A Research on the Causes of Ship Collision Accidents Based On Community Analysis

Jianzhou Liu*, Ying Lu, Yingjie Jia, Zhigang Han, Jinsong Li, Jingyang Lu

Binzhou Polytechnic University, Binzhou, 256603, China
*2922218851@qq.com

Abstract. In the analysis of ship collision accidents, it is of great significance to identify the key causes of ship collision accidents in time to reduce the incidence of accidents. In this paper, a network model of ship collision accident causation combined with fast Newman community detection algorithm is proposed. The ship collision accident reports from 2020 to 2023 are collected, the causal factors are extracted and the causal network is constructed. The fast Newman algorithm is used to identify the causal community, and the entropy weight method is used to quantify the causal risk, so as to identify the key causal community and the key causal. The results show that community 2 is the key cause risk community and human factors are the core causes of ship collision accidents.

Keywords: ship collision accidents; complex networks; community detection; risk quantification

1 Introduction

Maritime transport is a vital component of global trade, accounting for 90% of total cargo volume [1]. As the scale of global trade continues to expand, the total number of vessel voyages has increased significantly, and the probability of vessel collisions has risen accordingly. Vessel collisions have a significant impact on maritime safety. They can not only result in casualties, environmental pollution, and property damage, but may also disrupt shipping routes, causing supply chain disruptions and triggering a chain of negative effects on regional and even global trade order. Therefore, preventing vessel collisions is of great significance for ensuring the safety of maritime navigation and promoting the development of global trade.

Identifying the key causes of ship collision accidents and implementing preventive measures can effectively reduce the occurrence of such incidents. Currently, various methods are applied to the analysis and research of ship collision accidents. Yaqing Shu et al. [2] proposed a new method that combines ontology models with Bayesian networks to effectively analyze the allocation of liability in ship collision accidents. Mingyang Guo et al. [3] used 343 ship collision accident reports as a sample to construct the HFACS-SCA model based on HFACS. By combining the prior method, entropy weighting method, and BP-DEMATEL method to analyze risk factors, they found that organizational factors, “excessive speed,” and deficiencies in safety management sys-

L. W. Koon et al. (eds.), *Proceedings of the 2026 7th International Conference on Management Science and Engineering Management (ICMSEM 2026)*, Advances in Economics, Business and Management Research 399,
https://doi.org/10.2991/978-94-6239-758-3_28

tems were the primary causes. Xiyu Zhang et al. [4] proposed a framework combining deep learning and Bayesian prediction models for ship collision analysis; the analyses generated by this framework were consistent with actual accident reports in terms of accident analysis and liability allocation. Ivana Jovanović et al. [5] employed fault tree analysis (FTA) and Bayesian network (BN) methods to thoroughly investigate the influence and mechanisms of human factors in the occurrence of cargo ship accidents.

Complex networks first emerged in the 1980s [6]. In recent years, complex networks have been applied to the analysis of ship collision accidents. Xinsheng Zhang et al. [7] collected 320 ship collision accident reports and used complex network theory to construct a ship collision risk network comprising 66 nodes and 211 edges, effectively identifying key causal factors. Jun Ma et al. [8] addressed issues in complex network analysis of ship collision accidents—such as the lack of an intelligent foundation for causal chain extraction, inaccurate node importance ranking, and poor correlation between robustness and importance analysis—by utilizing 322 accident reports. They employed a semi-automated method to extract RIF propagation paths and construct a SCARIFCN network, proposed the 3-WLR algorithm to evaluate node importance, and validated its advantages through comparative analysis, ultimately providing targeted recommendations for vessel management, policy formulation, and maritime supervision. Qiaoling Du et al. [9] utilized grounded theory and the Human Factors Analysis and Classification System (HFACS) to identify and classify causal factors from 152 vessel collision accidents, construct corresponding causal networks, and determine key risk factors through network topological feature analysis.

However, the current research focuses more on the identification of the key causes of the accident, and pays less attention to the aggregation of the cause factors. It ignores that the ship collision accident is not caused by the independent action of a single factor, but the result of the synergy of multiple cause factors. Based on this, this paper uses the complex network community analysis method combined with the correlation mechanism between accident causes to identify the core cause communities and key causes, and proposes targeted prevention and control measures to provide theoretical support for maritime supervision and ship safety management.

