas of leaked LNG and air was combusted when encountering an open flame.
Indirect causes	Improper isolation of valves; failure of instrumentation engineers to implement operating procedures; insufficient confirmation of working conditions for fire operation; inadequate awareness of risk and risk control is not in place; inadequate management of contractors.

Table 2
Extracted factors from Rpt_1.

Category	Meaning
Human factors	Violation of operating procedures
	Weak safety awareness
Equipment factors	Safety device is missing or damaged
Environmental factors	High concentration of flammable gas in the air
	Open flame
Management factors	Inadequate safety management of outsourcing
	Inadequate special operation management

rule mining analysis does not need to specify dependent and independent variables in advance. Such advantages can help people discover valuable relationships in data. Using different rule generation criteria, there are different types of association rule analysis algorithms, such as the generalized rule induction algorithm and the Apriori algorithm. The Apriori algorithm is the commonly used association rule mining algorithm more widely used than other methods. This study also uses the Apriori algorithm to analyze the accident data of storage tank area. The Apriori algorithm uses an iterative method of layer-by-layer search, including two steps: the first step is to iteratively search the frequent itemsets in the database by scanning the database. Then, in the second step, strong association rules are generated from the frequent itemsets.

$I = \{i_1, i_2, ..., i_m\}$ is a set of items, Transaction T is a non-empty subset of I, that is, $T = \{i_j, \ldots, i_k\} \subseteq I$, $1 \leq j, k \leq m$. All transactions form database D. An association rule is an implication of the form X→Y, where X and Y are two itemsets such that X∩Y = ∅. X is called the left hand side (LHS) and Y is called the right hand side (RHS) (Agrawal and Srikant, 1994). Three important indicators are widely used to evaluate discovered association rules: "support", "confidence", and "lift" (Montella, 2011; Xu et al., 2018).

The "support" of an association rule refers to the ratio of the transactions containing both itemset A and itemset C to the total number of transactions in the database D, as shown in Eq. (1). Support can be understood as the frequency of occurrence of the association rule.

$$Support(A \rightarrow C) = P(A \cap C) = \frac{|A \cup C|}{|D|} \tag{1}$$

where, |D| represents the number of transactions in D, $|A \cup C|$ represents the number of transactions that contain both itemsets A and C.

The support of a rule is symmetrical. Therefore, the support of the rule A→C and that of rule C→A are equivalent.

The "confidence" of an association rule refers to the ratio of the transactions that contain itemsets A and C at the same time to the number of transactions that contain itemset A, as shown in Eq. (2). Confidence reflects the probability of occurrence of the RHS item C given that the LHS item A occurs.

$$Confidence(A \rightarrow C) = \frac{Support(A \rightarrow C)}{Support(A)} = \frac{|A \cup C|}{|A|} \tag{2}$$

where, $|A|$ means the number of occurrences of only itemset A.

The "lift" A→C reflects the degree of association between itemset A and itemset C. If the lift is greater than 1, it indicates that the positive correlation between itemsets A and C, and the greater the lift, the greater the positive correlation; if the lift is less than 1, it indicates that the negative correlation between itemsets A and C, and the smaller the lift, the greater the negative correlation If the lift value equals 1, it means that itemset A has no correlation with itemset C. The calculation of the lift is shown in Eq. (3).

$$Lift(A \rightarrow C) = \frac{Support(A \rightarrow C)}{Support(A) \times Support(C)} \tag{3}$$

2.4. Complex network analysis

Association rules reveal the associations between accident factors, and the associations between many factors form an association network. Further, the complex network can be used to analyze this factor association network to reveal the importance of each factor in the formation of the accident.

