LIVERPOOL
JOHN MOORES
UNIVERSITY

LJMU Research Online

Xin, X, Liu, K, Loughney, S, Wang, J, Li, H and Yang, Z

Graph-based ship traffic partitioning for intelligent maritime surveillance in complex port waters

http://researchonline.ljmu.ac.uk/id/eprint/20119/

Article

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

Xin, X, Liu, K, Loughney, S, Wang, J, Li, H and Yang, Z (2023) Graph-based ship traffic partitioning for intelligent maritime surveillance in complex port waters. Expert Systems with Applications, 231. ISSN 0957-4174

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/

Expert Systems With Applications 231 (2023) 120825

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Graph-based ship traffic partitioning for intelligent maritime surveillance in complex port waters

Xuri Xin [a,b], Kezhong Liu [a,c], Sean Loughney [b], Jin Wang [b], Huanhuan Li [b], Zaili Yang [b,*]

[a] School of Navigation, Wuhan University of Technology, Wuhan, Hubei, China
[b] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK
[c] Hubei Key Laboratory of Inland Shipping Technology, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Maritime safety
Intelligent maritime surveillance
Maritime knowledge extraction
Graph partitioning
AIS data

ABSTRACT

Maritime Situational Awareness (MSA) is a critical component of intelligent maritime traffic surveillance. However, it becomes increasingly challenging to gain MSA accurately given the growing complexity of ship traffic patterns due to multi-ship interactions possibly involving classical manned ships and emerging autonomous ships. This study proposes a new traffic partitioning methodology to realise the optimal maritime traffic partition in complex waters. The methodology combines conflict criticality and spatial distance to generate conflict-connected and spatially compact traffic clusters, thereby improving the interpretability of traffic patterns and supporting ship anti-collision risk management. First, a composite similarity measure is designed using a probabilistic conflict detection approach and a newly formulated maritime traffic route network learned through maritime knowledge mining. Then, an extended graph-based clustering framework is used to produce balanced traffic clusters with high intra-connections but low inter-connections. The proposed methodology is thoroughly demonstrated and tested using Automatic Identification System (AIS) trajectory data in the Ningbo-Zhoushan Port. The experimental results show that the proposed methodology 1) has effective performance in decomposing the traffic complexity, 2) can assist in identifying high-risk/density traffic clusters, and 3) is sufficiently generic to handle various traffic scenarios in complex geographical waters. Therefore, this study makes significant contributions to intelligent maritime surveillance and provides a theoretical foundation for promoting maritime anti-collision risk management for the future mixed traffic of both manned and autonomous ships.

1. Introduction

Maritime traffic safety management plays a vital role in human life security, economic development, and ocean environmental protection. Proactive maritime surveillance using modern and intelligent transportation technologies is one of the essential components of maritime traffic safety management (Liang et al., 2022; Liu et al., 2022). To enhance maritime surveillance and make port transportation more efficient, a great variety of new technologies and systems have been developed and deployed, including but not limited to Artificial Intelligence (AI), Big Data, Internet of Things (IoT), Intelligent Situational Awareness System (ISAS), and Automatic Identification System (AIS) (Filom, Amiri, & Razavi, 2022). They have different technical specifications and functionalities, enabling the realisation of maritime traffic monitoring and Maritime Situational Awareness (MSA) from diverse aspects. Nevertheless, the broadness of the surveillance areas and the diversity of ship motion activities (e.g., sailing, berthing, anchoring, and refuelling) bring significant challenges to their practical applications. Particularly, with respect to the transport demand growth, the application of super large-scale ships, the development of emerging technologies (e.g., autonomous ships), and the impact of non-classical risks (e.g., COVID-19), maritime traffic situations have become increasingly complicated and sophisticated, particularly in complex waters (e.g., ports). These challenges hinder the effectiveness of the currently used MSA systems (Fang, Yu, Ke, Shaw, & Peng, 2018; Shi & Weng, 2021; Yu, Fang, Murray, & Peng, 2019). As a result, new advanced MSA techniques and tools have to be developed and implemented urgently to cope with the ever-growing complexity of maritime traffic situations for better maritime traffic surveillance and ship collision risk management to aid the establishment of the intelligent transportation system.

* Corresponding author.
E-mail addresses: xrxin@whut.edu.cn (X. Xin), kzliu@whut.edu.cn (K. Liu), S.Loughney@ljmu.ac.uk (S. Loughney), j.wang@ljmu.ac.uk (J. Wang), h.li2@ljmu.ac.uk (H. Li), Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.eswa.2023.120825
Received 23 February 2023; Received in revised form 27 May 2023; Accepted 10 June 2023
Available online 15 June 2023
0957-4174/© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Based on the demand to enhance the operational monitoring over busy waters of interest, a variety of techniques and approaches have been proposed to undertake collision risk quantification and estimation (e.g., Chen, Huang, Mou, & van Gelder, 2019; Huang, Chen, Chen, Negenborn, & van Gelder, 2020; Yu, Liu, Chang, et al., 2020). They offer insights on providing a quantitative foundation for maritime surveillance and issuing early collision warnings to support anti-collision decision-making (Gilbert, Petrovic, Pickering, & Warwick, 2021; Liu, Zhang, Yan, & Soares, 2022; Öztürk, Boz, & Balcisoy, 2021). With the rapid development of AIS and the increased accessibility of a vast amount of vessel movement information (i.e., AIS data), accurate collision risk assessment and characterisation of maritime traffic have become possible and further attracted widespread attention in recent years. Apart from the usage for maritime collision analysis, advanced applications of AIS data in maritime safety-related studies have been witnessed, including maritime traffic pattern extraction (Li et al., 2020; Li et al., 2022), vessel trajectory anomaly detection (Iphar, Ray, & Napoli, 2020; Rong et al., 2019; 2022), vessel motion prediction (Li, Jiao, & Yang, 2023; Li & Yang, 2023; Zhang et al., 2023), and ship path optimization (Duan, Fan, Chen, Chen, & Ma, 2021; Tavakoli, Najafi, Amini, & Dashtimansh, 2021; Yu et al., 2021; Zhang, Liu, Hirdaris, Zhang, & Tian, 2022). While the applications of AIS data contribute to the improved analysis and modelling of ship motion behaviours, increasingly complex traffic situations associated with multi-ship interactions and the growth of ship spatiotemporal movement uncertainty remained unaddressed, requiring new intelligent MSA methods to be found. More specifically, an intelligent MSA method requires continuous assessment of the maritime geographical features, the spatiotemporal dynamics and uncertainty of ship motion, and the multi-dependent interrelationships among ships in a collective manner. Therefore, developing such an MSA method is very challenging in theory when so many constraints are embedded in complex dynamic waters, especially considering the emergence of future hybrid traffic encounter situations. One of the realistic solutions is to develop a practical maritime traffic partitioning approach to partition the whole maritime traffic into several interpretable traffic clusters, in which the ships in the same cluster have high spatiotemporal interrelationships while the ships between different clusters have low spatiotemporal interactions. Undoubtedly, it can reduce the difficulty of understanding the whole traffic situation and facilitate the identification of potential high-risk traffic clusters.

This study aims to propose an optimal ship traffic partition methodology to partition the regional ship traffic into several compact, scalable, and interpretable groups. The first step to partition maritime traffic is to rationally interpret its pattern complexity and the multiple interactions (e.g., spatiotemporal proximity and potential conflict) among ships. Compared to the investigations in urban transportation networks (Gu & Saberi, 2019; Ji & Geroliminis, 2012; Saeedmanesh & Geroliminis, 2016), traffic partitioning studies in maritime transportation is in infancy partially because of its traffic uniqueness in the sector, which requires one to generate traffic clusters with a guarantee of both conflict connectivity and spatial compactness. Both guarantees have some theoretical implications that have yet been well addressed in the current literature. For instance, the conflict calibration needs to incorporate the ship movement uncertain features in a dynamic traffic situation, whereas the spatial compactness measure requires extending the shortest path search approach based on a maritime traffic route network obtained by maritime trajectory knowledge extraction. On this basis, these two indices have to be integrated into a composite similarity measure model through an effective combination way, in which the weights assigned to the two indices also need to be determined based on sensitivity analysis as a trade-off parameter. Only by then, the similarity measure result can be fed into a robust graph-based clustering framework to produce traffic clusters with a balanced size where the intra-cluster similarity can be maximized but the inter-cluster similarity minimized. Despite the high demand on research efforts and resources, the success of this work will make significant contributions both in theory and in practice. Along with the main contributions of supporting intelligent MSA by decomposing the traffic complexity in regulatory waters to guide ship collision avoidance risk management, the originality and other methodological contributions of this work are given as follows:

1) Two optimization criteria, i.e., conflicting criticality and spatial distance, are integrated and incorporated into the traffic partitioning process to assist in generating both conflict-connected and spatial compact traffic clusters. To combine these criteria effectively, a linear combination function is utilized and the weights assigned to each criterion are determined through a sensitivity analysis to achieve a balanced trade-off between them.
2) The proposed approach makes use of historical AIS data to generate a data-driven representation of maritime traffic route network. It contributes toward capturing the traffic clusters with real spatial compactness by using the length of the shortest path of ship pairs on the network instead of the traditional physical distance, ensuring the adaptation to the traffic scenarios in any restricted geographical waters.
3) With respect to traffic partitioning, a graph-based clustering framework known as Symmetric Nonnegative Matrix Factorization (SNMF) is extended by employing the Newton-like algorithm to produce ideal traffic clusters with balanced sizes. It is flexible and scalable to handle various traffic scenarios beyond the maritime sector by optimizing the graph clustering objectives.

The remainder of the paper is organized as follows. Section 2 reviews the literature related to ship collision risk evaluation and estimation and AIS data applications in maritime traffic surveillance. The details of the developed modelling methodology are explained in Section 3, including the similarity measure model, graph partitioning algorithm, and metrics development. Application performance, validation, and implications are given in Section 4. Conclusions are presented in Section 5.

2. Literature review

2.1. Ship collision risk evaluation and estimation

Collision risk assessment and estimation as an indispensable part of safety management is fundamental to reducing accidents in the maritime transportation domain. There is a strong understanding and extensive literature on maritime collision risk modelling and prediction associated with new concepts and definitions, such as near-miss (Zhang, Goerlandt, Montewka, & Kujala, 2015), collision candidates (Chen, Huang, Mou, & van Gelder, 2018), and traffic conflict (Weng, Meng, & Qu, 2012; Weng & Shan, 2015). Due to the critical mass of the relevant publications, a comprehensive and detailed survey has been documented in such papers as Chen et al. (2019) and Huang et al. (2020). Briefly speaking, the relevant research can be classified as ship domain-based and synthetic index approaches. The former evaluates and estimates the collision risk in terms of the violation or overlap of the encountering ships' domain areas, whereas the latter quatifies the potential collision probability and severity by synergizing the indicators that characterise the approaching ships' spatiotemporal proximity.

Regarding ship domain-based approaches, many ship domain models have been developed to detect collision candidates. The critical concern for these models consists of the identification of the key influential factors (e.g., vessel attributes and knowledge of navigators (Liu, Feng, Li, Wang, & Wen, 2016; Wang & Chin, 2016)), the definition of the domain shapes (e.g., elliptical, polygonal, and fuzzy domains (Szlapczynski & Szlapczynska, 2016; Wang, 2013; Wang & Chin, 2016)), and the specification of the model training methods (e.g., knowledge-based, empirical, and analytical methods (Szlapczynski & Szlapczynska, 2017)). These models have been used to address different issues and are capable of estimating waterway capacity, capturing near-miss hot-spots, and deriving knowledge about primary factors related to near-collision clusters (Liu et al., 2016; Liu, Yuan, Xin, Zhang, & Wang, 2021; Rong, Teixeira, & Soares, 2021). Unfortunately, the model training outcomes heavily rely on the traffic characteristics in the investigated water areas when historical traffic trajectory records are utilized (Kulkarni, Goerlandt, Li, Banda, & Kujala, 2020; Wang & Chin, 2016). Additionally, it is unfeasible to directly apply these models for real-time Conflict Detection (CD) implementation. As a result, the ship domain-based collision estimation by incorporating refined trajectory prediction models has been put forward to address this deficiency.

With respect to synthetic index methods, the Distance to Closest Point of Approach (DCPA) and Time to Closest Point of Approach (TCPA) are the two most commonly employed indicators to characterise potentially dangerous encounter events. They were initially synthesised based on such techniques as linear regression (Chin & Debnath, 2009), weighting coefficient (Zhen, Riveiro, & Jin, 2017), and fuzzy theory (Lee & Rhee, 2001), and performed well in noticing potential collisions at open sea. Within this context, further extensions and enhancements are conducted. Examples include incorporating more detailed influential factors (e.g., relative speed, stability conditions, and ship manoeuvrability (Gil, Montewka, Krata, Hinz, & Hirdaris, 2020; Gil, 2021; Zhang, Montewka, Manderbacka, Kujala, & Hirdaris, 2021; Zhao, Li, & Shi, 2016)), deploying advanced index synthesis techniques (e.g., Analytical Hierarchical Process (Zhao et al., 2016), and Dempster-Shafer evidence theory (Li & Pang, 2013)), and considering their effectiveness to all kinds of encountering scenarios (Goerlandt, Montewka, Kuzmin, & Kujala, 2015). For easy implementation of the methods, a fundamental hypothesis is set as that ships keep a linear motion during their encounter process. Such a strong assumption significantly degrades their practical applicability in highly dynamic traffic situations. For instance, a ship may need to turn frequently due to the influence of waterway topography. It means that improvement is possible by exploring the ship motion dynamic characteristics associated with the actual collision risk in complicated waters.

