Geoinformatica manuscript No.
(will be inserted by the editor)

Understanding Evolution of Maritime Networks from Automatic Identification System Data

Emanuele Carlini · Vinicius Monteiro de Lira · Amilcar Soares · Mohammad Etemad · Bruno Brandoli · Stan Matwin

Received: date / Accepted: date

Abstract Recent studies on maritime traffic model the interplay between vessels and ports as a graph, which is often built using automatic identification system (AIS) data. However, only a few works explicitly study the evolution of such graphs and, when they do, generally consider coarse-grained time intervals. Our goal is to fill this gap by providing a conceptual framework for the fine-grained systematic study of maritime graphs evolution. To this end, this paper presents the month-by-month analysis of world-wide graphs built using a 3-years AIS dataset. The analysis focuses on the evolution of several topological graph features, as well as their stationarity and statistical correlation. Results have revealed some interesting seasonal and trending patterns that can provide insights in the world-wide maritime context and be used as building blocks toward the prediction of graphs topology.

Keywords Automatic Identification System · Graph Analysis · Time-series

1 Introduction

Maritime transportation represents 90% of international trade volume and plays a paramount role in today's economy, in terms of cargo shipping, passen-

Emanuele Carlini, Vinicius Monteiro de Lira
Institute of Information Science and Technologies
National Research Council (CNR)
Pisa, Italy E-mail:{emanuele.carlini, vinicius.monteirodelira}@isti.cnr.it

Amilcar Soares
Department of Computer Science, Memorial University of Newfoundland
St. John's, Canada
E-mail:amilcarsj@mun.ca

Mohammad Etemad, Bruno Brandoli, Stan Matwin
Institute for Big Data Analytics, Dalhousie University
Halifax, Canada
E-mail:{etemad,brunobrandoli}@dal.ca, stan@cs.dal.ca

ger transportation, leisure navigation, and fishing operation [36]. Globalization and multiple modal transportation of goods in the shipping industry resulted in a massive extension of the maritime vessel route network. The study of vessel movements is a well-established source of information to understand the role of maritime routes and ports in economic, social, and environmental contexts. These studies include maritime traffic control and prediction [30], human migration flows [16], bioinvasion [20] and maritime piracy [35]. However, such a role cannot be adequately unraveled by looking at ports and routes in isolation; instead, they must be put in relation to one another. This allows the study of the interplay of all the components in the complex maritime network, and it is even more important for understanding the evolution over time of those interactions.

A central concept for the analytical study of vessel routes is the Global Shipping Network (GSN), in which nodes are ports and edges are the routes between ports of cargo ships (Figure 1). Since the automatic identification system (AIS) for vessels was made mandatory in 2004 [13], there has been a surge of studies on the GSN and other maritime networks that use such data. Many works have studied GSN-like network according to graph theory [37,34, 33], but only a few of them analyzed the network in terms of its evolution over the years [26,28]. Also, those works which studied the network evolution used private data and performed exciting but high-level and coarse-grained analysis, such as in [12].

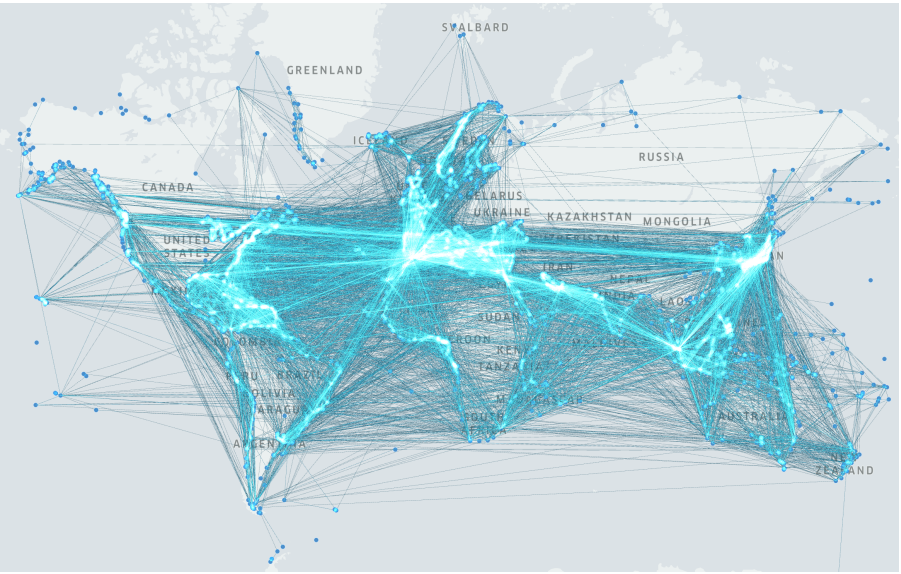


Fig. 1: World wide cargo routes in 2019 extrapolated from the dataset used in the paper. The nodes represent ports and the edges are voyages between two ports

The main goal of the analysis in this paper is to provide a systematic study of the evolutionary aspects of GSN-like networks, with the purpose of identifying recurrent patterns in their evolution. The analysis is based on a dataset provided by an agreement with ExactEarth [15] and with a defined and well-documented data model. This aspect is fundamental for the reproducibility of our analysis and its expansion and updating of results when new data arrives. Also, it considers the two necessary dimensions of *time* and *layers* (i.e., the evolution of the network can be observed for multiple types of vessels, such as cargo and passengers).

Such an ambitious objective has some inherent challenges that must be tackled. First, the size of AIS datasets is usually large. For example, ExactEarth alone claims to consistently track 165,000 vessels and over 7,000,000 AIS messages daily[1]. Analyzing such data over a long period of time typically requires large storage spaces and high processing capabilities. Second, the purpose of any network analysis is to abstract the complexity of a system in order to extract meaningful information that is not directly available when the individual components are examined separately. Therefore, the definition of a network that encompasses time information is a complex task. Suitable approaches need to be carefully selected to study the evolving network.

The contributions of this work can be outlined as the following:

- We propose an approach that uses AIS data to extract connections between ports derived from the vessels' movements. From these connections (or *voyages*), we build GSN-like networks in which the vertices correspond to the ports, whereas the edges or links correspond to the vessel voyage between two different ports. In addition, each edge has a semantic defined by the vessel types.
- We applied the aforementioned approach to a dataset containing 3 years of world-wide AIS data provided by ExactEarth [15];
- We study several topological properties of the temporal graphs generated from vessels' movements and how these features evolve over time. Specifically, we investigate features relative to graphs dimension, ability to form clusters, and geographical spatiality.
- We investigate the aspect of stationarity of the time series of the topological properties of the vessels' voyage networks over the time and discuss the obtained insights.
- We perform a correlation analysis of the topological properties extracted from the graphs generated by the different vessel types (e.g., cargo, passenger, fishing, tanker) routes. This study allows the identification of graph-based properties that are correlated among the different vessel type routes.

In our previous work [5], we analyzed an open source AIS dataset provided by MarineCadastre.gov and we focused on presenting the graph modeling and building process. This work extends the analysis by using a much larger database containing world-wide AIS messages and provides a deeper analysis

[1] https://www.exactearth.com/products/exactais

of the graph topological features, including their potential stationarity and correlation. The rest of the paper is organized as follows. Section 2 discusses related works. Section 3 defines some concepts used through the paper and describes our approach for deriving time-series of topological properties from graphs based on vessels' visits to ports.

We perform the analysis of the graphs time-series in Section 4 and their stationarity and correlation in Section 5. Section 6 draws conclusions and discusses future works.

2 Related Works

The work done in [22] is one of the first to study the concept of GSN as a complex network. They use AIS information about the itineraries of 16363 ships of three types (bulk dry carriers, container ships, and oil tankers) during 2007 to build a network of links between ports. The work of [22] shows that the three categories of ships differ in their mobility patterns and networks. Their results show that container ships follow regularly repeating paths, whereas bulk dry carriers and oil tankers move less predictably between ports. They also show that the network of all ship movements possesses a heavy-tailed distribution for the connectivity of ports and the loads transported on the links with systematic differences between ship types [22].

The work of [26] also uses a sample of the Lloyds database with the world container ship fleet movements from Chinese ports from the years of 2008 to 2010. Their work aims to look at changes in the maritime network before and after the financial crisis (2008-2010) and analyze the extent to which large ports have seen their position within the network change. The authors show how the global and local importance of a port can be measured using graph theory concepts. They also show that the goods transportation network was contracted concerning port throughput, but no contraction in the main hub ports' distribution capacity was found [26]. Finally, the authors show that there are new port regions in the entrance and exit of the Panama Canal, and there are several significant business opportunities in that region.