2 The Proposed Model

2.1 Model Assumptions

Assumption 1: In a causal network for ship collision accidents, nodes represent causal factors, and edges represent the relationships between them. The causal factors are derived from accident reports; if two causal factors appear in the same accident report, they are considered to have a relationship, meaning there is an edge between them. The weight of an edge is the number of times the two causal factors appear together in accident reports.

Assumption 2 : The network of causes underlying ship collision accidents exhibits distinct community characteristics. Generally speaking, accidents are not caused by a single factor but result from the interaction of multiple factors. These factors do not exist in isolation; rather, they form a network through complex interconnections. Factors that share similarities or similar mechanisms of action spontaneously cluster together, forming relatively independent community modules that reflect synergistic effects in specific accident scenarios.

Assumption 3: The network of causes of ship collision accidents is an undirected network. Typically, there is a domino effect among the causes of accidents, meaning that one causal factor leads to another. However, the association structure among factors is determined by their similarity. Therefore, the network of causes of ship collision accidents can be regarded as an undirected network.

2.2 Model Building

Step 1: Data Collection and Processing

Data on vessel collision accidents was obtained from the official website of the China Maritime Safety Administration. The vessel collision accident information on the website is presented in the form of accident reports, covering key details such as the time of the accident, the cause of the accident, the sequence of events, and expert analysis. This study collected vessel collision accident reports from 2020 to 2023. Using the expert analysis in the accident reports as the primary basis, we systematically identified various causal factors leading to vessel collisions and treated each causal factor as a node.

Step 2: Determining Edges and Edge Weights

If two causal factors are both present in the same incident, they form an edge; the number of incidents caused by these two factors constitutes the weight of the edge.

Step 3: Constructing the Causal Network of Vessel Collision Accidents

Complex networks are typically represented by adjacency matrices. In this study, the causal network of ship collision accidents is represented by an adjacency matrix A, where a_{ij} reflects the connection between causal factors i and j. If the two factors cooccur in the same accident, then $a_{ij} = 1$; if they do not co-occur, then $a_{ij} = 0$, and the specific value of a_{ij} serves as the weight of the edge.

Step 4: Community Structure Detection

Community detection algorithm is used to detect community structure in complex networks. Fast Newman algorithm is an efficient algorithm for community structure detection, which is often used to analyze the community structure of complex networks [10]. The main principle is to identify the community structure in the network by Q-modality. The formula for calculating the Q-modality is as follows:

$$Q = \frac{1}{2m} \sum_{ij} \left(W_{ij} - P_{ij}\right) \delta\left(C_i, C_j\right) \tag{1}$$

where m is the total number of edges in the network; W_{ij} is the number of edges between nodes i and j; P_{ij} is the expected number of edges between nodes i and j; if nodes i and j belong to the same community, then $\delta\left(C_i, C_j\right)$ is 1; otherwise, it is 0.

Step 5: Calculating Node Importance

1. The selection of evaluation index

To systematically evaluate key communities and key nodes within the causal network, the following three core metrics were selected: degree, weighted clustering coefficient, and weighted closeness centrality. These metrics comprehensively reflect the structural status and role of nodes within the network from three dimensions: local connectivity, local clustering, and global reachability.

(1) Degree : Measures the number of direct connections a node has with other nodes, reflecting its level of activity and influence within the network.

$$k_i = \sum_{j \in N(i)} a_{ij} \tag{2}$$

where k_i is the degree of node i, and a_{ij} is the edge between node i and node j.

(2) Weighted clustering coefficient : The weighted clustering coefficient is used to measure the degree of cohesion between a node and its neighborhood, reflecting the clustering properties of the local structure surrounding the node.

$$CC_i^w = \frac{1}{s_i(k_i-1)} \sum_{j,k \in N(i)} \left(\frac{w_{ij}+w_{ik}}{2}\right) a_{ij} a_{ik} a_{jk} \tag{3}$$

Where CC_i^w is the weighted clustering coefficient of several i, and k_i is the degree of node i ; s_i is the strength of node i, which is the sum of the weights of its connected edges ; w_{ij} is the weight of edge (i, j) ; aij is the element of adjacency matrix.

