applied sciences

MDPI

Article

Association Diffusion and Critical Causal Factors in Ship Self-Sinking Accidents: A Hybrid HFACS–Association Rule Mining–Complex Network Approach

Yuqing Ren [1,†], Yucheng Chen [1,†], Lili Zhou [1,2] and Yingbang Huang [2,3,*]

[1] School of Navigation and Naval Architecture, Dalian Ocean University, Dalian 116023, China; renyuqing@dlou.edu.cn (Y.R.); 15841259023@163.com (Y.C.); zhou01212026@163.com (L.Z.)

[2] South China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou 510300, China

[3] College of Ocean and Meteorology, Guangdong Ocean University, Zhanjiang 524088, China

* Correspondence: hkbone99@163.com

† These authors contributed equally to this work.

Abstract

Ship self-sinking accidents threaten maritime safety, human life, property, and the marine environment, and understanding their causal-factor associations is essential for developing effective preventive measures. This study aims to identify the multi-level factors, recurrent association patterns, and critical structural nodes involved in ship self-sinking accidents. A hybrid framework integrating grounded theory, the Human Factors Analysis and Classification System (HFACS), FP-growth association rule mining, and complex network analysis was applied to 150 accident investigation reports released by the China Maritime Safety Administration between 2014 and 2024. Findings suggest that adverse weather and sea conditions, inadequate ship safety management, and crew incompetence are the most frequent factors. Thirty causal factors were identified and classified into four HFACS levels, and 229 association rules were generated to construct a directed weighted causal-factor association network with 19 nodes and 229 edges. Network results indicate that inadequate ship safety management, crew incompetence, ship unseaworthiness, insufficient maintenance of hull weathertight integrity, and improper or untimely emergency measures occupy critical positions in the association structure. This research offers insight into ship self-sinking accidents and identifies priority intervention points for more targeted maritime supervision, safety management and accident prevention.

Keywords: ship self-sinking accident; HFACS; causal coupling; grounded theory; association rule mining; complex network analysis; maritime safety

Academic Editors: Andrea Maternová, Andrej Dávid, Nermin Hasanspahic and Lucia Švábová

Received: 12 May 2026
Revised: 18 June 2026
Accepted: 20 June 2026
Published: 23 June 2026

1. Introduction

A ship self-sinking accident refers to a sinking event that occurs without direct external impacts such as collision, grounding, or stranding, but instead results from the combined effects of water ingress, insufficient stability, improper cargo stowage, equipment failure, human error, and environmental disturbances. Compared with collisions and groundings, ship self-sinking accidents are usually more concealed and progressive. They often evolve through a continuous process of risk accumulation, deterioration of ship conditions, failure of safety barriers, and delayed or ineffective emergency responses. Once these latent risks are not identified and controlled in time, they may rapidly escalate into severe casualties, property losses, and environmental consequences. Therefore, systematically identifying the causal factors of ship self-sinking accidents and revealing their coupling relationships and association mechanisms are essential for improving maritime safety governance.

In recent years, scholars both domestically and internationally have conducted extensive research on the causes of maritime accidents. Existing studies indicate that maritime accidents are typically not triggered by a single factor, but rather result from the interaction of multiple factors, including unsafe human behavior, organizational management deficiencies, abnormal technical conditions of the ship, and external environmental disturbances. Relevant research has gradually shifted from traditional empirical induction toward systematic, structured, and data-driven analysis. For example, Qiao et al. [1] analyzed human factors in maritime accidents using a dynamic fuzzy Bayesian network, effectively addressing the uncertainty, dynamism, and hierarchical dependencies inherent in accident causation; Zohorsky et al. [2] applied the HFACS framework to the analysis of commercial fishing ship accidents, demonstrating that HFACS possesses good adaptability and scalability in maritime scenarios; Youssef et al. [3] further explored the issue of accident causation identification from the perspective of maritime human factors analysis models, emphasizing the systematic transmission process from organizational management and operational environments to frontline operational errors. Meanwhile, Xu et al. [4] conducted a systematic review of maritime transport safety management research, noting that maritime safety research is evolving toward more refined risk identification, scientific management decision-making, and integrated methodologies; Maternová et al. [5] further validated the dominant role of human factors in the formation of maritime accidents by examining the relationship between human error and casualties. The aforementioned studies indicate that the analysis of maritime accident causes has gradually shifted from empirical judgment toward a systematic research paradigm characterized by multi-level and multi-method integration.

As research continues to deepen, the application of complex network analysis, data mining, and probabilistic modeling methods in maritime accident analysis is increasing. Some studies have utilized complex network methods to analyze the association diffusion characteristics of major maritime accidents in China [6], finding that accident factors do not exist in isolation but exhibit distinct interconnections and chain-like transmission patterns. Other scholars have employed data mining techniques to identify correlations among maritime accident causes [7], demonstrating that combinations of high-frequency factors are more effective than single factors in revealing the mechanisms underlying accident occurrence. Regarding shipwreck or sinking accidents, some studies have proposed quantitative analysis frameworks from the perspectives of human reliability and safety barriers [8,9], providing new pathways for identifying critical failure points and evaluating risk prevention and control capabilities. Other studies have utilized machine learning methods to identify factors influencing the consequences of waterborne traffic accidents [10], further expanding the technical tools for accident severity prediction and risk warning analysis. Overall, existing research has not only enriched the methods for identifying the causes of maritime accidents but has also laid the foundation for subsequent work on accident factor correlation mining, network modeling, and association-based diffusion analysis.

Building on this, Liu et al. [11] used machine learning methods to reveal the nonlinear and complex interactive characteristics of the causes of maritime accidents along China's coastline; Deng et al. [12] analyzed the coupling characteristics of maritime accident risks based on the N-K model, demonstrating that the accident formation process exhibits strong systemic interdependence and interactive effects. Fan et al. [13] and Fu et al. [14] demonstrated, from the perspectives of Bayesian networks and quantitative frameworks, respectively, that maritime accidents are often the result of long-term coupling among multi-level factors. Ma et al. [15–17] conducted a series of studies on human factor modeling, contribution identification, and complex network construction in maritime accidents, systematically revealing the differences in the roles of human factors across different accident types and their network structural characteristics; Wang et al. [18] further analyzed the hierarchical transmission relationships between human and organizational factors in collision accidents. Subsequently, Liu et al. [19], Shi et al. [20], and Zhang et al. [21] identified key causal factors in collision accidents from a complex network perspective, while Cao et al. [22], Su et al. [23], and Chen et al. [24,25] expanded the scope to include network topology, knowledge graphs, and risk factor extraction, thereby shifting accident causation research from “factor identification” toward “relationship analysis” and “structural understanding.” Guo et al. [26], Yin et al. [27], Li et al. [28], and Zhou et al. [29] have advanced the development of maritime accident research from the perspectives of spatio-temporal characteristics, association chain identification, global data-driven Bayesian networks, and risk model construction, respectively, further illustrating that maritime accident research is shifting toward multi-source data fusion, cross-method coupling, and global risk cognition.

Despite these advances, several research gaps remain. First, most existing studies have focused on general maritime accidents or typical accident types such as collisions, groundings, and container ship accidents, whereas ship self-sinking accidents have received limited attention. Such accidents have distinctive characteristics, including concealed development, progressive water ingress, deterioration of hull watertight integrity, and delayed emergency response. Second, previous HFACS-based studies have mainly emphasized hierarchical classification of human and organizational factors, but paid less attention to recurrent cross-level factor combinations. Third, association rule mining can reveal statistical co-occurrence and implication patterns but cannot independently explain the hierarchical safety logic behind these factors. Fourth, complex network analysis can identify key nodes and structural associations but existing network-based studies have rarely examined the specific association structure of ship self-sinking accidents. Therefore, the interactions among organizational deficiencies, unsafe supervision, ship technical conditions, environmental disturbances, and emergency response failures in ship self-sinking accidents remain insufficiently clarified.

To address these gaps, this study uses ship self-sinking accident investigation reports from 2014 to 2024 as samples and develops a hybrid analytical framework integrating grounded theory, HFACS, FP-growth association rule mining, and complex network analysis. First, causal factors are extracted from accident investigation texts through grounded-theory coding and organized into a hierarchical causation system based on HFACS. Second, the FP-growth algorithm identifies recurrent association patterns among causal factors. Third, the extracted association rules are transformed into a directed weighted causal-factor association network to examine the association structure, key nodes, and potential association diffusion characteristics of ship self-sinking accidents. This hybrid framework differs from previous single-method studies by combining factor extraction, hierarchical classification, statistical association mining, and network structural analysis within one analytical process. Association rules reveal statistically supported co-occurrence and implication relationships rather than strict physical causality. Therefore, the constructed network is interpreted as a causal-factor association network rather than a deterministic association network. The findings provide theoretical support and practical implications for targeted accident prevention, maritime safety supervision, ship inspection, crew training, maintenance prioritization, and emergency decision-making.

2. Materials and Methods

As shown in Figure 1, this study develops a hybrid framework that integrates qualitative coding, hierarchical classification, association discovery, and network topology analysis. The research process consists of three stages. First, a database of ship self-sinking accidents is established, and accident investigation reports are coded using grounded theory to extract causal factors. These factors are then classified according to the HFACS framework to construct a hierarchical causation system. Second, the structured causal-factor data are converted into a transaction dataset, with each accident treated as one transaction. The FP-growth algorithm is applied to mine association rules among causal factors. Third, the valid association rules are mapped onto a directed weighted causal-factor association network in which causal factors are represented as nodes, association rules are represented as directed edges, and rule confidence is used as the edge weight. Network indicators, including node degree, average path length, clustering coefficient, betweenness centrality, PageRank, and robustness, are then used to identify key nodes and association diffusion structures.

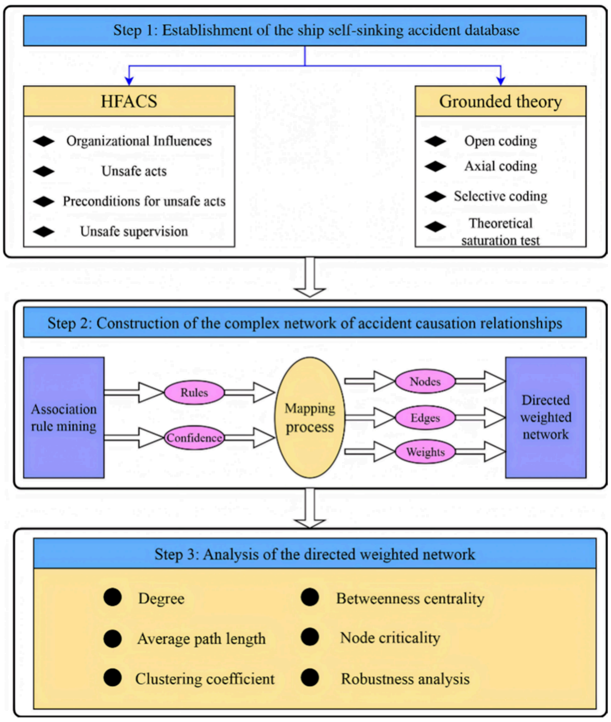

Figure 1. Overview of the methodology.