A study of topological changes in the maritime trade network is shown in [25]. The authors propose two new measures of network navigability called *random walk discovery* and *escape difficulty*. Their results show that the maritime network evolves by increasing its navigability while doubling the number of active ports. The authors suggest that unlike in other real-world evolving networks studied in the literature up to date, the maritime network does not densify over time, and its effective diameter remains constant [25].

In [12], the author investigates the degree of overlap among the different layers of circulation composing global maritime flows. His work uses several methods from complex network analysis to understand the dynamics affecting the evolution of ports and shipping. The results show a strong and path-dependent influence of multiplexity on traffic volume, range of interaction, and centrality from various perspectives (e.g., matrices correlations, homophily,

assortativity, and single linkage analysis) [12]. When growing the network and concentrating the analysis around large hubs over time, results show that the traffic distribution is place-dependent due to the reinforced position of already established nodes [12].

The work of [37] builds a GSN using the 2015 AIS data of the world with multiple spatial levels. Their process mainly consists of five steps, where the first three generate the network nodes, and the last two create the network links. The work of [37] applies the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to detect where ships stop and cross this information with terminal candidates of ports. A directed GSN is generated with the trip statistics between two nodes as the edges. Their work evaluates features such as average degree and betweenness centrality of each node, average shortest path length between any two nodes, and community clusters of the GSNs.

Following a similar idea of building GSNs, but with focus on anomaly detection, the work of [34] provides a mechanism that classifies vessel behavior in normal and abnormal, using historical information about similar vessels that operate in a particular area. In [34], the authors identify waypoints (i.e., a region of interest for a given application) that characterizes the operations and the sort of movement patterns that they follow (i.e., the nodes). As edges, the work of [34] uses the subtrajectories that link two waypoints, using the extracted features of those subtrajectories for analysis. They identified each edge by the subtrajectory that links two ways points. Features of each edge are generated using a trajectory mining library introduced by [14]. Their analysis tries to detect outliers from the subtrajectory features (e.g., course over ground, speed over ground, etc.) and using transition probabilities as the edges of the network. In a similar way, [10] presented an approach to learn automatically and represent compactly commercial maritime traffic in form of a graph, whose nodes represent clusters of waypoints, which are connected together by a network of navigational edges. The main objective of the work in [10] is representing the traffic motion using graphs and evaluating how graphs could be utilised for motion prediction.

In the work [39], the authors present a novel approach to extract maritime routes from AIS data automatically. Their method simplifies single AIS trajectory data using the Douglas-Peucker algorithm to compress redundant information and find graph nodes where vessels perform relevant direction changes. In a further step, their algorithm determines the connectivity of nodes using the vessel trajectories, linking nodes to form a coherent chain modeled as a directed graph. Their results show a study case in the Qiongzhou Strait (China), where the raw and the simplified strategies are compared. The simplified version of the trajectories extracts the most relevant direction change actions and declutters the view of the traffic in the Qiongzhou Strait. The method presented in [38] also proposes a maritime route extraction method based on AIS data. Their method transforms the vessel's trajectory into a ship trip semantic object (STSO). The STSO is further integrated into the nodes and edges

of a directed maritime traffic graph to understand the shipping routes. Their evaluation analysis is restricted to a local region of the globe.

From a computational perspective, recent works propose to use big data and streaming analytics frameworks (such as Spark and Flink) to find routes [40], extract high-level representations and evaluate local maritime traffic [17], and integrate AIS with other environmental data [31,32]. For example, [40] proposes a novel algorithm named ROTA that uses historical AIS positional data and port geometries to obtain maritime "patterns of life" at a global scale. In [17], the authors propose a parallelized method for the automatic reconstruction of a network reflecting the maritime traffic using AIS data that can be used in vessel routing and voyage planning. The framework named SPARTAN is described in [31] with the objective of performing real-time semantic integration of big mobility data with other data sources. SPARTAN's main goal is to provide enriched trajectories in the RDF (Resource Description Framework) format that can be exploited by higher-level analysis tasks, such as link discovery between the data sources. In a similar way, the platform called CRISIS shows an agile data architecture for real-time data representation, integration, and querying situations over heterogeneous data streams using RDF. Its goal is to improve knowledge interoperability and they apply the framework to the maritime ship traffic domain to discover real-time traffic alerts by querying and reasoning across multiple streams. Differently from SPARTAN, CRISIS does not use parallel processing frameworks to query RDF data. These approaches yield huge advancement in terms of the processed amount of data and performance of on-line and streaming with respect to traditional data management techniques. However, the scope of this paper is to statically analyse trends and patterns in a series of snapshots generated from historical data. Therefore, we have used in-memory computation techniques to simplify the implementation of the processing pipeline and focus on the analysis of the results.

In summary, the related works can be categorized into four aspects regarding the use of graph theory for the analysis of vessel movement patterns. The first aspect considers the data source provider. All works found either use Lloyd's database or AIS data. The second aspect lists the focus of the paper. We also identified whether the works evaluated the graph evolution over time. Finally, the data scope (local or global data) of the data used was listed. Table 1 summarizes how state of the art in the graph analysis with vessel data.

Differently from [22,26,25,12], our work use AIS data to determine vessel routes. Differently from [37], which uses stop points as nodes to evaluate centrality, shortest-path, and communities like, or using waypoints as nodes and being focused on anomaly detection, like [34], we use the ports as nodes, and we evaluate the evolution of the network as our primary task. However, our approach is different from what is done in [10] since their nodes represent clusters of waypoints and they do not focus in an analysis of the network over time. We also focus on evaluating edges generated by several trips obtained using AIS message, instead of creating graphs for single trips as in [39]. The works [40,17] do not focus in the analysis over time and we focus in global

Reference	Data source	Distributed	Focus	Analysis over time	Data scope
Kaluza et. al, 2010 [22]	Lloyd's	✗	Network analysis	✗	Global
González Laxe et al., 2012 [26]	Lloyd's	✗	Network analysis	✗	Local
Bartholdi et al., 2016 [2]	Not specified	✗	Connectivity index	✗	Global
Kosowska-Stamirowska et al., 2016 [25]	Lloyd's	✗	Network analysis	✓	Global
Ducruet, 2017 [12]	Lloyd's	✗	Network analysis	✓	Global
Zhang et al., 2018 [39]	AIS	✗	Route extraction	✗	Local
Coscia et al., 2018 [10]	AIS	✗	Network Analysis	✗	Local
Soares et al., 2019 [31]	AIS	✗	Data Integration	✗	Local
Varlamis et al., 2019 [34]	AIS	✗	Anomaly detection	✗	Local
Wang et al., 2019 [37]	AIS	✗	Network analysis	✗	Global
Zissis et al., 2020 [40]	AIS	✓	Data Integration and Link prediction	✗	Local
Santipantakis et al., 2020 [31]	AIS	✓	Data Integration	✗	Local
Filipiak et al., 2020 [17]	AIS	✓	Network Analysis	✗	Local
Yan et al., 2020 [38]	AIS	✗	Route extraction	✗	Local

Table 1: A summary of the related works regarding the four evaluated aspects.

instead of local analysis. The objectives of [31,32] is data integration using RDF, and in the case of [31], using distributed and parallel frameworks, while ours focus is analysing the evolution of networks of voyages of vessels between ports. We also do not focus on parallel and distributed processing, although our methods are parallelizable. To the best of our knowledge, this work is the first to use AIS and graph evolution analysis to evaluate worldwide vessel traffic information over time.

3 Definitions and Methodology

Vessels report their location through AIS messages while navigating. A vessel sends AIS messages with a frequency that varies from a few seconds to a few minutes, depending on the type of message, the vessel position, and the vessel activity. When they are underway, they may send AIS messages every 2 to 10 seconds, while when they are at anchor, the time window can increase to 3 minutes [37]. Positional information extracted from AIS messages can be interpreted as a representation of the spatial-temporal movement of a travelling vessel. We are interested in this spatial information with the intent of understating when a vessel is visiting a port.