(3) Weighted proximity centrality : The reciprocal of the weighted average distance from a node to all other nodes in the network, reflecting its efficiency in information propagation and its strategic position within the overall network.

$$WCC_i = \frac{1}{\sum_{j \neq i} d_{ij}^w} \tag{4}$$

Where WCC_i is the weighted closeness centrality of several i, d_{ij}^w is the weighted shortest path length from node i to j.

2. Node importance evaluation based on entropy weight method

Identifying key causal nodes is a core component of measuring community importance and analyzing the role of individual nodes within the overall network. In complex network research, degree, weighted clustering coefficient, and weighted closeness centrality are all commonly used metrics for identifying key nodes. However, these three metrics differ in their evaluation perspectives and focal points, and relying on a single metric to assess node importance has certain limitations. Therefore, this paper employs the entropy weighting method to comprehensively weight these three types of metrics, thereby achieving a more comprehensive quantitative evaluation of node importance.

(1) Data Standardization

To eliminate biases in subsequent calculations caused by differences in the units and values of various indicators, each indicator must be normalized prior to calculation, mapping its values uniformly to the interval [0, 1].

$$x_{ij}^* = \frac{x_{ij} - min(x_j)}{max(x_j) - min(x_j)} \tag{5}$$

where x_{ij}^* is the normalized value, x_{ij} is the value of node i for the jth indicator, and $min(x_j)$ and $max(x_j)$ are the minimum and maximum values, respectively, for the jth indicator.

(2) Calculation of Information Entropy

Information entropy is calculated as follows:

$$e_j = -\sum_{i=1}^{n} p_{ij}\log(p_{ij}) \tag{6}$$

$$p_{ij} = \frac{x_{ij}^*}{\sum_{i=1}^{n} x_{ij}^*} \tag{7}$$

where e_j represents information entropy, satisfying the condition $e_j \in [0, 1]$, n is the number of nodes, and p_{ij} is the normalized value of the i th node for the jth metric.

(3) Calculation of the weights for each indicator

Based on information entropy, the weight of each indicator can be calculated using the following formula:

$$w_j = \frac{1-e_j}{\sum_{j=1}^{m}(1-e_j)} \tag{8}$$

where w_j is the weight of the jth indicator, and m is the number of indicators.

(4) Calculation of Node Importance

$$D_i = \sum_{j=1}^{m} w_j x_{ij}^* \tag{9}$$

Where D_i represents the importance of node i; the larger the value of D_i, the more important node i is.

3 Case Study

3.1 Data Collection and Processing

We collected 300 reports of vessel collision accidents from 2020 to 2023 from the official website of the China Maritime Safety Administration and used the method described in Section 2.1 to identify 98 contributing factors, as shown in Table 1.

Table 1. Causes of Vessel Collision Accidents

Cause ID	Cause name
1	Improper look-out
2	Inappropriate assessment of the situation of the risk of collision
…	…
97	The shipping company does not fully grasp the navigation management regulations which is important to ship safety
98	Defects in Bridge Resource Management

3.2 Community Structure Analysis

Based on the FN algorithm, the network of accident causes was clustered, ultimately identifying seven clusters; the specific results are shown in Table 2. Calculations show that the network's modularity (Q) value is 0.5623. Since modularity values range from 0 to 1, and values closer to 1 indicate a clearer cluster structure, while the modularity of real-world networks typically falls within the range of 0.3 to 0.7. Therefore, these results indicate that the ship collision accident causation network constructed in this study exhibits distinct clustering characteristics. To ensure the statistical validity of the clustering, clusters with at least five nodes were considered valid clusters for subsequent analysis.

Table 2. Results of the Classification of Social Group Structures

Number of community	nodes contained
1	96, 93, 90, 89, 88, 83, 82, 58, 53, 4, 31, 23
2	98, 97, 91, 9, 87, 86, 85, 81, 80, 8, 79, 78, 77, 76, 75, 74, 73, 71, 70, 68, 67, 66, 65, 64, 63, 62, 61, 60, 59, 57, 56, 55, 54, 51, 50, 48, 47, 45, 43, 42, 41, 40, 39, 38, 30, 29, 27, 25, 22, 20, 2, 19, 15, 10
3	95, 92, 49, 44, 3, 26, 24, 46, 18, 17, 21, 36, 37, 84
4	94, 69, 5, 34, 32
5	7, 6, 52, 33, 28, 16, 11
6	72, 12, 1
7	35, 14, 13

An analysis of the internal factors within each organization leads to the following main conclusions:

(1) Each cluster structure contains more than one type of factor.

