



LJMU Research Online

Feng, Y, Liu, Z, Jiang, Z, Xia, G, Cao, Y, Wang, X and Wang, H

Analysis of factors affecting ship collisions based on association rule mining and complex network theory

<http://researchonline.ljmu.ac.uk/id/eprint/23829/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Feng, Y, Liu, Z, Jiang, Z, Xia, G, Cao, Y, Wang, X and Wang, H (2023) Analysis of factors affecting ship collisions based on association rule mining and complex network theory. Dalian Haishi Daxue Xuebao/Journal of Dalian Maritime University. 49 (3). pp. 31-44. ISSN 1006-7736

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

引用格式:冯胤伟,刘正江,蒋子怡,等. 基于关联规则挖掘和复杂网络理论的船舶碰撞事故影响因素分析[J].大连海事大学学报,2023,49(3):31-44.

基于关联规则挖掘和复杂网络理论的 船舶碰撞事故影响因素分析

冯胤伟¹,刘正江¹,蒋子怡¹,夏国庆¹,曹宇皓²,王新建^{*1},王焕新¹

(1.大连海事大学 航海学院,辽宁 大连 116026;2.利物浦约翰摩尔斯大学 LOOM 研究所,英国 利物浦 L3 3AF)

摘要:为更加科学地分析船舶碰撞事故影响因素的交互关系,揭示船舶碰撞事故的演变机理,以全球船舶碰撞事故报告为依据,建立包含人为因素、船舶因素、管理因素、环境因素及事故时间等五类影响因素的船舶碰撞事故数据库;应用 Apriori 关联规则挖掘算法识别船舶碰撞事故影响因素间的频繁模式、关联、共现或因果关系,运用复杂网络理论将关联规则挖掘的结果可视化;应用拓扑特征分析方法,基于互信息理论的重要节点排序算法和基于边介数中心性的边排序算法实现关键影响因素和边的识别,对事故影响因素交互网络进行鲁棒性分析。结果表明,大部分船舶碰撞事故的影响因素较活跃且影响因素交互网络联系紧密,船舶吨位、船龄、航行水域等影响因素在交互信息传递时较为重要。

关键词:水路运输;影响因素;关联规则;碰撞事故;复杂网络;机器学习

中图分类号:U698 文献标志码:A

文章编号:1006-7736(2023)03-0031-14

doi:10.16411/j.cnki.issn1006-7736.2023.03.004

Analysis of factors affecting ship collisions based on association rule mining and complex network theory

FENG Yinwei¹, LIU Zhengjiang¹, JIANG Ziyi¹,
XIA Guoqing¹, CAO Yuhao², WANG Xinjian^{*1},
WANG Huanxin¹

(1. Navigation College, Dalian Maritime University,

Dalian 116026, China; 2. LOOM Research Institute,

Liverpool John Moores University, Liverpool L3 3AF, UK)

Abstract:In order to analyze the interactive relationship between the influencing factors of ship collision accidents more scientifically and reveal the evolution mechanism of ship collision accidents, a ship collision accident database was established based on the global ship collision accident report, which included five types of influencing factors: human factors, ship factors, management factors, environmental factors, and accident time. The Apriori association rule mining algorithm was used to determine frequent patterns, associations, co-occurrences, and causal relationships among these influential factors. Visual representations of these results were obtained by using complex network theory. The topological analysis methods, important node sorting algorithm based on mutual information theory and edge sorting algorithm based on the centrality of edge mediations were used to identify critical influential factors and edges within the network, and evaluate their robustness. The results indicate that the influencing factors of most ship collision accidents are relatively active and the interaction network of influencing factors is closely connected, and factors such as ship tonnage, age, and navigation water area are more important in interactive information transmission.

Key words:water transportation; influencing factors; association rules; collision accidents; complex networks; machine learning

收稿日期:2023-06-04;修回日期:2023-07-25

基金项目:国家自然科学基金青年科学基金项目(52101399);中央高校基本科研业务费专项资金资助项目(3132023138);大连海事大学博联科研基金项目(3132023617)

作者简介:冯胤伟(1999-),男,硕士生,E-mail:fyinwleo@dlmu.edu.cn;王新建*(1988-),男,博士,副教授,E-mail:wangxinjian@dlmu.edu.cn

0 引言

随着全球经济一体化的发展,作为海运贸易的主要载体,船舶承担着大量的货物运输,航运贸易量不断增加,导致船舶航次增加,船舶事故发生率也随之增加。中国海事局事故报告统计显示,2013—2022年,我国共发生806起水上交通事故,其中,碰撞事故占比高达39.08%。因此,对船舶碰撞事故的影响因素进行分析,对于有效保障船舶安全具有重要意义。

船舶碰撞事故是多种因素耦合作用的结果,而导致船舶碰撞事故的影响因素具有复杂性和不确定性的特点。国内外学者针对船舶碰撞事故的影响因素做了大量研究。司东森等^[1]使用核密度估计和加权平均相结合的贝叶斯网络模型分析集装箱船碰撞事故中的影响因素,得出碰撞事故发生的致因链,并为航运企业及海事管理部门提出合理化建议。王海燕等^[2]利用基于管理疏忽与危险树的模型对“桑吉”轮和“水晶”轮因碰撞导致的溢油事故进行分析,结果表明,船员的违规操作、责任心不强、专业水平低是主要的事故影响因素。Hwang等^[3]应用数据驱动的文本分析方法分析事故调查报告,寻找碰撞事故和由碰撞引起其他事故之间的联系,挖掘碰撞事故向其他事故演化的影响因素,防止事故风险进一步升级。Hasan等^[4]从大量事故报告提取影响碰撞事故的因素,应用故障树分析法确定人为失误是船舶碰撞事故最重要的影响因素,并从船员方面提出相应建议。付姗姗等^[5]利用文本挖掘技术深入分析长江口水域船舶交通事故报告,并构建了人、船、环境和组织等多因素耦合作用下贝叶斯网络模型,研究表明,船舶碰撞事故通常是由环境等客观因素影响人员行为而引起的。廖诗管等^[6]采用贝叶斯层次log-logistic方法对船舶碰撞致因与碰撞频率之间的关系进行分析,结果表明,交通流密度高的区域船舶碰撞概率大,天气和昼夜差异对船舶碰撞概率也有一定影响。Li等^[7]建立了基于关联规则挖掘的贝叶斯网络模型,分析了环境因素和船舶因素对人为错误的影响,研究表明,在环境因素中,对人为错误造成较大影响的主要是能见度、航行水域及时间,在船舶因素中,总吨位、船舶类型及不安全航速对人为错误的影响较为显著。

Weng等^[8]考虑船舶抵进角度、船型、相对速度及昼夜条件对船舶碰撞事故的影响,提出一种船舶动态领域模型来评估船舶碰撞概率,研究结果表明,航道宽度较窄、恶劣天气条件、黑夜等因素对船舶碰撞概率的影响较大。赵志奎等^[9]利用故障树对LNG船碰撞事故严重性进行分析,研究结果表明,LNG船碰撞严重性在社会风险可接受范围内。上述研究采用了文本挖掘技术、故障树理论及贝叶斯网络等方法对船舶碰撞事故的影响因素进行分析。但这些方法还存在一定的局限性,如故障树理论能够分析已知的故障模型,但无法发现新的故障模式;文本挖掘技术能够分析影响因素出现频率,但并未考虑因素之间的相互作用;贝叶斯网络能够分析影响因素之间的耦合关系,但其存在对数据质量要求高、部分先验概率依赖专家知识等缺点。