Our methodology is depicted in Figure 2. We build the sequence of vessels' voyages by merging subsequent vessel visits to ports. From those sequences, we create multiple non-overlapping snapshot graphs (or networks), each considering a specific time window (e.g. one month). By extracting several topological features from each snapshot graph, we create a set of time-series to be able to study the evolution of the graphs using complex network concepts. Even if studying the snapshot graphs of the entire dataset has some interest, we

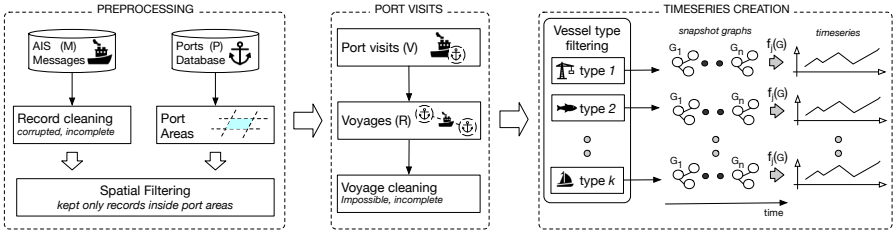


Fig. 2: Data analysis process: from raw data to graphs

have chosen to create multiple sets of snapshot graphs, one for each most represented type of vessel. In this way, we can highlight particular trends or patterns for specific vessel types that instead would be hidden if considering the entire dataset.

Graphs have some properties useful to unravel interesting information about the interaction and dynamism between two and more entities. In particular, in the context of a voyages graph, the topological properties of the graph can help us identify relevant characteristics within a network that would not emerge if the individual entities were examined separately [12, 25]. Topological properties can be applied to the network as a whole or to individual nodes and edges. In particular, for our study, we are interested in global network properties and their evolution.

Building the set of voyage graph snapshots directly from the original AIS data would be possible, but also very impractical. AIS data inevitably contains noise due to many reasons, including malfunctions, errors in transmission, and malicious use. In our context, such noise and mistakes translate into incorrect voyages, which requires a phase of data cleaning. Therefore, we applied an incremental approach to process the data which has the following advantages: (i) graph building is very fast, as the set of *voyages* is basically an edge list; (ii) the costly cleaning process is done only once, and from the clean collection of *voyages* it is possible to build multiple graphs; and (iii) it might be interesting to study *visits* and *voyages* without transforming them into a graph. The next sections explain more in detail our methodology.

3.1 AIS data pre-processing

At first we employ a pre-processing of the AIS messages with the aim of obtaining those records that have happened inside a port.

Definition 1 (*AIS Message*): An AIS Message m is a tuple (e, x, y, t, c) that represents the GPS coordinates (x, y) at a time stamp t assigned to a vessel e of type c. We define M_{ais} as the set of all original AIS messages.

In Definition 1, we consider the Maritime Mobile Service Identity (MMSI) as the vessel identifier e. However, AIS datasets usually have a lot of noise and

much information is redundant. Therefore, from the original set M_{ais} we create a new set M_{clean} from which we remove duplicates, incomplete and *incorrect* messages. Incorrect messages are those syntactically valid but with invalid semantics in relevant fields (typically position or vessel type). For example, several incorrect entries had a vessel identifier whose value is composed only by zeroes, which may indicate a placeholder for missing MMSI and thus prevent a correct identification of the vessel.

Definition 2 (*Port*): Given P as the set of all worldwide ports, a sea port $p \in P$ is represented as a tuple (id, x, y), where x and y are the latitude and longitude coordinates of its geographical center, and id is the code that identifies the port. We also define the spatial function $buffer(p, r)$ that defines a circular area of radius r centered on the coordinates of port p.

Depending on r, there could be overlapping port areas such that the same AIS record results transmitted inside multiple ports. In these cases, we discriminate by clustering the ports whose regions overlap and assign this cluster an unique port identifier.

Messages in M_{clean} are then spatial filtered with the clustered ports regions, in order to create a new set M_{port} that contains only those messages transmitted inside the port areas defined by $buffer(p, r), \forall p \in P$. Also, we add the indication of the port identifier to each message in M_{port}.

3.2 Vessel voyages

From M_{port} we then compute the set of visits V. We assume M_{port} to be sorted by time according to the time stamp of the AIS messages.

Definition 3 (*Visit*): Let Z be the set of all sub-sequences of consecutive messages $z_{e,p} = m_1 \dots m_k$ and $z_{e,p} \subset M_{port}$ such that each message $m \in z_{e,p}$ refers to the same vessel identifier e and port p. We then define a visit $v \in V$ for each $z_{e,p} \in Z$ as the tuple $(p, e, t_{start}, t_{end})$ where t_{start} and t_{end} are respectively the minimum and maximum timestamp for all messages in $z_{e,p}$.

From V we extract the set of voyages. The underlying assumption is that given a sorted set of visits, we record a voyage from p_o to p_t for a vessel e if the vessel is seen at the port p_o at time t_0 and at the port p_f at time t_1 (with $t_1 > t_0$) and there were no other visits to other ports in the meantime.

A possible limitation is that we register a visit when a vessel is passing through the buffer area of a port. Also, we do not put any limit on the time of a visit. However, the ports clustering, discussed above, mitigates these problems: if a vessel passes through several nearby ports, it counts as a single visit.

Definition 4 (*Voyage*): Given R as the set of all vessel voyages, a voyage $r \in R$ is a pair (v_1, v_2) of consecutive visits in V for the same vessel e.

The ports of v_1 and v_2 are called respectively origin and destination ports. The duration of a voyage is the time of the last visit of e in the origin port and the time of first visit in the destination port. We set the length of a voyage by computing the orthodromic distance (the minimum distance between two points on a sphere) in kilometres between the starting and arrival ports. We then removed those voyages whose "virtual" speed exceeds 60 knots (which is still very high speed, but we left some margin to cope with a possible degree of approximation in the data). Such invalid voyages are generated when the same MSSI is registered for different vessels, for instance due to errors in reception or sending of the signal. We did not remove the slow speed voyages as we cannot estimate (using the available data) how long would take the actual maritime route between two ports with respect to the orthodromic distance.

It is important to mention that we do not detect loops (i.e. when a vessel leaves and returns back to the same port). In fact, we consider a loop as a (possibly long) visit. We designed the procedure in this way because the time and computational power needed to detect loops is high and loops detection would not have brought any relevant insight for our studies. Indeed, loops are not interesting for our analysis as we focus on the interplay between different ports and the global movements of vessels around a large geographical area, whereas loops are relevant for studying local behaviours. Further, as a matter of fact, detecting loops would mean that also the area outside ports must be counted as a 'port', actually forcing the processing of all AIS messages. Such a huge task would have increased the processing time dramatically and made our entire analysis not feasible.

3.3 Time-series creation

From the clean set of voyages we build a graph by considering ports as nodes and the voyages as edges.

Definition 5 (*Voyage Graph*): A voyage graph is a graph $VG = (N, L)$ built according to a set of voyages $R' \subset R$, in which N contains all the ports in R' and L contains a single directed edge for each unique pair of ports in R'. With each edge $l \in L$ is associated a positive number w defined as the count of all unique voyages between the ports of l.

The resulting graph is then a directed graph built by essentially collapsing a multi-graph into a directed weighted graph. By using the above definitions it is possible to create different graphs by tuning the content of R'. In this paper, we create set of graphs for consecutive non-overlapping time windows and filtered by vessel types.

A time-series is a collection of observations made sequentially over the time [7]. In this paper we are interested in building time-series of topological features of voyage graphs created with the methodology above. The idea is to define a set of consecutive, non-overlapping time intervals to create the list of

graphs on which to compute topological features, which in turn represents our time-series.

Definition 6 (*Voyage Graph Features Time-series*) (VGT)

Let J be a set of topological features that can be applied on a directed graph, and f_j be the function that compute the feature $j \in J$. We discretize the time into n equal disjoint intervals and create one graph g for each of such intervals, such to have a set of snapshot graphs $S = \{g_1 \dots g_n\}$. Then, $T_j = \{ f_j(g) \mid g \in S \}$ is the set of timeseries corresponding to the given feature j. Finally, the set of the voyage graph features time-series is defined as $VGT = \{ T_j \mid j \in J \}$.

3.4 Dataset

To build the graph we have used three years (2017-2019) of worldwide AIS data provided by ExactEarth [15]. The full dataset contains around 2.5 Terabytes of AIS messages and around 20 billions of records stored in a relational database. Vessels visits (Definition 3) have been extracted from the database, by running a spatial query. We used Python to develop several in-memory scripts to compute graphs topological features. Graphs are stored in memory as edge-lists. Since our focus is not on the performance of the processing, we have not performed a formal analysis for robustness or scalability. The VGT have been build with a time interval of a solar month, resulting in a total of 36 VGT for each considered vessel type and each topological feature. To model the area of the ports we have used the World Port Index dataset [29] that contains spatial information, including latitude and longitude, of all known seaports in the world. The radius of the buffer function in Definition 2 has been set to be 3 nautical miles (around 5 km). This value was largely used to define country's territorial waters limit [3]. However, regardless of legal implications, our objective here is to define a reasonable area that could approximate the real visits of vessels to a given port.