The rationale for combining these four methods is that each method addresses a different analytical limitation. Grounded theory enables causal factors to be inductively extracted from unstructured investigation reports, while HFACS provides a standardized hierarchical structure for organizing these factors from unsafe acts to organizational influences. FP-growth association rule mining is then used to identify recurrent factor combinations that cannot be captured by frequency statistics alone. Finally, complex network analysis transforms these associations into a directed weighted structure, making it possible to examine node importance, structural connectivity, and potential propagation pathways. Therefore, the proposed framework is not intended to replace detailed physical accident reconstruction, but to provide a reproducible and data-driven way to identify statistically supported causal-factor associations in ship self-sinking accidents.

2.1. Data Collection

The research sample consisted of investigation reports on ship self-sinking accidents publicly released by the China Maritime Safety Administration from 2014 to 2024. An initial dataset of 154 reports was collected. To ensure data quality and analytical reliability, each report was cross-checked using key information, including accident number, occurrence date, location, ship name, and accident type. Reports were excluded if they were duplicate records, lacked key accident information, contained incomplete accident descriptions, or provided insufficient evidence for causal-factor identification. After screening, 150 accident reports were retained for analysis. These accidents resulted in 312 deaths or missing persons in total. The annual distribution of accidents and casualties is shown in Figure 2.

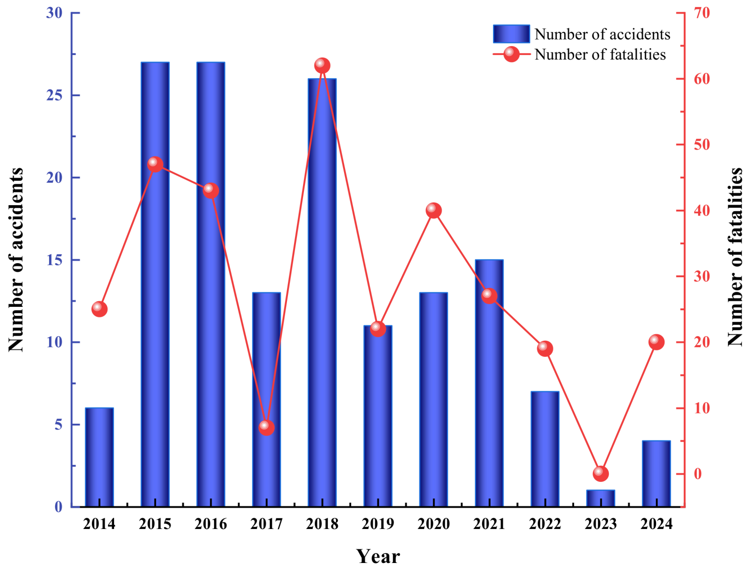

Figure 2. Annual distribution of ship self-sinking accidents and casualties from 2014 to 2024.

2.2. Grounded Theory and HFACS

Grounded theory is a qualitative research method that generates a theoretical framework from raw data through constant comparison and inductive analysis [30]. The research process typically involves the formulation of research questions, data collection, coding and analysis, preliminary theory development, and theoretical saturation testing [31]. As shown in Figure 3, this study addresses the causal factors of ship self-sinking accidents. Based on the collected accident investigation reports, open coding, axial coding, and selective coding were sequentially conducted to progressively extract causal concepts, develop causal categories, and identify core categories. The coding results were then repeatedly compared, revised, and refined through theoretical saturation testing until no new significant concepts or categories emerged, thereby enabling the construction of a causal model for ship self-sinking accidents [32]. To enhance the standardization and traceability of the coding process, NVivo 15 was employed for text management, annotation, and classification analysis. To improve reproducibility and reduce subjectivity, a coding protocol was developed before formal coding. The protocol defined the coding unit, inclusion and exclusion criteria, HFACS-level assignment rules, and examples of typical causal descriptions. During the reliability assessment, 30 accident reports, accounting for 20% of the full sample, were randomly selected and independently coded by three researchers with backgrounds in maritime safety, navigation, and accident analysis. For each report, the coders judged whether each of the 30 identified causal factors was present or absent, resulting in 900 binary coding decisions for each coder. Pairwise Cohen's kappa was then calculated for the three coder pairs. Disagreements were resolved through discussion with reference to the original investigation texts and the coding protocol. After the coding criteria were refined, the remaining reports were coded following the same protocol.

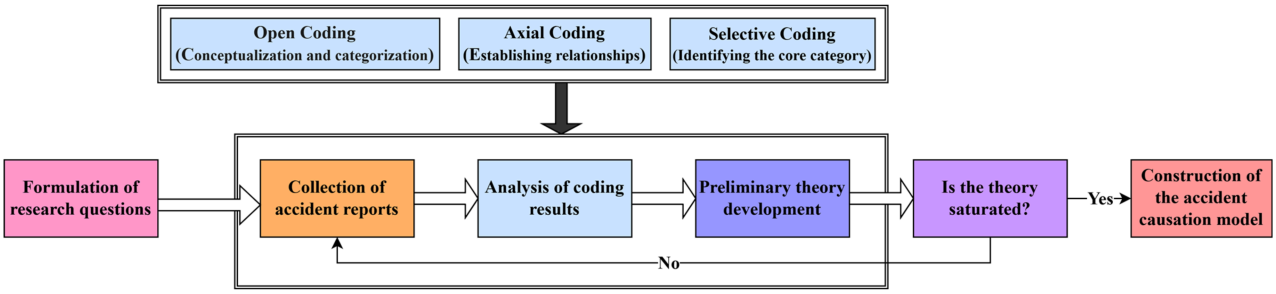

Figure 3. Grounded theory analysis process.

Human Factors Analysis and Classification System was originally developed for aviation safety research [33] and is particularly valuable for revealing, from a system perspective, the hierarchical relationships among human error, supervisory deficiencies, and organizational influences in accidents. With its continued development, HFACS has been progressively extended and widely applied to accident causation analysis in a variety of sectors, including maritime transport [34], coal mining [35], construction [36], and railways [37]. Compared with approaches that focus solely on direct errors, HFACS places greater emphasis on the hierarchical, systemic, and cascading characteristics of accident causation, making it especially suitable for the analysis of complex accidents. In maritime accident research, the classic HFACS framework is commonly adapted to the specific context by classifying accident causes into four interrelated levels: unsafe acts, preconditions for unsafe acts, unsafe supervision, and organizational influences. Unsafe acts mainly involve operational errors and improper emergency responses; preconditions for unsafe acts generally include factors related to personnel, equipment, cargo, and the environment; unsafe supervision primarily concerns deficiencies in navigation management and monitoring; and organizational influences are mainly reflected in ship safety management and cargo safety management. These four levels are not independent of one another, but rather constitute a causation structure characterized by top-down transmission and interaction across levels. Therefore, after extracting and coding the accident causal factors, this study further classifies and integrates these factors based on the HFACS framework, thereby constructing a hierarchical causation system for ship self-sinking accidents and laying the foundation for subsequent association rule mining and complex network analysis.

2.3. Association Rules

Association rule mining (ARM) is a data-driven data mining method that reveals latent associations in data by identifying frequent item sets [38]. Common algorithms include the FP-growth algorithm [39], the Apriori algorithm [40], and the Eclat algorithm [41], among others. The FP-growth algorithm only requires two scans of the database and does not require the generation of candidate sets, thereby significantly reducing computational complexity and improving mining efficiency. Therefore, this study selects the FP-growth algorithm to perform association analysis on the causal factors in ship self-sinking accident reports to support the subsequent construction of association networks.

In association rule mining, let $I = i_1, i_2, \ldots, i_n$ be the set of items, and let $D = d_1, d_2, \ldots, d_m$ be the transaction database, where each is a non-empty subset. An association rule is represented as $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \varnothing$. Here, X denotes the left-hand side (LHS) of the rule, and Y denotes the right-hand side (RHS). The three metrics used to evaluate association rules are support, confidence, and lift. The formulas for these indicators are as follows:

$$Support(X \rightarrow Y) = P(X \cup Y) \tag{1}$$

$$Confidence(X \rightarrow Y) = P(Y|X) = \frac{Support(X \rightarrow Y)}{Support(X)} \tag{2}$$

$$Lift(X \rightarrow Y) = \frac{P(Y|X)}{P(Y)} = \frac{Confidence(X \rightarrow Y)}{Support(Y)} \tag{3}$$

In the formulas, support refers to the probability that the dataset contains both items and confidence refers to the probability that the dataset contains an item, given that it already contains another item. If a rule satisfies the minimum support and minimum confidence thresholds, it is considered a strong association rule; if the lift is greater than 1, the rule is considered valuable.

This study employs the FP-growth algorithm to mine frequent item sets. The algorithm scans the transaction database twice: first to generate a frequent item header table, and then to construct an FP-tree; subsequently, it extracts the conditional pattern base and conditional FP-tree to recursively mine all frequent item sets. Based on this, support, confidence, and lift are further calculated to filter out association rules that meet the threshold requirements. Figure 4 presents the flowchart of the FP-growth algorithm.

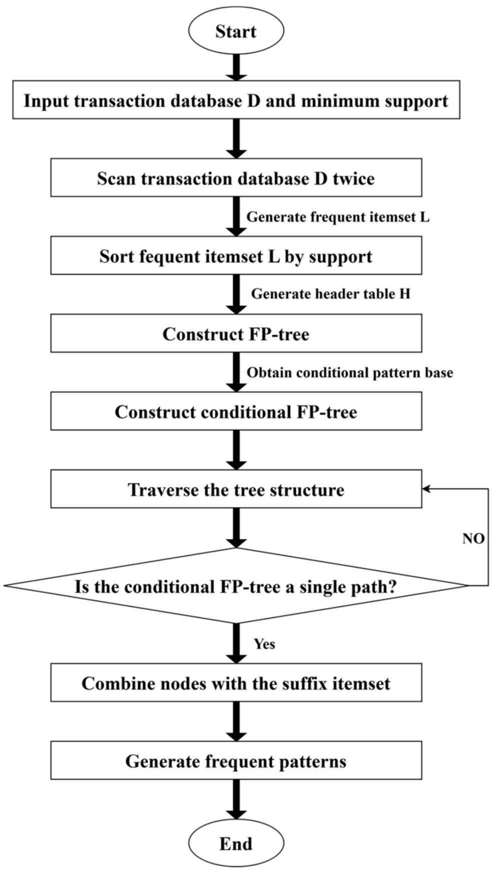

Figure 4. Flowchart of the FP-growth algorithm.

2.4. Complex Network

2.4.1. Construction of Directed Weighted Complex Networks Based on ARM

Complex networks provide an effective means of representing association structures and potential propagation relationships among accident-related factors. To uncover the coupling patterns among causal factors of ship self-sinking accidents, the association rules obtained from FP-growth mining were mapped onto a directed weighted causal-factor association network, denoted as $G = (V, E, W)$. In this network, $V = \{v_1, v_2, \ldots, v_n\}$ denotes the set of nodes representing accident causal factors, $E = \{e_1, e_2, \ldots, e_m\}$ denotes the set of edges representing associations between factors, and $W = \{w_1, w_2, \ldots, w_m\}$ denotes the set of edge weights representing the strength of these associations. Because the degree of association differs among causal factors, a directed weighted network is constructed based on the ARM results, whereby the LHS and RHS of each association rule are mapped to nodes v_i and v_j, respectively, and the confidence of the rule is taken as the edge weight w_{ij} [42]. Accordingly, the edge weights reflect the strength of association between causal factors rather than simple co-occurrence frequency. It should be noted that association rules reveal statistically recurrent co-occurrence and implication patterns rather than strict causal relationships in the experimental sense. Therefore, the constructed network is interpreted as a causal-factor association network based on accident investigation evidence. The network can further be expressed as a weighted adjacency matrix $A = (a_{ij})_{n \times n}$, which is defined as follows:

$$a_{ij} = \begin{cases} w_{ij}, (v_i, v_j) \in E \\ 0, (v_i, v_j) \notin E \end{cases} \tag{4}$$

where a_{ij} resents the connection weight between nodes v_i and v_j. If an association edge directed from v_i to v_j exists, then a_{ij} takes the corresponding value of w_{ij}; otherwise, $a_{ij} = 0$. As illustrated in Figure 5, the mapping from ARM results to a directed weighted complex network provides the foundation for subsequent analysis of network topological characteristics.