复杂网络在网络可视化、分析影响因素之间复杂的交互关系、识别关键因素方面较文本挖掘技术、事故树及贝叶斯网络等传统方法更有效。在复杂网络模型建立过程中,边的定义决定了复杂网络能否真实反映出复杂系统的本质特征和规律,从而影响复杂网络的有效性。目前,复杂网络边的构建方法主要有事故链、共现率^[10]及关联规则挖掘^[11-12]等三种,其中,事故链是通过分析事故报告寻找影响因素作用的先后顺序从而构建网络,该方法需要消耗大量人力且引入了过多的主观因素;共现率主要是利用Jaccard指数计算影响因素共同出现的频率,当影响因素共同出现时影响因素间存在边,该方法剔除了主观因素,但共同出现的影响因素并不代表两者之间存在有意义的联系,故也存在一定缺陷;关联规则挖掘技术引入了提升度这一指标,从而剔除了无意义的规则,提高了关联规则的有效性,另外,该方法为纯数据驱动的方法,剔除了主观因素的影响,数据不需要满足特定的分布条件和假设检验条件。近年来,陈伟等^[10]应用关联规则挖掘影响因素之间的耦合关系,构建起重机事故的影响因素复杂网络,通过网络的拓扑性指标识别关键节点,结果表明,控制“相关专业资质不足”、“现场人员滞留/混乱”等关键节点可以有效减少事故发生路径。Deng等^[12]应用关联规则和扎根理论建立地铁运营事故影响因素网络,研究发现,为员工提供必要的培

训有助于减少事故发生。孙家庆等^[11]利用复杂网络理论对水上交通事故的致因因素进行分析,研究发现,“疏忽瞭望”、“判断失误”及“安全意识不足”等是关键的影响因素。

以上研究表明,复杂网络理论可以更有效地将影响因素之间的交互关系可视化,揭示事故的演化规律;关联规则挖掘可以有效挖掘数据之间强有力的关联关系,从而更有效地建立复杂网络模型。然而,综合运用事故数据关联规则挖掘和复杂网络理论建模方法较少应用到水上交通事故研究中,特别是船舶碰撞事故的研究中。为此,本文以全球船舶碰撞事故数据为例,建立涵盖人为因素、船舶因素、管理因素、环境因素及事故时间等信息的船舶碰撞事故数据库,运用基于关联规则挖掘的复杂网络模型对船舶碰撞事故各影响因素间的关系进行深入分析,研究结果可为有关部门制定降低船舶碰撞事故发生率的决策提供理论支持。

1 数据说明与数据预处理

1.1 数据说明

本文选取中国海事局,美国联邦调查局海上事故调查机构,日本交通安全委员会、国家安全运输委员会,加拿大交通安全委员会、海洋事故调查部门及澳大利亚交通安全委员会等全球7家海上调查机构数据库中2000—2019年的事故报告作为主要数据来源。

对上述数据库的调查报告分析发现,由于事故记录是由不同国家上传的,事故记录的详略程度不同,部分数据不准确和/或不完整,因此,按照数据真实性和完整性原则需要删除数据不完整的事报告,在过滤并删除重复事故报告后,得到最初的1294份事故报告。由于本文聚焦船舶碰撞事故,因此,进一步剔除了非碰撞事故数据^[13-14],最终,本文建立的数据库包含从2000—2019年的共322条船舶碰撞事故数据。

1.2 数据预处理

本文数据库中的322条船舶碰撞事故报告提供了每起船舶事故的详细信息。虽然事故调查报告的格式和结构因调查机构不同而有所区别,但每份事故报告均包含事件摘要、事件信息、分析、

结论和建议等信息,其中,事件信息包括事故叙述、船舶信息、船员信息、天气海况信息及公司或监管机构信息等^[13-14]。借鉴安全系统工程中事故致因的分类方法,本文提取事故报告中的人为因素、船舶因素、环境因素、管理因素及事故时间作为数据库的一级特征。对事故报告分析发现,船员自身状况是造成事故的人为因素的主体部分,如在船舶执行任务时,船员的疲劳程度、心理状态、理论知识与经验水平及操作规范性等均会对船舶航行安全产生影响。对船舶因素分析发现,船舶自身存在的缺陷会对船舶航行安全产生较大影响。在环境因素方面,本文主要考虑船舶航行水域的航道状况和自然环境,其中,航道环境包括水域位置、通航密度及吃水型深比等,自然环境则更多是指风、浪、流及能见度情况。在管理因素方面,船公司和监管机构的管理缺陷也会在一定程度上引发船舶碰撞事故。如船公司缺乏安全文化会影响船员的工作态度,从而导致船舶碰撞事故的发生。事故时间是船舶碰撞事故的基本信息,可在时间维度上反映发生船舶碰撞事故的特征。

在数据处理过程中,本文直接应用现有的统一分类标准,如船舶类型的定义和分类均出现在事故报告中,并且该分类标准在行业内得到了广泛应用。但是对于某些不完全一致的分类指标,需要对信息进行整合和标准化处理,如事故报告中并未对“船员心理和生理状态”等进行统一定义,本文利用相关文献^[13,15]和专家意见确定因素的状态等级,如当事故报告中出现船员疲劳、抑郁、嗜睡等描述时,本文将“船员心理和生理状态”定义为“不佳”。同样,当船员在事故中出现沟通困难的问题时,船员的沟通情况被视为“沟通不畅”。除此之外,借鉴相关文献^[13-14]的分级标准,将具有连续特性的变量离散化为相互独立的子集,如“船龄”(1~10年、10~20年、20~30年及大于30年)、“强风等级”(6~7级,8~9级及10~12级)。

在利用文本分析和专家判断方法确定影响因素指标后,为满足关联规则挖掘数据要求,本文建立包含16项影响因素的船舶碰撞事故数据库。各影响因素类别、序号及描述见表1。

表 1 船舶碰撞影响因素数据库
Tab. 1 Ship collision effect factors database

| 类别 | 序号 | 描述 | 类别 | 序号 | 描述 | | |
|-------|----|-----------------------|------------|------------|--------------------------|--------------|-----------|
| 人为因素 | 1 | 船员心理和生理状态不佳 | 管理因素 | 32 | 规则制度不完善 | | |
| | 2 | 船员教育培训不足 | | 33 | 监管不力 | | |
| | 3 | 船员缺乏经验 | | 34 | 安全管理制度不完善 | | |
| | 4 | 沟通不畅 | | 35 | 未及时解决 | | |
| | 5 | 船员操纵失误 | | 36 | 公司缺乏安全文化 | | |
| | 6 | 船员违反规则 | | 37 | 缺乏船上训练 | | |
| 船型 | 7 | 散货船 | 38 | 未按计划进行日常训练 | | | |
| | 8 | 集装箱船 | 39 | 内河水域 | | | |
| | 9 | 油轮 | 40 | 港口水域 | | | |
| | 10 | 客船 | 41 | 沿岸水域 | | | |
| | 11 | 化学品船 | 42 | 开放水域 | | | |
| | 12 | 普通货船 | 能见度 | 43 | 能见度 0.5 n mile ~2 n mile | | |
| | 13 | 渔船 | | 44 | 能见度<0.5 n mile | | |
| | 14 | 游艇和帆船 | 环境因素 | 45 | 6~7 级强风 | | |
| | 15 | 拖船 | | 46 | 8~9 级强风 | | |
| | 16 | 其他船型 | | 47 | 10~12 级风暴 | | |
| 船舶因素 | 船龄 | 17 | 船龄 0~10 年 | 海况 | 48 | 不良海况 6~7 级 | |
| | | 18 | 船龄 10~20 年 | | 49 | 非常不良海况 8~9 级 | |
| | | 19 | 船龄 20~30 年 | | 水流 | 50 | 水流流速≥4 kn |
| | | 20 | 船龄大于 30 年 | | | 51 | 交通拥挤 |
| 总吨位 | 21 | 总吨位<500 t | 吃水型深比 | 52 | 吃水型深比<1.2 | | |
| | 22 | 总吨位 500~3000 t | | 53 | 1.2≤吃水型深比<1.5 | | |
| | 23 | 总吨位>3000 t | | 54 | 1.5≤吃水型深比<3 | | |
| 发动机功率 | 24 | 发动机功率<750 kW | 55 | 吃水型深比≥3 | | | |
| | 25 | 发动机功率 750 kW ~3000 kW | 事故时间 | 56 | 事故时间为 0000—0400 | | |
| | 26 | 发动机功率>3000 kW | | 57 | 事故时间为 0400—0800 | | |
| 航行数据 | 27 | 船舶证书不完整或无效 | | 58 | 事故时间为 0800—1200 | | |
| | 28 | 船舶配员不足 | | 59 | 事故时间为 1200—1600 | | |
| | 29 | 海员证不完整或无效 | | 60 | 事故时间为 1600—2000 | | |
| | 30 | 适航性差 | | 61 | 事故时间为 2000—2400 | | |
| | 31 | 未明确 PSC/FSC 检查 | | | | | |