4 World-wide VGT Analysis

This section discusses our empirical analysis of the computed VGT. In the first subsection, we provide an analysis of the amount of data for each vessel type (i.e. *layers*) and its distribution over time. The output of this analysis serves to limit the amount of vessels type considered in the remainder of the manuscript. The second subsection discusses the size and the geographical extension of the VGT. The third and final subsection studies the behaviour of several topological features of VGTs to assess their connectivity properties.

vessel type	unique vessels count	unique voyages count
cargo	37.03% (28.7K)	6.6% (1.87M)
tanker	15.76% (12.2K)	3.36% (951K)
special	9.83% (7.63K)	13.2% (3.74M)
tug tow	9.92% (7.7K)	13.65% (3.86M)
other	6.26% (4.86K)	12.1% (3.42M)
fishing	6.0% (4.66K)	47.93% (13.6M)
passenger	3.84% (2.98K)	3.16% (893K)
sailing	3.74% (2.9K)	0.32% (89.2K)
military	0.99% (765)	0.18% (51.1K)
high-speed	0.83% (641)	0.68% (192K)
dredger	0.65% (504)	0.05% (13.2K)
wing-in-ground	0.5% (389)	0.32% (91.8K)
Total	100% (77.6K)	100% (28.3M)

Table 2: Percentage of unique vessels and voyages by vessel type in all the dataset. Sorted by unique vessels value

4.1 Network layers

Diverse types of vessels transmit AIS data, and it is natural to assume that the network of distinct types (*layers*) of vessels would be different. To identify the vessel type, we used the *type* field of the AIS data, and their associated description has been taken from the marinetraffic.com website[2] (with minor modifications). The statistics on the amount of data available for each vessel type can be found in Table 2.

Interestingly, for cargo and tanker, a high percentage of unique vessels corresponds to a lower percentage of total routes, while fishing vessels are the opposite. Passenger vessels can be observed to have roughly the same percentage of unique vessels and voyages. Such trends can be explained by the fact that cargo and tanker vessels perform fewer but longer voyages concerning fishing and passengers, mostly moving from and to nearby ports.

For the remainder of this study, we have considered only those vessel types (*layers*) having a relevant amount of unique vessels and voyages count, namely: (i) Cargo; (ii) Tanker; (iii) Passenger; and (iv) Fishing. We did not consider the *special* or the *other* types as they contain many different types of vessels; similarly, we did not consider the *tug tow* type as they usually perform very short voyages between nearby ports, and therefore are not interesting in a global world-wide analysis.

4.2 Network Dimensions

Networks order (number of nodes) and size (number of edges) are stable over time for both LRVs and SRVs (Figure 4 and 5). It is interesting to notice how the size and order of SRV networks is comparably smaller than LRV networks, although they account for almost half of the total unique trips (Table 2). Still,

[2] `https://help.marinetraffic.com/`

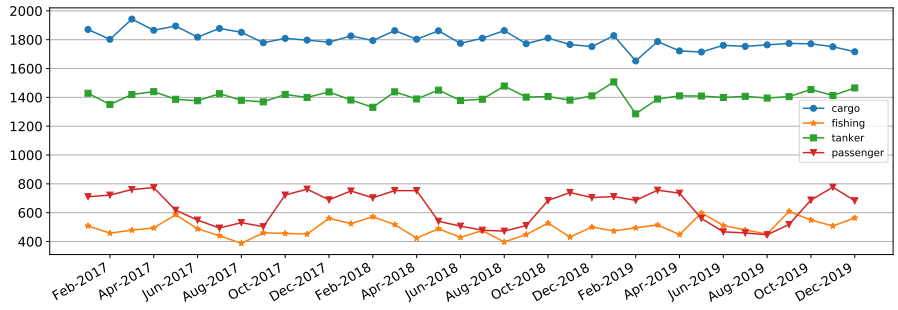


Fig. 3: Average orthodromic distance in kilometers between nodes connected by an edge

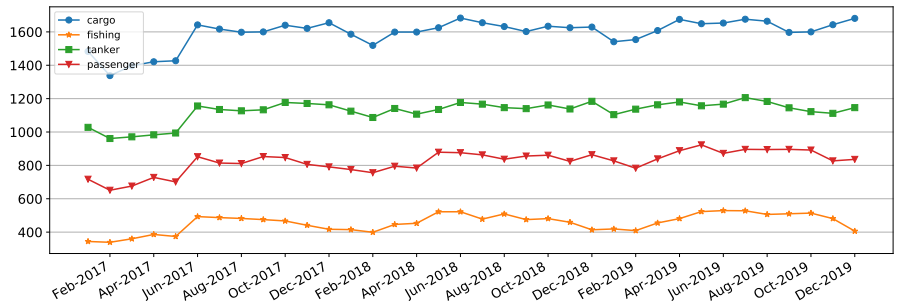


Fig. 4: Order (number of nodes)

this confirms the fact that SRVs tend to perform shorter routes but with a higher frequency.

Another essential metric for the maritime networks is the extension of their geographical spatiality, as the distance between ports can be linked with various cost aspects such as fuel consumption, maintenance rates, and insurance costs [12]. In addition, such a metric can give insights on whether certain vessel types are more oriented toward short or long routes. The average orthodromic distance between all the edges of the graph (Figure 3), as similarly observed in [12], confirms the above hypothesis: cargo and tanker perform longer voyages with respect to passenger and fishing vessels. Following these considerations, in the context of this section, we refer to cargo and tanker vessels as long-range vessels (LRV) due to their high average distances that variate few over time. In contrast, we refer to fishing and passenger vessels as short-range vessels (SRV) due to their low average distances that also show some variability. Such variability is quite evident in passenger vessels, for which in the summer months, we notice a neat decrease of the average distance.

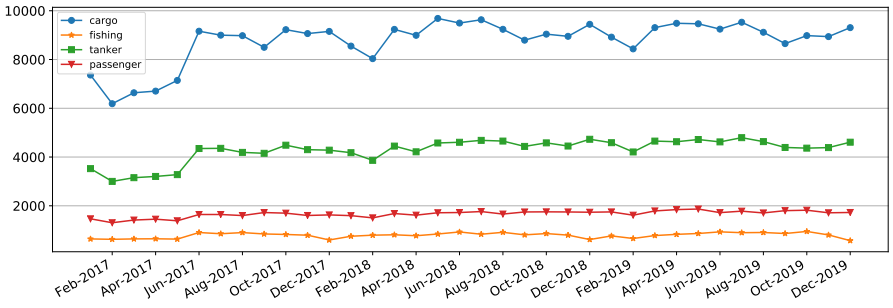


Fig. 5: Size (number of edges)

4.3 Networks Connectivity

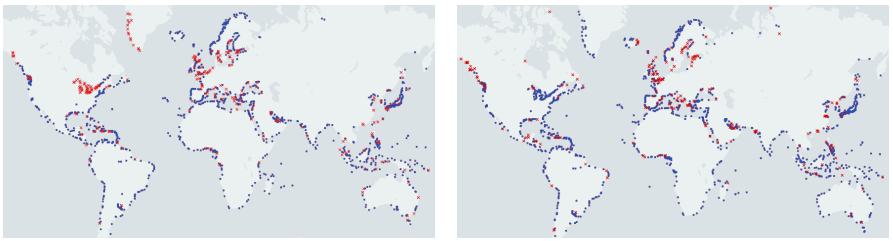
Connectivity properties of a network are commonly used to evaluate a network's resilience when removing nodes or edges. In terms of vessel networks, analyzing the connectivity properties and their evolution can help characterise and compare networks composed of different vessel types and can serve as a baseline for the accurate modeling of these networks. Our analysis focuses on the comparison of different network topologies values (rather than their study in isolation) in terms of different vessel types, by using complex network tools.

A relevant aspect in identifying cohesive subgroups of ports is the identification of those ports that share a strong tie in the traffic for a particular vessel type. The number of Strongly Connected Components (SCCs) is the number of subgraphs in which any node is reachable by any other nodes, and which is not connected to another subgraph [21]. Ideally, the number of SCC indicates how much the graph represents a global scale activity (low SCC number), rather than composed by a set of not connected and local activities (high SCC number).

Surprisingly the average number of SCCs for the SRV and LRV networks in the 3-years period is not so different (cargo: 187; fishing: 178; tanker: 168; passenger: 161). However, LRV networks are composed of a giant SCC that accounts for most of the nodes (>80%) on average over time, accompanied by many small components often composed by just two nodes (see Figure 7). As expected, nodes are more evenly distributed among the SCCs for SRV networks, in which the largest connected components account for just around 30% of the nodes on average.