In summary, ship collision risk assessment and estimation remain an active research subject. The existing methods are not applicable and new methods beyond state of the art have to be proposed due to the increasingly sophisticated traffic situations and dynamic ship motion behaviours. On the one hand, the influence of ship motion dynamics and uncertainty on collision risk estimation is rarely explored, while it is imperative to support accurate collision warning and provide further assistance in anti-collision decision-making. Indeed, this high and valuable research need is evident through the relevant studies in the air transportation field, which have already emphasised the necessity of involving traffic dynamics and uncertainties when implementing Conflict Detection and Resolution (CDR) (Hao, Zhang, Cheng, Liu, & Xing, 2018; Mitici & Blom, 2018; Prandini, Hu, Lygeros, & Sastry, 2000). On the other hand, multiple ship encounters in complex waters are frequent and riskier, particularly given the emerging autonomous ships. The risk assessment of a maritime navigational scenario is highly correlated with the dependent conflicts among multiple ships. Hence, much attention should be paid to the co-behaviour of multiple vessels relating to their spatiotemporal interactions. Although several applications have been presented to evaluate the collision risk of multiple ships or regional traffic complexity (e.g., Liu, Wu, & Zheng, 2019; Xin, Yang, Liu, Zhang, & Wu, 2022; Zhang, Zhang, Fu, Kujala, & Hirdaris, 2022), the research capturing the high-risk traffic clusters from a whole regulatory area is largely missing in the reported literature. Hence, incorporating the spatiotemporal dynamics and interaction effects among multiple ships is a crucial and promising avenue for advancing the development of an effective technique for detecting traffic clusters.

2.2. AIS data applications in maritime traffic surveillance

Because of its broad spatial coverage, high transmission frequency, and data availability, AIS has become the primary reliable information source for maritime traffic surveillance. The AIS data is supportive for maritime safety management, but also maritime traffic behaviour analysis and modelling. Many research communities have made great efforts to its promising practical applications, involving a great diversity of outputs in maritime research (Xiao, Fu, Zhang, & Goh, 2019; Yang, Wu, Wang, Jia, & Li, 2019; Zhou, Daamen, Vellinga, & Hoogendoorn, 2019). Among the existing studies, maritime traffic pattern mining is one of the most widely investigated research topics. It is dedicated to maritime traffic knowledge extraction and traffic characteristic analytics and exploitation, thereby serving as a prerequisite for intelligent maritime monitoring and surveillance.

Existing methods for maritime traffic pattern mining are conducted from three perspectives: grid-based, vector-based, and statistics-based methods (Rong et al., 2021; 2022; Xiao et al., 2019). The grid-based methods are devoted to segmenting the maritime traffic regions into a series of spatially indexed grids, in which each cell is characterised by the necessary traffic property statistics (Ristic, 2014; Vettor & Soares, 2015). The maritime knowledge extracted from the individual cells can then be applied for many purposes, such as supporting human-interpretable maritime traffic visualization, early threat recognition, and collision avoidance. The vector-based methods formulate the maritime traffic route network by abstracting the waypoints including entrance, exits, stationary spaces (e.g., ports and offshore platforms), and turning sections into nodes and the routes into navigational legs to connect nodes (Arguedas, Pallotta, & Vespe, 2017; Rong, Teixeira, & Soares, 2022). These methods contribute to presenting a light and structured representation of maritime traffic, facilitating the MSA in traffic prediction, route planning, and anomaly detection. The statistics-based methods focus on revealing traffic behaviours and patterns by quantitatively modelling and analysing the traffic characteristics, with the aim of estimating the maritime traffic capacity, supporting maritime traffic decision-making, and characterising the waters' traffic properties (Kang, Meng, & Liu, 2018; Xin, Liu, Yang, Yuan, & Zhang, 2019; Yu, Liu, Teixeira, et al., 2020). Working towards MSA, all three categories of approaches provide essential prior knowledge for monitoring, analysing, and understanding the maritime traffic situation. They therefore show great potential for tackling challenging traffic scenarios in complex waters and assisting in maritime traffic surveillance and management. As a result, this study establishes a maritime geographical network through maritime traffic pattern mining from historical AIS data to facilitate the recognition of real spatial compact traffic clusters.

Nevertheless, there remains much potential for making advanced use of AIS-based trajectory data to conduct MSA. The literature on partitioning maritime traffic and identifying real-time high-interaction connected multi-ship encounters in complex water areas is extremely limited, despite its crucial role in decreasing the difficulty of situational awareness and further guiding ship anti-collision risk management. The most relevant research (Liu, Wu, & Zheng, 2019; Zhen et al., 2017; Zhen, Shi, Shao, & Liu, 2021) in the maritime domain applied the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm to detect clusters of encounter vessels and filter out the relatively safe vessels. These works suffer from the following drawbacks:

1) Only the spatial distance among ships is considered when conducting maritime traffic clustering, which is insufficient to reveal the complex dependencies of encounter ships. To reflect different aspects of ship traffic interactions simultaneously, it is of paramount importance to identify the encountering traffic clusters by fully considering the multiple dependent interrelationships (e.g., spatiotemporal proximity and conflict severity) among ships.
2) These works detected traffic clusters or multi-ship encounters based on traditional Euclidean/physical distances among ships but yet considered the influence of restricted geographical waters on the spatial distance calculation. Evidently, it is problematic to overlook the effects of water topography on spatial distance measurement

because the two spatially close ships (of a short distance) may be blocked by obstacles, e.g., islands and skerries, especially in complex port waters.

3) The adopted clustering techniques (i.e., DBSCAN) of these works reveal challenges in figuring out various interrelationships (e.g., conflict severity) among ships as well as discovering the traffic clusters in waters with varying traffic densities. The high complexity of ship spatiotemporal distribution, the unpredictability of ship motion behaviours, and the restricted geographical waters jointly create difficulties in pioneering an effective traffic clustering model. These unique and stochastic characteristics in maritime traffic justify the deficiencies of direct applications of arbitrary clustering techniques, as they often produce error-prone clustering solutions.

To address the above issue, recent advances in urban transportation network partitioning using graph-based clustering techniques offer valuable insights (Gu & Saberi, 2019; Ji & Geroliminis, 2012; Saeedmanesh & Geroliminis, 2016). Specifically, the road network partitioning focuses on segmenting a heterogeneous traffic network into several spatially connected, homogeneous, and compact-shaped sub-regions in terms of indices like link speed and density. For instance, Saeedmanesh and Geroliminis (2016) developed a three-step clustering framework to segment the heterogeneous road networks into several homogeneous sub-regions. In the framework, a snake-based similarity measure was developed to account for the spatial relations between links and a non-linear graph-based optimization technique was employed to assign the links to proper clusters. These studies have shown much attractiveness in decomposing traffic network complexity and identifying the network congested regions. Despite that, to the authors' best knowledge, there has not been any maritime traffic partitioning using graph-based clustering techniques based on multiple spatiotemporal interactions among ships. Therefore, this study attempts to combine the graph-based clustering framework with a multi-attribute interrelation measure model, as a hybrid pioneer, to investigate the maritime traffic partitioning.

3. Methodology: Optimal maritime traffic partitioning

As aforementioned, the development of an optimal ship traffic partition methodology should aid to achieve the goals of 1) extracting the traffic clusters that have shown high conflict connectivity to detect real traffic conflict patterns; and 2) generating the traffic clusters that are spatially compact to ease the design and deployment of traffic management strategies. Based on these two goals, this study involves constructing an undirected graph for ship traffic partition, in which each ship is modelled as a node and their neighbouring relationships (i.e., edges) are built based on their conflict criticality and spatial distance. By doing so, the traffic partitioning problem is transformed into a graph cut problem. It is dedicated to separating the network into several subgraphs. The proposed partitioning methodology consists of the following major steps. Firstly, a composite similarity model that considers conflict connectivity and spatial compactness is introduced. The conflict relations are quantified by a probabilistic conflict detection approach, which can precisely compute the conflict criticality between ship pairs by incorporating the ship motion dynamic and uncertain characteristics. The spatial compact relations, on the other hand, are measured based on a maritime traffic knowledge extraction technique. It extracts the real spatial distance between ship pairs from a derived ship traffic route network. Based on the constructed similarity model, a graph clustering mathematical framework is further utilized to assign the ships with high conflict criticality and spatial compactness into one cluster. Additionally, four metrics are adopted to evaluate and check the performance of the proposed traffic partitioning framework. Fig. 1 provides the associated methodological framework. The important supporting techniques embedded into each step are explained in the following subsections.

The symbols and explanations used in the optimal traffic partitioning model are provided in Table 1.

3.1. Similarity measures and models

The key issue of graph partitioning is how to define a similarity/adjacent measure to describe the connections/interactions between each pair of ships. This study is devoted to developing a similarity model to enable the simultaneous consideration of both the conflict relation and spatial distance of ship pairs. The similarity model comprises the following elements: 1) a probabilistic conflict criticality evaluation model to reflect the conflict relation in Section 3.1.1, 2) a real spatial distance identification model to define the spatial relation in Section 3.1.2, and 3) the developed composite similarity model in Section 3.1.3.

Table 1
The symbols used in the optimal traffic partitioning model.

Symbol	Definition and Explanation
W^c_{ij}	the spatial compactness similarity between ships i and j
W^c_{ij}	the conflict connectivity similarity between ships i and j
W_{ij}	the composite similarity between ships i and j
G	maritime traffic network with node set V and edge set E
W	the similarity matrix for G, where $W \in \mathbb{R}^{N \times N}$
N	the number of data samples in G
D	the diagonal matrix with $D_{ii} = \sum_{j=1}^{N} W_{ij}$
L	the Laplacian matrix with $L = D\text{-}W$
k	the number of clusters
X^T	the transpose of matrix X
H	the clustering membership matrix
$\boldsymbol{Tr}$(W)	the trace of matrix W
$\widetilde{W}$	the normalized W, where $\widetilde{W} = D^{-1/2} W D^{-1/2}$
A	the whole dataset for G
$\{A_c\}_{c=1}^{k}$	the partitioned traffic clusters for G
$vol(A_o)$	the number of samples in subset A_o
NS_k	the average $NS_k(A_o)$ of partitioning results
NC	the Normalized Cut ($Ncut$)
f_1	the conflict connectivity measure of partitioning results
f_2	the spatial compactness measure of partitioning results

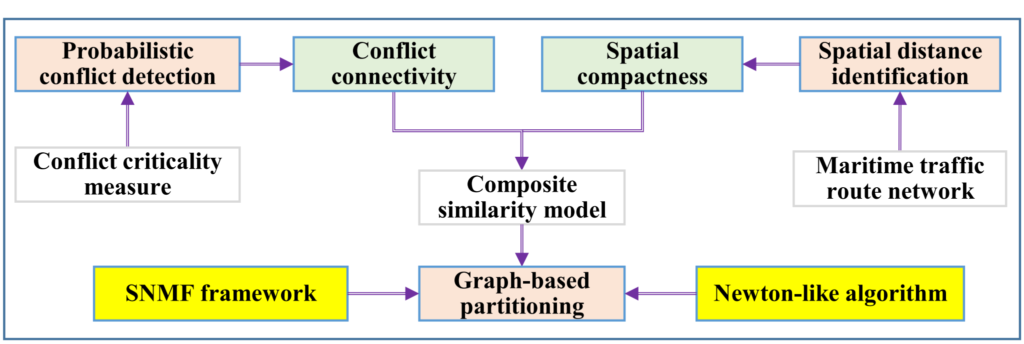

Fig. 1. The research framework.

3.1.1. Probabilistic conflict detection

Collision risk quantification is an integral part of the detection of conflicting traffic clusters. Here, the conflict criticality is measured in a probabilistic manner to ensure the adaptation to traffic scenarios with high movement uncertainty.

In general, a conflict is defined as a situation in which the minimum safe separation between two ships is violated over a finite prediction horizon (Hernandez-Romero, Valenzuela, & Rivas, 2019; Mitici & Blom, 2018). This study uses a classical ship domain model (Fujii & Tanaka, 1971) adopted in restricted waters to characterise the conflict between ships. The ship pairs are in conflict if the following inequality holds during the CD horizon (see Fig. 2).

$$Dist_{AB}(t) \leq SD_A(t) + SD_B(t) \tag{1}$$

where $Dist_{AB}$ denotes the distance between ships A and B at time t, and SD_A and SD_B are the distances from each ship's centre to their domain boundaries, respectively. The distance SD_A can be calculated based on the following equation:

$$SD_A(t) = \left(\frac{1 + tan^2(\beta_{AB}(t) - \varphi_A(t))}{\frac{1}{R_{L,A}{}^2} + \frac{tan^2(\beta_{AB}(t) - \varphi_A(t))}{R_{S,A}{}^2}} \right)^{1/2} \tag{2}$$

where $\beta_{AB}(t)$ denotes the relative course of ship B's position over ship A' position, $\varphi_A(t)$ represents the ship A's heading course, and $R_{L,A}$ and $R_{S,A}$ represent the length of the semi-major and semi-minor axis of ship A's elliptical domain, which are equal to $3L_A$ (L_A is the Length Overall (LOA) of ship A) and $0.8L_A$ (refer to (Fujii & Tanaka, 1971)), respectively.