注：风力等级的单位为“蒲福风级”；海况等级的单位为“道格拉斯标度”。

2 基于关联规则复杂网络模型构建

基于关联规则的复杂网络模型构建分为 3 个阶段：第 1 阶段为利用关联规则挖掘技术对船舶碰撞事故的影响因素进行分析，得到各影响因素之间的关联关系；第 2 阶段则在各影响因素间关联关系的基础上构建复杂网络，并分析网络的拓扑特性，以及节点与边的重要程度；第 3 阶段为网络鲁棒性分析。模型构建的具体流程如图 1 所示。

2.1 关联规则挖掘

关联规则挖掘是一种机器学习算法，通过提取频繁项集发现数据中潜在的关系^[6]。关联规则挖掘具有多种算法，如 FP-Growth、Apriori、Eclat 和 PrefixSpan 等。通常意义上来说，不同的关联规则挖掘算法只是在运行时间和挖掘模式上有所区别，但其最终得到的频繁项集均是相同的，其

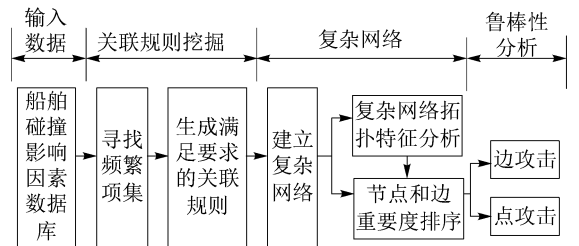


图1 基于关联规则的复杂网络模型构建

Fig.1 Complex network models construction based on association rules

中,Apriori 是目前运用最广泛、使用最简便的算法,尽管 Apriori 算法需要多次遍历数据集,从而导致运算时间较其他算法长,但是,鉴于本文数据库只有 322 条数据,算法实际运算时间只需要 36.2 ms±284 μs(在 MacBook Pro M2 系统下运行 10 次,每次运行循环 100 次 Python 3.0 程序),因此,本文选择 Apriori 算法对船舶碰撞事故影响因素的关联规则进行挖掘。

Apriori 算法是一种基于最小支持度和最小置信度逐层搜索频繁项集的迭代算法^[7]。首先,扫描数据库,计算每个项出现的频数,收集满足最小支持度的项,得到频繁项集 L_1 ; 然后,通过频繁项集 L_1 找出频繁项集 L_2 , 利用频繁项集 L_2 找出频繁项集 L_3 , 不断迭代,直至无法再找到频繁项集 L_k ; 最后,根据得到的所有频繁项集计算各项集间的关联规则,选择大于最小支持度和最小置信度且提升度大于 1 的关联规则,并将其作为有价值的关联规则,输出结果。关联规则的计算公式为:

$$Supp(X) = \frac{N_X}{N} \quad (1)$$

$$Con(X \Rightarrow Y) = \frac{Supp(XY)}{Supp(X)} \quad (2)$$

$$Lift(X \Rightarrow Y) = \frac{Con(X \Rightarrow Y)}{Supp(Y)} \quad (3)$$

式中: $Supp(X)$ 为 X 事务的支持度; N 为项目集中事务总数; N_X 为项目集中 X 事务出现的次数; $X \Rightarrow Y$ 为事务 X 作为前项事务、事务 Y 作为后项事务的关联规则; $Con(X \Rightarrow Y)$ 为 $X \Rightarrow Y$ 的置信度; $Lift(X \Rightarrow Y)$ 为 $X \Rightarrow Y$ 的提升度。

2.2 复杂网络模型

复杂网络是一种特殊的网络结构,是将复杂系统中的元素抽象为节点、元素间的关系抽象为边的网络结构模型,已被广泛用于铁路交通事故、道路交通事故影响因素之间的关系分析^[6-7]。本文将复杂网络应用于船舶碰撞事故影响因素分析,利用各影响因素间的关联规则建立加权的有向复杂网络。

2.2.1 复杂网络构建

设船舶碰撞事故影响因素交互网络为 $DiG = (N, DiE)$, 其中, $N = \{n_1, n_2, \dots, n_p\}$ 为网络节点集合; $DiE = \{die_1, die_2, \dots, die_q\}$ 为网络边集合, 则网络的数学表达描述为邻接矩阵 G ,

$$G = \begin{matrix} & \begin{matrix} n_1 & n_2 & \cdots & n_p \end{matrix} \\ \begin{matrix} n_1 \\ n_2 \\ \vdots \\ n_p \end{matrix} & \begin{bmatrix} g_{11} & g_{12} & \cdots & g_{1p} \\ g_{21} & g_{22} & \cdots & g_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ g_{p1} & g_{p2} & \cdots & g_{pp} \end{bmatrix} \end{matrix} \quad (4)$$

$$g_{ij} = w_{ij} e_{ij} \quad (5)$$

$$e_{ij} = \begin{cases} 1, & i \Rightarrow j \text{ 存在关联规则} \\ 0, & i \Rightarrow j \text{ 不存在关联规则} \end{cases} \quad (6)$$

式中: g_{ij} 为节点 n_i 到节点 n_j 的连接情况; w_{ij} 为网络中节点 n_i 到节点 n_j 的边的权重(即 $i \Rightarrow j$ 的置信度); e_{ij} 为节点 n_i 到节点 n_j 是否连接, 如果连接, e_{ij} 为 1, 反之为 0。

2.2.2 网络的拓扑特征

网络的静态几何量为给定网络微观量的统计分布和宏观统计的平均值,是分析复杂网络拓扑特征的基础。对于有向加权网络,其静态几何量主要包括节点度、节点强度、介数中心系数、接近中心系数、网络聚集系数等。

(1) 节点度

节点度为连接节点的边的总数。有向网络的度分为节点总度、节点出度及节点入度,节点 n_i 的总度为连接节点 n_i 的边数,节点 n_i 的出度为从节点 n_i 到其他节点的边数,节点 n_i 的入度为从其他节点到节点 n_i 的边数。在社交网络中,节点入度被视为声望,节点出度被视为合群性。计算公式为:

$$k_i^m = \sum_{j, j \neq i} e_{ji} \quad (7)$$

$$k_i^{out} = \sum_{j, j \neq i} e_{ij} \quad (8)$$

$$k_i^{total} = k_i^{in} + k_i^{out} \quad (9)$$

式中: k_i^{in} 为节点 n_i 的入度; k_i^{out} 为节点 n_i 的出度; k_i^{total} 为节点 n_i 的总度。

(2) 节点强度

节点强度为连接节点边的权重之和。有向网络的节点强度分为节点总强度、节点出强度及节点入强度,节点 n_i 的总强度为节点 n_i 各边的权重之和,节点 n_i 的出强度为从节点 n_i 到其他节点的权重之和,节点 n_i 的入强度为从其他节点到节点 n_i 的权重之和,具体计算公式为:

$$S_i^{in} = \sum_{j, j \neq i} w_{ji} e_{ji} \quad (10)$$

$$S_i^{out} = \sum_{j, j \neq i} w_{ij} e_{ij} \quad (11)$$

$$S_i^{total} = S_i^{in} + S_i^{out} \quad (12)$$

式中: S_i^{in} 为节点 n_i 的入强度; S_i^{out} 为节点 n_i 的出强度; S_i^{total} 为节点 n_i 的总强度。

(3) 节点介数中心系数

节点介数中心系数为一个网络中通过节点的最短路径数。介数中心系数反映了节点的枢纽性和传递性,计算公式为:

$$B_v = \sum_{\substack{\forall i, j \in N \\ i \neq v \neq j}} \left(\frac{\sigma_{ij}(v)}{\sigma_{ij}} \right) \quad (13)$$

$$BC_v = \frac{B_v}{(p-1)(p-2)} \quad (14)$$

式中: $\sigma_{ij}(v)$ 为经过节点 n_v 的节点 n_i 到节点 n_j 的最短路径数; σ_{ij} 为节点 n_i 到节点 n_j 的最短路径数; B_v 为节点 n_v 的介数; p 为网络中节点数目; BC_v 为节点 n_v 的介数中心系数。