From a geographic perspective, the LRV giant component spans worldwide. Those ports that remain out of the giant component show a seasonal trend with a clear difference from winter and summer periods (see Figure 6).

The number of bidirectional edges (i.e. given the nodes u and v, there exist both the edges $[u, v]$ and $[v, u]$) can be used as an indication about network connectivity. A large fraction of bidirectional edges in a vessel network means tight interactions between ports, indicating vessels inter-exchange from most ports pairs. In LRV networks we notice a lower fraction of bidirectional edges,



(a) January 2018 (b) July 2018

Fig. 6: Ports in the giant connected components (blue circles) vs ports outside it (red crosses). During winter periods (left) several north-most areas are cut out from the giant component, such as in the Greenland or the Great Lakes of North America, whereas they are present during summer(right)

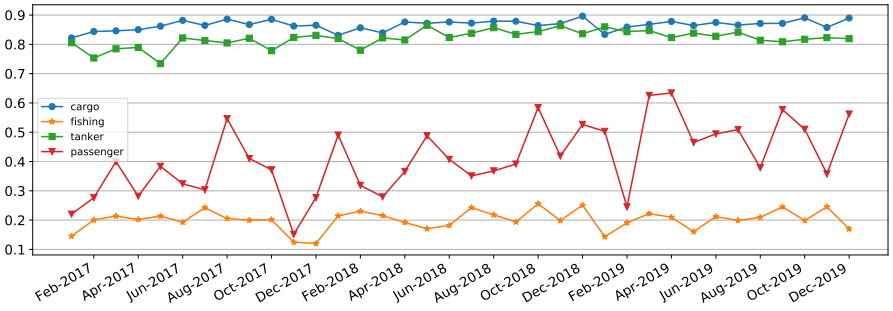


Fig. 7: Fraction of nodes in the largest strongly connected component

with around 70% of the ports connected only in one direction. By comparison, the SRV networks have a large fraction and are more variable (around 40% on average, see Figure 8). It is also interesting to notice how the values for the passenger networks form valleys during springs and autumns, while it peaks during summers and winters, indicating a seasonal change in the traffic patterns. LRV networks show a low fraction of bidirectional edges but a giant connected component: this suggests that LRVs are likely returning to the same set of ports but not directly, i.e., visiting other ports beforehand. This suggests that LRV traffic is mostly composed of unidirectional routes organised in 'circular' patterns. These findings correspond with the results obtained by similar research works [22]. By comparison, in SRV networks we observe many SSCs with an even distribution of vessel, and a higher symmetry, suggesting clusters of small local networks of predefined routes that are not connected to each other.

The average shortest path in a graph is the minimum number of edges to traverse from a node origin to a node destination averaged on all pairs of nodes. The average shortest path is generally to measure the density and robustness of networks. In the vessel network, a lower average shortest path

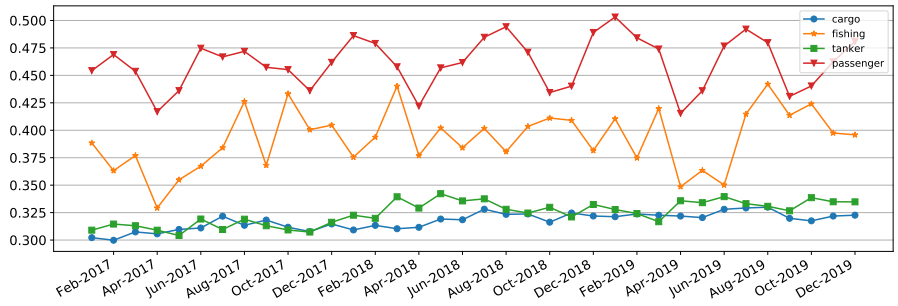


Fig. 8: Fraction of bidirectional edges

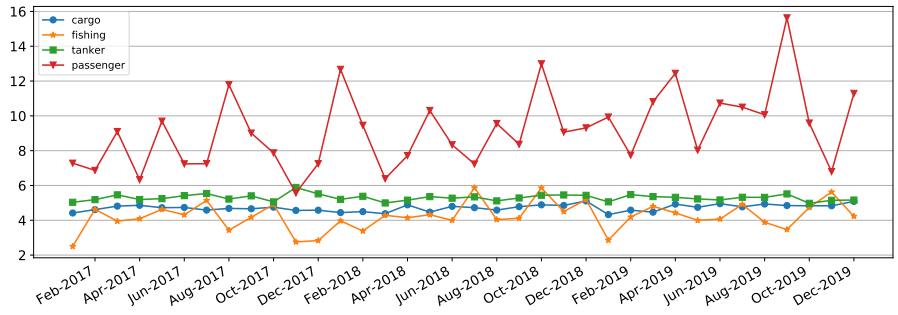


Fig. 9: Average shortest path. As a matter of comparison, for similar size random networks, we measured the following average shortest path: 4 for cargo, 2.5 for fishing, 3.7 for tanker, and 3.2 for passenger.

reveals more dense port connections. The average shortest path (computed on the giant connected component) of LRV networks is around 5 for tankers and 4 for cargo and is stable over time (see Figure 9). For SRV networks, the average shortest path is low (around 3) for fishing vessels, and relatively high and variable for passenger vessels, indicating a low-density graph affected by seasonal trends. However, the largest component in fishing networks is generally small compared to the number of nodes, so that such low values can be a direct consequence of that.

The Average Clustering coefficient, is the average of local clustering coefficients of all nodes. The local clustering of each node in the graph is the fraction of triangles (set of 3 vertices such that any two of them are connected by an edge) that exist over all possible triangles in its neighborhood [21]. In other words, this coefficient represents the tendency that two neighbours of a port are neighbours themselves, and can serve to evaluate how many voyages happen around the same set of ports. The results of the average clustering coefficient of the network observed in Figure 10 show that cargo, tanker and passenger networks create networks of higher density with respect to fishing networks. The clustering coefficient variability is high for all the type of ves-

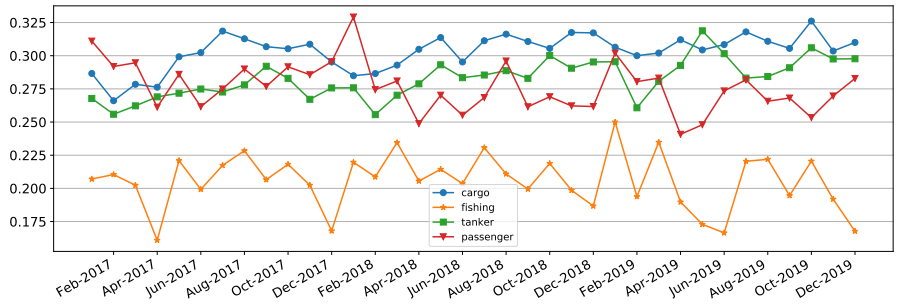


Fig. 10: Average clustering. As a matter of comparison, for similar size random networks, we measured the following average shortest path: 0.01 for cargo, 0.05 for fishing, 0.02 for tanker, and 0.02 for passenger.

sels, but larger for SRV networks, and there is no noticeable pattern. Such variability indicates that most of the connections are indeed volatile and their existence can depend on specific local factors.

5 VGT Stationarity and Correlation

In general, time series analysis accounts for the fact that data points taken over time may have an internal structure, such as auto-correlation, trend or seasonal variation. In this section, we first study the VGT values focusing on assessing the presence of stationary or seasonal patterns. Afterwards, we go further in analyzing relationships between different VGTs exploring their correlations.

Hereby we describe the experiments to observe potential stationary behaviour and correlation of VGT values. We perform such analysis for the types of vessels considered in Section 4, and for the following topological features: (i) *order*: number of nodes; (ii) *size*: number of edges; (iii) *avg. clustering*: the average clustering coefficient; (iv) *avg. degree*: the average node degree; (v) *avg. shortest path*: average shortest path between all pairs of nodes (vi) *avg. distance*: the average orthodromic distance between all pair of nodes considering the center of the associated port. (vii) *symmetry*: the percentage of symmetric links in the graph; (viii) *#cc*: the number of strongly connected components; (iv) *size largest cc*: the number of nodes in the larges strongly connected component.

The experiments conducted aim to answer the following research questions comprehensively:

RQ1. Are the VGT stationary? A time series is said to be stationary if its statistical properties do not change over time. In other words, it has constant mean and variance, and covariance is independent of time. In this research question, we focus on the analysis of the VGT values' in what concerns the presence of characteristics such as stationarity, trending, or seasonality in

the time-series values. To address this research question, we use a statistical test designed to comment on whether a time series is stationary explicitly. Section 5.1 addresses this research question.