Complex networks consist of nodes and edges. Therefore, using a complex network to construct an accident causal network in a petrochemical storage tank farm, the key is to clarify the nodes and edges in the network, so that the relationship between each node can be reasonably expressed. The construction steps of the accident-causing network in the petrochemical tank farm based on association rules are mainly divided into the following three steps:

(1) Determine the set of association rules. Generally, the association rules that the user is interested in are selected by setting the minimum support and minimum confidence of the association rules.
(2) Build the adjacency matrix corresponding to the association rules. According to the LHS items and RHS items constituting the association rules, a relationship table between items can be established, and the relationship a_{ij} between items i and j is obtained according to the relationship Table. 1f there is a relationship between items i and j, then $a_{ij} = 1$, if there is no relationship between them, then $a_{ij} = 0$.
(3) Draw the network according to the adjacency matrix, where nodes correspond to items and edges correspond to relationships.

Fig. 2 illustrates the process of establishing the complex network of accident factors.

After the complex network of accident factors is built, the factors can be analyzed by using the network topology characteristic analysis methods of the complex network, mainly including the following indicators:

(i) Degree of nodes

The degree of a node refers to the number of nodes in the network that are directly connected to the node. The degree of a node can reflect the importance of the node in the network to a certain extent. The greater the degree of a node, the more nodes it can affect, indicating that the node is more important.

(ii) Betweenness. Betweenness of a node refers to the number of shortest paths going through the node in the network. It implies the importance of a node for the transfer of information or matter. The greater the betweenness of a node, the stronger its ability to influence risk transfer, thus the more critical its position in the network, and the more important it needs to be focused on.
(iv) Clustering coefficient. The clustering coefficient is an attribute describing the degree of clustering of nodes in a network, which indicates the possibility that there is also an association between nodes connecting to a certain node. The clustering coefficient of a single node is the ratio of the number of connections between all its adjacent nodes to the maximum possible number of connections.
(v) Network diameter and average shortest path length. For any node i and node j in a network, the number of edges in the shortest path from node i to node j is the distance between them, and the maximum distance between all node pairs in the network is the network diameter. The average shortest path length refers to the average distance between all pairs of nodes in the network, which reflects the closeness of the connections between the nodes in the network.

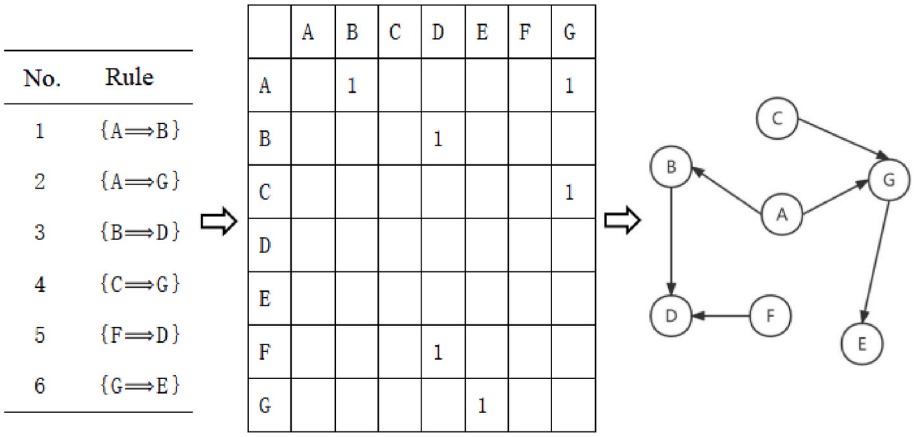

Fig. 2. Construction process of the complex network of accident factors.

3. Analysis results

3.1. Association rules mining

The Apriori algorithm was used to extract the association rules, and some rules whose lift value is less than 1 were eliminated, and finally 75 association rules were obtained.

Table 3 lists the top 5 association rules in terms of the value of support, and the support of these 5 association rules ranges from 0.160 to 0.292. The great support of the association rule indicates that the combination of the LHS factor and the RHS factor of the rule frequently occurs in accidents. For example, “operator violates the operating procedures” and “flammable gas concentration is high”, “electrostatic sparks” and “flammable gas concentration is high”, “open flame” and “flammable gas concentration is high”, etc. These factors often appear at the same time in accidents, indicating that their co-existence is likely to cause petrochemical storage tank system accidents.

