

A Topological Perspective of Port Networks From Three Years (2017-2019) of AIS Data

Emanuele Carlini¹, Vinicius Monteiro de Lira¹, Amilcar Soares², Mohammad Etemad³, Bruno Brandoli³ and Stan Matwin³

¹*Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Pisa, Italy*

²*Department of Computer Science, Memorial University of Newfoundland, St. John's, Canada*

³*Institute for Big Data Analytics, Dalhousie University, Halifax, Canada*

Abstract

Complex network analysis is a fundamental tool to understand non-trivial aspects of graphs and networks and is widely used in many fields. In this paper, we apply complex network techniques to study port networks, in which nodes are ports and edges are maritime lines between ports. In particular, we study the temporal evolution of several topological features of a network of ports, including connected components, shortest paths, and clustering coefficients. We built the network with three years of Automatic Identification System data from 2017 to 2019. We highlight several interesting trends and behaviors that differentiate long-range vessels from short-range vessels.

Keywords

Automatic Identification System, Graph Analysis, Time series

1. Introduction

The analysis of maritime data is a well-established source of information to understand the role of maritime routes in economic, social, and environmental contexts. Recent works propose to use data analytics and complex network tools to find routes [1], extract high-level representations and evaluate local maritime traffic [2], and integrate maritime data with other environmental data [3]. Such studies often employ the modelling of the relationship between vessels and ports as a network. Such a network, which we call Port Network, represents sea ports as nodes and edges as maritime lines connecting two ports. Since the introduction of the Automatic Identification System (AIS) for vessels, there has been a surge of studies on the Port Networks [4, 5]. Port Networks are usually analyzed with tools that are typical of complex networks, which allow to compute non-trivial topological features of the network. However, only few works study the network in terms of its evolution [6, 7]. Also, those works that studied the network evolution used private data and performed interesting but high-level and coarse-grained analysis, such as in [8].


SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy

✉ emanuele.carlini@isti.cnr.it (E. Carlini); vinicius.monteirodelira@isti.cnr.it (V.M. d. Lira); amilcarsj@mun.ca (A. Soares); etemad@dal.ca (M. Etemad); brunobrandoli@dal.ca (B. Brandoli); stan@cs.dal.ca (S. Matwin)

🆔 0000-0003-3643-5404 (E. Carlini); 0000-0002-7580-1756 (V.M. d. Lira); 0000-0001-5957-3805 (A. Soares); 0000-0002-3770-180X (M. Etemad); 0000-0001-6629-8434 (S. Matwin)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Reference	Data source	Analysis over time	Data scope
<i>Kaluza et. al, 2010 [11]</i>	Lloyd's	✗	Global
<i>González Laxe et al., 2012 [6]</i>	Lloyd's	✗	Local
<i>Kosowska-Stamirowska et al., 2016 [12]</i>	Lloyd's	✓	Global
<i>Ducruet, 2017 [8]</i>	Lloyd's	✓	Global
<i>Coscia et al., 2018 [13]</i>	AIS	✗	Local
<i>Wang et al., 2019 [4]</i>	AIS	✗	Global

Table 1

A summary of the related works regarding the four evaluated aspects.

The main goal of our analysis is to provide an overview on the evolution and trends of several Port Network topological features, by consider 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). Networks can be analysed by looking at the local features of their main components (i.e., nodes and edges), or by looking at features of the network as a whole. In this paper we are interested in the latter. We analyzed 3 years of world-wide AIS data to investigate Port Networks in terms of topological features such as connected components, shortest paths, and clustering coefficients. From our analysis we observed that long range vessels, such as cargoes and tanker vessels, tend to form well-connected large networks that are relatively stable over time, while short range vessel (i.e., passenger, fishing) form small well-connected networks with a lot of variability over time.

The content of this discussion paper is based on other manuscripts [9, 10] already published by the same authors, which contains an extensive related work, a detailed methodology, and a correlation and stationarity analysis of the network measures. With respect to [9] however, this paper provide the analysis on a cleaner and improved AIS dataset.

2. Related Work

Table 1 summarizes how state of the art in the graph analysis with vessel data, in terms of source data (AIS or Lloyd's database), whether the works evaluated the graph evolution over time, and the scope (local or global).

The work done in [11] is one of the first to study the concept of Port Networks 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. They show that the three categories of ships differ in their mobility patterns and networks. The work of [6] 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

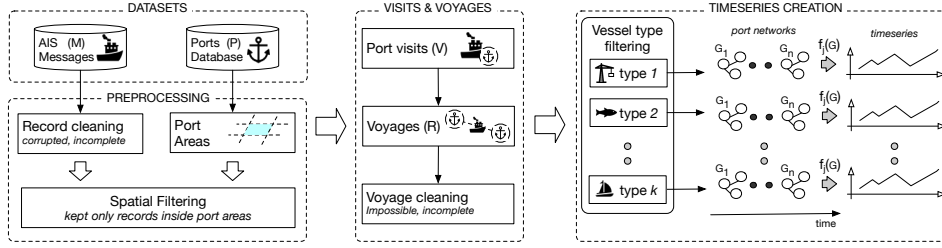


Figure 1: Methodology for the creation of Port Networks and Time-series from AIS data

the global and local importance of a port can be measured using graph theory concepts.