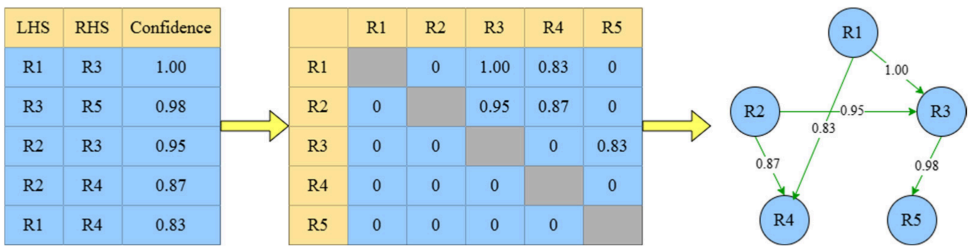

Figure 5. The mapping process from ARM to directed weighted complex networks.

2.4.2. Complex Network Analysis

Every network possesses specific topological characteristics that describe its connectivity, interactions, and evolutionary processes. Analyzing the topological characteristics of a constructed complex network helps to grasp the structural properties of the network as a whole, reveal its organizational patterns, and further identify key nodes and their dynamic interaction mechanisms [43]. Based on this, this study selects the following metrics to perform a topological analysis of complex networks.

(1) Node degree

The number of edges connecting node i to other nodes is called the degree of node i, denoted by $d(i)$. In a directed weighted network, degree can be divided into in-degree and out-degree:

$$d(i) = \sum_{j \in V} a_{ji} + \sum_{j \in V} a_{ij} \tag{5}$$

$$d_{in}(i) = \sum_{j \in V} a_{ji} \tag{6}$$

$$d_{out}(i) = \sum_{j \in V} a_{ij} \tag{7}$$

(2) Clustering coefficient

The clustering coefficient measures the degree of connectivity among a node's neighbors, specifically the extent to which these neighbors form closed triangles.

$$C(i) = \frac{2E(i)}{k_i(k_i - 1)} \tag{8}$$

In the formula, $E(i)$ denotes the number of edges between the neighbors of node i, and k_i denotes the number of edges connected to node i. The average of the clustering coefficients (CCs) of all nodes in the network is referred to as the average clustering coefficient (ACC). For directed weighted complex networks, the weighted clustering coefficient is further introduced to characterize the influence of edge weights, and its calculation formula is as follows:

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk} \tag{9}$$

(3) Average path length

The average path length is the average of the shortest paths between all possible pairs of nodes and is used to measure the degree of separation in a network:

$$APL = \frac{1}{n(n-1)} \sum_{i<j} l_{i,j} \tag{10}$$

(4) Betweenness centrality

Betweenness centrality refers to the proportion of the shortest paths between all pairs of nodes in a network that pass through a given node. It is used to assess the extent to which a node acts as a "bridge" within the network, reflecting its hub-like nature and connectivity. The formula for calculating it is:

$$B_i = \sum_{j \neq i \neq k \in N} \frac{N_{jk}(i)}{N_{jk}} \tag{11}$$

In the formula, N_{jk} represents the total number of the shortest paths from node j to node k, while $N_{jk}(i)$ represents the number of those shortest paths that pass through node i.

Betweenness centrality is the normalized form of betweenness, and its formula is as follows:

$$BC_i = \frac{2B_i}{(N-1)(N-2)} \tag{12}$$

(5) PageRank

When conducting risk analysis using node centrality (CN), the identification and evaluation of node centrality are widely recognized as critical steps. As a classic method for identifying node centrality in directed networks, the PageRank algorithm is based on the core principle of determining node centrality through the in- and out-degree relationships between nodes [44]. The advantages of PageRank are as follows: first, it is highly adaptable and suitable for processing directed, weighted, and sparse networks; second, by introducing a damping factor, it effectively resolves issues such as "infinite loops" or "trap nodes," enhancing the algorithm's robustness and ensuring stable convergence; third, it uses an iterative approach to gradually update the PageRank values of each node until convergence to a stable result is achieved, thereby enabling the ranking of node importance. The calculation formula is:

$$PR_i = \gamma \sum_{j \in V} \frac{a_{ji}}{d_{out}(j)} PR_j + \frac{1-\gamma}{N} \tag{13}$$

In the formula, PR_i represents the PageRank value of node i, $Y = [0, 1)$ denotes the damping factor and is usually set to 0.85, and N denotes the total number of nodes in the network.

$$PR_i = \gamma \sum_{j \in V} \frac{w_{ji}}{s_{out}(j)} PR_j + \frac{1-\gamma}{N} \tag{14}$$

In the formula, w_{ji} represents the weight of the edge from node j to node i, and $S_{out}(j)$ is the sum of the weights of all outgoing edges of node j.

(6) Robustness analysis

The robustness of a complex network refers to its ability to maintain overall connectivity and functionality when some nodes or edges fail due to faults or attacks. It reflects the network's capacity to withstand external disturbances and is an important indicator for evaluating network stability and continuity. In a causation network, weakening network robustness can effectively disrupt network connections, thereby blocking the dynamic evolution chain of causal factors. In the context of maritime accidents, weakening or damaging the interaction structure of the accident-related causation network and cutting off propagation paths are crucial for preventing accident escalation, reducing losses, and promoting accident prevention.

In robustness studies, attacks can generally be classified into two types: random attacks, which randomly remove or destroy nodes or edges in the network without any specific target, resulting in relatively dispersed damage; and intentional attacks, which selectively remove or destroy nodes or edges with high importance in the network, usually focusing on hub nodes or critical paths. In this study, global efficiency is adopted to analyze network robustness. Global efficiency reflects the development and propagation speed of causal factors in the network. By evaluating the connections and interactions among causal factors, it can reveal the propagation paths of causal factors and their potential impacts. The calculation formula is as follows:

$$E_g = \frac{1}{N(N-1)} \sum\nolimits_{i \neq j} \frac{1}{d_{ij}} \tag{15}$$

In the formula, N represents the total number of nodes in the network, and d_{ij} represents the length of the shortest path from node i to node j.

3. Results

3.1. Identification of Accident Causal Factors by Integrating Grounded Theory and the HFACS Model

In this study, 150 investigation reports on ship self-sinking accidents were selected as research samples to identify accident causal factors and explore their interaction mechanisms. NVivo 15 was employed to manage and code the accident texts. Grounded theory was applied to identify the causal factors, which were subsequently mapped onto the HFACS model for classification analysis. Before the final causal-factor classification was established, inter-rater reliability was assessed. The pairwise Cohen's kappa values were 0.893, 0.869, and 0.885, with an average value of 0.882. These results indicate a high level of inter-rater agreement and suggest that the causal-factor coding was reliable. The detailed results are shown in Table 1.

Based on the grounded-theory coding and HFACS classification, 30 causal factors associated with ship self-sinking accidents were ultimately identified and categorized into four groups: organizational influences, unsafe supervision, preconditions for unsafe acts, and unsafe acts. The detailed classification results are presented in Table 2.

Table 1. Pairwise Cohen's kappa results for inter-rater reliability assessment.

Coder Pair	Observed Agreement	Expected Agreement	Cohen's Kappa	Interpretation
coder 1 vs. coder 2	0.966	0.678	0.893	high agreement
coder 1 vs. coder 3	0.958	0.677	0.869	high agreement
coder 2 vs. coder 3	0.963	0.679	0.885	high agreement
Average	0.962	0.678	0.882	high agreement

Table 2. HFACS model of causal factors in ship self-sinking accidents.

HFACS Level	Subcategory	Code	Causal Factor	Frequency (%)
Organizational Influences	Ship safety management issues	O1	Inadequate safety management system	4.7
		O2	Insufficient safety education and training	16.7
		O3	Inadequate ship safety management	48.7
	Cargo safety management issues	O4	Inadequate cargo operation safety system	4.7
		O5	Insufficient understanding of the ship's loading condition	4.7
Unsafe Acts	Emergency response errors	UA1	Improper or untimely emergency measures	35.3
		UA2	Improper or untimely rescue measures	4.7
		UA3	Risk-taking navigation	22.7
	Personnel operational errors	UA4	Improper operation by the officer on watch	21.3
		UA5	Inadequate watchkeeping and inspection by crew members	10.7
Preconditions for Unsafe Acts	Ship equipment factors	P1	Water ingress into the hull	28.0
		P2	Insufficient reserve buoyancy and stability	16.7
		P3	Poor technical condition of the ship	33.3
		P4	Insufficient maintenance of hull weathertight integrity	40.0
	Environmental factors	P5	Adverse weather and sea conditions	53.3
		P6	Complex navigation environment	6.7
	Cargo factors	P7	Cargo unsuitable for shipment	4.0
		P8	Improper cargo stowage	28.7
		P9	Cargo overload	24.7
		P10	Improper cargo securing	17.3
	Personnel factors	P11	Poor physiological condition affecting safety	1.3
		P12	Poor safety habits	4.7
		P13	Weak safety awareness	25.3
		P14	Lack of safety knowledge	18.0
Unsafe Supervision	Inadequate ship navigation management	US1	Ship unseaworthiness	32.7
		US2	Crew incompetence	47.3
		US3	Insufficient understanding of the navigation environment	4.7
	Inadequate ship monitoring and management	US4	Deliberate shutdown of the AIS signal	8.7
		US5	Inadequate monitoring of ship equipment condition	23.3
		US6	Failure to maintain effective shore-based communication with the ship	20.0

As shown in Figure 6, the frequency statistics indicate that adverse weather and sea conditions, inadequate ship safety management, and crew incompetence were the most frequently identified factors. These high-frequency factors suggest that ship self-sinking accidents are often associated with the combined effects of external environmental disturbance, organizational management deficiencies, and crew-related operational vulnerabilities.