Table 4 lists the top 5 association rules with confidence. Their confidence ranges from 0.773 to 0.960. Association rules with high confidence indicate that the LHS factor of the rule occurs, and the RHS factor of the rule also appears or occurs with a high possibility. Among them, Rule 1 indicates that 11.3% of the accidents involve both “bad weather” and “high concentration of flammable gas in the air”, and in accidents caused by “bad weather”, there is a 96% chance that the factor “high concentration of flammable gas in the air” appears. Similarly, Rule 2 indicates that 9.9% of the accidents involve “inadequate safety management of outsourcing” and “high concentrations of flammable gas”, and 87.5% of the accidents caused by “inadequate safety management of outsourcing” involves “high concentration of flammable gas in the air”. The correlation between accident factors is extracted through historical data. When the LHS factor occurs, special precautions should be taken to prevent the occurrence of the RHS factor, thereby reducing the possibility of accidents.

Table 3
Top 5 association rules with support.

No.	LHS	RHS	Support	Confidence	Lift
1	{ Violation of operating procedures }	{ High concentration of flammable gas in the air }	0.292	0.756	1.233
2	{ Spark of static electricity }	{ High concentration of flammable gas in the air }	0.288	0.469	1.631
3	{ Open flame }	{ High concentration of flammable gas in the air }	0.245	0.400	1.631
4	{ Lack of safety supervision and inspection }	{ High concentration of flammable gas in the air }	0.184	0.829	1.353
5	{ Safety management is not implemented }	{ Violation of operating procedures }	0.160	0.708	1.831

Table 4
Top 5 association rules with confidence.

No.	LHS	RHS	Support	Confidence	Lift
1	{ Bad weather }	{ High concentration of flammable gas in the air }	0.113	0.960	1.566
2	{ Inadequate safety management of outsourcing }	{ High concentration of flammable gas in the air }	0.099	0.875	1.427
3	{ Lack of safety supervision and inspection }	{ High concentration of flammable gas in the air }	0.184	0.829	1.353
4	{ Illegal command }	{ High concentration of flammable gas in the air }	0.113	0.800	1.305
5	{ Inadequate special operation management }	{ High concentration of flammable gas in the air }	0.160	0.773	1.260

3.2. Complex network analysis

(i) Accident factors network

Using the approach presented in Section 3.2, an association network of accident factors is established based on the obtained association rules, which is shown as Fig. 3. The meanings of the nodes are listed in Table 5.

The node degree can reflect the influence and status of the accident factor in a network. The degree of each node (factor) in the network of the petrochemical tank farm accident factors is shown in Fig. 4. It can be seen from the figure:

“Violation of operating procedures” (H01) has the greatest node degree, with a node degree of 14, followed by “High concentration of flammable gas in the air (En01)”, with a node degree of 12. The factor with greater node degree is more likely to interact with other accident factors, and is a more important accident factor in the network.

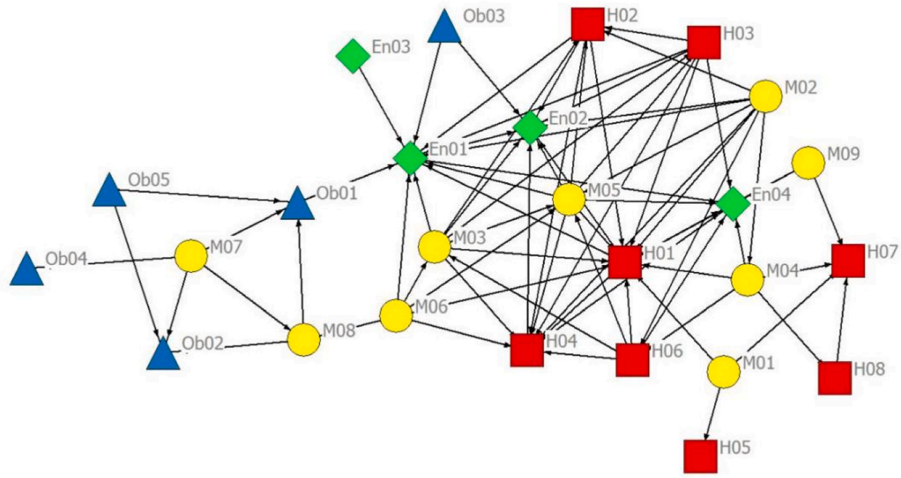

Fig. 3. Petrochemical tank farm accident cause complex network.