Since the future positions and heading courses of the encountering ships are uncertain due to various disturbances, the variables $Dist_{AB}(t)$, $SD_A(t)$ and $SD_B(t)$ are uncertain as well. Furthermore, the event that the encountering ships will violate the minimum separation distances in the future time t (i.e., Eq. (1)) becomes a probabilistic question to answer. Hence, the conflict probability between ships at a given time t is calibrated by using the following expression:

$$PC(t) = \Pr[L(t) \leq 0] = \int_{-\infty}^{0} f_{L(t)} dL(t) \tag{3}$$

where $f_{L(t)}$ denotes the probability density function of the loss of minimum safe separation between two ships, in which $L(t) = Dist_{AB}(t)\text{-}SD_A(t)\text{-}SD_B(t)$.

Note that Eq. (3) represents the instantaneous conflict probability at a given time. Here the conflict criticality over the CD horizon is quantified by considering both the maximum $PC(t)$ with $0 < t \leq T_{CDH}$ (T_{CDH} is the CD horizon) and its corresponding occurrence moment. The first indicator reveals the highest intensity of a potential conflict, while the second indicator represents the urgency of a traffic case needing immediate conflict resolution actions. Indeed, these two indices play equally significant roles and are equivalent to the commonly used indices (i.e., DCPA and TCPA) in maritime traffic navigation (Cho, Han, & Kim, 2020; Hu et al., 2019). Therefore, an exponential function (Hu et al., 2019; Wang, Song, & Wen, 2018) is utilized to synthesise the outlined indices as follows:

$$C(\gamma) = MPC^{1+\left(\frac{t_{MPC}}{T_{CDH}}\right)} \tag{4}$$

where MPC represents the maximum conflict probability over the CD horizon, and t_{MPC} denotes the occurrence moment of the maximum conflict probability. Note that the CD horizon (i.e., T_{CDH}) is set to be 15 min in terms of the work presented by Bakdi, Glad, and Vanem (2021), as this study pays attention to the CD in the medium-term time horizon.

After defining the conflict criticality measure model from a probabilistic viewpoint, the probabilistic CD can be further implemented. The probabilistic CD methodology consists of two important elements: one is the uncertain trajectory prediction, which serves as a prerequisite for potential collision detection and evaluation; the other is the conflict probability computation, which concerns the development and use of techniques and approaches with which the actual conflict criticality can be estimated accurately and efficiently. In the authors' previous research (Xin, Liu, Yang, Zhang, & Wu, 2021), uncertain trajectory forecasting is conducted by modelling the ship motion as a deterministic motion correlated with the ship navigation plan plus a stochastic component given by various perturbations. Regarding conflict probability computation, a two-stage Monte Carlo (MC) simulation model is deployed to achieve a fast and accurate estimation of the conflict criticality. More details and discussion on the probabilistic CD procedure are found in Xin et al. (2021).

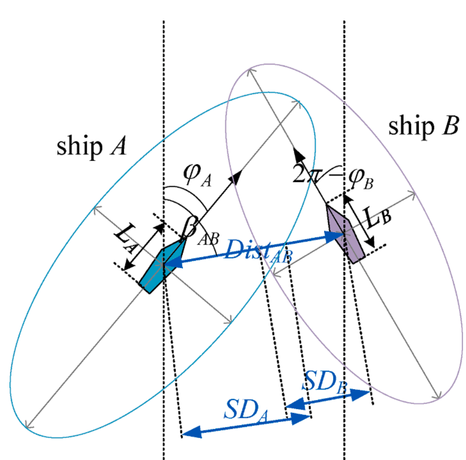

Fig. 2. Definition of a ship conflict.

3.1.2. Real spatial distance identification

Maritime traffic partitioning requires guaranteeing the spatial compactness of the produced traffic clusters to ease the design of collision risk management strategies. The traditional measure of spatial compactness is conducted in terms of the Euclidean distance (or called linear distance) among ships (e.g., Liu et al., 2019; Xin, Liu, et al., 2022; Zhen et al., 2021). However, in complex and restricted waters, the two ships spatially adjacent may not be reachable from each other. For instance, the obstacles (e.g., small islands) between the ships often block them away. Hence, the traditional linear distance measure is not applicable to describe the spatial compactness of traffic scenarios in complex waters involving restricted geographical features.

A practical solution to this issue is to find the shortest distance between ship pairs with reference to the maritime traffic route network as their actual spatial distance. However, in contrast to road networks, there are only a limited number of conventional transportation routes and traffic lanes in complex port waters, which makes it difficult to accurately measure the real spatial distance between any pair of ships. As a result, some recent studies have resorted to identifying ship traffic motion patterns through the extraction of maritime traffic knowledge to establish a complete and accurate maritime traffic route network. Establishing such a network by extracting nodes and legs is a common solution, as stated in Section 2.2 (Arguedas et al., 2017; Rong et al., 2022). However, these methods become challenging when applied to complex traffic waters where traffic motion behaviours are difficult to categorize (Xiao et al., 2019). In particular, these methods can detect high-density waypoints but ignore low-density waypoints. This makes it difficult to extract all ship motion patterns and further determine the real spatial distance between ship pairs under any situation.

To overcome these difficulties, this study proposes using image processing technique as an effective solution. This approach captures the main skeleton of the navigable waters as the traffic route network, helping to accurately capture the reasonable spatial distance on the network. Firstly, the Kernel Density Estimation (KDE) is applied to distinguish the navigable and unnavigable water areas. It estimates the spatial probability distribution of ship traffic based on real AIS data using the following formula:

$$g(x) = \frac{1}{K'}\sum_{i=1}^{K'}\phi_h(x - x_i) = \frac{1}{K'h}\sum_{i=1}^{K'}\phi_h\left(\frac{x - x_i}{h}\right) \tag{5}$$

where ϕ_h is a kernel function satisfying $\phi_h(x)\rangle 0$ and $\int \phi_h(x)dx = 1$, h denotes a bandwidth parameter larger than 0, and K' represents the number of samples to be investigated within the bandwidth h. The entire investigated water area is divided into a series of grids. For each grid, if its spatial probability distribution value of ship traffic is larger than a defined threshold, it represents a navigable area. Otherwise, it is unnavigable.

Leveraging on the probability distribution results obtained using the KDE, the whole investigated water area can be transformed into a binary image comprising of grids with 1 representing the navigable area and 0 representing the unnavigable area. The image processing operation is then applied to the binary image to extract the image skeleton (Lam, Lee, & Suen, 1992). Compared with the approaches that perform maritime traffic network abstraction based on node and edge extraction, it is more easily implemented by using the morphological algorithms in the MATLAB toolbox. Through the execution of the morphological algorithms, a network skeleton that provides a compact, structured, and precise traffic route description can be built.

After obtaining the traffic network representation, it can be employed to identify the real spatial distance between ship pairs. The procedure implementation comprises the following steps. First, several points (e.g., 10) are evenly sampled on the connection lines between the ship pairs to identify whether they fall into the navigable areas. If all these points are in the navigable areas, the real spatial distance between ship pairs is calculated in terms of the Euclidean distance; otherwise, the nearest point on the traffic route network that each ship is close to is searched for. Then, Dijkstra's algorithm is applied to calculate the shortest path distance between the two points. In this way, the procedure offers the potential to support generating actual spatial compact traffic clusters. The flowchart demonstrating the traffic network representation learning and real spatial distance identification procedure is presented in Fig. 3.

3.1.3. *Composite similarity measure model*

Furthermore, the conflict relation and distance relation measures can be merged to fulfil the spatial compactness and conflict connectivity requirements simultaneously. In this study, the two measure indices are combined through a linear combination method for clustering purposes. It provides a simple yet powerful way to describe the relationships between ship pairs when the two indices are presented by the same value range. Note that the value ranges of conflict criticality between ship pairs fall within [0, 1] (i.e., Eq. (4)). Hence, the conflict connectivity similarity W_{ij}^c between ships i and j can be defined as equal to their conflict criticality. However, the real distance between ship pairs varies significantly (e.g., tens of nautical miles). A compactness similarity W_{ij}^d that allows its value range to be in line with W_{ij}^c is therefore defined by transforming the real distance between a ship pair as follows:

$$W_{ij}^d = \begin{cases} 1, Dist_{ij} \le D_1 \\ \left(\dfrac{D_1}{Dist_{ij}}\right)^{\beta}, D_1 < Dist_{ij} < D_2 \\ 0, Dist_{ij} \ge D_2 \end{cases} \tag{6}$$

where $Dist_{ij}$ is the real spatial distance between the two ships, β is a scaling parameter, and D_1 and D_2 are two user-specified parameters that put the spatial compact relations into three categories, i.e., high, medium, and negligible compact relations. According to Eq. (6), if the real spatial distance falls below D_1, the ship pairs are regarded as high compact and W_{ij}^d is set to be 1. If the real spatial distance is between D_1 and D_2, the compact similarity is monotonically decreasing based on an exponential mathematical expression. If the real spatial distance exceeds the threshold D_2, the compact relation between ship pairs is negligible. Overall, Eq. (6) has the following properties: 1) exhibits a normalization effect to ensure that W_{ij}^d falls within [0, 1]; 2) offers flexibility to control the relations between W_{ij}^d and $Dist_{ij}$ by using β (e.g., a larger β results in a high decline rate, and vice versa); and 3) produces a sparse similarity matrix to simplify the optimization complexity of graph partitioning by setting 0 similarities for ship pairs with extremely large spatial distance.

A composite similarity measure is further defined to put different weights for W_{ij}^d and W_{ij}^c through a linear combination way, as follows:

$$W_{ij} = W_{ij}^c \cdot \alpha + W_{ij}^d \cdot (1 - \alpha) \tag{7}$$

where W_{ij} denotes the similarity between ships i and j, and α is a trade-off weighting coefficient. This model explicitly considers the above two similarity measures and helps systematically describe the multi-interrelationships among ships in the whole investigated water. However, the conflict connectivity and spatial compactness indices may conflict with each other because the conflict criticality between ship

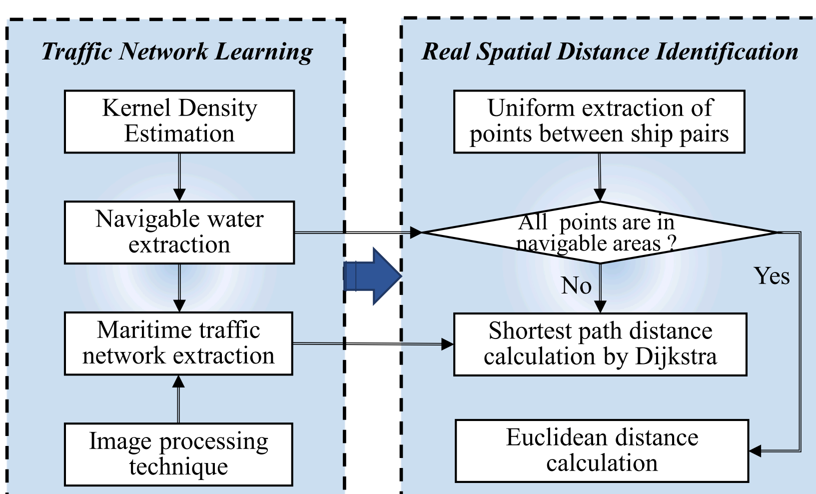

Fig. 3. Flowchart of traffic network representation learning and real spatial distance identification.

pairs is not totally dependent on their real spatial distance. The indices such as ship size, speeds, and spatial approaching rate of encountering ships also have an impact on the conflict relations. This means that the set of weighting coefficient α is crucial, playing an important role in achieving a trade-off between the two indices. For instance, a higher α that puts more weight on conflict connectivity may result in spatially noncompact clusters. Consequently, the tuning/optimization of α is investigated and discussed based on the sensitivity analysis in the experimental section (see Section 4.3).

3.2. Graph partitioning: Symmetric Non-negative matrix Factorization

Spectral clustering represents a widely used type of clustering algorithm for solving graph partitioning issues. Different from other classes of clustering algorithms (e.g., prototype-based, and density-based clustering) that focus on the dataset itself, spectral clustering assigns the data samples into proper clusters in terms of the similarity between each pair of data samples and makes no assumptions on the form of the clustering datasets. Nevertheless, the successful applications of spectral clustering heavily rely on the properties of the leading eigenvectors and eigenvalues of the Laplacian matrix L. When the eigengap between the kth and $(k+1)$th leading eigenvalues of matrix L is not sufficiently large, the application of spectral clustering could fail because the k-dimensional subspace spanned by the leading k eigenvectors of L is unstable (Ng, Jordan, & Weiss, 2001).

To address this problem, an extended and more competitive mathematical formulation, i.e., SNMF, was presented for graph clustering. It is a variant of Nonnegative Matrix Factorization (NMF) and distinguishes different clusters by performing the nonnegative lower rank approximation for a graph similarity matrix. According to the comprehensive study in (Kuang, Ding, & Park, 2012; Kuang, Yun, & Park, 2015), SNMF has the following unique features of: 1) being adaptive to more general cases by offering the flexibility to define any similarity measure that describes the dataset structure well; 2) being capable of achieving higher accuracy and quality compared with other clustering algorithms including the standard forms and variations of spectral clustering, k-means, and NMF for graph clustering. These merits make SNMF appealing for graph partitioning applications. It has therefore been successfully applied in a diversity of research fields, such as community detection (Chunaev, 2020) and traffic network partitioning (Saeedmanesh & Geroliminis, 2016).