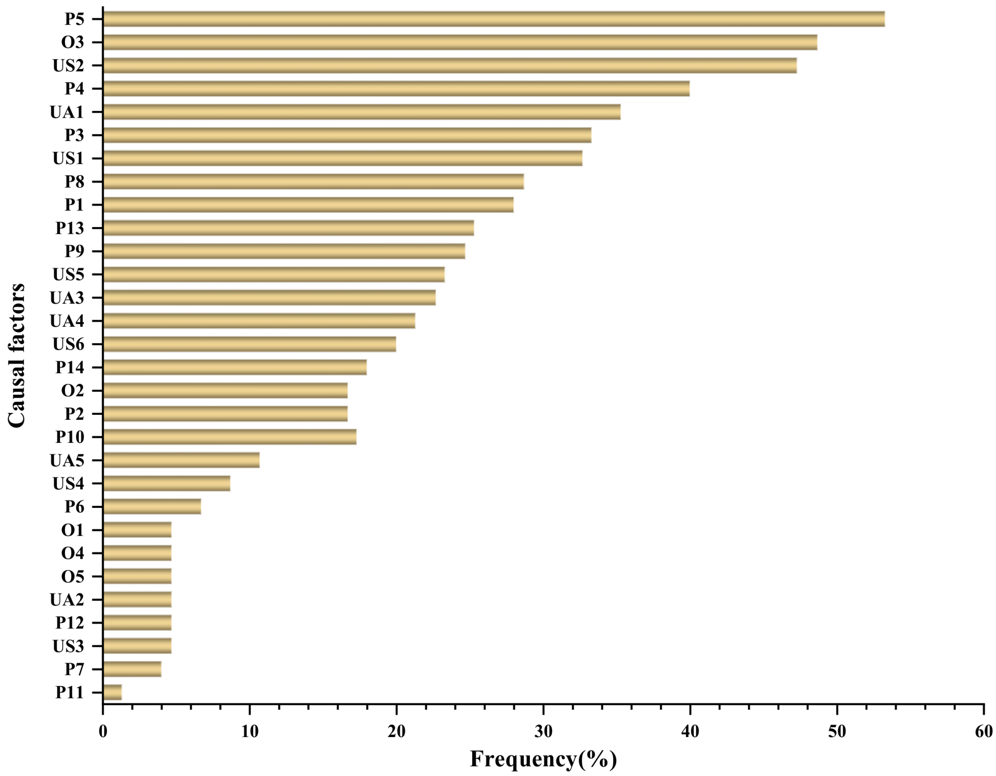

Figure 6. Frequency distribution of the 30 identified causal factors.

Figure 6 presents the frequency distribution of the 30 identified causal factors. The three most frequent factors were P5 (adverse weather and sea conditions), O3 (inadequate ship safety management), and US2 (crew incompetence). This result indicates that ship self-sinking accidents are commonly formed under the combined pressure of environmental disturbance, organizational management deficiencies, and inadequate crew competence.

Adverse weather and sea conditions can directly reduce reserve safety margins by increasing wave impact, accelerating water ingress, and aggravating stability loss. Inadequate ship safety management weakens routine maintenance, safety training, risk identification, and emergency preparedness. Crew incompetence further reduces the effectiveness of operational control and emergency decision-making. Therefore, these high-frequency factors should not be interpreted as isolated causes, but as interacting components in the progressive evolution of self-sinking accidents.

During the open coding stage, accident-causation-related texts were analyzed and labeled line by line to extract initial concepts and summarize them into initial categories. During the axial coding stage, these initial categories were systematically integrated and regrouped into broader main categories. During the selective coding stage, the internal logical relationships among the main categories were further examined, and the core category capable of integrating the overall framework was ultimately identified. Table 3 presents the specific coding process and examples of accident causation analysis based on grounded theory.

Table 3. Specific example of grounded theory results.

Original Statement	Initial Concept	Initial Category	Main Category	Core Category
The sea area was affected by a gale force of 7–8, with Adverse weather and sea conditions.	Adverse weather and sea conditions	Adverse weather and sea conditions	Environmental factors	Preconditions for unsafe acts

To verify the stability and adequacy of the coding results, an additional 15 accident reports (approximately 10% of the total sample) were selected for a theoretical saturation test. The results showed that no new concepts or categories emerged, indicating that the constructed coding system had reached theoretical saturation. On this basis, the distribution characteristics of the 30 identified causal factors across the accident samples were further statistically analyzed, as shown in Figure 7.

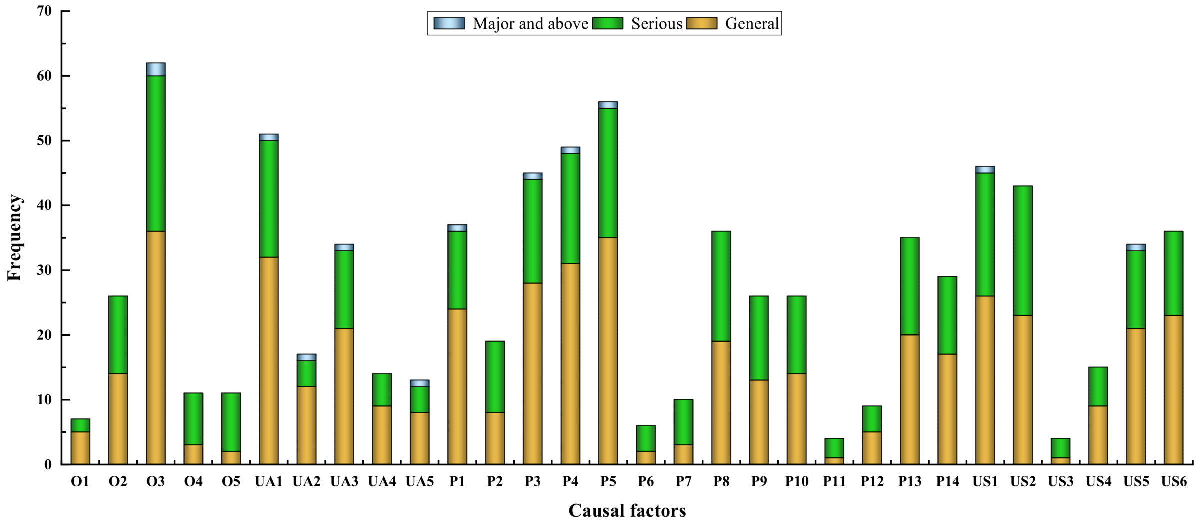

Figure 7. Distribution of 30 causal factors in accidents.

3.2. Analysis of Association Rule Mining Results

In association rule mining, the setting of the threshold has a significant impact on the results of rule extraction. When the threshold is set high, it effectively filters out rules with high support and strength, yielding results that are more concise and reliable; however, this may lead to the omission of some low-frequency associations that are potentially meaningful. Conversely, a lower threshold helps uncover more potential rules and enhances the comprehensiveness of the results, but it also tends to generate a large number of redundant and weakly associated rules, thereby increasing the complexity of interpreting the results. Relevant research on association rule mining indicates that there is currently no unified standard for setting minimum support and minimum confidence thresholds [45]. Researchers have employed a trial-and-error approach to propose reasonable threshold combinations for related studies [46].

The selection of minimum support and confidence thresholds strongly influences both the number of association rules generated and their interpretability. To ensure a rigorous justification for the chosen thresholds, a sensitivity analysis was conducted by varying support from 0.01 to 0.14 and confidence from 0.10 to 0.80. The number of rules generated under each combination is illustrated in Figure 8.

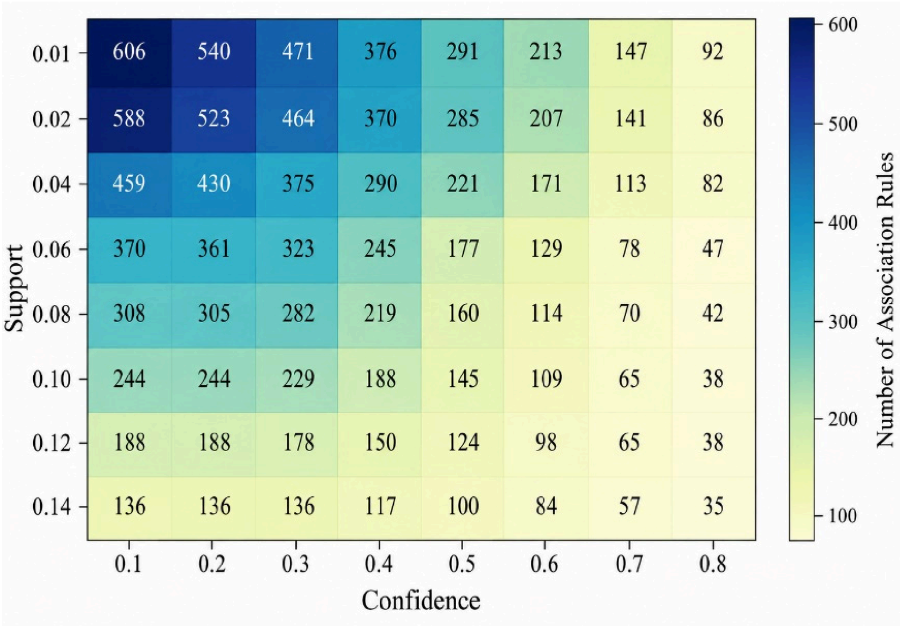

Figure 8. Sensitivity analysis of association rule numbers under different support and confidence thresholds.

As observed, lower thresholds, such as support = 0.01 and confidence = 0.10, produced a very large number of rules (606), many of which were weak or redundant. These excessive rules could complicate subsequent network construction and interpretation. In contrast, very high thresholds, such as support = 0.20 and confidence = 0.80, yielded only 22 rules, insufficient to construct a meaningful causal-factor association network. This trend highlights the trade-off between rule coverage and interpretability.

Considering both empirical reliability and interpretability, support = 0.10 and confidence = 0.30 were selected as the final thresholds. At this setting, 229 rules were retained, which are sufficient for building a robust causal-factor network while avoiding excessive redundancy. Given the 150 accident reports in the dataset, a support of 0.10 ensures that each retained rule appears in at least 15 cases, providing a solid empirical foundation. Furthermore, a confidence threshold of 0.30 is suitable for ship self-sinking accidents, which are complex, multi-factor events rather than deterministic one-to-one causal relationships.

To maintain clarity in the network, the maximum length of each rule was limited to two items, ensuring that each node corresponds to a single causal factor. Rules were generated using the Python "FP-growth" package (PyCharm 2024.1.7). Figure 9 shows a 3D bubble chart illustrating the relationship among support, confidence, and lift for the generated rules. Each point represents a rule, with the color intensity corresponding to lift values. Most rules have a confidence between 0.4 and 0.7, while support values primarily fall between 0.1 and 0.15, consistent with the selected thresholds and confirming their suitability.

The top 10 rules ranked by confidence are listed in Table 4. These rules reveal several recurrent causal-factor combinations in ship self-sinking accidents. Among them, four rules involve P5 (adverse weather and sea conditions), indicating that adverse environmental conditions frequently co-occur with other operational, supervisory, and emergency-response factors. Furthermore, analysis of the rule with the highest confidence (Rule 1) shows that in all samples where US5 (Inadequate monitoring of ship equipment condition) occurred, P4 (Insufficient maintenance of hull weathertight integrity) also occurred simultaneously. This indicates that this rule has extremely high reliability, and the two exhibit a highly consistent association in the samples.