(4) 节点接近中心系数

节点接近中心系数反映了节点在网络中居于中心的程度,是衡量节点的中心性指标之一,计算公式为:

$$CC_i = \frac{p-1}{\sum_{j=1, j \neq i}^p l_{ij}} \quad (15)$$

式中: CC_i 为节点 n_i 的接近中心系数; l_{ij} 为节点 n_i 到 n_j 的最短路径长度。

(5) 网络平均聚集系数

聚集系数表示网络节点的聚合程度。网络平均聚集系数(局部聚集系数)计算公式为:

$$C_i = \frac{2E_i}{k_i^{total}(k_i^{total} - 1)} \quad (16)$$

$$AC = \frac{\sum_i C_i}{p} \quad (17)$$

式中: E_i 为节点 n_i 邻居节点的边数; C_i 为各节点的聚集系数; AC 为网络平均聚集系数。

(6) 网络平均最短路径长度

网络平均最短路径长度表示整个网络任意两节点间最短路径长度的平均值,其计算公式为:

$$AP = \frac{\sum_{i=1}^p \sum_{j=1}^p l_{ij}}{p(p-1)} \quad (18)$$

式中: AP 为网络平均最短路径长度。

(7) 网络传递性

网络传递性(全局聚集系数)是指网络闭合三角形占所有可能三角形的比例,其计算公式为:

$$CT = \frac{3\Delta}{\Phi} \quad (19)$$

式中: CT 为网络传递性; Δ 为网络中三角形个数; Φ 为网络中具有公共顶点的边的对数。

2.2.3 网络节点和边的重要度排序

网络节点和边的重要度排序对识别关键影响因素、降低船舶碰撞事故发生率具有重要意义。

(1) 节点重要度排序

许多学者从不同角度提出了多种算法对节点重要度进行精准排序,这些算法一般均是通过计算获得每个节点的得分,对节点按照得分从高到低进行排序。常见的节点重要度排序算法有 PageRank 算法、Weight LeaderRank (WLR) 算法等,但是这些算法的迭代次数多,计算复杂度高。

本文采用一种基于互信息理论的节点重要度排序算法^[17],该方法既考虑了网络节点间的连通强度,也考虑了网络的拓扑特性,相比其他算法,该方法更加准确且计算复杂度更低。节点信息量计算公式为:

$$I_{ij} = \begin{cases} \ln \frac{S_i^{out}}{S_j^{in}}, & e_{ij} = 1 \\ 0, & e_{ij} = 0 \end{cases} \quad (20)$$

$$I(i) = \sum_{m \neq i} I_{im} - \sum_{n \neq i} I_{ni} \quad (21)$$

式中: I_{ij} 为节点 n_i 到节点 n_j 的互信息; $I(i)$ 为节点 n_i 的信息量; m 和 n 分别为网络中非节点 n_i 的节点 n_m 和 n_n 。

(2) 边重要度排序

相较节点重要度排序,边重要度排序的研究较少。在边重要度排序研究中,边的介数是一种最直接的度量方法。文献[18-19]研究表明,通过一条边的最短路径越多,则该边获得的信息量就越大,同时该边也越重要。因此,本文采用边介数中心系数作为衡量边的重要性指标,其公式表示为:

$$EB_v = \sum_{\substack{\forall i, j \in N \\ v \neq (i, j), (j, i)}} \left(\frac{E\sigma_{ij}(v)}{\sigma_{ij}} \right) \quad (22)$$

$$EBC_v = \frac{EB_v}{(p-1)(p-2)} \quad (23)$$

式中: $E\sigma_{ij}(v)$ 为经过边 die_v 的节点 n_i 到节点 n_j 的最短路数目; EB_v 为边 die_v 的介数; EBC_v 为边 die_v 的介数中心系数。

2.3 鲁棒性

复杂网络鲁棒性是指当网络中的一部分节点或边出现故障时,整个网络还能保持连通的能力。鲁棒性可在一定程度上反映网络的稳定性和持续性。本文将复杂网络的可达性指标作为评价网络鲁棒性的标准。在对鲁棒性研究过程中,对船舶碰撞事故影响因素交互网络使用随机攻击和蓄意攻击两种攻击方式。在随机攻击时,采用蒙特卡洛模拟算法分别随机产生失效节点和失效边,且失效节点和边的数量从1逐步增加至全部节点和边;重复 K 次实验得到复杂网络鲁棒性变化。在蓄意攻击时,利用式(21)和式(23)得到节点和边的重要度排序结果,分别对网络节点和边进行攻击。可达性计算公式为:

$$RP = \begin{bmatrix} RP_{11} & \cdots & RP_{1p} \\ \vdots & \vdots & \vdots \\ RP_{p1} & \cdots & RP_{pp} \end{bmatrix} \quad (24)$$

$$RP_{ij} = \begin{cases} 1, & \text{节点 } n_i \text{ 可以到达节点 } n_j \\ 0, & \text{节点 } n_i \text{ 不能到达节点 } n_j \end{cases} \quad (25)$$

$$R = (\text{ones})_{1 \times p} \cdot RP \cdot (\text{ones})_{p \times 1} \quad (26)$$

$$NR = \frac{R_{\text{attack}}}{R_o} \quad (27)$$

式中: RP 为可达矩阵; RP_{ij} 为节点 n_i 到 n_j 的可达状态; R 为网络可达性; $(\text{ones})_{1 \times p}$ 和 $(\text{ones})_{p \times 1}$ 分别为元素全为1的1行 p 列和 p 行1列矩阵; R_{attack} 为被攻击后网络可达性; R_o 为被攻击前网络可达性; NR 为网络鲁棒性。

3 船舶碰撞事故影响因素分析

3.1 影响因素的关联规则

为找到各影响因素之间的潜在联系,需要对数据进行关联规则分析。关联规则挖掘阈值的选择对于关联规则的质量有很大影响。本文采用试错法寻找合适的阈值。首先,选择最小支持度和最小置信度分别为0.01和0.1及0.2和0.5的两

组阈值来判断船舶碰撞事故的关联规则。在使用第一组阈值时,1210条关联规则被挖掘出来;在使用第二组阈值时,只有63条关联规则被挖掘出来。经过上述两次实验,发现这些结果并未提供合适的关联规则,这会使得探索影响因素间的相互作用变得困难。为关注强大的关联规则,经过多次试验,本文最终将最小支持度阈值和最小置信度阈值分别设置为0.1和0.3。此外,为使构建的复杂网络的每一个节点只代表一个影响因素,最大限制长度设置为2。

本文使用 Apriori 算法挖掘了 293 条关联规则。置信度大小反映了关联规则的强弱,也反映了复杂网络中边的权重。置信度排名越靠前,复杂网络中相邻两影响因素之间的交互作用越强。各关联规则的分布如图2所示,其中,置信度排名前10的关联规则如表2所示。由图2可知,关联规则主要分布在支持度为0.1~0.2、置信度为0.3~0.5的范围内。同时,提升度较高的关联规则大都分布在稀疏区域。由表2可知,排名靠前的关联规则几乎均与船舶因素有关,如吨位大的船舶,发动机功率高,说明网络中船舶因素之间的交互作用是最强的;对于发生碰撞事故的集装箱船舶,其发动机功率大于3000 kW的概率高达93.48%;发生碰撞事故的渔船大多为小型船舶;发生碰撞事故时,船员教育培训不足与船员缺乏经验、未按计划进行日常训练与缺乏船上训练具有很强的关联关系。此外,由表2还可以发现一个潜在的关联关系,即当船舶处于吃水型深比小于1.2的条件下,89.06%的碰撞事故与船员缺乏经验有关。

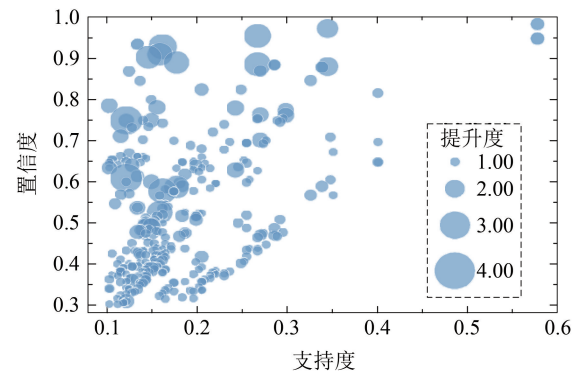


图2 船舶碰撞事故影响因素关联规则分布

Fig.2 Distribution of effect factors association rules for waterborne traffic collisions