Accidents are not caused by a single category of contributing factors, but rather exhibit a complex interplay of multiple factors, and the patterns of factor combinations vary significantly across different accidents. Taking Cluster 1 as an example, it consists solely of management and human factors; in contrast, Cluster 2 encompasses all four major categories of causal factors: management, vessel, human, and environmental. This phenomenon reveals deep intrinsic connections among the subsystems within the ship navigation system; different types of causal factors do not operate in isolation but form tight coupling and interaction mechanisms within each cluster. Each cluster essentially corresponds to a highly interconnected risk subsystem within the system, and the structural differences between them reflect the specificity of causal combination patterns and transmission pathways under different accident scenarios. Therefore, in ship operation management, a single-perspective approach must be abandoned in favor of a systems-theory perspective for comprehensive governance.

(2) The Prevalence of Human Factors in Organizational Structures

The analysis revealed that human factors permeate all five communities. This indicates that human factors do not exist in isolation but rather serve as a central connecting link, forming a stable symbiotic relationship with heterogeneous factors such as management, vessels, and the environment, and running through the construction of various community structures. This finding further confirms the central role of crew behavior in the evolution of accidents. Therefore, strengthening the supervision and control of crew behavior should be regarded as a critical component of the ship collision prevention system.

(3) The Marginalization of Vessel Factors in Networks

Vessel-related factors primarily manifest as equipment-level failures, such as main engine malfunctions, steering gear failures, and other equipment anomalies. Data indicates that such factors account for a relatively small proportion of the causal system and, following cluster analysis, are concentrated solely within Cluster 2. This distribution pattern suggests that while vessel equipment failures may serve as triggers for accidents, their role in driving overall accident occurrence is relatively limited, classifying them as non-essential factors within the network. Consequently, in the allocation of resources for routine vessel safety management and accident prevention, these factors should not be treated as the primary focus; instead, efforts should be concentrated on dominant factors such as human error and management. Of course, routine maintenance and inspection of vessel equipment must still be carried out systematically to eliminate the possibility of such failures becoming accident triggers.

3.3 Risk Quantification Analysis of Causal Factors

The magnitude of risk associated with each causal factor can be quantified by node importance; a higher risk value indicates that the causal node plays a more significant role in the network. Based on this, this paper employs the node importance calculation method described in Section 1.2 to compute the comprehensive importance of each causal node and uses this as the risk value for the corresponding factor. The risk values for each causal factor are shown in Table 3, and the risk values for each community are presented in Table 4.

Table 3. Risk Values for Each Node

node	Degree	Weighted clustering coefficient	weighted closeness centrality	Risk
1	93	0.033911018	0.42173913	0.882582526
2	90	0.048192043	0.040450375	0.934667828
3	75	0.029810086	0.461904762	0.732867258
...	...	...	...	...
96	35	0.026446469	0.407563025	0.430310216
97	14	0.012604163	0.434977578	0.183694396
98	59	0.015475259	0.453271028	0.512748969

Table 4. Risk Values for 7 Communities

community	Risk	community	Risk
1	6.996993762	5	3.991101261
2	18.42939541	6	2.295618308
3	4.345305268	7	1.548301654
4	3.22714381		

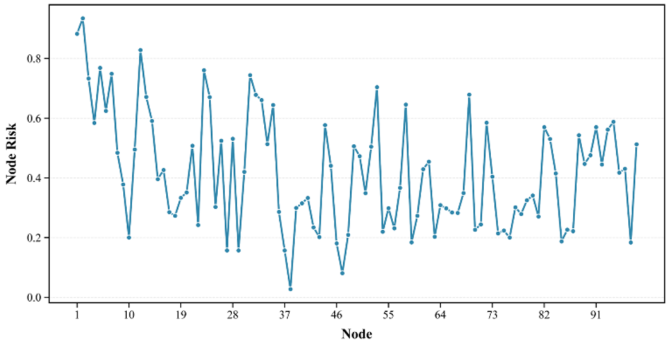

Fig. 1. Risk values for each node

Figure 1 shows the risk profile of each node before the communities were clustered. It reveals that the accident causation network exhibits an overall non-uniform risk distribution, with significant differences in risk values among nodes representing different causes. This indicates that ship collision accidents are not triggered by a single factor, nor are they jointly induced by multiple uniformly distributed factors; rather, they are dominated by a small number of high-risk cause nodes, which thus become key elements in accident prevention.