In theory, SNMF and spectral clustering are two highly relevant approaches according to the graph clustering objective but adopt fundamentally distinct ways to optimize the objective function. For a typical graph partitioning problem, the objective function is inherently consistent, which is mathematically equivalent to a trace maximization formulation (Kuang, Yun, & Park, 2015), as follows:

$$\begin{aligned} &\max_{H\geq 0, H^T H = I} \boldsymbol{Tr}\left(H^T W H\right) \Leftrightarrow \\ \min_{H\geq 0, H^T H = I} &\boldsymbol{Tr}\left(W^T W\right) - 2\boldsymbol{Tr}\left(H^T W H\right) + \boldsymbol{Tr}\left(H^T H\right) \Leftrightarrow \\ &\min_{H\geq 0, H^T H = I} \|W - HH^T\|_F^2 \end{aligned} \tag{8}$$

where $W\in\mathbb{R}^{N\times N}$, and $H\in\mathbb{R}^{N\times k}$ subject to $H\geq 0, H^T H = I$. It is an NP-hard problem to find the optimal solution to minimize the graph clustering objective (Eq. (8)) due to the two constraints on H. As a result, spectral clustering and SNMF attempt to loosen one of the constraints on H to obtain a tractable formulation. More concretely, SNMF keeps the nonnegativity constraint, while spectral clustering retains the orthogonality constraint. These two relaxed versions result in significantly different approaches for solving the optimization problems in Eq. (8). The orthogonality constraint in spectral clustering requires that each data sample falls into one cluster only. In contrast, by removing the orthogonality constraint in SNMF, the data sample can be assigned to several clusters with different membership values. In Zass and Shashua (2005), it was verified that the nonnegativity constraint on H plays a more crucial role than the orthogonality constraint. Additionally, Ding, He, and Simon (2005) have pointed out that keeping the nonnegativity constraint by SNMF brings about a near orthogonality approximation of the columns in matrix H. This property is beneficial and promising for SNMF to effectively figure out the graph partitioning problem.

To partition maritime ship traffic into balanced groups with similar sizes, the commonly used objective function termed Normalized Cut (*Ncut*) (Shi & Malik, 2000) is adopted to produce the proper clusters. In Bach and Jordan (2006), it was proved that the normalized cut can be expressed as follows:

$$Ncut = k - \boldsymbol{Tr}\left[H^T\left(D^{-\frac{1}{2}}WD^{-\frac{1}{2}}\right)H\right] = \boldsymbol{Tr}\left(H^T D^{-\frac{1}{2}}LD^{-\frac{1}{2}}H\right) \tag{9}$$

In terms of the derivation in Eq. (8), the minimization of *Ncut* can be achieved by using the normalized similarity matrix $\widetilde{W}$ to replace the W in the third formula in Eq. (8). Consequently, given the normalized similarity matrix $\widetilde{W}\in\mathbb{R}_+^{N\times N}$, the desired number of clusters k, and the nonnegativity constraints on $H\in\mathbb{R}_+^{N\times k}$, the graph clustering optimal model for SNMF can be formulated as follows:

$$\underset{H\geq 0}{\mathrm{argmin}}\|\widetilde{W} - HH^T\|_F^2 \tag{10}$$

The purpose of SNMF is to search for a nonnegative lower rank approximation H for the matrix $\widetilde{W}$. For an optimal matrix H, each column can be considered as the membership degree of the clustering data samples belonging to one cluster. Accordingly, the clustering assignments of the data samples can be directly identified in terms of the largest entry in each row in the low-rank matrix H.

Different optimization approaches can be contemplated for solving the minimization problem described in Eq. (10), such as the Newton-like algorithm (Gu & Saberi, 2019; Kuang et al., 2012), Alternating Nonnegative Least Squares (ANLS) algorithm (Kuang et al., 2015), and interior-point theory (Saeedmanesh & Geroliminis, 2016). In this study, the optimization problems are directly solved by implementing the Newton-like algorithm. It is suitable for small-size issues (e.g., $N <$ 3000) and can produce accurate solutions with higher-quality clustering results. Despite that, it may encounter a local minimum solution due to its sensitivity to the initialization of H. Regarding this issue, the Newton-like algorithm is performed many times with the randomly sampled initial H to find a global minimal solution or at least guarantee a near-global minimum.

3.3. Metrics development

The performance evaluation is crucial to ensure the effectiveness of the proposed methodology. Therefore, four metrics are introduced to evaluate and compare the traffic partitioning results.

The first adopted metric is 'NcutSilhouette' (*NS*) (Ji & Geroliminis, 2012), which is expressed as follows:

$$NS_k\left(A_o, A_p\right) = \frac{\sum_{u\in A_o}\sum_{v\in A_p}\left(1 - W_{uv}\right)^2}{vol\left(A_o\right)\cdot vol\left(A_p\right)} \tag{11}$$

where A_o and A_p represent two cluster subsets, $vol(A_o)$ represents the of samples in subset A_o, and k denotes the number of clusters. Here $NS_k\left(A_o, A_p\right)$ measures the average quadratic dissimilarity between clusters A_o and A_p.

On this basis, whether the ships in one cluster are properly grouped is measured using the following metric:

$$NS_k\left(A_o\right) = \frac{NS_k\left(A_o, A_o\right)}{NSN_k\left(A_o, A_q\right)} \tag{12}$$

where $NSN_k\left(A_o, A_q\right) = \min\left\{NS_k\left(A_o, A_p\right) \middle| A_p\in A,\ A_p\neq A_o\right\}$, and A_q

denotes the most relevant cluster with A_o. This metric measures the ratio of intra-cluster dissimilarity ($NS_k(A_o, A_o)$) over inter-cluster dissimilarity ($NS_k(A_o, A_q)$). Evidently, $NS_k(A_o)\langle 1$ indicates that cluster A_o is properly separated. Furthermore, the overall performance for a given traffic partitioning can be evaluated in terms of the average $NS_k(A_o)$ of all partitioned clusters, as follows:

$$NS_k = \frac{\sum_{o=1}^{k} NS_k(A_o)}{k} \tag{13}$$

A small NS_k value implies that the overall traffic scenario is well partitioned.

Additionally, the graph-based measure, *Ncut* (i.e., Eq. (9)), is employed to evaluate the comprehensive partitioning quality. This metric also considers both the similarity between different clusters and the similarity within the cluster. It is subsequently expressed as *NC*. The smaller the value of *NC* is, the better quality the partitioning scheme has.

Note that the above two comprehensive metrics are highly dependent on the designed similarity model (i.e., Eq. (7)). They cannot directly examine the spatial compactness and conflict connectivity of partitioning results because of the influence of the super parameters (e.g., α) in the similarity model. Therefore, two specific metrics associated with these two criteria are further presented.

With respect to the conflict connectivity, it can be calibrated based on the degree that the ship pairs with conflicts are segmented into different clusters, as follows:

$$f_1 = \sum_{i=1}^{N_{vc}} C(\gamma)_i \tag{14}$$

where N_{vc} denotes the number of ship pairs with conflicts that are arranged into different clusters, and $C(\gamma)_i$ represents the conflict criticality of the ith ship pair. A smaller f_1 value suggests that more ship pairs with conflicts are effectively clustered into the same group.

Regarding spatial compactness, NS_k can still be applied by using the real spatial distance to replace the dissimilarity in Eq. (11), expressed as f_2 in the following experimental section. The smaller the f_2 value, the smaller the spatial distance within the clusters is while the larger the spatial distance between the clusters is. This suggests that the spatial compactness of the traffic partitioning is well fulfilled.

4. Case study: Implementation and results

In this section, the effectiveness of the proposed traffic partitioning methodology is evaluated and discussed. Section 4.1 shows the investigated water area and the AIS data pre-processing. Section 4.2 illustrates the offline training results of the ship traffic route network. In Section 4.3, sensitivity analyses on the super parameters in the proposed methodology are performed. Section 4.4 illustrates how the proposed methodology assists in intelligent MSA and capturing high-risk traffic clusters. Section 4.5 conducts model comparison and validation to highlight the reliability of the proposed methodology. Furthermore, insights and implications are analysed in Section 4.6.

4.1. Study area and data description

The illustrative case to be analysed in this study is the Ningbo-Zhoushan Port. It is a unique deep-water port with some of the densest traffic in the world in terms of cargo throughput. There are more than 620 production berths, including approximately 170 large-scale berths above 10,000 tons and more than 100 super large-scale deep-water berths above 50,000 tons. The restricted geographical regions, various vessel types, diversified movement behaviour, and the presence of complex environmental conditions expose it as a highly sophisticated traffic situation. It therefore provides typical challenging scenarios for traffic partitioning tests. The port areas for capturing the traffic clusters cover longitudes from 121°52′E to 122°22′E and latitudes from 29°43′N to 30°02′N (see Fig. 4). Additionally, one month's AIS data records from 01/11/2018 to 30/11/2018 are collected from the area.

Although the AIS data provides a reliable source of information for maritime traffic research, it is not immune to data errors due to various technical malfunctions and failures. Therefore, it is essential to eliminate the possible errors/noises before conducting the experimental analysis. The AIS data pre-processing consists of a sequential process of four steps, which are the noise removal for each ship property (Kang et al., 2018), trajectory extraction and segmentation, trajectory consistency verification (Zhao, Shi, & Yang, 2018), and trajectory linear interpolation (Zhang, Meng, & Fwa, 2019). By doing so, it ensures the reliability and effectiveness of AIS data for maritime knowledge extraction and traffic partitioning analysis.

4.2. Ship traffic route network extraction and analysis

Based on the procedure in Section 3.1.2, Fig. 4 illustrates the identified unnavigable regions and the derived ship traffic route network. The dark red areas represent the unnavigable regions, while the blue lines indicate the traffic route network. It is found that the blue lines effectively describe the skeleton of the navigable areas, which reveals its goodness-of-fit. Particularly, this precise and structured representation of the maritime traffic network allows the real spatial distance between ship pairs to be measured. To enlighten the use of the derived network for real spatial distance computation, an example of how to identify the spatial relations of ship pairs based on the graph-based topology is presented in Fig. 5. In the figure, ships *B* and *C* are separated from *A* by obstacles. The real spatial distance (RSD) based on the route network and the linear spatial distance (LSD) between ships *A* and *B* are 6.43 and 4.51 nautical miles (nm), respectively. It is evident that the distance of a ship pair should be better measured by the length of their shortest path on the route network instead of using the physical distance because of the obstacles between them. The route network contributes to identifying the real spatial distance in complex waters as a first step toward recognizing actual spatial compact traffic clusters.

4.3. Sensitivity analyses on different design parameters

According to the methodology section, four super parameters need to be determined to obtain the optimal traffic partitioning results, which are D_1, D_2, β, and α. The first three come from the compactness similarity measure model (i.e., Eq. (6)), while the last one is the trade-off weighting coefficient used to balance the spatial compactness and conflict connectivity (i.e., Eq. (7)). Their optimal values are respectively confirmed based on the following sensitivity analysis.

4.3.1. Distance measure parameters

In the maritime field, when the distance between ship pairs is larger

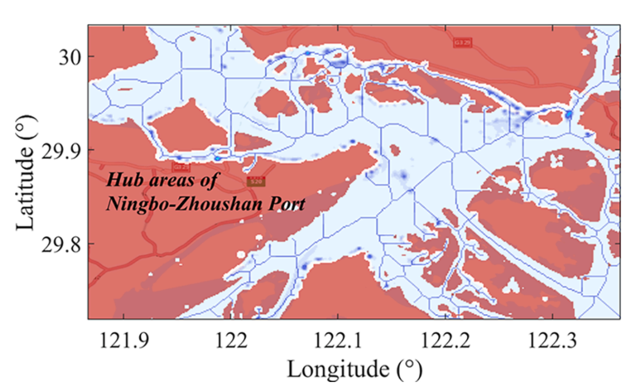

Fig. 4. The identified unnavigable regions based on kernel density estimation and extracted ship traffic route network based on image processing technique in Ningbo_Zhoushan port.

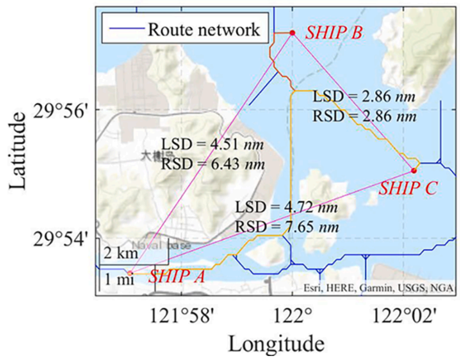

Fig. 5. An example of real spatial distance computation based on the formulated traffic route network.

than 6 nautical miles, they are not considered to be in an encountering situation (Cho et al., 2020; Zhang, Goerlandt, Kujala, & Wang, 2016). Hence, D_2 is directly set to be 6 nautical miles to distinguish the encountered and non-encountered ship pairs. On the other hand, the alarm procedure is normally activated when the predicted distance between ship pairs is smaller than 1 nautical mile in open sea (Hu et al., 2019). Considering the high density and greater tolerance for a small distance in complex port waters, four values within 1 nautical mile are selected for D_1, which are 0.125, 0.25, 0.5, and 1. As for β, it is set as 0.5, 1, 2 and 3 to control the decline rate of W_{ij}^d with $Dist_{ij}$. Fig. 6 illustrates the average f_1 and f_2 of the tested traffic scenarios when using different combinations of D_1 and β. Based on the Pareto principle in multiple objective optimizations that one is not dominated by others if at least one objective is better, three optimal combinations of (0.125, 2), (0.125, 3), and (0.25, 3) that are non-dominated by any other combination constitute the Pareto front. It is also observed that both D_1 and β have profound impacts on the partitioning quality in terms of the change degree of f_1 and f_2, indicating the necessity to perform sensitivity analysis to find the optimal combinations. Notably, f_1 is more sensitive to these two parameters than f_2 in terms of its higher fluctuations.