表 2 置信度排名前 10 的关联规则

Tab.2 Top 10 association rules ranked by confidence values

| 排序 | 前项事务 | 后项事务 | 支持度 | 置信度 | 提升度 |
|----|-----------------------|-----------------------|--------|--------|--------|
| 1 | 22 (总吨位 500 t~3000 t) | 26 (发动机功率>3000 kW) | 0.5776 | 0.9841 | 1.6168 |
| 2 | 38 (未按计划进行日常训练) | 37 (缺乏船上训练) | 0.3447 | 0.9737 | 2.4883 |
| 3 | 24 (发动机功率<750 kW) | 21 (总吨位<500 t) | 0.2671 | 0.9556 | 3.1721 |
| 4 | 26 (发动机功率>3000 kW) | 22 (总吨位 500 t~3000 t) | 0.5776 | 0.9490 | 1.6168 |
| 5 | 8 (集装箱船) | 26 (发动机功率>3000 kW) | 0.1335 | 0.9348 | 1.5357 |
| 6 | 13 (渔船) | 24 (发动机功率<750 kW) | 0.1615 | 0.9286 | 3.3222 |
| 7 | 13 (渔船) | 21 (总吨位<500 t) | 0.1584 | 0.9107 | 3.0232 |
| 8 | 2 (船员教育培训不足) | 3 (船员缺乏经验) | 0.1460 | 0.9038 | 3.0003 |
| 9 | 52 (吃水型深比<1.2) | 3 (船员缺乏经验) | 0.1770 | 0.8906 | 2.9565 |
| 10 | 21 (总吨位<500 t) | 24 (发动机功率<750 kW) | 0.2671 | 0.8866 | 3.1721 |

3.2 影响因素的复杂网络

基于 3.1 节挖掘出的关联规则,本文利用式

(4)~(6)建立由 34 个节点和 293 条有向边组成的加权网络(图 3)。

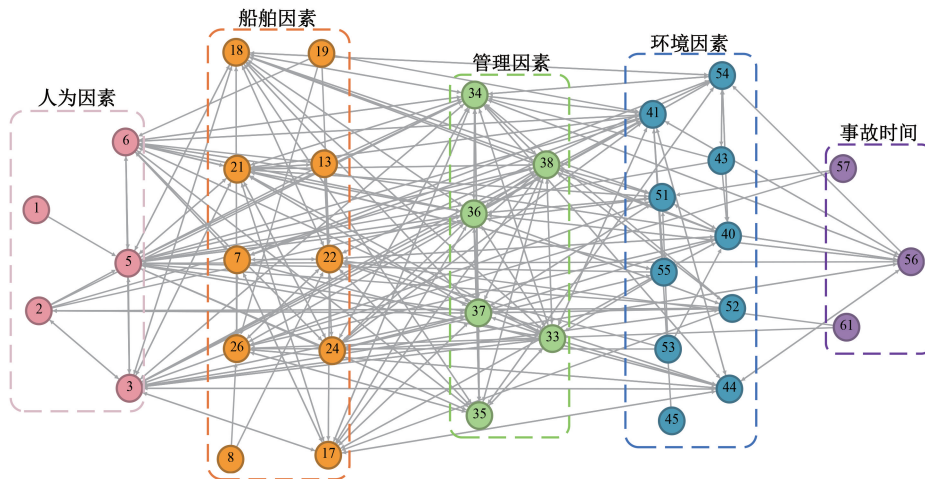


图 3 船舶碰撞事故影响因素的交互网络

Fig. 3 An interactive network of effect factors for waterborne traffic collisions

为探索船舶碰撞事故影响因素间的交互特性,对已建立的复杂网络的拓扑特性进行分析。表3为该网络基本的拓扑特征。由表3可以看

表3 影响因素交互网络的拓扑特征

Tab.3 Topological characteristics of effect factor interaction networks

| 拓扑指标 | 结果 | 拓扑指标 | 结果 |
|---------------------|-------|--------------|-------|
| 节点数 | 34 | $\max(BC_i)$ | 0.065 |
| 边数 | 293 | $\min(BC_i)$ | 0 |
| $\max(k_i^{total})$ | 36 | $\max(CC_i)$ | 0.69 |
| $\min(k_i^{total})$ | 1 | $\min(CC_i)$ | 0 |
| $\max(S_i^{total})$ | 21.39 | AP | 1.49 |
| $\min(S_i^{total})$ | 0.65 | CT | 0.62 |
| AC | 0.57 | | |

出,网络的平均聚集系数(AC)和传递效率(CT)远高于同等规模随机网络的平均聚集系数(约为 0.26)和传递效率(约为 0.26),其中,约 38%的节点聚集系数为 0.6~1.0,表明网络具有较高的聚集性,各节点间联系程度较为紧密,即节点间更容易快速产生交叉影响。由表 3 还可以发现,网络的平均最短路径长度(AP)小于同等规模随机网络(同等规模随机网络的相关参数为代码生成的 1000 个随机网络的参数平均值)的平均最短路径长度(约为 1.81),表明网络中任意不相邻的两点之间均可能在很短的距离内产生联系,即不直接交叉的影响因素之间也可通过较短的连锁反应间接地交叉影响,这样联系紧密且传播效率高的网

络给降低船舶碰撞事故率带来了巨大挑战。

3.2.1 网络基本拓扑指标分析

在复杂网络中,入度和入强度反映了节点被影响的程度,出度和出强度反映了某节点对其他节点造成影响的程度,总度和总强度反映了节点在整个复杂系统中的中心程度,即中心程度越高,节点对复杂系统的影响越大。介数中心系数和接近中心系数反映了节点在复杂系统中对信息传递的枢纽作用,枢纽作用越大的因素,在网络中越处于中心位置。本文主要从节点度和强度、节点介数中心系数和接近中心系数两方面对网络基本拓扑指标进行分析。

(1) 节点度和强度

节点度表示影响因素之间交互的多样性。由图 4(a)可以看出,网络中大部分影响因素均是较为活跃的,其中,总度值最大的影响因素为 34(安全管理制度不完善),然后依次为 3(船员缺乏经验)、36(公司缺乏安全文化)、5(船员操纵失误)及 37(缺乏船上训练),表明这些影响因素会与其他大多数影响因素产生交互影响。总度值最小的两个影响因素为 1(船员心理和生理状态不佳)和 45(6~7 级强风),这些因素在网络中只有一个邻接点。除节点总度值外,有向网络的度值还分为入度和出度两部分,入度值较高的影响因素有 34(安全管理制度不完善)、5(船员操纵失误)及 36(公司缺乏安全文化),其值均 ≥ 17,表明其会直接接收网络中一半以上影响因素所传递的交互信

息。入度值为 0 的影响因素为 1(船员心理和生理状态不佳)、45(6~7 级强风)、57(事故时间为 0400—0800)、61(事故时间为 2000—2400)、8(集装箱船)、19(船龄 20~30 年)及 53(1.2 ≤ 吃水型深比 < 1.5),这些影响因素不会接收其他影响因素传递的交互信息。出度值较高的影响因素有 3(船员缺乏经验)、34(安全管理制度不完善)、37(缺乏船上训练)及 36(公司缺乏安全文化),这些影响因素的出度值均 ≥ 15,说明其能直接向网络中大部分节点传递交互信息。出度值最低的影响因素为 1(船员心理和生理状态不佳)和 45(6~7 级强风),这两种影响因素的出度值和总度值均为 1,说明其在网络中只会向一个影响因素传递信息,且不会接收其他网络传递的信息。