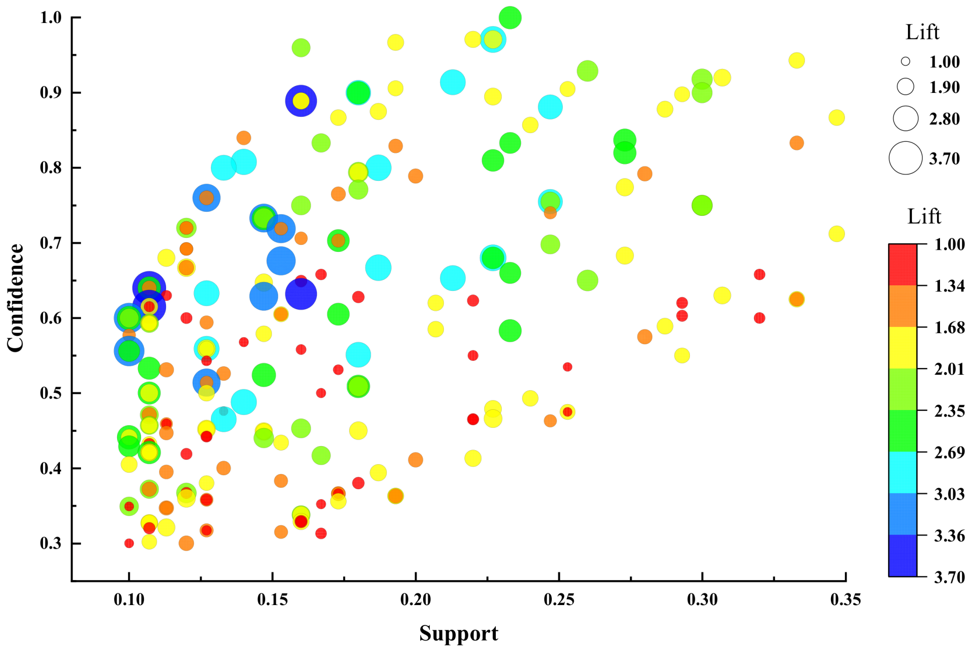

Figure 9. 3D bubble chart of association rules.

Table 4. Top 10 association rules by confidence.

Rule Number	Antecedent	Consequent	Support	Confidence	Lift
1	US5	P4	0.233	1.000	2.500
2	US5	P5	0.227	0.971	2.914
3	US5	O3	0.227	0.971	1.996
4	UA3	P5	0.220	0.971	1.820
5	US6	P5	0.193	0.967	1.813
6	O2	US2	0.160	0.960	2.028
7	UA1	P5	0.333	0.943	1.769
8	P1	P4	0.260	0.929	2.321
9	P3	O3	0.307	0.920	1.890
10	US1	P4	0.300	0.918	2.296

Although 30 causal factors were identified through grounded-theory coding and HFACS classification, only 19 factors appeared as nodes in the final network. This is because the network included only association rules that met the predefined support, confidence, and lift thresholds. The remaining 11 factors were retained in the frequency analysis but did not form statistically strong association rules under the selected thresholds. Their absence from the network does not imply that they are unimportant in individual accidents; rather, it indicates that they did not exhibit sufficiently recurrent association patterns in the present dataset. Therefore, frequency analysis and network analysis should be interpreted as complementary: the former reflects the occurrence distribution of all identified factors, whereas the latter highlights factors involved in stable association structures.

The high-confidence rules can be further interpreted as three typical association chains. First, the ship-condition deterioration chain is reflected by rules such as US5 $\rightarrow$ P4, P1 $\rightarrow$ P4, and US1 $\rightarrow$ P4, indicating that inadequate equipment monitoring, water ingress, and ship unseaworthiness are closely associated with insufficient maintenance of hull weathertight integrity. Second, the environmental-trigger chain is represented by rules such as UA3 $\rightarrow$ P5, US6 $\rightarrow$ P5, and UA1 $\rightarrow$ P5, suggesting that adverse weather and sea conditions often appear together with risk-taking navigation, poor shore-based communication, and improper emergency response. Third, the management-competence chain is reflected by O2 → US2 and P3 → O3, indicating that insufficient safety education and poor technical ship condition are closely related to crew incompetence and inadequate ship safety management.

3.3. Construction of Complex Networks and Analysis of Topological Features

Based on the 229 association rules generated in Section 3.2, causal factors were mapped as nodes, association rules were mapped as directed edges, and rule confidence was used as the edge weight. Accordingly, a directed weighted causal-factor association network comprising 19 nodes and 229 directed edges was constructed. In Figure 10, yellow nodes represent organizational influences, blue nodes represent unsafe acts, green nodes represent unsafe supervision, and red nodes represent preconditions for unsafe acts, all classified according to the HFACS framework.

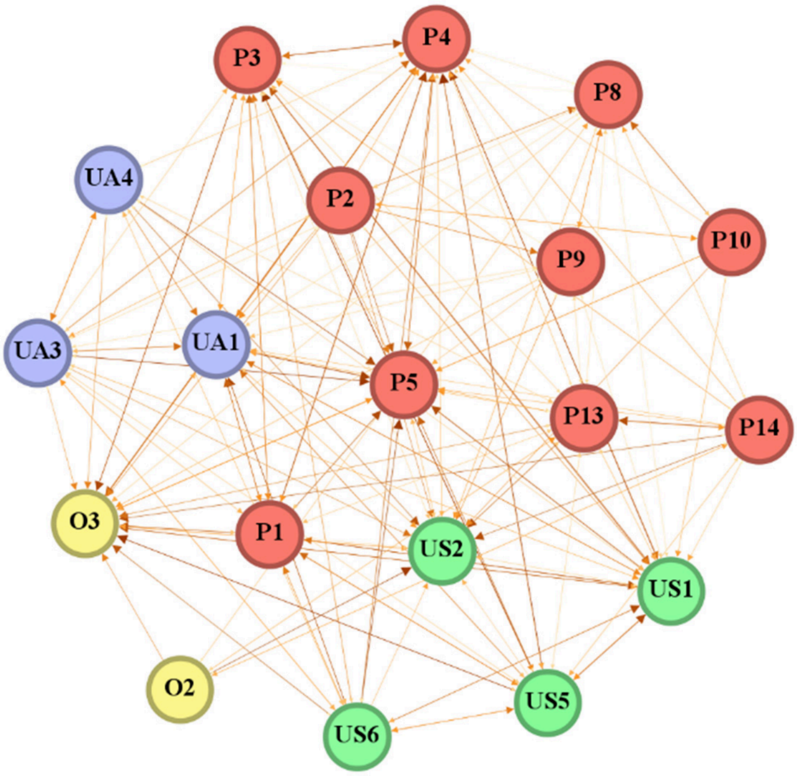

Figure 10. Complex network model of causes for ship self-sinking accidents.

(1) Node degree

In complex networks, a node's degree is used to measure its level of connectivity within the network. In an association network, a node's degree reflects the interactive relationships among various causal factors. A node's in-degree indicates the extent to which it is influenced by other nodes; its out-degree indicates the extent to which it influences other nodes; and its total degree represents the node's direct influence within the network. The in-degree, out-degree, and total degree values for the association network of ship self-sinking accidents are shown in Figure 11. P5 (adverse weather and sea conditions), O3 (inadequate ship safety management), and UA1 (improper or untimely emergency measures) showed relatively high in-degree, out-degree, and all degree values. This indicates that these factors have extensive direct associations with other factors in the network. From a prevention perspective, controlling these highly connected factors may help reduce the overall coupling intensity of the accident causation system.

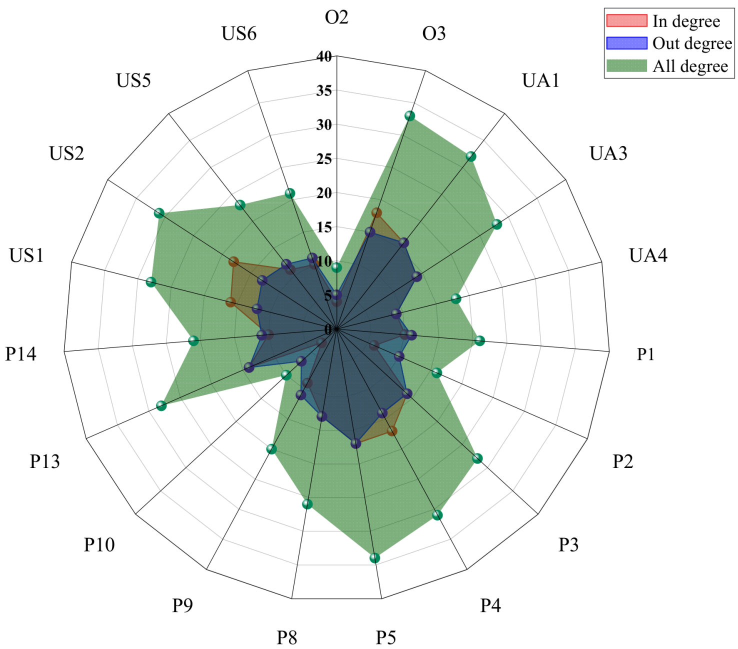

Figure 11. Node degree.

(2) Average path length

In an association network, the average path length reflects the efficiency with which causal factors propagate through the network. The shorter the average path, the more easily causal factors can spread rapidly throughout the network. The average path length of the network was 1.33, indicating that most causal factors can be connected through very short paths. This suggests that the network has high propagation efficiency, and changes in key factors may rapidly affect other related factors through direct or indirect associations. This indicates that changes in key factors can rapidly trigger a chain reaction, thereby significantly influencing the progression of an incident. Therefore, in incident prevention and management, it is essential to focus on high-impact nodes and their interconnections.

(3) Clustering coefficient

The clustering coefficient measures the likelihood of nodes forming triangular relationships within a network, i.e., the degree to which a node is closely connected to its neighboring nodes. The average clustering coefficient of the network was 0.802, indicating a high level of local clustering among causal factors. This means that several factors tend to form closely connected groups, within which risk-related associations may reinforce each other. Such a structure suggests that local factor combinations, rather than isolated factors, play an important role in the formation of ship self-sinking accidents. This study employs the weighted clustering coefficient to characterize the degree of clustering among causal factors in the network; the results are shown in Figure 12.

In the association network, the node with the highest clustering coefficient is O2 (insufficient safety education and training), with a value of 1.0, followed by O3 (inadequate ship safety management), P1 (water ingress into the hull), and P10 (improper cargo securing). The results indicate that these nodes are closely interconnected with their adjacent nodes. Any change in these nodes could trigger a chain reaction among adjacent nodes, potentially leading to accidents on a larger scale. Therefore, proactive control over these nodes must be strengthened to effectively reduce accident risks.

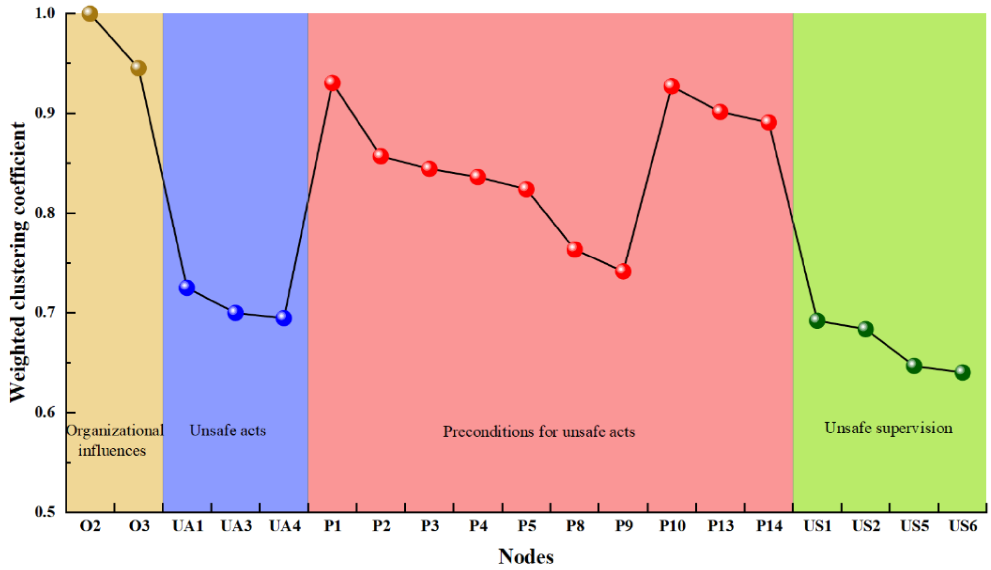

Figure 12. Clustering coefficient.