Table 5
Node meaning of the complex network of petrochemical storage tank farm accident factors.
(ii)Network topology characteristic analysis
(1) Node degree

Category	Node	Meaning
Human factors	H01	Violation of operating procedures
	H02	Operator is not certified
	H03	Illegal command
	H04	Hazards are not identified
	H05	Improper emergency response
	H06	Weak safety awareness
	H07	Operation error
	H08	Lack of experience and professional skills
Equipment factors	Ob01	Rupture damage of tank or pipeline
	Ob02	Equipment failure
	Ob03	Safety device is missing or damaged
	Ob04	Runaway reaction
	Ob05	Defects in equipment design or quality
Environmental factors	En01	High concentration of flammable gas in the air
	En02	Open flame
	En03	Bad weather
	En04	Spark of static electricity
Management factors	M01	Work plan is not good enough
	M02	Inadequate safety management of outsourcing
	M03	Inadequate special operation management
	M04	Lack of safety education and training regulations
	M05	Lack of safety supervision and inspection
	M06	Safety management is not implemented
	M07	Potential hazards are not eliminated in time
	M08	Equipment is not regularly maintained and repaired
	M09	Poor operating procedure

A large in-degree indicates that the accident factor is easily induced by other factors. The five risk factors with the largest in-degree are “violation of operating regulations (H01)”, “High concentration of flammable gas in the air (En01)”, “Hazards are not identified (H04)”, “open flame (En02)”, “Spark of static electricity (En04)”. It can be seen that the accident factors with larger in-degree are mostly human factors and environmental factors. These two types of accident factors are more easily affected by other factors, which directly lead to the occurrence of accidents. A large out-degree indicates that the accident factor is likely to induce other accident factors. The five factors with the largest out-degree are “Inadequate safety management of outsourcing (M02)”, “Inadequate special operation management (M03)”, “Lack of safety supervision and inspection (M05)”, “Safety management is not implemented (M06)”, and “illegal command (H03)”. Most or all of the accident factors with larger out-degree are management factors, which can reflect that management errors are the root cause of accidents, and management errors will induce other factors to cause accidents.

The betweenness reflects the degree of control of accident factors on hazard transmission in the network, and the factor with a larger betweenness plays a more critical role in the transmission of risk in the network. Fig. 5 shows the betweenness of each accident factor in the network. The five factors with the largest betweenness are “High concentration of flammable gas in the air (En01)”, “violation of operating regulations (H01)”, “rupture damage of tank or pipeline (Ob01)”, “weak safety awareness (H06)”, and “inadequate special operation management (M03)”. Accident factors with larger betweenness play a key role in the connection between various factors in the network. By avoiding the occurrence of these accident factors, the connection of the causal network can be cut off and the occurrence of accidents can be prevented.

The clustering coefficient reflects the aggregation degree of accident factors in the network. The larger the clustering coefficient, the stronger the connection between the various factors in the network. The clustering coefficients of each risk factor are shown in Fig. 6. The top 5 risk factors with the largest clustering coefficient are “lack of experience and professional skills (H08)”, “safety device is missing or damaged (Ob03)”, “Operator is not certified (H02)”, “illegal command (H03)”, and “Lack of experience and professional skills (H06)”. The factors with large clustering coefficients are mostly human factors. Compared with other types of factors, human factors have stronger aggregation characteristics in the network, and their neighbor factors are closely related. Once human unsafe behavior occurs, it may easily lead to its adjacent factors also occurring, thereby easily causing accidents.