RQ2. Within the different spectrum of the vessel's type, are the VGT values correlated? The idea is to perform correlation analysis to verify when two series of the same VGT feature are correlated or inversely correlated. We use a statistical test to assess the relationship of the two different VGTs. This research question is discussed in Section 5.2.

5.1 RQ1: Stationary behavior analysis

In this section, we investigate the (non) stationary behavior of the VGTs. Stationarity is an important concept in the field of time series analysis with tremendous influence on how the data is perceived and predicted [27]. From a visual perspective, time series that do not show trends or seasonality can be considered stationary. A common assumption when forecasting or predicting the future values in time-series is that each point is independent [18]. The best indication of this is when the values of past instances are stationary. This means that there is no seasonal or trending behaviour observed in the data. In other words, a time series is stationary if they do not have a trend or seasonal effects. This means that the statistics calculated on the time series, such as the mean, variance, and auto-correlation of the observations are consistent over time [4]. Most statistical forecasting methods are based on the assumption that the time series can be modeled approximately stationary through the use of mathematical transformations [24]. Stationary time series are easier to model since they represent a broader family of existing models of reality. Our objective is to study which VGTs represent a stationary process or show any trending or seasonal behaviour.

Different methods can be used to verify whether a time series is stationary or not. In particular, statistical tests can be used to analyze if the requirements of stationary are met or have been violated. Here, we adopted the Augmented Dickey-Fuller Test [11,6] (ADF) which is widely used in literature to assess the stationary property of time-series. The ADF test, also known as the "unit root test", uses an autoregressive model and optimizes an information criterion across multiple different lag values [8]. The augmented Dickey–Fuller (ADF) statistic, used in the test, is a negative number. The more negative it is, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence. A unit root is a stochastic trend in a time series. Therefore, if a time series has a unit root, it shows a non stationary pattern, having an unpredictable behaviour.

In the ADF Test, the null hypothesis assumes that the time series has an unit root (i.e. then not stationary), therefore showing some degree of time dependency. The alternative hypothesis (rejecting the null hypothesis) is that the time series is stationary, without time dependency. ADF statistic values lower than a critical value (such as 1% or 5%) suggest rejecting the null hy-

	ADF	**crit. 5%**	**p-value**	**ADF**	**crit. 5%**	**p-value**
	Cargo			**Tanker**		
avg. clustering	-2.880	-2.949	0.048	-2.886	-2.949	0.047
avg. degree	-2.561	-2.949	0.101	-2.387	-2.951	0.146
avg. shortest path	-2.451	-2.951	0.128	**-6.312**	-2.949	0.000
avg. distance	-1.261	-2.954	0.647	**-7.681**	-2.949	0.000
#cc	**-6.508**	-2.949	0.000	**-5.423**	-2.949	0.000
largest cc	**-5.280**	-2.949	0.000	-2.502	-2.951	0.115
order	**-3.462**	-2.951	0.009	-2.413	-2.949	0.138
symmetry	-2.626	-2.949	0.088	-1.451	-2.951	0.558
size	**-3.936**	-2.986	0.002	-2.619	-2.986	0.089
	Fishing			**Passenger**		
avg. clustering	**-6.080**	-2.949	0.000	**-4.820**	-2.949	0.000
avg. degree	**-4.547**	-2.951	0.000	**-4.012**	-2.949	0.001
avg. shortest path	**-6.461**	-2.949	0.000	-0.964	-2.961	0.766
avg. distance	**-4.705**	-2.949	0.000	-1.102	-2.986	0.714
#cc	**-3.950**	-2.951	0.002	**-3.904**	-2.957	0.002
largest cc	**-5.755**	-2.949	0.000	-1.115	-2.961	0.709
order	-2.910	-2.986	0.044	-2.002	-2.986	0.285
symmetry	**-3.132**	-2.954	0.024	**-5.871**	-2.954	0.000
size	-2.619	-2.986	0.089	-1.809	-2.981	0.376

Table 3: Augmented Dickey-Fuller (ADF) Test. Bold values indicate stationarity

pothesis, i.e., the time-series is stationary. The ADF statistic above the critical values suggests not rejecting the null hypothesis, meaning that the time-series is non-stationary.

We report the ADF test results on the VGTs in Table 3. The table reports the ADF-Statistic, the p-value, and the critical value for each VGT associated to different vessel types. The bold lines represent the ones with the ADF statistic lower than the critical value (5%), suggesting then that the time-series has a stationary behavior. The lower is the ADF statistic and the lower is the p-value, the more likely we have a stationary time-series.

By looking at the results, we can observe that for all the type of vessels, the feature *#cc* seems to have stationary behavior. On the other hand, both the VGT of *size* and *order* show a non-stationary behavior for most of the vessel types. A trending possibly explains this on the number of active ports and voyages performed over time. We also notice that for Passenger vessels, both the VGT of *avg. distance* and *avg. shortest path* present very high p-values demonstrating to have relevant time-dependency, 0.714 and 0.766 respectively. We recall that when we have a high p-value, we fail to reject the null hypothesis, the data has a unit root and is then non-stationary. This could indicate the presence of some trend or seasonality for the *avg. distance* and *avg. shortest path* of the passenger vessels voyages.

We studied these two time-series more in-depth by analyzing the auto-correlation of these two time-series individually. Auto-correlation analysis is perhaps one of the most compelling aspects to uncover hidden patterns in time-series data and represents the similarity between observations as a func-

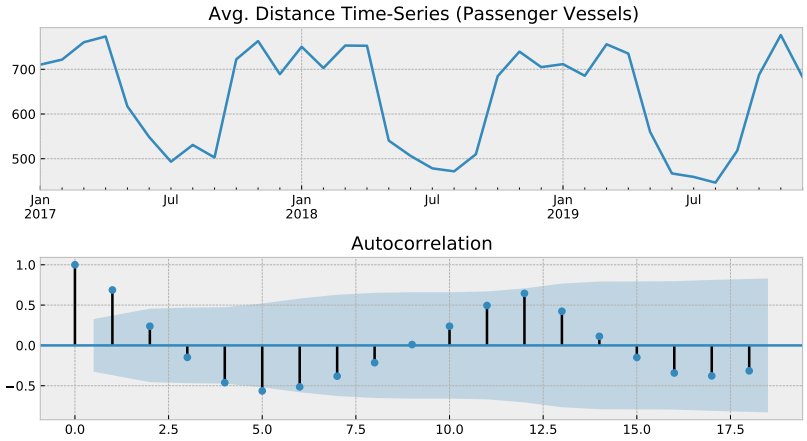


Fig. 11: Autocorrelation of the VGT of *Avg. distance* from Passenger vessels

tion of the time lag between them [19]. Figure 11 shows, within 95% confidence interval (represented by the solid gray line), the values of the AutoCorrelation Function (ACF) for the *avg. distance* time-series with lags ranging from 1 month to 18 months. Interestingly, we can observe a seasonal period of 12 months for this VGT. In turn, for the VGT of *avg. shortest path*, we applied a seasonal decomposition using a moving average to identify any presence of trend in the time-series [9]. Figure 12 shows the components of the decomposition. The first panel shows the observed values of the VGT in question, the second panel exhibits the trend component, and finally the third panel shows the seasonal component. From the trend component, we can notice a positive trending for the VGT of *avg. shortest path* corroborating its non-stationary characteristic.

Finally, regarding the stationary behaviour of the VGTs for the different vessel types (i.e., *cargo*, *fishing* , *passenger* and *tanker*), our analysis suggests a quite uniform binomial distribution (stationary/non-stationary). Indeed, this characteristic regards half of the investigated VGTs (i.e 18 of the cases over the 36 time-series analysed).

5.2 RQ2: Correlation Analysis

In this section, we tackle our second research question concerning the correlation between the VGT. To this end, we performed a correlation analysis to gain insights into how the VGT values derived from different vessel types correlate. For example, we investigate how the networks created by the cargo vessels are correlated to the ones made by passenger's vessels. Also, we inves-

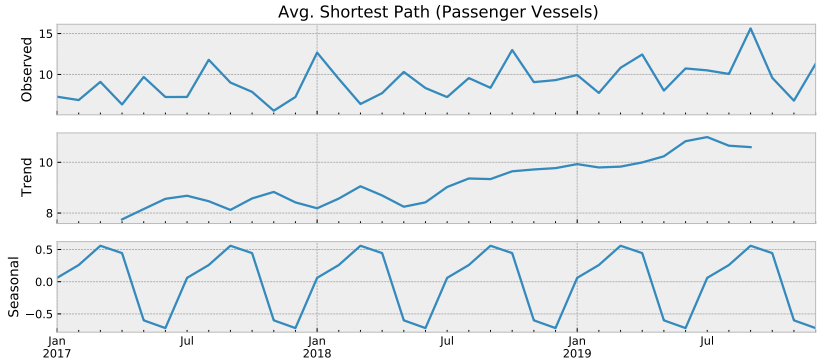


Fig. 12: Seasonality decompositon using moving average of the VGT of *Avg. shortest path* from Passenger vessels

tigate how the topological features of the networks correlates to each other for networks of different vessel types. From a topological perspective, these insights can help us understand the correlation between the voyages done by the different types of vessels over time.