(4) Betweenness centrality

Betweenness centrality reflects a node's role as a "bridge" within the network. A higher value indicates that the node occupies a critical transmission position between different causal factors and plays a pivotal role in the evolution of accidents. As shown in Figure 13, there are significant differences in the Betweenness centrality of each node, indicating that the propagation of risks in ship self-sinking accidents does not occur uniformly but primarily relies on a few key nodes to facilitate cross-level transmission.

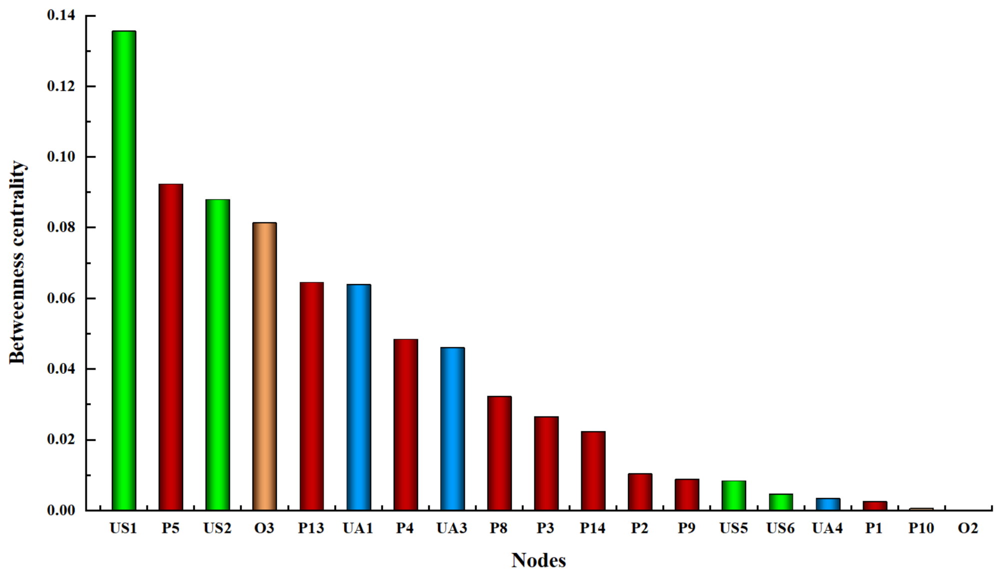

Figure 13. Betweenness centrality.

Among these, US1 (ship unseaworthiness) has the highest Betweenness centrality, indicating that it plays the strongest bridging role in the association network and serves as a crucial hub connecting multiple risk factors; it was followed by P5 (adverse weather and sea conditions), US2 (crew incompetence), and O3 (inadequate ship safety management), suggesting that environmental, human, and management factors exert a strong transmission effect in the evolution of accidents. At the same time, the Betweenness centrality of UA1 (improper or untimely emergency measures) and P4 (insufficient maintenance of hull weathertight integrity) is also relatively high, indicating that they played a significant role in driving the accident from the accumulation of hidden dangers to a state of uncontrolled development. In contrast, the mediating centrality of some nodes is close to zero, suggesting that they are primarily local triggering factors with limited bridging effects on the overall transmission pathway. Therefore, in the prevention and control of ship self-sinking accidents, interventions should prioritize key nodes such as US1, P5, US2, and O3 to effectively break the chain of association diffusion pathways.

(5) Analysis of node criticality

To identify critical nodes in the association network, this study employs the PageRank algorithm to evaluate the importance of each node. Compared to degree metrics, which reflect only local connectivity characteristics, PageRank provides a more comprehensive reflection of a node's overall influence on association diffusion processes. As shown in Figure 14, the five nodes with the highest PageRank values were O3 (inadequate ship safety management), US2 (crew incompetence), US1 (ship unseaworthiness), P4 (insufficient maintenance of hull weathertight integrity), and UA1 (improper or untimely emergency measures). These nodes represent the most influential factors in the network because they are not only directly connected to many other factors but also connected to other important nodes. This result indicates that the core risk structure of ship self-sinking accidents is concentrated in the coupling among safety management, crew competence, ship seaworthiness, hull weathertight integrity, and emergency response.

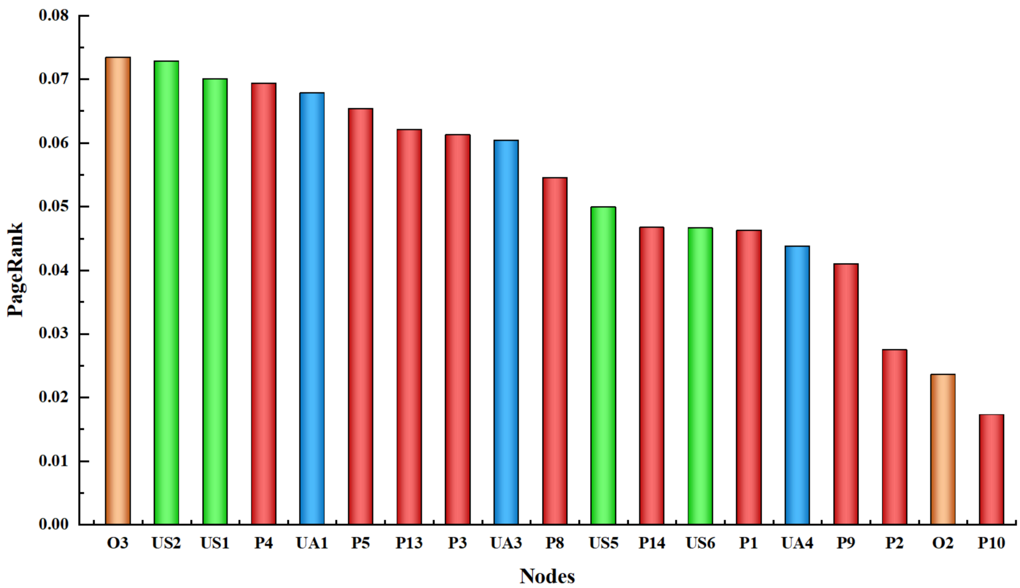

Figure 14. PageRank values.

Among these, O3 ranks first, indicating that inadequate ship safety management is the core node in the association network and exerts a fundamental influence on accident formation through aspects such as training implementation, equipment maintenance, and emergency preparedness. The high rankings of US2 and US1 indicate that crew incompetence and unseaworthiness of the ship have a significant risk-amplifying effect; the prominence of P4 and UA1 suggests that insufficient weathertightness of the hull and improper emergency response are key factors driving accidents from the accumulation of hazards to a loss of control. Overall, ship self-sinking accidents exhibit a progressive chain of events characterized by "management deficiencies—deteriorating conditions—improper conduct," and high PageRank nodes serve as the pivotal points in this chain.

(6) Robustness analysis

https://doi.org/10.3390/app16136307

Robustness analysis is used to evaluate a network's ability to maintain structural connectivity and functional integrity following node failures or attacks; it is a key method for measuring the stability and disturbance resistance of association networks. This study employs three attack strategies: random attacks, PageRank-based targeted attacks, and betweenness-centrality-based targeted attacks. By comparing changes in the network's global efficiency under different attack scenarios, we analyze the degree of dependence of the association network of ship self-sinking accidents on critical nodes. We identify critical nodes whose effective control can significantly weaken the accident's propagation capability, thereby providing priority intervention directions for risk governance.

As shown in Figure 15, the network's initial global efficiency is 0.483734. As the number of removed nodes increases, the network efficiency continues to decline under all three attack strategies, indicating that the network's propagation capability gradually weakens as nodes fail. Compared to random attacks, the two types of targeted attacks cause a faster and more significant drop in efficiency, suggesting that the network is highly dependent on a small number of critical nodes and is not a homogeneous structure.

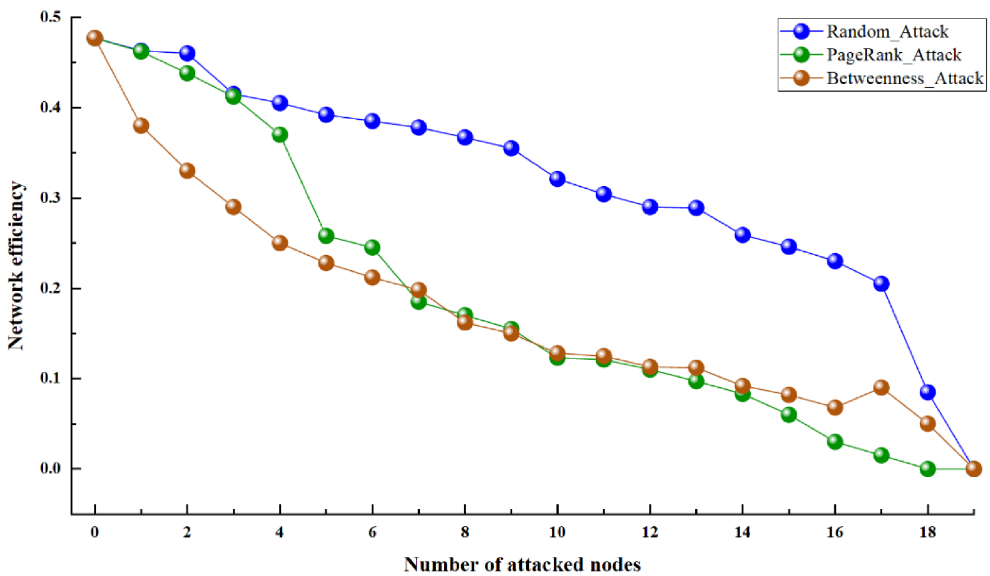

Figure 15. Network robustness simulation results.

When 10 nodes were removed, the global efficiency under random attack decreased to 0.321, representing a reduction of approximately 32.2% from the initial value. In contrast, the global efficiency decreased to 0.123 and 0.128 under PageRank-based and betweenness-based targeted attacks, respectively, corresponding to reductions of approximately 74.6% and 73.2%. The identified key nodes represent structurally dominant factors within the available dataset and should be interpreted as preliminary evidence rather than universally generalizable causal determinants. These results indicate that the network is highly dependent on a small number of critical nodes, and targeted control of these nodes can more effectively weaken association diffusion effects compared with random or average interventions. This indicates that once critical nodes are prioritized for weakening, the network structure rapidly collapses, and its association diffusion capability also declines markedly. As the number of removed nodes further increases, network efficiency gradually approaches 0. When 18 nodes are removed, the network is close to complete collapse, indicating that under extreme destructive conditions, the original causal coupling structure and propagation functions are essentially lost. Overall, the robustness of the association network for ship self-sinking accidents continues to decline as the scale of attacks increases, and targeted attacks are significantly more effective than random attacks. This indicates that, compared to average interventions, prioritizing precise interventions targeting key causal factors is more effective in severing association diffusion chains and enhancing accident prevention and control outcomes.