节点强度表示网络中影响因素产生交互影响的概率。由图 4(b)可以看出,节点度和节点强度的分布趋势大致相同,只是在部分节点上可能会有区别。在节点总强度方面,强度较大的影响因素有 34(安全管理制度不完善)、5(船员操纵失误)、36(公司缺乏安全文化)、37(缺乏船上训练)及 3(船员缺乏经验),其中所包含的影响因素与总度较大的影响因素一样,但顺序发生了改变,说明度值相似的影响因素所能发生交互影响的概率也会有较大差别,例如,3(船员缺乏经验)和 5(船员操纵失误)的总度值相同,但是 3(船员缺乏经验)的总强度小于 5(船员操纵失误),表明 3(船员缺乏经验)在网络中的影响力大于 5(船员操纵

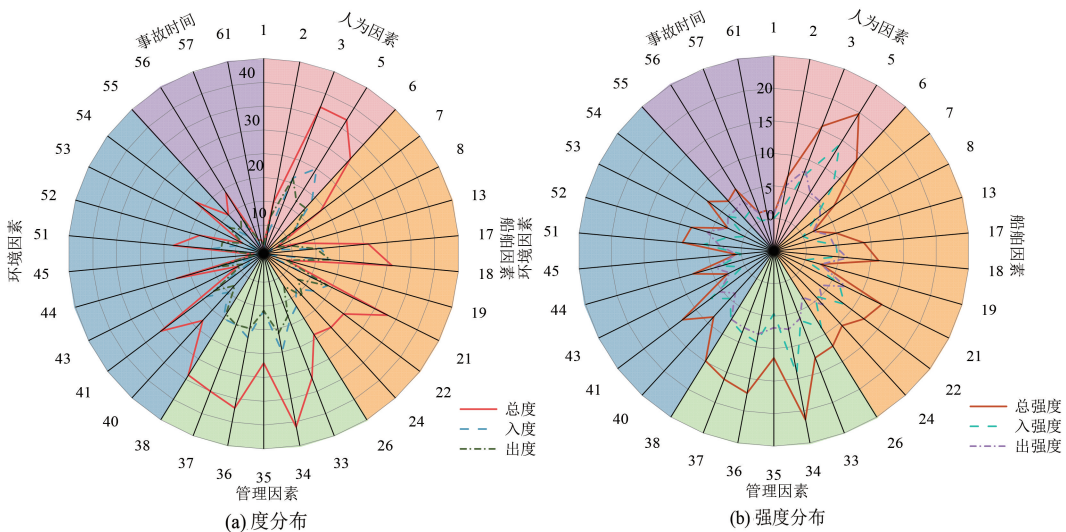


图 4 网络节点度和强度分布

Fig. 4 Network nodes degree and intensity distribution

失误)。在节点入强度方面,26(发动机功率>3000 kW)的入度小于41(沿岸水域),但26(发动机功率>3000 kW)的入强度大于41(沿岸水域),表明虽然26(发动机功率>3000 kW)只能直接接收较少影响因素所传递的信息,但其接收到交互信息的概率更大。在节点出强度方面,52(吃水型深比<1.2)的出度小于6(船员违反规则),但52(吃水型深比<1.2)的出度强度大于6(船员违反规则),表明虽然能直接接收52(吃水型深比<1.2)所传递的交互信息的因素较少,但其所传递的信息被直接接收的概率更大。

综合节点度和强度来看,人为因素和管理因素的度和强度均较大,表明人为因素和管理因素与其他影响因素联系紧密,影响力较大。虽然这两类影响因素占比较小,但在网络中占据主要地位,其失效后对整个网络的影响也较大。因此,加强对人为因素和管理因素的防控是必要的。

(2) 节点介数中心系数和接近中心系数

图5为节点介数中心系数和接近中心系数分布。由图5可知,在船舶碰撞事故影响因素网络中,34(安全管理制度不完善)的介数中心系数最大,说明该影响因素发挥着最重要的信息传递枢纽作用。当该影响因素的状态改变时,网络中信息的传递效率会发生较大幅度的改变,影响网络的稳定性。在船舶实际航行过程中,安全管理制度不完善是连接多种影响因素的桥梁,会增加船

员失误、违反规则等影响因素发生的概率,也会加速事故风险的信息传递,从而导致船舶碰撞事故的发生。

由图5还可以发现,接近中心系数较大的影响因素有5(船员操纵失误)、34(安全管理制度不完善)、36(公司缺乏安全文化)、3(船员缺乏经验)、18(船龄10~20年)、37(缺乏船上训练)、38(未按计划进行日常训练)及21(总吨位<500 t),说明这几个影响因素到其他影响因素的距离很短,处于网络中心位置且传递的信息能快速到达其他影响因素,其中,3(船员缺乏经验)和5(船员操纵失误)属于人为因素,以节点3为例,当船舶遇到极端天气时,缺乏经验的船员不能熟练应对各种风险,从而会导致船舶碰撞事故发生。大量研究表明,80%的水上交通事故均与人为因素有关^[16],这佐证了人为因素在船舶碰撞事故影响因素网络中的重要性。34(安全管理制度不完善)、36(公司缺乏安全文化)、37(缺乏船上训练)及38(未按计划进行日常训练)属于管理因素,以36为例,当公司缺乏安全文化时,出现船员教育培训不足、船员违反规则及缺乏船上训练等情况的概率就会增加,从而导致船舶碰撞事故发生。18(船龄10~20年)及21(总吨位<500 t)为船舶因素,以21为例,当船舶总吨位小于500 t时,船舶尺度小,遇到天气恶劣、船员失误等情况时,小型船舶更容易发生碰撞事故^[13]。

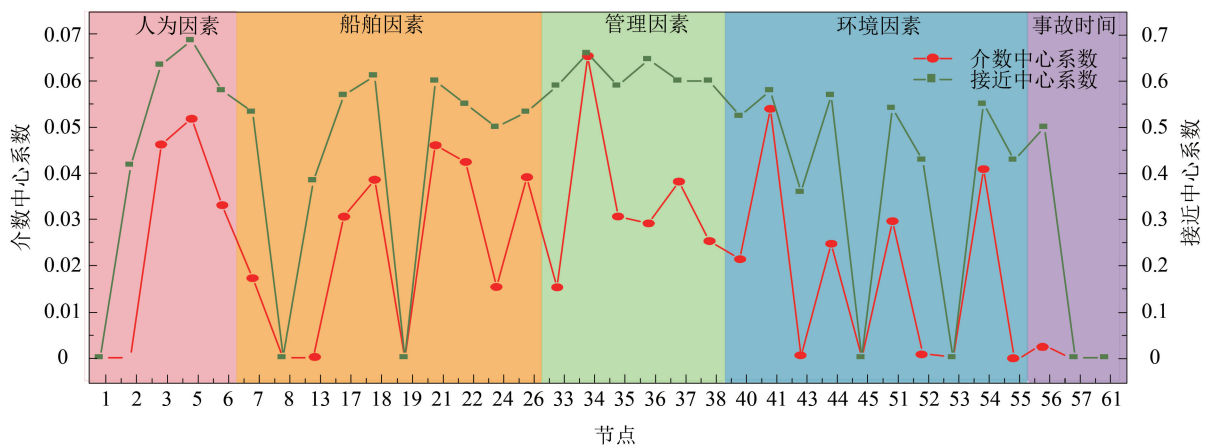


图5 节点介数中心系数和接近中心系数分布

Fig.5 Distribution of nodal between-center coefficients and near-center coefficients

3.2.2 网络重要节点和边分析

为预防事故及快速有效地阻止风险演化,需要对网络中关键节点和边进行分析。在复杂系统

中,信息量越大的节点越重要,而信息量需要通过置信度计算。对于网络中的边,越有助于系统信息传递的边越重要。

(1)重要节点

本文基于互信息理论计算各节点的信息量,并根据信息量对各节点进行排序,排序结果如表 4 所示。由表 4 可知,管理因素的信息量均排名前 13,说明所有管理因素在船舶碰撞事故影响因素网络中均占据重要的地位。对管理因素进行防

控,对降低碰撞事故发生率有较好的效果。事故发生时间均排在末尾,表明时间在影响因素网络中占据较不重要的地位。因此,在夜晚 20 点至清晨 8 点间加强安全管理,便能有效延缓风险演化,降低事故发生率。