Figure 2 shows the risk distribution among the internal nodes of Cluster 2, which has the highest risk value after clustering. It reveals that there is a distinct risk gradient within Cluster 2 as well. High-risk nodes constitute the core set of causal factors for this cluster, medium-risk nodes serve as important associated triggers, and low-risk nodes act as secondary contributing factors. A comparison with the overall network risk level reveals that the risk values of the high-risk nodes in Cluster 2 fall within the mid-to-high range of the global distribution. This confirms that Cluster 2 is the core causative cluster for ship collision accidents, and its internal factors play a critical driving role in the occurrence of such accidents.

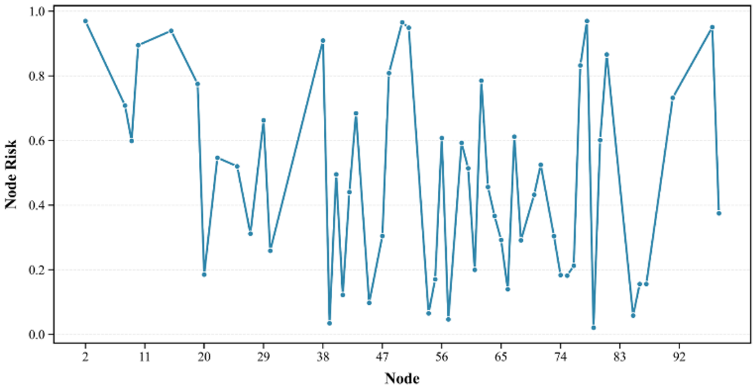

Fig. 2. Risk values for each node in Community 2

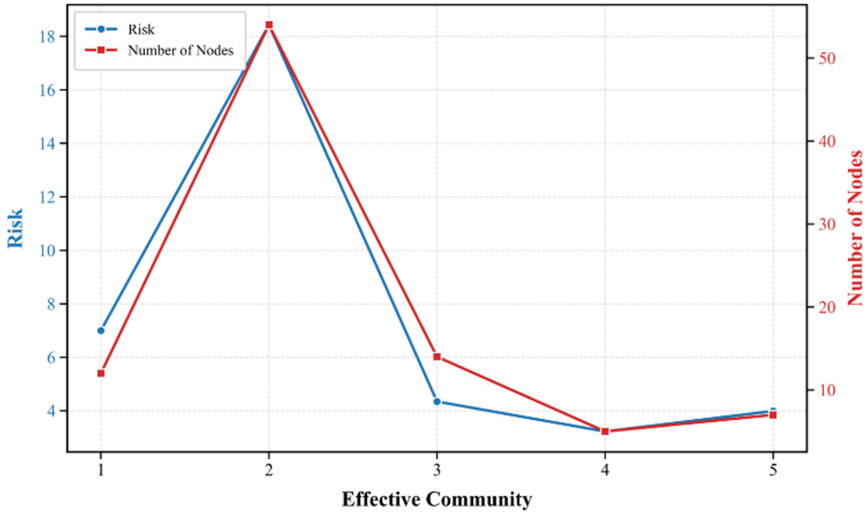

Fig. 3. Risk and Number of Nodes

Figure 3 illustrates the relationship between risk and the number of nodes across the five communities. As shown in the figure, there is not a perfect positive correlation between the number of nodes and risk values in the five valid communities. Community 2 has the highest number of nodes, significantly exceeding that of the other communities, and it also has the highest risk value; Community 1 has 12 nodes, fewer than the 14 in Community 3, yet its risk value is significantly higher than that of Community 3.

This phenomenon—where a community with fewer nodes exhibits higher risk—indicates that the magnitude of risk is not solely determined by the number of causal factors, but is also related to the type of causal factors, the extent of their impact, and their interconnections. Community 1 is primarily composed of human and management-related factors, such as deficiencies in shipping company management, unqualified crew members, insufficient manning, and improper watch arrangements. These factors constitute root causes of accidents; they are highly transmissible, have a broad scope of impact, and a high probability of triggering incidents. Therefore, even with a smaller number of nodes, this community still exhibits a higher risk level. In contrast, Community 3 is dominated by on-site execution factors such as navigation operations, equipment usage, and overtaking rules. Although it contains a larger number of nodes, these are mostly localized operational behaviors with shorter impact chains, resulting in a weaker risk aggregation effect compared to Community 1.