4. Discussion

4.1. Evolution Mechanism of Ship Self-Sinking Accidents

The results show that ship self-sinking accidents are not caused by a single direct factor but evolve through the interaction of multiple HFACS levels, including organizational influences, unsafe supervision, preconditions for unsafe acts, and unsafe acts. This multi-level structure indicates that self-sinking accidents are systemic events characterized by progressive risk accumulation and cross-level transmission. By combining the results of the HFACS classification, association rule mining, and complex network analysis, it can be seen that the occurrence of ship self-sinking accidents exhibits clear characteristics of hierarchical transmission and multi-factor coupling. In essence, such accidents represent a process in which latent vulnerabilities continuously accumulate and eventually lead to a concentrated loss of control under specific conditions. In this context, the proposed framework is designed for exploratory structural analysis of maritime accident causation rather than predictive benchmarking. Accordingly, its validation emphasizes internal consistency, structural coherence, and network robustness, rather than performance comparison with alternative predictive or machine learning models.

From the frequency statistics, P5 (adverse weather and sea conditions), O3 (inadequate ship safety management), and US2 (crew incompetence) occurred most frequently, indicating that the combined effects of external environmental stress, internal management failure, and insufficient personnel competence constitute an important background to accident occurrence. P5 (adverse weather and sea conditions) often acts as a triggering condition, reducing the ship's safety margin and aggravating risks such as water ingress and instability. By contrast, O3 (inadequate ship safety management) and US2 (crew incompetence) reflect deeper underlying factors, which continuously weaken the ship's risk prevention and control capacity through lax implementation of regulations, insufficient training, and biased risk identification.

The association rule results further indicate that causal factors do not exist in isolation, but are strongly interrelated. A significant association was identified between US5 (inadequate monitoring of ship equipment condition) and P4 (insufficient maintenance of hull weathertight integrity), suggesting that failures in equipment monitoring often occur together with a decline in the hull's protective capability. The association among US1 (ship unseaworthiness), P4 (insufficient maintenance of hull weathertight integrity), and UA1 (improper or untimely emergency measures) further suggests that ship self-sinking accidents typically involve a continuous process of hazard accumulation, risk escalation, and failure of response. The complex network analysis, which shows a relatively short average path length and a relatively high clustering coefficient, also indicates high transmission efficiency and close interconnections among factors. Once a critical link fails, risk may spread rapidly and trigger a chain reaction. Therefore, ship self-sinking accidents are essentially systemic accidents characterized by progressive destabilization. Their key implication lies not only in the final consequence of sinking, but also in the highly coupled fragile structure formed by multi-level factors before the accident occurs.

4.2. Implications for the Governance of Key Factors

From the perspective of accident prevention and control, frequency statistics reveal the common weak links in ship self-sinking accidents, while the identification of critical nodes in the complex network further shows which factors exert stronger amplifying and transmitting effects during accident evolution. Taken together, these findings indicate that the prevention and control of ship self-sinking accidents should not be limited to the general governance of high-frequency issues, but should also focus on the key factors occupying central positions in the network structure, so as to optimize priorities in risk prevention and control.

According to the PageRank and betweenness centrality results, O3 (inadequate ship safety management), US2 (crew incompetence), US1 (ship unseaworthiness), P4 (insufficient maintenance of hull weathertight integrity), and UA1 (improper or untimely emergency measures) are the most representative key factors. More specifically, the identified risk structures can be incorporated into existing Safety Management Systems (SMSs) as operational control items. High-centrality factors, such as inadequate ship safety management, crew incompetence, ship unseaworthiness, poor hull weathertight integrity, and improper emergency response, may be translated into company risk registers, inspection checklists, crew training priorities, and emergency preparedness procedures. When associated risks occur simultaneously, such as adverse weather combined with poor watertight integrity or insufficient crew competence, the SMS can trigger enhanced pre-voyage assessment, targeted inspection, additional crew briefing, or navigation postponement. Among them, O3 (inadequate ship safety management) occupies a core position in the network, indicating that it is not only a high-frequency issue but also an important source of the accumulation and spread of other risks. Many problems that appear on the surface as equipment defects, operational errors, and inadequate emergency response are often closely related to the failure to implement safety responsibilities, lax enforcement of regulations, and weak routine management. Therefore, improving the level of safety management can not only reduce the occurrence of specific risk factors, but also weaken the coupling strength of the risk network as a whole.

US2 (crew incompetence) and US1 (ship unseaworthiness) are likewise of major governance significance. US2 (crew incompetence) indicates deficiencies in risk identification, operational control, and emergency judgment. When facing complex sea conditions or sudden dangerous situations, crew members may easily miss the optimal window for control because of misjudgment and delayed response. US1 (ship unseaworthiness) indicates that the ship no longer possesses the basic technical conditions required for safe navigation; under such circumstances, accident risk may rise rapidly even when external disturbances are limited. Meanwhile, P4 (insufficient maintenance of hull weathertight integrity) and UA1 (improper or untimely emergency measures) reflect the critical links through which accidents progress from hazard accumulation to loss of control. The former concerns the hull's ability to resist seawater ingress and maintain reserve buoyancy, while the latter determines whether the situation can be brought under control in time and whether accident escalation can be delayed once a dangerous situation occurs.

The robustness analysis further shows that, compared with random interventions, targeted interventions on critical nodes are more effective in significantly weakening the propagation capacity of the network. Therefore, the prevention and control of ship self-sinking accidents should shift from decentralized and averaged governance to critical-chain governance, with limited resources prioritized toward the core links most likely to trigger chain reactions. In particular, greater attention should be paid to strengthening the implementation of safety management, ensuring crew competence, verifying seaworthiness, maintaining weathertight structures, and improving the quality of emergency training. Only by moving the prevention and control threshold forward to the stage of accident accumulation and carrying out targeted governance around critical nodes can accident evolution pathways be effectively interrupted and the precision and effectiveness of ship safety governance be enhanced.

4.3. Influence of Data Scope and Methodological Choices on Factor Associations

The identified structure and associations of causal factors are inevitably shaped by both the scope of the dataset and the methodological choices adopted in the analysis. This study uses 150 accident investigation reports released by the China Maritime Safety Administration, providing a detailed empirical basis for examining association diffusion patterns in ship self-sinking accidents. Therefore, the findings mainly reflect the reporting practices, regulatory context, vessel types, and operating environments covered by these reports. Although the dataset enables an in-depth analysis of self-sinking accidents within the Chinese maritime context, the observed distribution and coupling of causal factors may partly reflect region-specific operational practices, regulatory priorities, and accident-reporting conventions.

Methodological choices may also influence the association structures identified. Restricting the association-rule analysis to two-factor combinations facilitates the identification and interpretation of the most recurrent and robust interactions. However, this approach may underestimate higher-order relationships involving multiple interacting factors, particularly in accident scenarios characterized by nonlinear escalation and the progressive accumulation of risk. Consequently, the resulting association network should be interpreted as a representation of dominant pairwise dependencies rather than a complete reconstruction of all possible causal configurations.

Compared with studies on maritime collisions, groundings, and container vessel accidents, the present study focuses specifically on ship self-sinking accidents, which are characterized by progressive water ingress, deterioration of hull weathertight integrity, delayed emergency response, and the interaction between ship safety management and technical conditions. These distinctive characteristics indicate that the association structures observed in this study are shaped not only by general maritime accident mechanisms but also by the specific evolutionary dynamics of self-sinking scenarios. The findings nevertheless reveal both common patterns of latent–active risk interaction and accident-specific pathways through which management deficiencies, technical deterioration, and frontline actions become interconnected.

Accordingly, the generalizability of the results to other countries, regulatory systems, or maritime operational contexts should be treated with caution. Differences in fleet composition, vessel operation, safety governance, investigation procedures, and reporting standards may lead to different factor distributions and association patterns. Further validation using accident datasets from other jurisdictions, vessel categories, and operational environments is therefore required to assess the robustness and transferability of the identified relationships.

Despite these limitations, the findings suggest that the prevention of ship self-sinking accidents should not focus solely on isolated frontline errors but should also address the governance of critical factor combinations and the pathways through which risks accumulate and propagate. Maritime safety supervision should therefore place greater emphasis on ship safety management, crew competence, seaworthiness inspection, maintenance of hull weathertight integrity, and emergency response training. More broadly, this study demonstrates how data scope and analytical design influence the identification of critical factors and their interdependencies while providing a methodological and empirical foundation for future research based on broader datasets and higher-order association structures.

5. Conclusions

This study analyzed 150 ship self-sinking accident investigation reports released by the China Maritime Safety Administration between 2014 and 2024. By integrating grounded theory, HFACS, FP-growth association rule mining, and complex network analysis, a hybrid framework was developed to identify causal factors, reveal recurrent factor combinations, and examine the propagation structure of ship self-sinking accidents.

The results show that ship self-sinking accidents are not directly triggered by a single causal factor; rather, they evolve gradually as complex systemic accidents under the combined effects of multi-level factors, exhibiting clear coupling and propagation characteristics. A total of 30 causal factors were identified and classified into four HFACS levels. Frequency analysis showed that P5 (adverse weather and sea conditions), O3 (inadequate ship safety management), and US2 (crew incompetence) were the most frequent factors. Association rule mining further revealed stable factor combinations related to equipment monitoring, hull weathertight integrity, ship seaworthiness, and emergency response.

The directed weighted causal-factor association network constructed from the association rules contained 19 nodes and 229 directed edges. Topological analysis showed that the network was highly interconnected and exhibited high propagation efficiency. O3 (inadequate ship safety management), US2 (crew incompetence), US1 (ship unseaworthiness), P4 (insufficient maintenance of hull weathertight integrity), and UA1 (improper or untimely emergency measures) were identified as critical nodes in the network. Robustness analysis further demonstrated that targeted interventions on these critical nodes were more effective than random interventions in reducing the network's association diffusion capacity.

Several limitations should be acknowledged. First, the accident samples were limited to publicly available investigation reports released by the China Maritime Safety Administration. Therefore, the findings may reflect the reporting practices, regulatory context, and operational characteristics of Chinese maritime transportation, and their generalizability to other countries, vessel types, and maritime safety regimes requires further validation. Second, although association rules reveal stable co-occurrence and implication patterns among factors, they do not prove strict causal relationships. The identified links should therefore be interpreted as statistically supported causal-factor associations, rather than definitive physical association chains. Third, the constructed network is static and cannot fully capture the temporal evolution of accident development. Fourth, the framework does not yet provide a direct prediction of accident probability or a complete error model for factor classification. Future studies can integrate Bayesian networks, system dynamics, temporal accident-chain modeling, expert elicitation, near-miss data, and operational monitoring data to validate causal dependencies, quantify uncertainty, and develop predictive risk-warning tools.