表 4 基于互信息理论的节点重要度排序

Tab. 4 Nodes importance ranking based on mutual information theory

| 排序 | 节点 | 信息量 | 排序 | 节点 | 信息量 | 排序 | 节点 | 信息量 |
|----|----|-------|----|----|--------|----|----|--------|
| 1 | 34 | 56.92 | 13 | 35 | 17.47 | 25 | 56 | -23.43 |
| 2 | 5 | 49.99 | 14 | 18 | 16.71 | 26 | 8 | -24.49 |
| 3 | 36 | 45.08 | 15 | 24 | 12.52 | 27 | 2 | -30.33 |
| 4 | 37 | 40.99 | 16 | 51 | 5.24 | 28 | 61 | -33.16 |
| 5 | 3 | 37.78 | 17 | 17 | 4.46 | 29 | 55 | -33.24 |
| 6 | 38 | 35.91 | 18 | 40 | 3.53 | 30 | 57 | -33.58 |
| 7 | 6 | 31.62 | 19 | 44 | -4.36 | 31 | 13 | -46.70 |
| 8 | 21 | 29.00 | 20 | 54 | -4.86 | 32 | 45 | -46.94 |
| 9 | 33 | 25.76 | 21 | 52 | -8.94 | 33 | 1 | -52.18 |
| 10 | 41 | 25.73 | 22 | 53 | -12.00 | 34 | 43 | -95.30 |
| 11 | 22 | 22.10 | 23 | 7 | -15.54 | | | |
| 12 | 26 | 20.00 | 24 | 19 | -15.76 | | | |

(2)重要边

本文根据边介数中心系数对网络中风险演化路径(即边)进行排序,排名前 10 的边如表 5 所示。由表 5 可以看出,较重要的边为 54(1.5 ≤ 吃水型深比 < 3)至 43(能见度 0.5~2 n mile)、1(船员心理和生理状态不佳)至 5(船员操纵失误)、45

(6~7 级强风)至 41(沿岸水域)及 21(总吨位 < 500 t)至 13(渔船),说明网络中大多数信息在传递时均会经过这些边。当阻断这些边时,网络结构会发生较大变化。因此,当某些因素不能直接防控时,可以通过阻断相应的边来达到预防船舶碰撞事故发生的目的。

表 5 介数中心系数排名前 10 的边

Tab. 5 Top 10 edges of the between-center coefficient

| 排序 | 边 | 介数中心系数 | 排序 | 边 | 介数中心系数 |
|----|----------|--------|----|----------|--------|
| 1 | (54, 43) | 0.0296 | 6 | (57, 51) | 0.0151 |
| 2 | (1, 5) | 0.0241 | 7 | (3, 2) | 0.0149 |
| 3 | (45, 41) | 0.0241 | 8 | (22, 17) | 0.0148 |
| 4 | (21, 13) | 0.0209 | 9 | (8, 22) | 0.0140 |
| 5 | (34, 54) | 0.0189 | 10 | (61, 22) | 0.0140 |

3.3 鲁棒性分析

网络的鲁棒性是指当网络某些节点或边失效时,网络保持稳定的能力。破坏海上交通事故影响因素的交互网络、降低网络的稳定性可在一定程度上阻止事故的演化。本文分别攻击网络节点和边,并分析网络鲁棒性变化,旨在为降低船舶碰撞事故发生率提供决策支持。

3.3.1 点攻击

采用蒙特卡洛模拟随机攻击和按节点重要度排序的蓄意攻击两种策略,分别对网络节点进行攻击并记录网络鲁棒性变化(图 6)。由图 6 可知,不论是随机攻击还是蓄意攻击,随着失效节点数量增加,网络鲁棒性逐渐降低。对比可知,蓄意攻击和随机攻击对网络破坏程度不同,蓄意攻击对网络破坏的速度较随机攻击更快,说明有针对

性地防控关键影响因素对防止船舶碰撞事故发生更加有效。相较基于 WLR 和 PageRank 的蓄意攻击,基于互信息理论的蓄意攻击能够更快使网络完全崩溃,这也说明基于互信息理论的节点重要度排序方法更适用于船舶碰撞事故影响因素网络。

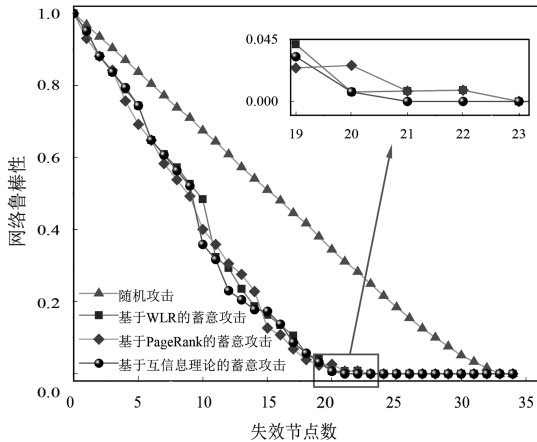


图 6 网络节点失效的鲁棒性变化

Fig. 6 Robust variation of network nodes failure

3.3.2 边攻击

采用蒙特卡洛模拟随机攻击和按边重要度排序的蓄意攻击两种策略,分别对网络的边进行攻击并记录网络鲁棒性变化(图 7)。由图 7 可知,当失效边数达到 150 时,随机攻击边的网络鲁棒性下降 10.18%,蓄意攻击边的网络鲁棒性下降 44.67%,此时,随机攻击时关键边失效的概率较低,只有当绝大部分边被攻击时,网络鲁棒性才会大幅度下降;虽然蓄意攻击能快速降低网络鲁棒性,但是,只有攻击所有边时才能使网络鲁棒性为零。这是因为本文的网络鲁棒性是以网络可达性作为指标,网络中只要存在边,可达性就大于零。

由图 7 还可以发现,蓄意攻击边的网络鲁棒性是呈阶梯性下降的,即失效的边数在一定范围时,网络的鲁棒性并不会发生改变;只有失效边数超出该范围,网络鲁棒性才会继续下降。为研究在网络鲁棒性未发生改变时网络结构的变化,本文计算了蓄意攻击时网络平均最短路径长度的变化情况(图 7 实线)。由图 7 可以发现,网络平均最短路径长度的变化与网络鲁棒性变化同步;但是,当网络鲁棒性维持不变时,网络平均最短路径长度开始增加。这是因为本文根据边介数中心系数的大小顺序对边进行蓄意攻击,而边介数根据

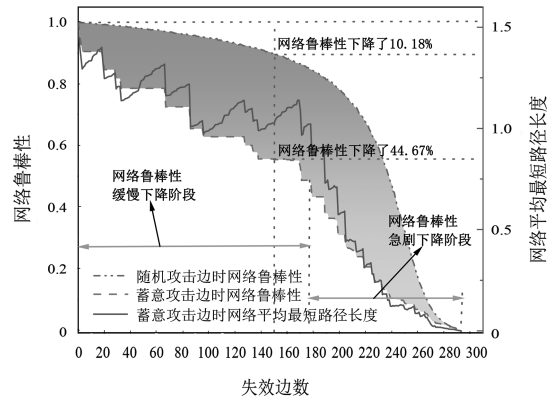


图 7 网络边失效的鲁棒性变化

Fig. 7 Robust variation of network edges failure

通过该边的最短路径数目计算,因此,边失效直接改变了各节点间的最短路径长度,而网络中节点间的可达性并不一定会立刻改变。

虽然对网络的边蓄意攻击不会直接使网络完全崩溃,但能使网络中各影响因素交互更加困难。因此,边攻击可在节点攻击措施较难直接实施时起到辅助作用,从而降低船舶碰撞事故发生概率。例如,44(能见度 <0.5 n mile) \rightarrow 7(散货船)只需一步便能发生交互作用,且两者交互的概率为 39.43%,即当能见度 <0.5 n mile 时,散货箱船发生碰撞事故的概率为 39.43%。如果不能直接对 44(能见度 <0.5 n mile)或 7(散货船)采取措施,但散货船在能见度 <0.5 n mile 条件下航行时,切断两者的连接,使 44(能见度 <0.5 n mile)与 7(散货船)之间的交互作用路径边变为 44(能见度 <0.5 n mile) \rightarrow 26(发动机功率 >3000 kW) \rightarrow 7(散货船),此时,两者间需要两步才能产生交互,同时,使船舶碰撞事故发生率降低至 29.79%。