To further identify the best D_1 and β, the performance of each pair of super parameter combinations is compared. For combinations A and B, their domination relations for each traffic scenario can be identified. Then the percentage that each one dominates another in all experimental traffic scenarios is calculated, and the one with a higher percentage is better than another one. By doing so, the number of times that each combination dominates other combinations can be counted. According to Fig. 7, the combination of $D_1 = 0.125$ and $\beta = 2$ dominates all other 15 combinations. These results enable us to determine the optimal parameter combination as 0.125 and 2 by observing the relevant turning points.

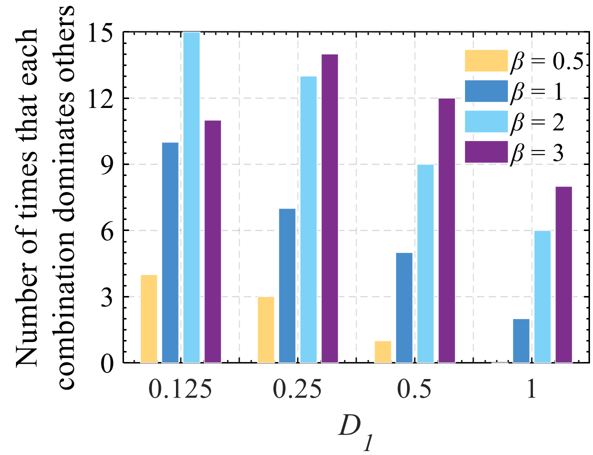

Fig. 7. Traffic partitioning performance comparison with respect to various combinations of D_1 and β.

4.3.2. Composite similarity weight coefficient

The weight coefficient α is fundamental to supporting a good trade-off between the two considered clustering criteria. Therefore, the traffic partitioning results with different α are analysed. In Fig. 8(a), an increasing α results in a decrease/improvement in f_1 and in an increase/deterioration in f_2, implying the conflicting relations between the conflict connectivity and spatial compactness. When α is lowered, more penalty is imposed on the compactness dissimilarity and vice versa. Therefore, an appropriate way is applied to determine the optimal α. It is based on the principle that the increase of α should lead to a more substantial improvement in one metric than the deterioration in another. The change degree of the metric improvement/deterioration from (m-1)th to mth α is calibrated using the following equation:

$$\Delta\delta^m = \left(\frac{\overline{f}_1^m - \overline{f}_1^{m-1}}{\left|\overline{f}_1^M - \overline{f}_1^1\right|} + \frac{\overline{f}_2^m - \overline{f}_2^{m-1}}{\left|\overline{f}_2^M - \overline{f}_2^1\right|}\right) \times 100\% \tag{15}$$

where $\overline{f}_1^m$ and $\overline{f}_2^m$ represent the average f_1 and f_2 for the mth α, $m = 1, 2, \ldots, M$. In Eq. (15), the first term measures the improvement degree (negative index) in f_1 while the second term measures the deterioration degree (positive index) in f_2. Additionally, the normalization is conducted by using the denominators to make the change degree of the two metrics comparable in the same scale. Hence, when $\Delta\delta^m < 0$, it implies a whole improvement gained by increasing α and vice versa. From Fig. 8 (b), f_1 starts to decline slowly while f_2 starts to rise rapidly when α reaches up to 0.6. More exactly, $\Delta\delta^m$ remains larger than 0 when $\alpha \geq 0.6$ (see the subfigure in Fig. 8(b)). It means that the increase in α from 0.6

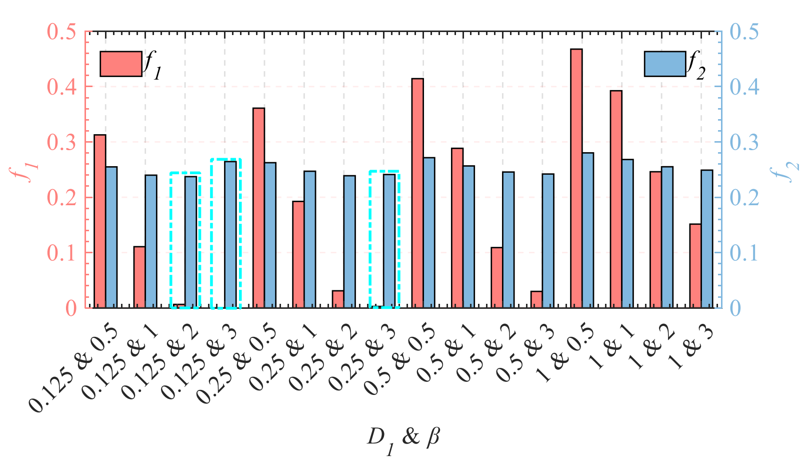

Fig. 6. Average f_1 and f_2 of the partitioned traffic scenarios with various combinations of D_1 and β.

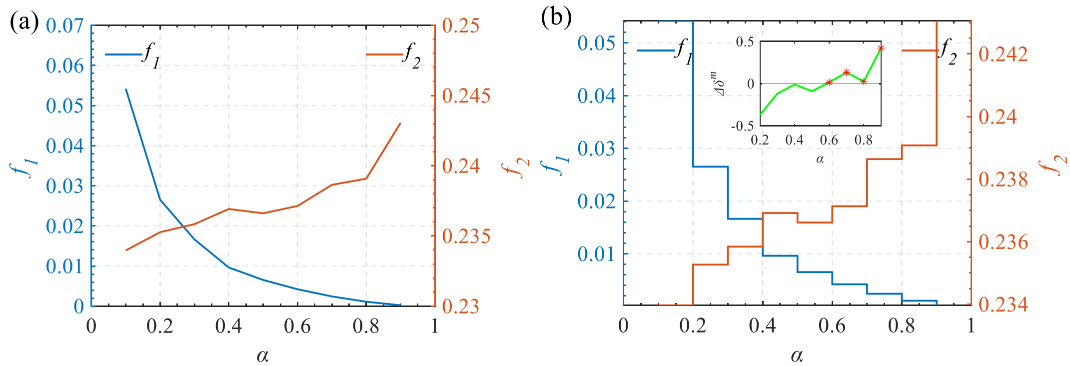

Fig. 8. Sensitivity analysis of composite similarity weight coefficient α: (a) average f_1 and f_2 with different α; (b) increase/decrease degree in f_1 and f_2 with the increase in α.

would not improve the whole partitioning performance. Therefore, a sensible balance between the two conflicting objectives is achieved by using $\alpha = 0.6$.

4.4. Application results and analysis

In this subsection, the effectiveness of the proposed traffic partitioning methodology is demonstrated based on the real cases. It started by highlighting the application effect on decomposing the whole traffic

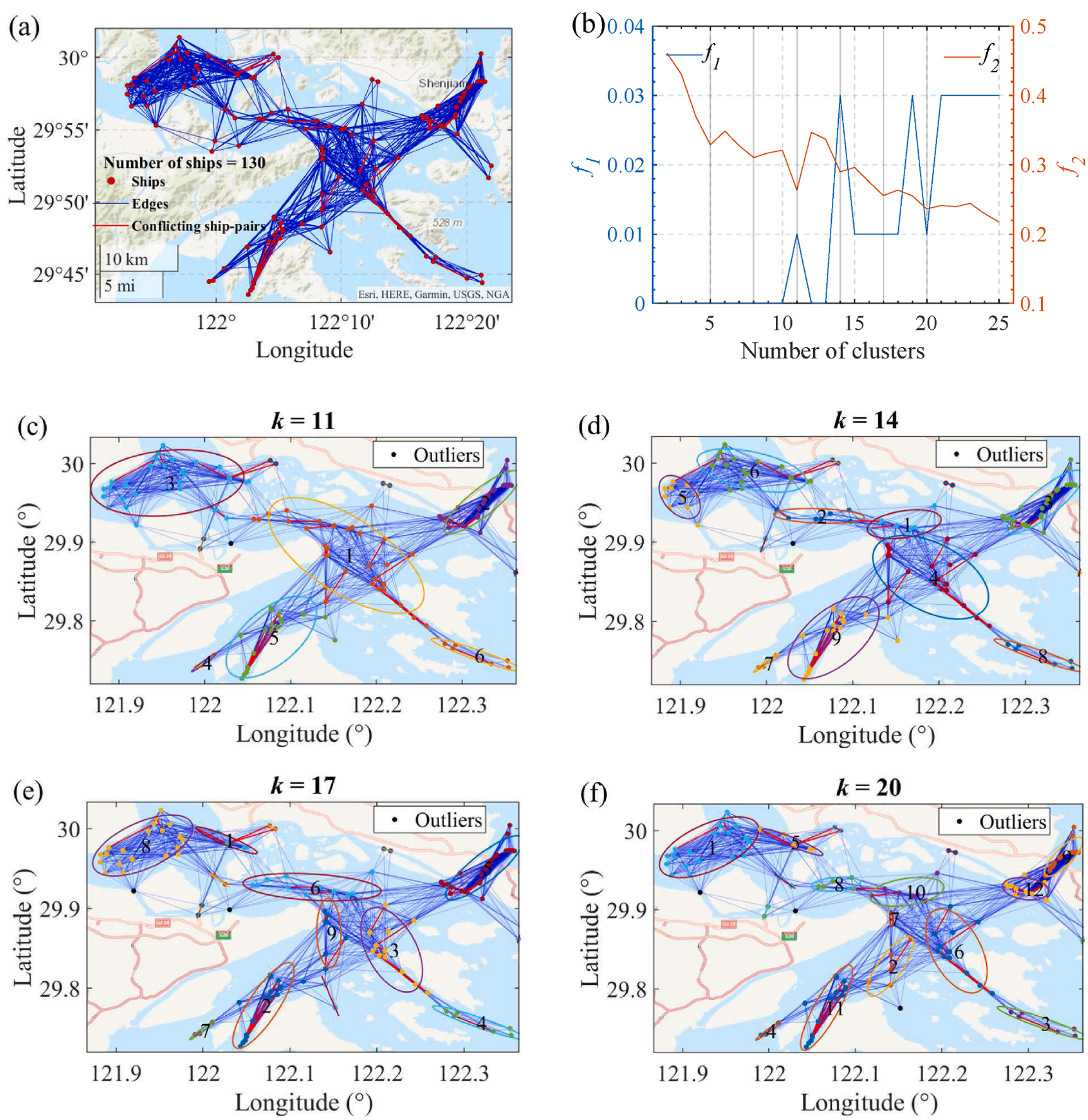

Fig. 9. Illustration of ship traffic partitioning results at one moment. (a) visualization of ship traffic network; (b) f_1 and f_2 with different numbers of clusters; (c)-(f) traffic partitioning results when the numbers of clusters are 11, 14, 17 and 20.

complexity through a specific maritime traffic scenario. Then a traffic evolution scheme is analysed to displayed how the proposed methodology sheds light on enhancing maritime traffic surveillance and guiding ship anti-collision risk management.

Fig. 9 illustrates the ship traffic partitioning results for a traffic scenario with high traffic density at one moment. In Fig. 9(a), the visualization of the graph representation of traffic relations is displayed. The red points represent the ships, the blue lines represent the similarities (i.e., W_{ij} greater than 0) between ship pairs, and the red lines indicate that the ship pairs have conflicting interactions. As the number of clusters for a clustering issue needs to be determined in advance, the values of f_1 and f_2 when performing clustering with different numbers of clusters are presented in Fig. 9(b). Four traffic partitioning results in terms of multiple troughs of orange polyline in Fig. 9(b) are exhibited (see Fig. 9(c-f)). It is evident from these figures that the produced traffic clusters are spatially compact, while at the same time most of the conflicting ship pairs are assigned to the same clusters, illustrating the good properties of the proposed methodology. In the meantime, it is found that there are complicated conflicting relations among ships (e.g., Clusters 2 and 3 in Fig. 9(e)), hence much attention should be paid to the spatiotemporal interactions of multiple ships instead of focusing on the interactions between ship pairs. Additionally, the clustering quality is robust with respect to different numbers of clusters, and more outliers (i.e., the produced clusters with one ship, which can be regarded as safe ships) tend to be filtered out with the increase in the number of clusters. This implies that instead of focusing on a single number of clusters, one can conduct a multi-view analysis by exploring clustering performance of a traffic scenario with different input numbers of clusters. Overall, the proposed methodology performs well in partitioning the whole ship traffic into several high spatial compact and conflict-connected clusters.