An important consideration for this experiment is that we are dealing here with time-series data. When exploring relationships between two time series, we want to observe whether variations in one series are correlated with variations in another or not. For a proper correlation analysis involving this type of data, trends should be removed. For this purpose, we use a non-parametric method called first differences, in which each point of the VGT is subtracted by its previous point.

Then, after applying the first difference, we represent the time-series as a vector with N-1 points, where N is the number of points of the time-series. We used both Pearson and Spearman correlation coefficients to perform the correlation analysis. Pearson seeks linear relationship [23], while Spearman benchmarks monotonic relationship [1]. We also recall that both correlation coefficients have values between -1 and +1, where -1 means an inverse relationship, +1 (perfectly related), and zero, indicating no correlation at all.

We report the VGTs with significant coefficient values displaying those with the most positive and most negative correlations. Among the VGTs, the *order*, *size*, and *average_degree* were the ones exhibiting higher positive correlation. Figure 13 shows the results for these three VGTs. As we can see, for all of them, the cargo and tanker series are the ones with higher correlations, with Pearson and Spearman coefficient higher than 0.79. This means that, over the years, the cargos and tankers' voyages show a linear and monotonic relationship for the number of reached ports and the number of unique origin-destination ports voyages. For the VGT for *order* and *size* we can also observe a relevant

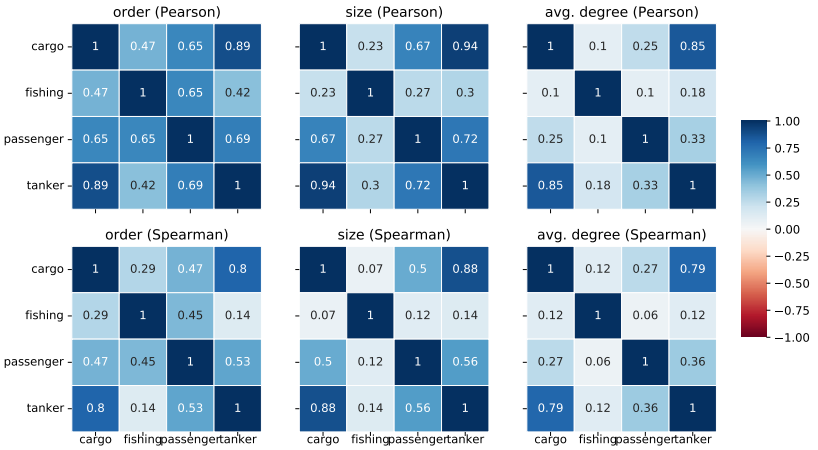


Fig. 13: Pearson (on the top) and Spearman (on the bottom) correlation coefficients for the VGT of *order*, *size* and *avg. degree*.

correlation between passenger and tanker vessels' voyages, exhibiting Pearson, and Spearman correlation higher than 0.53.

On the other side, when looking for negative coefficients, we did not observe any high inverse correlation among the VGT values. Figure 14 shows the correlation coefficients for the features *#cc* and *largest cc*, these two features were the ones having lowest coefficient values. The results show a negative Spearman coefficient equals to -0,48 for the *largest cc* obtained from the tankers' voyages and two other vessel types voyages, cargo, and fishing. This is an interesting finding indicating an inverse relationship between the length of the largest connected ports network created by tracing these types of vessels.

Figure 15 shows the pair-wise Pearson correlation for the topological features (Spearman correlation is not shown as the results are essentially the same). Most of the results are expected. The size of the larges connected components negatively correlates with the total number of connected components, with lower correlation values for the cargo network. This confirms that the cargo network, and to a minor extent also the tanker network, represents worldwide operations, compared to local operations of the passenger and fishing networks. Interestingly, the number of connected components also negatively correlates with the average degree, with lower values for the fishing networks. This reflects the poor connectivity properties of such a network, with vessels that infrequently move between ports.

We can conclude this Research Question by highlighting the statistical correlation between some VGT derived from the different vessels type. Regarding positive correlation, the features *size*, *order* and *avg. degree* show a relevant

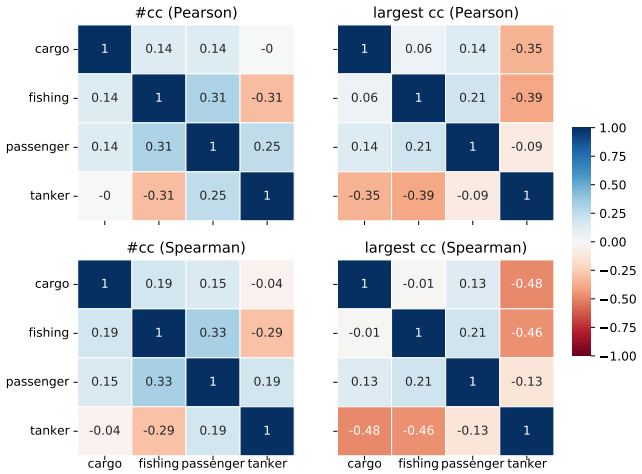


Fig. 14: Pearson (on the top) and Spearman (on the bottom) correlation coefficients for the VGT of *#cc* and *largest cc*

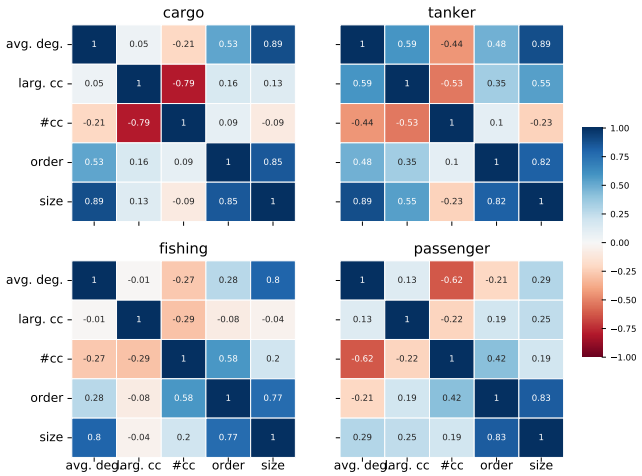


Fig. 15: Pearson correlation of several topological features for the various vessel types

similarity between cargos and tankers. While, the VGT values *largest cc*, for the same two vessel types, present a slight negative correlation.

6 Conclusion

This paper presented an analysis of the evolution of networks of voyages of vessels between ports, based on several topological features of the network (VGT). The networks were built in a bottom-up and data-driven fashion, considering three years of worldwide AIS data provided by ExactEarth. The empirical evaluation of the VGT shown that LRVs, such as cargos and tanker vessels, tend to form well-connected giant strongly connected components that are relatively stable over time; by comparison, the SRVs behaviour is more variable over time and the resulting networks are more fragmented, with each component well-connected even if small. The analysis of stationarity and correlation confirms these findings. In particular, among the topological features considered, we have observed that half VGTs present characteristics of non-stationary, therefore suggesting the presence of seasonal patterns. For example, we observed that the average distance of the networks formed by passenger vessels has a seasonal period of one year. Another interesting aspect addressed in this work refers to the correlation analysis of the VGT values between the different types of vessels. This study brought some insights into how the network built by the voyages of different types of vessels present some correlated VGT values in the years considered.

Several future directions can be considered to improve and expand upon this work. First, the definition of the spatial area of ports can be improved to increase the precision in voyages mining, in a way similar to the one performed in [37], or using other spatial division techniques, such as the one based on Voronoi partitions. Second, an in depth analysis of per-node graph features, in contrast with the per-graph features of this paper could provide additional elements to evaluate and model the evolution of maritime traffic. For example, an accurate study of nodes centrality and ego networks would extract additional insights on the role of specific ports in the network. Third, the performed analysis can guide the implementation of data-driven prediction mechanisms to forecast the evolution of the networks using the large amount of AIS data available.