The network diameter reflects the distance between two accident factors in the network, and the network diameter is 3 in the accident factor network of petrochemical storage tank farms, for example, from “Lack of safety education and training regulations (M04)" to “Illegal command (H03)", from “Defects in equipment design or quality (Ob05)" to “Open flame (En02)", from “Potential hazards are not eliminated in time (M07) to “Spark of static electricity (En04)”. There is an indirect connection between these factors, and through the transmission of factors, the former factor can have an impact on the latter.

The average shortest path length is the average of the distances between all accident factors, which reflects the strength of each risk factor’s ability to influence each other. The average shortest path length in the accident factor network of petrochemical storage tank farms is 1.442, indicating that one factor in the network can affect another factor in an average of only 1 to 2 steps. The average shortest path length is short, indicating the small average factor-to-factor distance, which is more likely to cause the spread of hazardous events and lead to

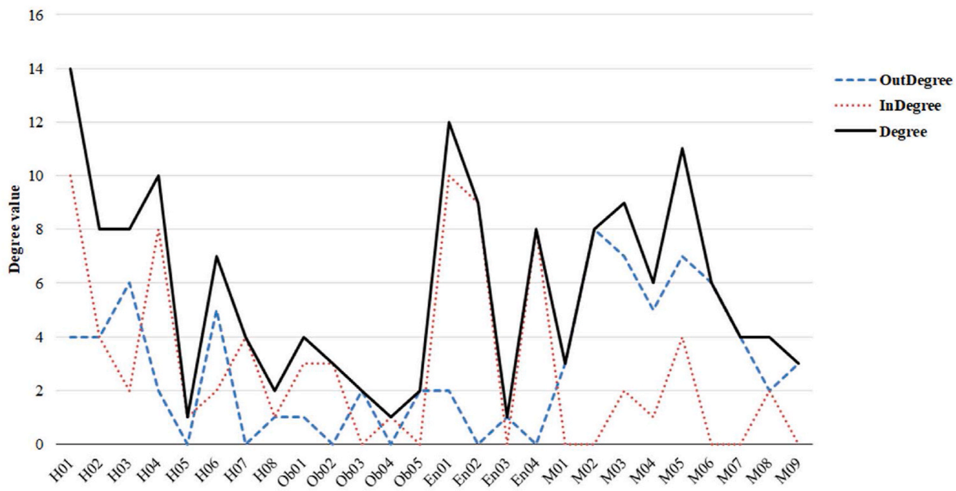

Fig. 4. The degree of each node in the causal network.

(2) Betweenness

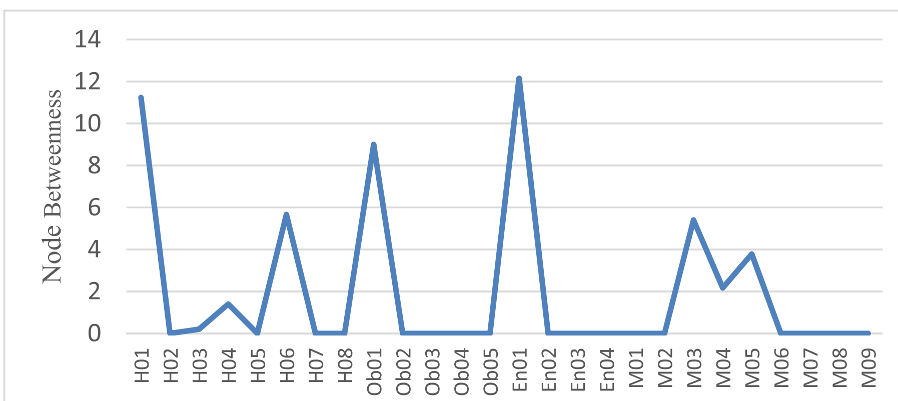

Fig. 5. The betweenness of each node in the causal network.