Author Contributions: Conceptualization, Y.R. and Y.C.; Methodology, Y.C.; Formal analysis, Y.C.; Data curation, Y.C.; Writing—original draft preparation, Y.C.; Writing—review and editing, Y.R. and L.Z.; Funding acquisition, Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Agriculture and Rural Affairs of China grant number B050102.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy and ethical restrictions associated with some of the accident records. To maximize transparency, the official source used to compile and verify the dataset is publicly accessible at https://www.msa.gov.cn/6bbe386ee9ed4805a54c57154040de3f/index.jhtml (accessed on 11 May 2026), where relevant accident information can be retrieved.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Qiao, W.L.; Liu, Y.; Ma, X.X.; Liu, Y. Human factors analysis for maritime accidents based on a dynamic fuzzy Bayesian network. *Risk Anal.* **2020**, *40*, 957–980. [CrossRef] [PubMed]
2. Zohorsky, P.; Handley, H. Adapting the human factors analysis and classification system for commercial fishing vessel accidents. In Proceedings of the 13th AHFE International Conference on Human Error, Reliability, Resilience, and Performance, New York, NY, USA, 24–28 July 2022; pp. 1–7.
3. Youssef, E.A.; El Sayed, S.F.; Abdelkader, S. Maritime accidents analysis using maritime human factors and analysis model. *Marit. Res. Technol.* **2023**, *2*, 44–60. [CrossRef]
4. Xu, M.Q.; Ma, X.X.; Zhao, Y.L.; Qiao, W.L. A systematic literature review of maritime transportation safety management. *J. Mar. Sci. Eng.* **2023**, *11*, 2311. [CrossRef]
5. Maternová, A.; Materna, M.; Dávid, A.; Török, A.; Švábová, L. Human error analysis and fatality prediction in maritime accidents. *J. Mar. Sci. Eng.* **2023**, *11*, 2287. [CrossRef]
6. Qiao, W.L.; Guo, H.T.Y.; Deng, W.Y.; Huang, E.; Lin, G.Q.; Ma, X.X.; Lian, C.P. Complex network-based risk analysis for maritime heavy casualties in China during 2012–2021. *Ocean Eng.* **2024**, *308*, 118258. [CrossRef]
7. Ma, X.X.; Lan, H.; Qiao, W.L.; Han, B.; He, H.L. On the causation correlation of maritime accidents based on data mining techniques. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2024**, *238*, 905–919.
8. Qiao, W.L.; Yang, J.H.; Zhao, Y.L.; Deng, W.Y.; Ma, X.X. On the determination of the maritime-specific EPC values in reducing human factors based on maritime foundering accidents in China. *Ocean Eng.* **2024**, *307*, 118192. [CrossRef]
9. Deng, W.Y.; Ma, X.X.; Qiao, W.L. A novel methodology to quantify the impact of safety barriers on maritime operational risk based on a probabilistic network. *Reliab. Eng. Syst. Saf.* **2024**, *243*, 109884.
10. Qiao, W.L.; Huang, E.; Zhang, M.; Ma, X.X.; Liu, D. Risk influencing factors on the consequence of waterborne transportation accidents in China (2013–2023) based on data-driven machine learning. *Reliab. Eng. Syst. Saf.* **2025**, *257*, 110829. [CrossRef]
11. Liu, K.Z.; Yu, Q.; Yuan, Z.T.; Yang, Z.S.; Shu, Y.Q. A systematic analysis for maritime accidents causation in Chinese coastal waters using machine learning approaches. *Ocean Coast. Manag.* **2021**, *213*, 105859. [CrossRef]
12. Deng, J.; Liu, S.Y.; Xie, C.; Liu, K.Z. Risk coupling characteristics of maritime accidents in Chinese inland and coastal waters based on NK model. *J. Mar. Sci. Eng.* **2022**, *10*, 4.
13. Fan, S.Q.; Blanco-Davis, E.; Yang, Z.L.; Zhang, J.F.; Yan, X.P. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107070. [CrossRef]
14. Fu, S.S.; Yu, Y.R.; Chen, J.H.; Xi, Y.T.; Zhang, M.Y. A framework for quantitative analysis of the causation of grounding accidents in Arctic shipping. *Reliab. Eng. Syst. Saf.* **2022**, *226*, 108706. [CrossRef]
15. Ma, L.H.; Ma, X.X.; Lan, H.; Liu, Y.; Deng, W.Y. A data-driven method for modeling human factors in maritime accidents by integrating DEMATEL and FCM based on HFACS: A case of ship collisions. *Ocean Eng.* **2022**, *266*, 112699. [CrossRef]
16. Ma, L.H.; Ma, X.X.; Wang, T.; Zhao, Y.L.; Lan, H. A data-driven approach to determine the distinct contribution of human factors to different types of maritime accidents. *Ocean Eng.* **2024**, *295*, 116874. [CrossRef]
17. Ma, L.H.; Ma, X.X.; Wang, T.; Chen, L.G.; Lan, H. On the development and measurement of human factors complex network for maritime accidents: A case of ship groundings. *Ocean Coast. Manag.* **2024**, *248*, 106954.
18. Wang, H.; Chen, N.; Wu, B.; Soares, C.G. Human and organizational factors analysis of collision accidents between merchant ships and fishing ships based on HFACS-BN model. *Reliab. Eng. Syst. Saf.* **2024**, *249*, 110201.
19. Liu, J.Z.; Zhu, H.W.; Yang, C.X.; Chai, T. A network model for identifying key causal factors of ship collision. *J. Mar. Sci. Eng.* **2023**, *11*, 982. [CrossRef]
20. Shi, J.H.; Liu, Z.J.; Feng, Y.W.; Wang, X.J.; Zhu, H.W.; Yang, Z.L.; Wang, J.; Wang, H.X. Evolutionary model and risk analysis of ship collision accidents based on complex networks and DEMATEL. *Ocean Eng.* **2024**, *305*, 117965. [CrossRef]
21. Zhang, X.S.; Chen, P.F.; Mou, J.M.; Chen, L.Y.; Li, M.X. Critical causation factor analysis in ship collision accidents with complex network. *Ocean Eng.* **2025**, *315*, 119837.
22. Cao, Y.H.; Iulia, M.; Majumdar, A.; Feng, Y.W.; Xin, X.; Wang, X.J.; Wang, H.X.; Yang, Z. Investigation of the risk influential factors of maritime accidents: A novel topology and robustness analytical framework. *Reliab. Eng. Syst. Saf.* **2025**, *254*, 110636.
23. Su, W.; Lu, J.; Zhang, Q.R.; Liu, Y.; Ji, Y.; Fan, H.W. Risk factors identification and analysis for maritime accidents using complex network theory and improved Leader-Rank algorithm. *Int. J. Shipp. Transp. Logist.* **2025**, *20*, 245–270. [CrossRef]
24. Chen, J.H.; Zhuang, C.L.; Shi, J.; Jiang, H.Q.; Xu, J.Y.; Liu, J.T. Risk factors extraction and analysis of Chinese ship collision accidents based on knowledge graph. *Ocean Eng.* **2025**, *322*, 120536. [CrossRef]
25. Chen, M.Z.; Gao, L.; Zhu, J.H. Research on maritime accident complex network: A case of ship groundings. In *Proceedings of the International Conference on SmartRail, Traffic and Transportation Engineering (ICSTTE 2024)*; Springer Nature: Singapore, 2024; pp. 367–373.
26. Guo, X.R.; Zheng, Q.Q.; Guo, Y.M. Maritime accident causation: A spatiotemporal and HFACS-based approach. *Ocean Eng.* **2025**, *340*, 122329. [CrossRef]

27. Yin, X.Y.; Yan, Y.; Wang, J.H.; Zhao, H.Z.; Wu, Q.Y.; Xu, Q. Data-driven analysis of the causal chain of waterborne traffic accidents: A hybrid framework based on an improved human factors analysis and classification system with a Bayesian network. *J. Mar. Sci. Eng.* **2025**, *13*, 393. [CrossRef]
28. Li, H.H.; Ren, X.J.; Yang, Z.L. Data-driven Bayesian network for risk analysis of global maritime accidents. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108938.
29. Zhou, K.W.; Xing, W.B.; Wang, J.B.; Li, H.H.; Yang, Z.L. A data-driven risk model for maritime casualty analysis: A global perspective. *Reliab. Eng. Syst. Saf.* **2024**, *244*, 109925.
30. Turner, C.; Astin, F. Grounded theory: What makes a grounded theory study? *Eur. J. Cardiovasc. Nurs.* **2021**, *20*, 285–289. [CrossRef] [PubMed]
31. Al-Eisawi, D. A design framework for novices using grounded theory methodology and coding in qualitative research: Organisational absorptive capacity and knowledge management. *Int. J. Qual. Methods* **2022**, *21*, 1–13. [CrossRef]
32. Moura, C.O.D.; Silva, Í.R.; Silva, T.P.D.; Santos, K.A.; Crespo, M.D.C.A.; Silva, M.M.D. Methodological path to reach the degree of saturation in qualitative research: Grounded theory. *Rev. Bras. Enferm.* **2021**, *75*, e20201379.
33. Meng, B.; Lu, N. A hybrid model integrating HFACS and BN for analyzing human factors in CFIT accidents. *Aerospace* **2022**, *9*, 711. [CrossRef]
34. Kaptan, M.; Sarıalioğlu, S.; Uğurlu, Ö.; Wang, J. The evolution of the HFACS method used in analysis of marine accidents: A review. *Int. J. Ind. Ergon.* **2021**, *86*, 103225. [CrossRef]
35. Niu, L.X.; Zhao, J.; Yang, J.H. Risk assessment of unsafe acts in coal mine gas explosion accidents based on HFACS-GE and Bayesian networks. *Processes* **2023**, *11*, 554. [CrossRef]
36. Wang, Y.C.; Liu, C.F.; Xu, H.S.; Geng, X.X.; Wang, Y.R.; Liu, Y. Analysis of the causes of falling accidents on building construction sites in China based on the HFACS model. *Buildings* **2025**, *15*, 1412. [CrossRef]
37. Lee, Y.N. A new approach to identify critical causal factors and evaluate intervention strategies for mitigating major railway occurrences in Taiwan. *J. Rail Transp. Plan. Manag.* **2025**, *33*, 100507. [CrossRef]
38. Çakır, E.; Fışkın, R.; Sevgili, C. Investigation of tugboat accidents severity: An application of association rule mining algorithms. *Reliab. Eng. Syst. Saf.* **2021**, *209*, 107470. [CrossRef]
39. Qiu, Z.X.; Liu, Q.L.; Li, X.C.; Zhang, J.J.; Zhang, Y.Q. Construction and analysis of a coal mine accident causation network based on text mining. *Process Saf. Environ. Prot.* **2021**, *153*, 320–328. [CrossRef]
40. Li, S.; You, M.J.; Li, D.W.; Liu, J. Identifying coal mine safety production risk factors by employing text mining and Bayesian network techniques. *Process Saf. Environ. Prot.* **2022**, *162*, 1067–1081. [CrossRef]
41. Lan, H.; Ma, X.X.; Qiao, W.L.; Ma, L.H. On the causation of seafarers' unsafe acts using grounded theory and association rule. *Reliab. Eng. Syst. Saf.* **2022**, *223*, 108498. [CrossRef]
42. Ma, X.X.; Deng, W.Y.; Qiao, W.L.; Lan, H. A methodology to quantify the risk propagation of hazardous events for ship grounding accidents based on directed CN. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108334. [CrossRef]
43. Zhou, Y.; Li, C.S.; Ding, L.Y.; Sekula, P.; Love, P.E.D.; Zhou, C. Combining association rules mining with complex networks to monitor coupled risks. *Reliab. Eng. Syst. Saf.* **2019**, *186*, 194–208. [CrossRef]
44. Zhang, P.P.; Wang, T.D.; Yan, J. PageRank centrality and algorithms for weighted, directed networks. *Phys. A* **2022**, *586*, 126438. [CrossRef]
45. Geng, X.J.; Sun, Q.X.; Zhou, Z.J.; Jiao, L.M.; Ma, Z.F. Adaptive fuzzy-evidential classification based on association rule mining. *Inf. Sci.* **2024**, *665*, 120408. [CrossRef]
46. Lan, H.; Ma, X.X.; Qiao, W.L.; Deng, W.Y. Determining the critical risk factors for predicting the severity of ship collision accidents using a data-driven approach. *Reliab. Eng. Syst. Saf.* **2023**, *230*, 108934.