Acknowledgment

The authors acknowledge the support of the H2020 EU Project MASTER (Multiple ASpects TrajEctoRy management and analysis) funded under the Marie Skłodowska-Curie grant agreement No 777695.

References

1. Spearman Rank Correlation Coefficient, pp. 502–505. Springer New York, New York, NY (2008). DOI 10.1007/978-0-387-32833-1_379. URL https://doi.org/10.1007/978-0-387-32833-1_379
2. Bartholdi, J.J., Jarumaneeroj, P., Ramudhin, A.: A new connectivity index for container ports. Maritime Economics & Logistics **18**(3), 231–249 (2016)
3. Blake, G.H.: Maritime boundaries. In: The Oceans: Key Issues in Marine Affairs, pp. 63–76. Springer (2004)
4. Brockwell, P.J., Davis, R.A.: Introduction to time series and forecasting. springer (2016)
5. Carlini, E., de Lira, V.M., Soares, A., Etemad, M., Machado, B.B., Matwin, S.: Uncovering vessel movement patterns from ais data with graph evolution analysis. In: EDBT/ICDT Workshops (2020)
6. Cavaliere, G., Robert Taylor, A.: Time-transformed unit root tests for models with non-stationary volatility. Journal of Time Series Analysis **29**(2), 300–330 (2008)
7. Chatfield, C.: Time-series forecasting. Chapman and Hall/CRC (2000)
8. Cheung, Y.W., Lai, K.S.: Lag order and critical values of the augmented dickey–fuller test. Journal of Business & Economic Statistics **13**(3), 277–280 (1995)
9. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: Stl: A seasonal-trend decomposition. Journal of official statistics **6**(1), 3–73 (1990)
10. Coscia, P., Braca, P., Millefiori, L.M., Palmieri, F.A.N., Willett, P.: Multiple Ornstein–Uhlenbeck Processes for Maritime Traffic Graph Representation. IEEE Transactions on Aerospace and Electronic Systems **54**(5), 2158–2170 (2018). DOI 10.1109/TAES.2018.2808098. URL https://ieeexplore.ieee.org/document/8295117/
11. Dickey, D.A., Fuller, W.A.: Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association **74**(366a), 427–431 (1979)
12. Ducruet, C.: Multilayer dynamics of complex spatial networks: The case of global maritime flows (1977–2008). Journal of Transport Geography **60**, 47–58 (2017)
13. Edition, S.C.: Safety of life at sea (solas), consolidated edition. IMO, London (2004)
14. Etemad, M., Júnior, A.S., Matwin, S.: Predicting transportation modes of gps trajectories using feature engineering and noise removal. In: Canadian Conference on Artificial Intelligence, pp. 259–264. Springer (2018)
15. exactearth.com: ExactEarth (last accessed July 2020). URL https://www.exactearth.com/
16. Fagiolo, G., Mastrorillo, M.: International migration network: Topology and modeling. Physical Review E **88**(1), 012812 (2013)
17. Filipiak, D., We, K., Abramowicz, W.: Extracting Maritime Traffic Networks from AIS Data Using Evolutionary Algorithm. Bus Inf Syst Eng p. 17 (2020)
18. Franses, P.H.: Seasonality, non-stationarity and the forecasting of monthly time series. International Journal of forecasting **7**(2), 199–208 (1991)
19. Ghanavati, G., Hines, P.D., Lakoba, T.I., Cotilla-Sanchez, E.: Understanding early indicators of critical transitions in power systems from autocorrelation functions. IEEE Transactions on Circuits and Systems I: Regular Papers **61**(9), 2747–2760 (2014)
20. Gollasch, S., Hewitt, C.L., Bailey, S., David, M.: Introductions and transfers of species by ballast water in the adriatic sea. Marine Pollution Bulletin **147**, 8 – 15 (2019). DOI https://doi.org/10.1016/j.marpolbul.2018.08.054. URL http://www.sciencedirect.com/science/article/pii/S0025326X1830626X
21. Hernández, J.M., Van Mieghem, P.: Classification of graph metrics. Delft University of Technology: Mekelweg, The Netherlands pp. 1–20 (2011)
22. Kaluza, P., Kölzsch, A., Gastner, M.T., Blasius, B.: The complex network of global cargo ship movements. Journal of the Royal Society Interface **7**(48), 1093–1103 (2010)
23. Kirch, W. (ed.): Pearson's Correlation Coefficient, pp. 1090–1091. Springer Netherlands, Dordrecht (2008). DOI 10.1007/978-1-4020-5614-7_2569. URL https://doi.org/10.1007/978-1-4020-5614-7_2569
24. Kitagawa, G., Akaike, H.: A procedure for the modeling of non-stationary time series. Annals of the Institute of Statistical Mathematics **30**(2), 351–363 (1978)

25. Kosowska-Stamirowska, Z., Ducruet, C., Rai, N.: Evolving structure of the maritime trade network: evidence from the lloyd's shipping index (1890–2000). Journal of Shipping and Trade **1**(1), 10 (2016)
26. Laxe, F.G., Seoane, M.J.F., Montes, C.P.: Maritime degree, centrality and vulnerability: port hierarchies and emerging areas in containerized transport (2008–2010). Journal of Transport Geography **24**, 33–44 (2012)
27. McCabe, B.P., Tremayne, A.R.: Testing a time series for difference stationarity. The Annals of Statistics pp. 1015–1028 (1995)
28. Montes, C.P., Seoane, M.J.F., Laxe, F.G.: General cargo and containership emergent routes: A complex networks description. Transport Policy **24**, 126–140 (2012)
29. msi.nga.mil: World Port Index (last accessed July 2020). URL `https://msi.nga.mil/Publications/WPI`
30. Perera, L.P., Oliveira, P., Soares, C.G.: Maritime traffic monitoring based on vessel detection, tracking, state estimation, and trajectory prediction. IEEE Transactions on Intelligent Transportation Systems **13**(3), 1188–1200 (2012)
31. Santipantakis, G.M., Glenis, A., Patroumpas, K., Vlachou, A., Doulkeridis, C., Vouros, G.A., Pelekis, N., Theodoridis, Y.: SPARTAN: Semantic integration of big spatio-temporal data from streaming and archival sources. Future Generation Computer Systems **110**, 540–555 (2020). DOI 10.1016/j.future.2018.07.007. URL `https://linkinghub.elsevier.com/retrieve/pii/S0167739X17320319`
32. Soares, A., Dividino, R., Abreu, F., Brousseau, M., Isenor, A.W., Webb, S., Matwin, S.: Crisis: Integrating ais and ocean data streams using semantic web standards for event detection. In: 2019 International conference on military communications and information systems (ICMCIS), pp. 1–7. IEEE (2019)
33. Varlamis, I., Kontopoulos, I., Tserpes, K., Etemad, M., Soares, A., Matwin, S.: Building navigation networks from multi-vessel trajectory data. GeoInformatica pp. 1–29 (2020)
34. Varlamis, I., Tserpes, K., Etemad, M., Júnior, A.S., Matwin, S.: A network abstraction of multi-vessel trajectory data for detecting anomalies. In: Proceedings of the Workshops of the EDBT/ICDT 2019 Joint Conference (2019)
35. Vespe, M., Greidanus, H., Alvarez, M.A.: The declining impact of piracy on maritime transport in the indian ocean: Statistical analysis of 5-year vessel tracking data. Marine Policy **59**, 9–15 (2015)
36. Vespe, M., Visentini, I., Bryan, K., Braca, P.: Unsupervised learning of maritime traffic patterns for anomaly detection. In: 9th IET Data Fusion Target Tracking Conference (DF TT 2012): Algorithms Applications, pp. 1–5 (2012)
37. Wang, Z., Claramunt, C., Wang, Y.: Extracting global shipping networks from massive historical automatic identification system sensor data: a bottom-up approach. Sensors **19**(15), 3363 (2019)
38. Yan, Z., Xiao, Y., Cheng, L., He, R., Ruan, X., Zhou, X., Li, M., Bin, R.: Exploring ais data for intelligent maritime routes extraction. Applied Ocean Research **101**, 102271 (2020)
39. Zhang, S.k., Shi, G.y., Liu, Z.j., Zhao, Z.w., Wu, Z.l.: Data-driven based automatic maritime routing from massive ais trajectories in the face of disparity. Ocean Engineering **155**, 240–250 (2018)
40. Zissis, D., Chatzikokolakis, K., Spiliopoulos, G., Vodas, M.: A Distributed Spatial Method for Modeling Maritime Routes. IEEE Access **8**, 47556–47568 (2020). DOI 10.1109/ACCESS.2020.2979612. URL `https://ieeexplore.ieee.org/document/9028151/`