(3)Clustering coefficient

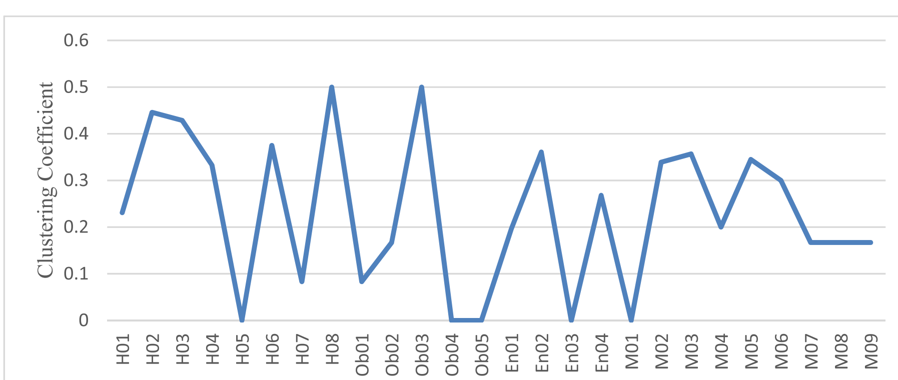

Fig. 6. The clustering coefficient of each node in the causal network.

(4)Network diameter and average shortest path length

accidents.

4. Conclusions

A large amount of dangerous substances are stored in the storage tank area of petrochemical enterprises, and may easily lead to great losses in the event of an accident. Understanding accident factors plays an important role in taking targeted measures to prevent accidents. Most accidents are caused by a combination of multiple factors. Therefore, this study uses the association rule mining method to analyze historical accidents to determine the main factors that cause accidents and their connections.

(1) Through the analysis of accident reports in petrochemical tank farms, a data set of accident factors of any petrochemical tank farm was obtained, and association rules were extracted from the factor data set based on the Apriori algorithm, from which rules with large support or large confidence are determined.
(2) Based on the analysis results of association rules, the accident cause network of a petrochemical storage tank area was constructed by using the complex network.The importance of the accident factors was analyzed by using the topological feature analysis methods of the network, e.g., "violation of operating regulations", "high concentration of flammable gas in the air", "lack of experience and professional skills", etc., are important factors. By avoiding the emergence of these factors the connection of the network can be effectively cut off and may thus prevent the occurrence of accidents.

Using the approach presented in this study in the safety management of tank farms, the factors that may associate together to cause accidents can be extracted, and the important factors can be identified. This can help to improve safety management and prevent accidents with specific aims. This study is mainly based on the analysis of accident investigation reports of the safety departments of petrochemical storage tank farms. Such accidents are usually characterized by low frequency but serious consequences, so the number of accidents collected in this study was 212. Although the main association factors can be obtained, as a method based on data analysis, more data can reveal more relationships and provide more comprehensive guidance for safety management. To better reflect the accident factors and their associations in the future, more accident cases need to be accumulated in future research.

Declaration of competing interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by National Natural Science Foundation of China (No. 71673060).