Besides, the properties of the generated clusters in Fig. 9(d) and (f) are examined and analysed in more detail. Here the clusters with a number of ships smaller than 3 are not labelled. The number of ships and $NS_k(A_o)$ (Eq. (12)) of each produced cluster are shown in Fig. 10(a) and (c). It is found that the values of $NS_k(A_o)$ of all produced clusters are smaller than 1, implying the traffic scenario is properly partitioned. Then Fig. 10(b) and (d) present each cluster's traffic density and sum of conflict criticality. The traffic density of each cluster is measured based on the average ship density in one cluster and one can refer to the work in Tan, Steinbach, and Kumar (2016) about the density definition. As shown in these figures, the clusters with high density/conflict severity can be easily found, e.g., Cluster 3 in Fig. 10(b) and Cluster 9 in Fig. 10(d). This indicates the necessity of decomposing the whole traffic instead of directly implementing MSA from a global/regional perspective. Regarding the practical application of the proposed methodology, one can check the risk/density indices of partitioned traffic clusters to assist surveillance operators in paying more attention to the critical traffic clusters. In this way, the proposed traffic partitioning methodology is supportive for improving situational awareness and identifying high-risk/density traffic clusters.

To illustrate how the proposed methodology enhances the maritime operational monitoring and provides vital support in anti-collision decision-making over the water areas of interest, the evolution of density and conflict criticality of both the whole ship traffic and generated traffic clusters is provided in Fig. 11. Here the maximum density and conflict criticality of traffic clusters generated at each moment are exhibited. The number of clusters adopted for partitioning all traffic scenarios at different time moments is 15. From the figure, two interesting findings are revealed. First, the whole traffic density is unfeasible to assist in comprehending the traffic situation due to its slight fluctuations with time. Based on this indicator, maritime operators may encounter difficulties in issuing timely warnings. By contrast, the density of generated traffic clusters varies over time, which can facilitate maritime operators and regulators in identifying which time moments are in high traffic complexity. Second, the conflict criticalities of the whole ship traffic and the traffic clusters show consistent trends. Notably, they are very close during some periods, e.g., 130–150 min. This implies that the conflicting ship pairs have a high probability of

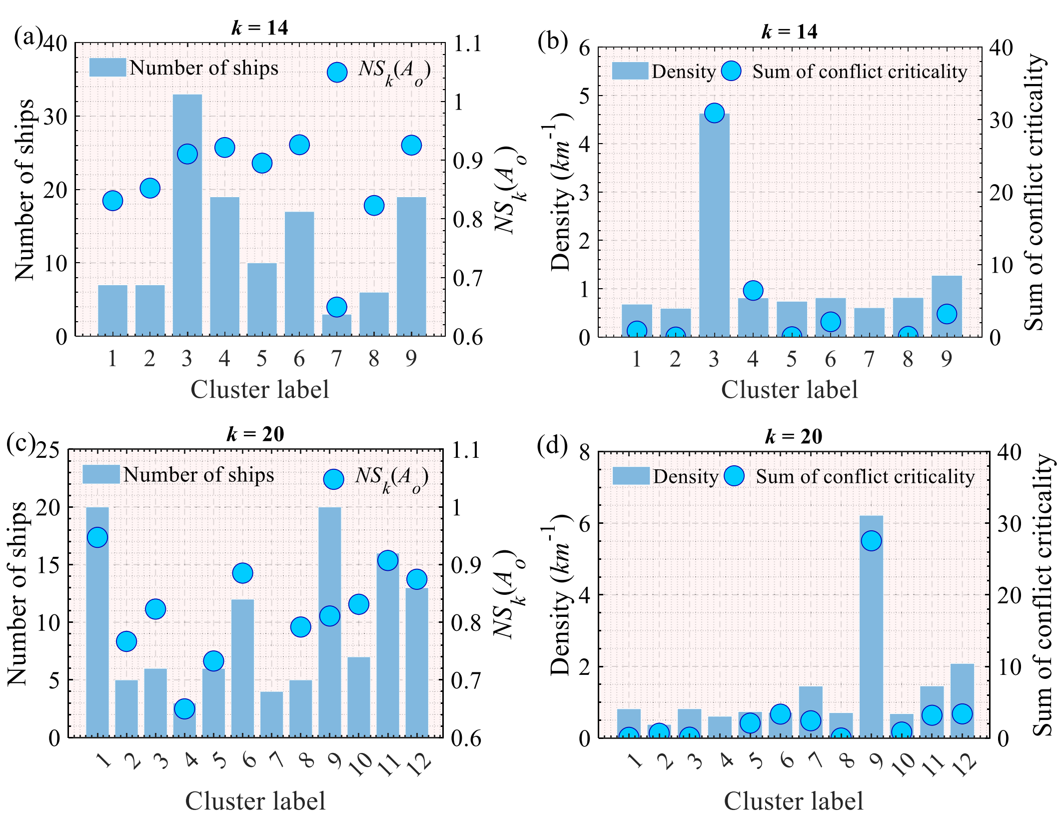

Fig. 10. Feature statistics of each cluster in Fig. 9(d) and (f), including number of ships, $NS_k(A_o)$, traffic density and sum of conflict criticality.

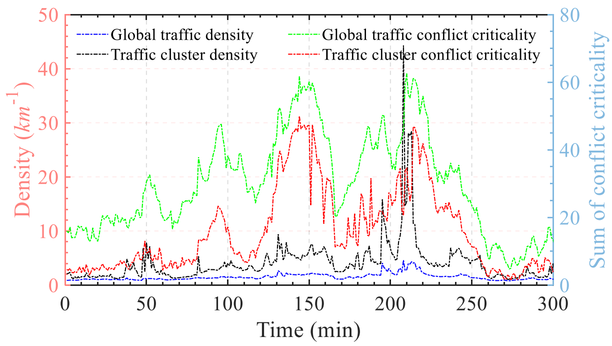

Fig. 11. Density and conflict criticality evolution of whole ship traffic and traffic clusters over five hours.

being in the same cluster; that is, the traffic partitioning approach can effectively group the ships with high conflict relations into one cluster, which further provides a practical foundation for maritime operators to devise and implement anti-collision risk control strategies. These observations highlight the necessity and effectiveness of the traffic partitioning approach in strengthening MSA and supporting collision risk control.

4.5. Model comparison and validation

The model comparison and validation are essential for the practical application of the modelling methodology. Therefore, the proposed methodology is first compared with the widely used graph-based algorithm (i.e., spectral clustering) to exhibit the superiority of the SNMF framework. Subsequently, the functionality and utility of the functional modules (i.e., the composite similarity model) are tested and examined.

Table 2 presents a comprehensive comparison between the proposed methodology and spectral clustering. As shown in the table, the overall performance of the proposed methodology outperforms that of spectral clustering in terms of multiple evaluation metrics. The reason is mainly because of the good properties of the SNMF framework and the fact that the orthogonality constraint has smaller influence on it. Note that other classes of clustering algorithms, like prototype-based and density-based clustering, are not considered for comparison because they focus on each data sample's features. For instance, k-means algorithm performs clustering based on the cluster centres, which is meaningless when the spatial distance between ships is measured by the length of their shortest path on the route network instead of Euclidean distance. The DBSCAN algorithm requires identifying the core samples and has difficulty of handling datasets with varying densities. Therefore, they are not feasible for traffic partitioning based on the interactions/similarities between ships. To further evaluate the generalization ability of the proposed model, extensive comparisons of the two approaches with different numbers of clusters and ships are conducted. As shown in Fig. 12, the proposed methodology remains superior to spectral clustering under all kinds of situations with respect to both the NS_k and NC. These results confirmed the stability and scalability of the traffic partitioning model.

As the designed composite similarity model is among the most critical methodological contributions in this work, the functionality and utility of key modules in the model are tested and analysed from the following two aspects. Firstly, a traffic scenario is displayed, where using the Euclidean/physical distance may encounter issues in ensuring good

Table 2
A comprehensive comparison between proposed traffic partitioning model and spectral clustering.

Clustering model	NS_k	NC	f_1	f_2
Proposed model	0.535	0.032	0.006	0.371
Spectral clustering	0.741	0.115	0.157	0.538

clustering quality (see Fig. 13(a)), while the proposed spatial distance measure model could be potentially better (see Fig. 13(b)). According to Fig. 13(a), ships i and j are surrounded by obstacles and have negligible interactions with other ships. However, they are grouped into clusters, indicating that the Euclidean distance is not appropriate for complex waters with restricted geographical characteristics. In contrast, by using the shortest path length on the derived traffic route network as the distance measure criteria, ships i and j in Fig. 13(b) can be identified as outliers. Besides, it is found that Group k in Fig. 13(b) is also well separated by using the newly proposed distance measure model. These comparisons reveal that the ship traffic should be more reasonably grouped based on their real spatial distance relations instead of their physical distance. As a result, the proposed spatial distance measure leads to a significant improvement in the traffic partitioning performance.

Another traffic scenario is used to examine the necessity of considering both the spatial compactness and conflict connectivity indices. Fig. 14 illustrates a clustering performance comparison in which one conducts clustering only based on the compactness similarity model, whereas another uses the composite similarity model. From Fig. 14(a), the generated traffic clusters are highly spatial compact, but the ships in conflict are not guaranteed to be assigned to the same clusters. For example, the ships with conflict relations in Circles i, j, and p are not well grouped, which is detrimental to discovering conflicting interaction patterns among ships. On the other hand, it is found from Fig. 14(b) that the conflicting ships are well clustered while the spatial compactness is maintained properly as well. Indeed, Fig. 14(a) provides an extreme scenario with the weighting coefficient $\alpha = 0$, while Fig. 14(b) makes a good trade-off between spatial compactness and conflict connectivity. It must be mentioned that conflict-based interactions among ships receive more attention from ship navigators and maritime operators than distance-based interactions. This is because high conflicting relations explicitly indicate the potentially dangerous situation, while high-density relations merely mean the traffic situation is busy and complicated. Therefore, both the conflict connectivity and spatial compactness indices are critical to improving the traffic partitioning quality. In summary, the designed bi-objective similarity model is desirable as it allows the two indices to be considered simultaneously.

5. Discussion, insights, implications, and limitations

This study conducts a comprehensive experimental analysis and validation for the proposed traffic partitioning methodology, covering from sensitivity analysis of super parameters and application case demonstration to model comparison as well as examination of key modules' functionality.

Based on the experimental analysis, three methodological insights can be drawn. Firstly, performing MSA in terms of global traffic situation evaluation indices is not constantly recommended for maritime operators because these indices will likely provide less insight into the comprehension of the traffic situation. By contrast, traffic partitioning could improve traffic pattern interpretability and facilitate the discovery of high-risk/density traffic clusters. Secondly, the incorporation of both spatial compactness and conflict connectivity as well as the influence of water topography features into traffic partitioning could help obtain a full understanding of the actual multi-attribute interrelationships among ships. Existing studies such as Liu et al. (2019) and Zhen et al. (2017; 2021) have not addressed either of these two issues. Thirdly, the proposed methodology has strong applicability and robustness to complex waters. Generally, complex waters refer to ports or channels that are exposed to highly intricate traffic situations, characterised by high traffic density, dynamic ship movements, and restricted geographical features, among others. The proposed approach effectively addresses these complexities by 1) utilizing a probabilistic conflict detection model that considers traffic motion dynamics and uncertainty to accurately identify potential collision danger; 2) adopting a traffic route

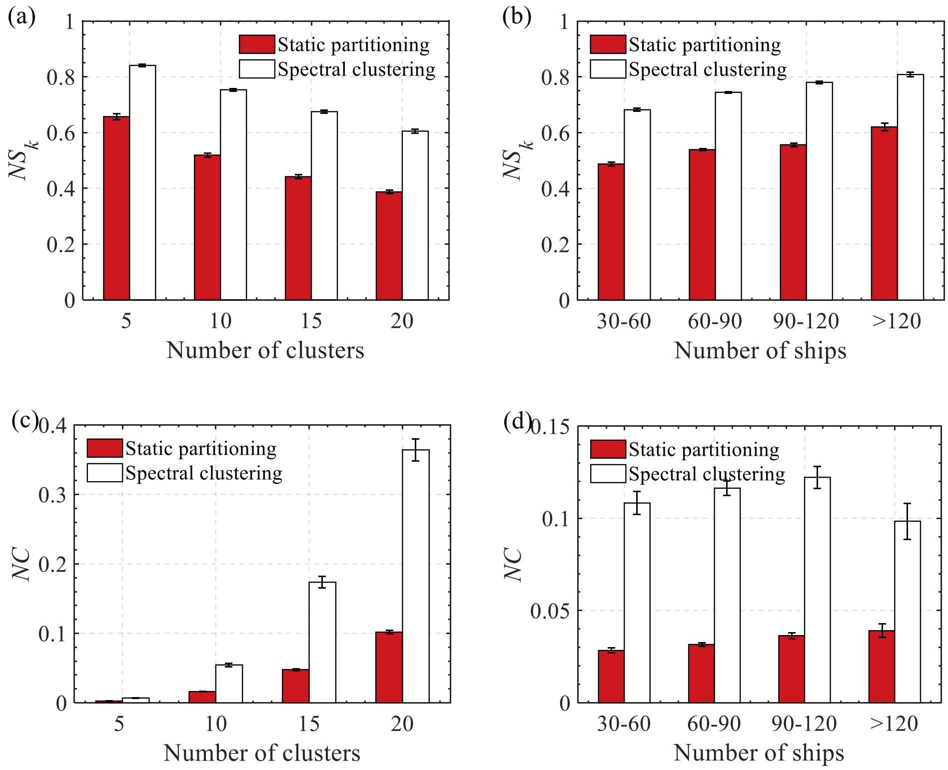

Fig. 12. Performance comparison between the proposed traffic partitioning model and spectral clustering with different numbers of ships and clusters: (a-b) NS_k comparisons; (c-d) *NC* comparisons.