4 结 论

本文首先根据全球事故报告分析船舶碰撞事故的影响因素;其次,利用关联规则挖掘技术挖掘各影响因素间的内在关联;然后,在关联规则的基础上构建船舶碰撞事故影响因素交互网络,并分析网络的基本拓扑特性和重要节点与边;最后,对网络进行鲁棒性分析。研究结论如下:

(1) 船舶碰撞事故各影响因素之间存在 293 条强有力的关联规则,其中,置信度排名前 10 的关联规则大部分与船舶因素有关。

(2) 船舶碰撞事故影响因素交互网络中,各

影响因素间平均只需要 1.49 步便能连接到其他影响因素,远小于同等规模的随机网络;网络聚集系数也远高于同等规模的随机网络,网络联系十分紧密,易发生连锁反应导致事故发生。因此,降低船舶碰撞事故发生率较为困难。

(3)船舶碰撞事故影响因素交互网络中,较多的影响因素拥有大量的连接,且交互影响的概率较大。例如,“船员缺乏经验”、“船员操纵失误”、“船员违反规则”、“船龄 0~10 年”、“船龄 10~20 年”、“安全管理制度不完善”、“公司缺乏安全文化”、“缺乏船上训练”、“未按计划进行日常训练”及“沿岸水域”等影响因素与其他影响因素之间联系较为紧密。因此,需要加强对这些因素的防控,防止发生连锁反应,阻断船舶碰撞事故的发生。

(4)船舶碰撞事故影响因素交互网络中,部分因素起着较强的枢纽作用,更接近网络中心。例如,“船员缺乏经验”、“船员操纵失误”、“总吨位<500 t”、“总吨位 500~3000 t”、“安全管理制度不完善”、“沿岸水域”及“6~7 级强风”等因素接近网络中心,有利于交互信息快速传递。因此,需要加强对这些因素的防控,降低各影响因素交互信息传递的效率,防止船舶碰撞事故的发生。

(5)船舶碰撞事故影响因素交互网络中,管理因素和人为因素普遍占据较为重要的位置。因此,需要强化安全管理,通过提高船员的安全意识和操船技能等来降低船舶碰撞事故的发生率。船公司、培训机构及监管部门需要加强对人为因素和管理因素的管控,建设良好安全文化,加大培训和监管力度。

(6)通过制定针对性的防控措施,能够有效破坏船舶碰撞事故影响因素交互网络,更有效地降低船舶事故发生率。

(7)相较于影响因素的交互路径进行防控,直接对船舶碰撞事故影响因素交互网络中关键影响因素进行防控,能够快速且有效地破坏网络结构。但是,对于某些不能直接防控的影响因素,例如,船舶类型、船舶尺度及船龄等,对交互影响的路径进行防控也能有效破坏事故演化的网络结构,阻断船舶碰撞事故的发生。

参考文献(References):

- [1] 司东森,张英俊,郎坤. 基于改进 BN 的集装箱船舶碰撞事故致因分析[J]. 中国安全科学学报, 2019, 29(10): 31-37.
SI D S, ZHANG Y J, LANG K. Causation analysis of container ship collision accidents based on improved BN [J]. China Safety Science Journal, 2019, 29(10): 31-37. (in Chinese)
- [2] 王海燕,钟小敏. 基于 MORT 的“桑吉”轮溢油事故致因分析[J]. 安全与环境学报, 2020, 20(5): 1624-1630.
WANG H Y, ZHONG X M. Causative analysis of oil spill accident of “Sanchi” oil carrier based on the management oversight and the risk tree[J]. Journal of Safety and Environment, 2020, 20(5): 1624-1630. (in Chinese)
- [3] HWANG T, YOUN I H. Latent-cause extraction model in maritime collision accidents using text analytics on Korean Maritime Accident Verdicts [J]. Applied Sciences, 2022, 12(2): 914.
- [4] HASAN U, ISMAIL C. Analysis and assessment of ship collision accidents using fault tree and multiple correspondence analysis[J]. Ocean Engineering, 2022, 245: 110514.
- [5] 付姗姗,张悦,席永涛,等. 多因素耦合下长江口水域交通事故致因链分析[J]. 中国安全科学学报, 2023, 33(3): 60-67.
FU S S, ZHANG Y, XI Y T, et al. Causal chain of maritime accidents in Yangtze River Estuary considering coupling effects of multi-risks factors[J]. China Safety Science Journal, 2023, 33(3): 60-67. (in Chinese)
- [6] 廖诗管,翁金贤. 基于贝叶斯时空 log-logistic 模型的船舶碰撞频率[J]. 中国航海, 2023, 46(1): 24-29.
LIAO S G, WENG J X. Ship collision frequency prediction with Bayesian spatiotemporal log-logistic model[J]. Navigation of China, 2023, 46(1): 24-29. (in Chinese)
- [7] LI G R, WENG J X, HOU Z Q. Impact analysis of external factors on human errors using the ARBN method based on small sample ship collision records[J]. Ocean Engineering, 2021, 236: 109533.
- [8] WENG J X, LIAO S G, WU B, et al. Exploring effects of ship traffic characteristics and environmental conditions on ship collision frequency[J]. Maritime Policy & Management, 2020, 47(4): 523-543.
- [9] 赵志奎,张闯. 基于事件树方法 LNG 船舶碰撞事故安全评估[J]. 大连海事大学学报, 2014, 40(3): 45-48.

- ZHAO Z L, ZHANG C. Risk assessment for LNG carrier collision accidents based on event tree [J]. *Journal of Dalian Maritime University*, 2014, 40(3): 45-48. (in Chinese)
- [10] 陈伟, 王利莹, 杨劼, 等. 基于关联规则的塔式起重机事故致因网络模型研究 [J]. *安全与环境学报*, 2023, 23(4): 1161-1168.
- CHEN W, WANG L Y, YANG J, et al. Research on the network model of tower crane accidents based on association rules [J]. *Journal of Safety and Environment*, 2023, 23(4): 1161-1168. (in Chinese)
- [11] 孙家庆, 李默涵, 修晓仪. 基于复杂网络理论的水上交通事故致因分析 [J]. *大连海事大学学报*, 2023, 49(2): 80-90.
- SUN J Q, LI M H, XIU X Y. Research on causation of water traffic accidents based on complex network theory [J]. *Journal of Dalian Maritime University*, 2023, 49(2): 80-90. (in Chinese)
- [12] DENG Y, ZHANG Y, YUAN Z, et al. Analyzing subway operation accidents causations: Apriori algorithm and network approaches [J]. *International Journal of Environmental Research and Public Health*, 2023, 20(4): 3386.
- [13] CAO Y H, WANG X J, WANG Y H, et al. An analysis of factors affecting the severity of marine accidents using a data-driven Bayesian network [J]. *Ocean Engineering*, 2022, 269: 113563
- [14] WANG H X, LIU Z J, WANG X J, et al. An analysis of factors affecting the severity of marine accidents [J]. *Reliability Engineering & System Safety*, 2021, 210: 107513.
- [15] FAN S Q, YANG Z L, BLANCOI-DAVIS E, et al. Analysis of maritime transport accidents using Bayesian networks [J]. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2020, 234(3): 439-454.
- [16] CORADDU A, ONETO L, NAVAS M B, et al. Determining the most influential human factors in maritime accidents: a data-driven approach [J]. *Ocean Engineering*, 2020, 211: 107588.
- [17] FENG Y W, WANG Z F, DONG Q B. How to analyze the urban rail transit network in the combination view of complex network and clustering: take Shenzhen as an example [C] // 2021 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). Shanghai: IEEE, 2021.
- [18] GIRVAN M, NEWMAN M E. Community structure in social and biological networks [J]. *Proceedings of the National Academy of Science*, 2002, 99(12): 7821-7826.
- [19] 蒋林承. 复杂网络节点和边重要性度量以及多源信息传播问题研究 [D]. 长沙: 国防科技大学, 2018.
- JIANG L C. Research on identifying node and edge importance and influence maximization in complex networks [D]. Changsha: National University of Defense Technology, 2018. (in Chinese)