References

Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules. In: IEEE Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487–499.
Argyropoulos, C.D., Christolis, M.N., Nivolianitou, Z., Markatos, N.C., 2012. A hazards assessment methodology for large liquid hydrocarbon fuel tanks. J. Loss Prev. Process. Ind. 25, 329–335.
Boyd, D.D., 2015. Causes and risk factors for fatal accidents in non-commercial twin engine piston general aviation aircraft. Accid. Anal. Prev. 77, 113–119.
Buncefield Major Incident Investigation Board, 2008. The Buncefield Incident 11 December 2005: the Final Report of the Major Incident Investigation Board Volume 1. Available at: https://www.icheme.org/media/13707/buncefield-miib-final-report-volume-1.pdf.
Chu, Z., You, Z., Wang, F., Chen, X., Zhang, B., 2020. A targeted risk prediction method based on statistical analysis of accidents in petroleum geophysical exploration. J. Petrol. Sci. Eng. 192, 107314.
Guo, X., Ji, J., Khan, F., Ding, L., Yang, Y., 2021. Fuzzy Bayesian network based on an improved similarity aggregation method for risk assessment of storage tank accident. Process Saf. Environ. Protect. 149, 817–830.
Hou, L., Wu, X., Wu, Z., Wu, S., 2020. Pattern identification and risk prediction of domino effect based on data mining methods for accidents occurred in the tank farm. Reliab. Eng. Syst. Saf. 193, 106646.
Ikwan, F., Sanders, D., Hassan, M., 2021. Safety evaluation of leak in a storage tank using fault tree analysis and risk matrix analysis. J. Loss Prev. Process. Ind. 73, 104597.
Jiang, F., Yuen, K.K.R., Lee, E.W.M., 2020. Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology. J. Saf. Res. 75, 292–309.
Kavsek, B., Lavrac, N., Jovanoski, V., 2008. APRIORI-SD: adapting association rule learning to subgroup discovery. In: International Symposium on Intelligent Data Analysis. Springer Berlin Heidelberg.
Leveson, N., 2004. A new accident model for engineering safer systems. Saf. Sci. 42, 237–270.
Li, K., Wang, S., 2018. A network accident causation model for monitoring railway safety. Saf. Sci. 109, 398–402.
Luo, T., Wu, C., Duan, L., 2018. Fishbone diagram and risk matrix analysis method and its application in safety assessment of natural gas spherical tank. J. Clean. Prod. 174, 296–304.
Ma, X., Deng, W., Qiao, W., Lan, H., 2022. A methodology to quantify the risk propagation of hazardous events for ship grounding accidents based on directed CN. Reliab. Eng. Syst. Saf. 221, 108334.
Mirabadi, A., Sharifian, S., 2010. Application of association rules in Iranian Railways (RAI) accident data analysis. Saf. Sci. 48, 1427–1435.
Montella, A., 2011. Identifying crash contributory factors at urban roundabouts and using association rules to explore their relationships to different crash types. Accid. Anal. Prev. 43 (4), 1451–1463.
Nivolianitou, Z., Konstandinidou, M., Michalis, C., 2006. Statistical analysis of major accidents in petrochemical industry notified to the major accident reporting system (MARS). J. Hazard Mater. A137, 1–7.
Ozaydin, E., Fiskin, R., Ugurlu, O., Wang, J., 2022. A hybrid model for marine accident analysis based on Bayesian Network (BN) and Association Rule Mining (ARM). Ocean Eng. 247, 110705.
Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27 (2), 183–213.
Ren, R., Zhou, H., Hu, Z., He, S., Wang, X., 2019. Statistical analysis of fire accidents in Chinese highway tunnels 2000-2016. Tunn. Undergr. Space Technol. 83, 452–460.
Savasere, A., Omiecinski, E., Navathe, S.B., 1995. An efficient algorithm for mining association rules in large databases. In: International Conference on Very Large Data Bases Morgan. Kaufmann Publishers Inc.
U.S. Chemical Safety and Hazard Investigation Board, 2015. Final investigation report: caribbean petroleum tank terminal explosion and multiple tank fires. Available at: https://www.csb.gov.
Wang, J., Li, H., Huang, J., Su, C., 2016. Association rules mining based analysis of consequential alarm sequences in chemical processes. J. Loss Prev. Process. Ind. 41, 178–185.
Wu, D., Chen, Z., 2016. Quantitative risk assessment of fire accidents of large-scale oil tanks triggered by lightning. Eng. Fail. Anal. 63, 172–181.
Xia, J., Liu, Y., Zhao, D., Tian, Y., Li, J., Zhong, Y., Roy, N., 2021. Human factors analysis of China's confined space operation accidents from 2008 to 2018. J. Loss Prev. Process. Ind. 71, 104480.
Xu, C., Bao, J., Wang, C., Liu, P., 2018. Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. J. Saf. Res. 67, 65–75.
Yu, M., Quddus, N., Kravaris, C., Mannan, M.S., 2020. Development of a FRAM-based framework to identify hazards in a complex system. J. Loss Prev. Process. Ind. 63, 103994.