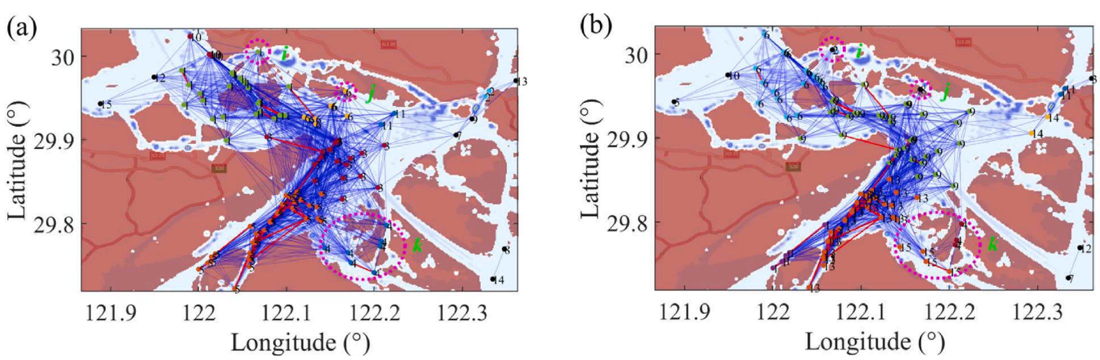

Fig. 13. A comparison of clustering results when using (a) Euclidean distance and (b) real spatial distance based on the formulated traffic route network.

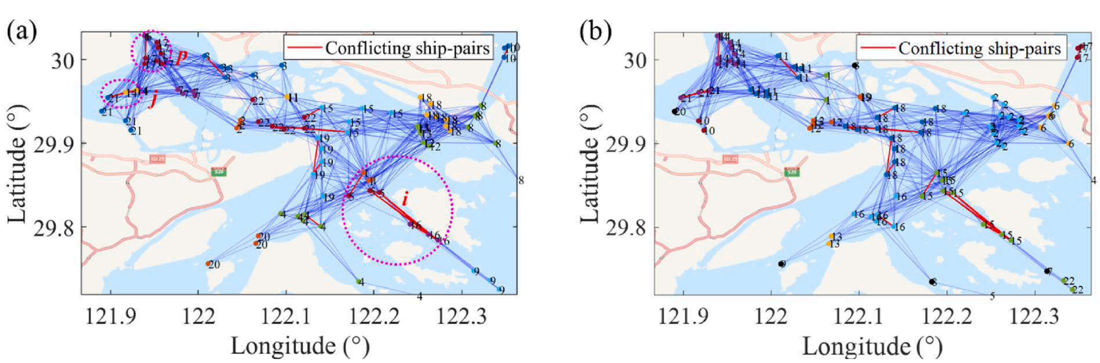

Fig. 14. A comparison of clustering results when using (a) compactness similarity model and (b) composite similarity model.

network to measure spatial distance reliably when ships are blocked by obstacles such as islands; and 3) designing a traffic partitioning model that focuses on the multi-ship interactions in heavy traffic waters rather than ship pair interrelationships. These desirable features of the proposed approach allow it to be easily tailored and applied to other port and waters.

The analytical discussion of the methodology and experimental results also provides practical implications to maritime surveillance operators, vessel navigators, and port stakeholders. With respect to the implications for maritime surveillance operators such as maritime management authorities and port safety-related departments, the proposed methodology is helpful for them to enhance maritime awareness capabilities by decomposing the complexity levels of the whole traffic situation. Such a methodology provides insightful knowledge concerning how maritime risks are distributed spatially and which traffic clusters need priority attention instead of focusing on the global traffic situation. Without the proposed methodology, maritime controllers would have to undertake MSA in complex waters based on their intuition and experience, which may result in them being subject to tremendous monitoring pressure and even impede the timely implementation of collision risk management strategies. This issue even becomes more worrisome once the traffic situations in ports are more complex due to the occurrence of mixed encounter situations involving both manned and autonomous ships. However, the proposed methodology will aid maritime surveillance operators in promoting maritime traffic safety management and proactively making timely and efficient decisions to control ship collision risks.

The most important implication of this study for vessel navigators is to provide useful guidelines for anti-collision risk management. Vessels tend to focus on their operations and situations rather than taking the traffic situation from a global/regional perspective. However, in a complicated encounter scenario, the measures taken by one ship to avoid a collision with another ship could pose a higher risk to others. This is because the navigational complexity of a scenario may be highly associated with multiple dependent conflicts, especially in high-traffic waters, which would lead to confusion in the anti-collision decision-making design. This work captures the traffic clusters with high intra-interactions by a new traffic partitioning methodology, thereby aiding the traffic conflicts to be resolved at a regional level instead of based on a local ship pair. In other words, it makes a ground-breaking development by shifting the anti-collision control from being dependent on the vessel navigator locally to taking strategical action so that the collision risks of multi-ship encounters can be better managed. On the other hand, the division of the whole traffic into small clusters by the traffic partitioning methodology can support to tackle each cluster's risk independently, which would not make the design of risk mitigation schemes too sophisticated. As a result, the proposed methodology would be particularly applicable in autonomous maritime anti-collision risk management and lay a solid foundation for the future coexistence of mixed manned and autonomous ships.

This study also brings significant benefits to strengthening port competitiveness and sustainability. Implementing the proposed approach in an intelligent transportation support system offers the potential to effectively manage port traffic. From the economic development point of view, intelligent traffic safety management means the enhancement of port performance and efficiency, which is seen as one of the key determinants of attracting port users and investment. Evidently, the ships will be more willing to give priority to the ports with high-end port services. This work therefore makes a significant contribution in achieving the competitive advantages of the port over its competitors.

Although the proposed analytical approach has demonstrated its superiority over traditional traffic cluster detection models, it still has limitations that could be addressed in future research. These limitations include the following:

1. The dynamic evolution characteristics of ship traffic clusters require further exploration, especially in the port areas where traffic experiences intense dynamic behaviour at various times of the day. The influence of traffic evolution over time on traffic partitioning should not be underestimated. It is essential to develop a dynamic traffic partitioning technique to produce temporally consistent partitioning results that are less sensitive to traffic evolution. This can facilitate the continuous implementation of anti-collision risk management strategies for any detected traffic cluster.
2. The effect of traffic topological properties on the collision risk of traffic clusters needs to be explored. This study determines and monitors the critical traffic clusters based on the sum of conflict criticality and traffic density. However, these indicators alone are insufficient to capture the complete interactions among multiple ships. Other indicators, such as Clustering Coefficient and K-shell Decomposition, in complex network theory, can measure the resolving difficulty of multiple ship conflicts. Therefore, advanced models that can measure the interactions among multiple ships from various perspectives could be developed, to assist maritime operators in reasonably determining which traffic clusters should be given prioritized attention.
3. In this study, the maritime traffic partitioning approach captures traffic clusters based on conflict and spatial distance relations among ships. However, more vessel motion interactions, such as converging/diverging trends of ship pairs and ship movement behaviour patterns, could be incorporated into the traffic partitioning process to better unveil complementary information related to ship traffic interactions.
4. A new conflict resolution approach that can coordinate and balance intra-cluster and inter-cluster collision risks deserves more attention. This approach could guide surveillance operators to devise multi-layered strategies for hierarchical risk control, to realize maritime traffic safety surveillance and management.

6. Conclusion

The development of advanced MSA techniques and tools is one of the essential components of emerging intelligent ports and autonomous ships. This study proposes an optimal ship traffic partitioning methodology that captures conflict-connected and spatial compact traffic clusters to enhance situational awareness and support collision risk management. The developed methodology has been embedded with several unique features: 1) the multi-attribute interrelationships between ships are considered, including their conflict relation and spatial distance; 2) it identifies the exact spatial distance based on maritime traffic knowledge extraction, enabling the methodology to be adaptive to complex geographical waters; and 3) a more competitive graph-based clustering formulation is employed to support robust traffic partitioning. Extensive numerical experiments with real AIS-based data are conducted to demonstrate the practicality and superiority of the proposed methodology. The experimental results show that the proposed approach works well in partitioning the whole ship traffic scenario into spatial compact and conflict-connected clusters, can help detect high-risk/density traffic clusters, and is supportive for collision risk control. Additionally, the model comparison and validation results indicate that the proposed model remains superior to the typical clustering techniques and has good generalization ability and stability under various traffic scenarios. It sheds valuable light on supporting intelligent maritime surveillance and promoting autonomous anti-collision risk management. Furthermore, it provides the possibility and applicability for the intelligent maritime safety management of both manned ships and autonomous ships as well as their hybrid traffic. Therefore, the proposed methodology could be applied in the maritime autonomous navigation system to aid in automatic situation awareness and update.

Potential future research can be conducted from the following aspects. First, the dynamic evolution characteristics of maritime traffic could be incorporated into the traffic partitioning process to produce temporally stable clusters and support the continuously implementation of risk management strategies for the detected traffic clusters. Second, a new evaluation model that can compare the collision risk between different traffic clusters is worth being developed to more reliably identify and monitor high-risk traffic clusters. Lastly, it would be valuable to explore conflict resolution strategies that can effectively control and balance the local and regional collision risks.

CRediT authorship contribution statement

Xuri Xin: Methodology, Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing, Validation. **Kezhong Liu:** Conceptualization, Methodology, Supervision, Funding acquisition, Project administration, Writing – review & editing. **Sean Loughney:** Project administration, Supervision, Writing – review & editing. **Jin Wang:** Project administration, Supervision, Writing – review & editing. **Huanhuan Li:** Investigation, Formal analysis, Writing – review & editing. **Zaili Yang:** Conceptualization, Methodology, Supervision, Project administration, Writing – original draft, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No. 52031009), and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 864724).

References

Arguedas, V. F., Pallotta, G., & Vespe, M. (2017). Maritime traffic networks: From historical positioning data to unsupervised maritime traffic monitoring. *IEEE Transactions on Intelligent Transportation Systems, 19*(3), 722–732.

Bach, F. R., & Jordan, M. I. (2006). Learning spectral clustering, with application to speech separation. *The Journal of Machine Learning Research, 7*, 1963–2001.

Bakdi, A., Glad, I. K., & Vanem, E. (2021). Testbed scenario design exploiting traffic big data for autonomous ship trials under multiple conflicts with collision/grounding risks and spatio-temporal dependencies. *IEEE Transactions on Intelligent Transportation Systems, 22*(12), 7914–7930.

Chen, P., Huang, Y., Mou, J., & van Gelder, P. (2018). Ship collision candidate detection method: A velocity obstacle approach. *Ocean Engineering, 170*, 186–198.

Chen, P., Huang, Y., Mou, J., & van Gelder, P. (2019). Probabilistic risk analysis for ship-ship collision: State-of-the-art. *Safety Science, 117*, 108–122.

Chin, H. C., & Debnath, A. K. (2009). Modeling perceived collision risk in port water navigation. *Safety Science, 47*(10), 1410–1416.

Cho, Y., Han, J., & Kim, J. (2020). Efficient COLREG-compliant collision avoidance in multi-ship encounter situations. *IEEE Transactions on Intelligent Transportation Systems, 23*, 1899–1910.

Chunaev, P. (2020). Community detection in node-attributed social networks: A survey. *Computer Science Review, 37*, Article 100286.

Ding, C., He, X., & Simon, H. D. (2005). On the equivalence of nonnegative matrix factorization and spectral clustering. In *Proceedings of the 2005 SIAM International Conference on Data Mining* (pp. 606–610).

Duan, G., Fan, T., Chen, X., Chen, L., & Ma, J. (2021). A hybrid algorithm on the vessel routing optimization for marine debris collection. *Expert Systems with Applications, 182*, Article 115198.

Fang, Z., Yu, H., Ke, R., Shaw, S.-L., & Peng, G. (2018). Automatic identification system-based approach for assessing the near-miss collision risk dynamics of ships in ports. *IEEE Transactions on Intelligent Transportation Systems, 20*(2), 534–543.

Filom, S., Amiri, A. M., & Razavi, S. (2022). Applications of machine learning methods in port operations–A systematic literature review. *Transportation Research Part E: Logistics and Transportation Review, 161*, Article 102722.

Fujii, Y., & Tanaka, K. (1971). Traffic capacity. *Journal of Navigation, 24*(04), 543–552. https://doi.org/10.1017/S0373463300022384

Gil, M. (2021). A concept of critical safety area applicable for an obstacle-avoidance process for manned and autonomous ships. *Reliability Engineering & System Safety, 214*, Article 107806.

Gil, M., Montewka, J., Krata, P., Hinz, T., & Hirdaris, S. (2020). Determination of the dynamic critical maneuvering area in an encounter between two vessels: Operation with negligible environmental disruption. *Ocean Engineering, 213*, Article 107709.

Gilbert, A., Petrovic, D., Pickering, J. E., & Warwick, K. (2021). Multi-attribute decision making on mitigating a collision of an autonomous vehicle on motorways. *Expert Systems with Applications, 171*, Article 114581.

Goerlandt, F., Montewka, J., Kuzmin, V., & Kujala, P. (2015). A risk-informed ship collision alert system: Framework and application. *Safety Science, 77*, 182–204.