In summary, the risk intensity of collision accident causation clusters is more significantly influenced by the nature of the causal factors. Human factors and management deficiencies contribute far more to the risk value than the number of nodes, which is the primary reason why the risk value of Cluster 1 is higher than that of Cluster 3.

4 Conclusion

This paper proposes a network model for analyzing the causes of ship collision accidents. Based on ship collision accident data from 2020 to 2023, a ship collision accident causation network was constructed, in which nodes represent accident causation factors and edges represent the relationships between different causation factors. The Fast Newman algorithm was used to partition the network into five effective clusters. The risk of accident causes was quantified using the entropy weighting method. The results indicate that Cluster 2 poses a higher risk. Therefore, in daily ship management, priority should be given to controlling the causal factors within Cluster 2. Additionally, since Cluster 2 contains a higher proportion of human factors, greater emphasis should be placed on personnel management and standardizing operational procedures in daily management.

Finally, it is worth considering that there are some limitations in the collection of ship collision accident reports in this study. The data collection channels are relatively single, only obtained through the official website of China Maritime Safety Administration, and the sample size collected is relatively limited. In the follow-up study, we will further broaden the data collection channels and expand the data scale to improve the universality and reliability of the research conclusions.

References

1. Thangalakshmi, S., and K. Sivasami. "Low-Emissions and Energy-Efficient Alternatives for Shipping." Journal of The Institution of Engineers (India): Series C (2026): 1–9, 2026, doi: 10.1007/s40032-026-01317-w.
2. Shu Y, Dong A, Ye B, Liu Chengyong et al. "Liability division for ship collision accidents based on ontology model and bayesian network." Ocean & Coastal Management, 2025, 269: 107824, 2025, doi: 10.1016/j.ocecoaman.2025.107824.
3. Guo M, Chen M, Yuan L, Zhang Z et al. " 4.Investigation of ship collision accident risk factors using BP-DEMATEL method based on HFACS-SCA. " Reliability Engineering & System Safety, 2025, 257: 110875, 2025, doi:10.1016/j.ress.2025.110875.
4. Zhang X, Gan L, Cui H, Shu Y et al. " A hybrid deep learning and large language models framework for ship collision accident analysis. " Reliability Engineering & System Safety, 2026: 112333, 2026, doi: 10.1016/j.ress.2026.112333.
5. Jovanović I, Perčić M, Vladimir N. " Assessment of human contribution to cargo ship accidents using Fault Tree Analysis and Bayesian Network Analysis. " Ocean engineering, 2025, 323: 120628, 2025, doi: 10.1016/j.oceaneng.2025.120628.
6. Merenda J V B S, Bruno O M. " Using deterministic self-avoiding walks as a small-world metric on Watts–Strogatz networks. " Physica A: Statistical Mechanics and its Applications, 2023, 621: 128713, 2023, doi: 10.1016/j.physa.2023.128713.
7. Zhang X, Chen P, Mou J, Chen L et al. " Critical causation factor analysis in ship collision accidents with complex network. " Ocean Engineering, 2025, 315: 119837, 2025, doi: 10.1016/j.oceaneng.2024.119837.
8. Ma J, Tian H, Xu L, Xu Ting et al. " Semi-automatic construction and analysis of complex networks for ship collision accidents. " Ocean & Coastal Management, 2025, 261: 107519, 2025, doi: /10.1016/j.ocecoaman.2024.107519.

9. Qiaoling Du, Xiaoxue Ma, Ruiwen Zhang, and Weiliang Qiao, "Analyzing the Causation of Collision Accidents Between Merchant and Fishing Vessels in China's Coastal Waters by Integrating Association Rules and Complex Networks," Journal of Marine Science and Engineering 13, no. 6 (2025): 1086, 2025, doi: 10.3390/jmse13061086.
10. Tokala, Enduri, Lakshmi, Hajarathaiah et al. "Evaluating community detection algorithms: A focus on effectiveness and efficiency." Journal of Scientometric Research 14.1 (2025): 62-74, 2025, doi: 10.5530/jscires.20250839.