Gu, Z., & Saberi, M. (2019). A bi-partitioning approach to congestion pattern recognition in a congested monocentric city. *Transportation Research Part C: Emerging Technologies, 109*, 305–320.

Hao, S., Zhang, Y., Cheng, S., Liu, R., & Xing, Z. (2018). Probabilistic multi-aircraft conflict detection approach for trajectory-based operation. *Transportation Research Part C: Emerging Technologies, 95*, 698–712.

Hernandez-Romero, E., Valenzuela, A., & Rivas, D. (2019). A probabilistic approach to measure aircraft conflict severity considering wind forecast uncertainty. *Aerospace Science and Technology, 86*, 401–414.

Hu, L., Naeem, W., Rajabally, E., Watson, G., Mills, T., Bhuiyan, Z., … Pekcan, C. (2019). A multiobjective optimization approach for COLREGs-compliant path planning of autonomous surface vehicles verified on networked bridge simulators. *IEEE Transactions on Intelligent Transportation Systems, 21*(3), 1167–1179.

Huang, Y., Chen, L., Chen, P., Negenborn, R. R., & van Gelder, P. (2020). Ship collision avoidance methods: State-of-the-art. *Safety Science, 121*, 451–473.

Iphar, C., Ray, C., & Napoli, A. (2020). Data integrity assessment for maritime anomaly detection. *Expert Systems with Applications, 147*, Article 113219.

Ji, Y., & Geroliminis, N. (2012). On the spatial partitioning of urban transportation networks. *Transportation Research Part B: Methodological, 46*(10), 1639–1656.

Kang, L., Meng, Q., & Liu, Q. (2018). Fundamental diagram of ship traffic in the Singapore Strait. *Ocean Engineering, 147*, 340–354.

Kuang, D., Ding, C., & Park, H. (2012). Symmetric nonnegative matrix factorization for graph clustering. In *Proceedings of the 2012 SIAM International Conference on Data Mining* (pp. 106–117).

Kuang, D., Yun, S., & Park, H. (2015). SymNMF: Nonnegative low-rank approximation of a similarity matrix for graph clustering. *Journal of Global Optimization, 62*(3), 545–574.

Kulkarni, K., Goerlandt, F., Li, J., Banda, O. V., & Kujala, P. (2020). Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus. *Safety Science, 129*, Article 104798.

Lam, L., Lee, S.-W., & Suen, C. Y. (1992). Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 14*(09), 869–885.

Lee, H.-J., & Rhee, K. P. (2001). Development of collision avoidance system by using expert system and search algorithm. *International Shipbuilding Progress, 48*(3), 197–212.

Li, B., & Pang, F.-W. (2013). An approach of vessel collision risk assessment based on the D-S evidence theory. *Ocean Engineering, 74*, 16–21.

Li, H., Jiao, H., & Yang, Z. (2023). AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. *Transportation Research Part E: Logistics and Transportation Review, 175*, 103152.

Li, H., Lam, J. S. L., Yang, Z., Liu, J., Liu, R. W., Liang, M., & Li, Y. (2022). Unsupervised hierarchical methodology of maritime traffic pattern extraction for knowledge discovery. *Transportation Research Part C: Emerging Technologies, 143*, Article 103856.

Li, H., Liu, J., Yang, Z., Liu, R. W., Wu, K., & Wan, Y. (2020). Adaptively constrained dynamic time warping for time series classification and clustering. *Information Sciences, 534*, 97–116.

Li, H., & Yang, Z. (2023). Incorporation of AIS data-based machine learning into unsupervised route planning for maritime autonomous surface ships. *Transportation Research Part E: Logistics and Transportation Review, 176*, 103171.

Liang, M., Liu, R. W., Zhan, Y., Li, H., Zhu, F., & Wang, F.-Y. (2022). Fine-grained vessel traffic flow prediction with a spatio-temporal multigraph convolutional network. *IEEE Transactions on Intelligent Transportation Systems*.

Liu, J., Feng, Z., Li, Z., Wang, M., & Wen, L. R. (2016). Dynamic ship domain models for capacity analysis of restricted water channels. *Journal of Navigation, 69*(3), 481–503.

Liu, J., Zhang, J., Yan, X., & Soares, C. G. (2022). Multi-ship collision avoidance decision-making and coordination mechanism in Mixed Navigation Scenarios. *Ocean Engineering, 257*, Article 111666.

Liu, K., Yuan, Z., Xin, X., Zhang, J., & Wang, W. (2021). Conflict detection method based on dynamic ship domain model for visualization of collision risk Hot-Spots. *Ocean Engineering, 242*, Article 110143.

Liu, R. W., Liang, M., Nie, J., Yuan, Y., Xiong, Z., Yu, H., & Guizani, N. (2022). STMGCN: Mobile edge computing-empowered vessel trajectory prediction using spatio-temporal multi-graph convolutional network. *IEEE Transactions on Industrial Informatics*.

Liu, Z., Wu, Z., & Zheng, Z. (2019). A cooperative game approach for assessing the collision risk in multi-vessel encountering. *Ocean Engineering, 187*, Article 106175.

Liu, Z., Wu, Z., & Zheng, Z. (2019). A novel framework for regional collision risk identification based on AIS data. *Applied Ocean Research, 89*, 261–272.

Mitici, M., & Blom, H. A. P. (2018). Mathematical models for air traffic conflict and collision probability estimation. *IEEE Transactions on Intelligent Transportation Systems, 20*(3), 1052–1068.

Ng, A., Jordan, M., & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems, 14*.

Öztürk, Ü., Boz, H. A., & Balcisoy, S. (2021). Visual analytic based ship collision probability modeling for ship navigation safety. *Expert Systems with Applications, 175*, Article 114755.

Prandini, M., Hu, J., Lygeros, J., & Sastry, S. (2000). A probabilistic approach to aircraft conflict detection. *IEEE Transactions on Intelligent Transportation Systems, 1*(4), 199–220.

Ristic, B. (2014). Detecting anomalies from a multitarget tracking output. *IEEE Transactions on Aerospace and Electronic Systems, 50*(1), 798–803.

Rong, H., Teixeira, A. P., & Soares, C. G. (2019). Ship trajectory uncertainty prediction based on a Gaussian Process model. *Ocean Engineering, 182*, 499–511.

Rong, H., Teixeira, A. P., & Soares, C. G. (2021). Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics. *Reliability Engineering & System Safety, 209*, Article 107463.

Rong, H., Teixeira, A. P., & Soares, C. G. (2022). Maritime traffic probabilistic prediction based on ship motion pattern extraction. *Reliability Engineering & System Safety, 217*, Article 108061.

Saeedmanesh, M., & Geroliminis, N. (2016). Clustering of heterogeneous networks with directional flows based on "Snake" similarities. *Transportation Research Part B: Methodological, 91*, 250–269.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 888–905.

Shi, K., & Weng, J. (2021). Impacts of the COVID-19 epidemic on merchant ship activity and pollution emissions in Shanghai port waters. *Science of The Total Environment, 790*, Article 148198.

Szlapczynski, R., & Szlapczynska, J. (2016). An analysis of domain-based ship collision risk parameters. *Ocean Engineering, 126*, 47–56.

Szlapczynski, R., & Szlapczynska, J. (2017). Review of ship safety domains: Models and applications. *Ocean Engineering, 145C*, 277–289.

Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.

Tavakoli, S., Najafi, S., Amini, E., & Dashtimansh, A. (2021). Ship acceleration motion under the action of a propulsion system: A combined empirical method for simulation and optimisation. *Journal of Marine Engineering & Technology, 20*(3), 200–215.

Vettor, R., & Soares, C. G. (2015). Detection and analysis of the main routes of voluntary observing ships in the North Atlantic. *The Journal of Navigation, 68*(2), 397–410.

Wang, H., Song, Z., & Wen, R. (2018). Modeling air traffic situation complexity with a dynamic weighted network approach. *Journal of Advanced Transportation, 2018*.

Wang, N. (2013). A novel analytical framework for dynamic quaternion ship domains. *Journal of Navigation, 66*(02), 265–281. https://doi.org/10.1017/S0373463312000483

Wang, Y., & Chin, H.-C. (2016). An empirically-calibrated ship domain as a safety criterion for navigation in confined waters. *Journal of Navigation, 69*(2), 257–276.

Weng, J., Meng, Q., & Qu, X. (2012). Vessel collision frequency estimation in the singapore strait. *Journal of Navigation, 65*(2), 207–221.

Weng, J., & Shan, X. (2015). Ship collision frequency estimation in port fairways: A case study. *Journal of Navigation, 68*(3), 602–618.

Xiao, Z., Fu, X., Zhang, L., & Goh, R. S. M. (2019). Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey. *IEEE Transactions on Intelligent Transportation Systems, 21*(5), 1796–1825.

Xin, X., Liu, K., Loughney, S., Wang, J., & Yang, Z. (2022). Maritime traffic clustering to capture high-risk multi-ship encounters in complex waters. *Reliability Engineering & System Safety*, Article 108936.

Xin, X., Liu, K., Yang, X., Yuan, Z., & Zhang, J. (2019). A simulation model for ship navigation in the "Xiazhimen" waterway based on statistical analysis of AIS data. *Ocean Engineering, 180*, 279–289.

Xin, X., Liu, K., Yang, Z., Zhang, J., & Wu, X. (2021). A probabilistic risk approach for the collision detection of multi-ships under spatiotemporal movement uncertainty. *Reliability Engineering & System Safety*, Article 107772.

Xin, X., Yang, Z., Liu, K., Zhang, J., & Wu, X. (2022). Multi-stage and multi-topology analysis of ship traffic complexity for probabilistic collision detection. *Expert Systems with Applications*, Article 118890.

Yang, D., Wu, L., Wang, S., Jia, H., & Li, K. X. (2019). How big data enriches maritime research–a critical review of automatic identification system (AIS) data applications. *Transport Reviews, 39*(6), 755–773.

Yu, H., Fang, Z., Murray, A. T., & Peng, G. (2019). A direction-constrained space-time prism-based approach for quantifying possible multi-ship collision risks. *IEEE Transactions on Intelligent Transportation Systems, 22*, 131–141.

Yu, H., Murray, A. T., Fang, Z., Liu, J., Peng, G., Solgi, M., & Zhang, W. (2021). Ship path optimization that accounts for geographical traffic characteristics to increase maritime port safety. *IEEE Transactions on Intelligent Transportation Systems, 23*, 5765–5776.

Yu, Q., Liu, K., Chang, C.-H., & Yang, Z. (2020). Realising advanced risk assessment of vessel traffic flows near offshore wind farms. *Reliability Engineering & System Safety, 203*, Article 107086.

Yu, Q., Liu, K., Teixeira, A. P., & Soares, C. G. (2020). Assessment of the influence of offshore wind farms on ship traffic flow based on AIS data. *The Journal of Navigation, 73*(1), 131–148.

Zass, R., & Shashua, A. (2005). A unifying approach to hard and probabilistic clustering. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 1*, 294–301.

Zhang, J., Liu, J., Hirdaris, S., Zhang, M., & Tian, W. (2022). An interpretable knowledge-based decision support method for ship collision avoidance using AIS data. *Reliability Engineering & System Safety*, Article 108919.

Zhang, L., Meng, Q., & Fwa, T. F. (2019). Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transportation Research Part E: Logistics and Transportation Review, 129*, 287–304.

Zhang, M., Montewka, J., Manderbacka, T., Kujala, P., & Hirdaris, S. (2021). A big data analytics method for the evaluation of ship-ship collision risk reflecting hydrometeorological conditions. *Reliability Engineering & System Safety, 213*, Article 107674.

Zhang, M., Taimuri, G., Zhang, J., & Hirdaris, S. (2023). A deep learning method for the prediction of 6-DoF ship motions in real conditions. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment*, 14750902231157852.

Zhang, M., Zhang, D., Fu, S., Kujala, P., & Hirdaris, S. (2022). A predictive analytics method for maritime traffic flow complexity estimation in inland waterways. *Reliability Engineering & System Safety*, Article 108317.

Zhang, W., Goerlandt, F., Kujala, P., & Wang, Y. (2016). An advanced method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering, 124*, 141–156.

Zhang, W., Goerlandt, F., Montewka, J., & Kujala, P. (2015). A method for detecting possible near miss ship collisions from AIS data. *Ocean Engineering, 107*, 60–69.

Zhao, L., Shi, G., & Yang, J. (2018). Ship trajectories pre-processing based on AIS data. *The Journal of Navigation, 71*(5), 1210–1230.

Zhao, Y., Li, W., & Shi, P. (2016). A real-time collision avoidance learning system for unmanned surface vessels. *Neurocomputing, 182*, 255–266.

Zhen, R., Riveiro, M., & Jin, Y. (2017). A novel analytic framework of real-time multi-vessel collision risk assessment for maritime traffic surveillance. *Ocean Engineering, 145*, 492–501.

Zhen, R., Shi, Z., Shao, Z., & Liu, J. (2021). A novel regional collision risk assessment method considering aggregation density under multi-ship encounter situations. *The Journal of Navigation*, 1–19.

Zhou, Y., Daamen, W., Vellinga, T., & Hoogendoorn, S. (2019). Review of maritime traffic models from vessel behavior modeling perspective. *Transportation Research Part C: Emerging Technologies, 105*, 323–345.