A study of topological changes in the maritime trade network is shown in [12]. 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 and the maritime network does not densify over time, and its effective diameter remains constant. In [8], the author investigates the degree of overlap among the different layers of circulation composing global maritime flows. 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). The work of [4] builds a Port Network using the 2015 AIS data of the world with multiple spatial levels. 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. In a similar way, [13] 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.

3. Creation of Port Networks and Timeseries

This section describes our methodology to generate Port Networks and the correspondent Time-series. A visualization of the various steps is depicted in Figure 1.

Datasets. To build the Port Networks we have used three years (2017-2019) of worldwide AIS data provided by ExactEarth [14]. The full dataset contains around 2.5 Terabytes of AIS messages and around 20 billions of records stored in a relational database. In all our analysis we consider the Maritime Mobile Service Identity (MMSI) as each vessel unique identifier. 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 scalability. To model the ports we have used the World Port Index dataset [15] that contains spatial information, including latitude and longitude, of all known seaports in the world. The radius of the port area has been set to be 3 nautical miles (around 5 km). This value has been used to define country's territorial waters limit [16].

AIS data pre-processing. The aim of the pre-processing step is twofold. First, extracting the AIS messages that happened inside the area of a port. Depending on the radius, 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 are then spatial filtered with the clustered ports regions, in order to create a new set that contains only those messages transmitted inside the port areas. Second, removing incorrect, duplicated, and noisy messages. Incorrect messages are those syntactically valid but with invalid semantics in relevant fields (typically position or vessel type).

Visits and Voyages. We define a visit by the continuous presence of a particular vessels in a port area. Multiple consecutive AIS messages in the same port area are considered as the same visit. By ordering the visits by time, we obtain a sequence of visits for each vessel. From the sequence of visits we extract the set of voyages. The underlying assumption is that given a sorted set of visits, we record a voyage from a origin to a destination port by observing the visit sequence of each vessel. The duration of a voyage is the time of the last visit of a vessel in the origin port and the time of first visit in the destination port. We then removed those voyages whose speed exceeds the capabilities of the vessels.

Port Network and Time-series. From the sequences of visits, we create multiple Port Networks, each considering a specific consecutive, non-overlapping time intervals. A Port Network is built by considering ports as nodes and the voyages as edges. The resulting network is a directed graph built by essentially collapsing a multi-graph into a directed graph. By extracting several topological features from each Port Network, we create a set of time-series to be able to study the evolution of the graphs using complex network concepts. The Time-series have been build with a time interval of a solar month, resulting in a total of 36 Time-series for each considered vessel type and each topological feature.

4. Port Networks Analysis

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](https://help.marinetraffic.com/) website¹ (with minor modifications). We have considered only those vessel types (*layers*) having a relevant amount of unique vessels and voyages count, namely: (i) Cargo (37% unique vessel count, 6.6% unique voyages count); (ii) Tanker (15.7%, 3.3%); (iii) Passenger (3.84%, 3.13%); and (iv) Fishing (6%, 47.9%). We did not consider the *special* or the *other* types as they contain many different types of vessels; similarly, we did not consider *tug tows* as they usually perform very short voyages between nearby ports, and therefore are not interesting in a global world-wide analysis.

The average orthodromic distance between all the edges of the graph (Figure 2), as similarly observed in [8], confirms that cargo and tanker perform longer voyages with respect to passenger and fishing vessels. Following these considerations we refer to cargo and tanker vessels as

¹<https://help.marinetraffic.com/>

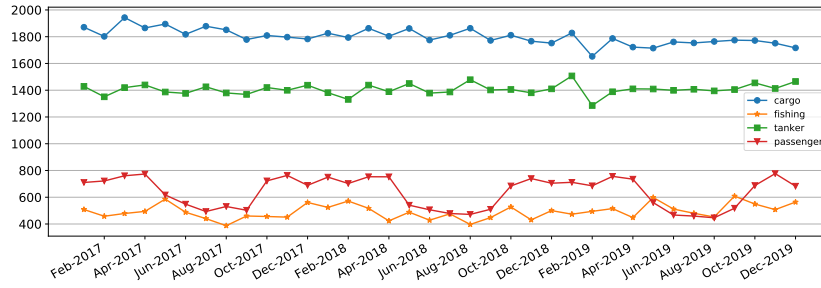


Figure 2: Average orthodromic distance in kilometers between nodes connected by an edge

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.

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 [17]. 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). The average number of SCCs for the SRV and LRV networks in the 3-years period confirm this trend, with LRV having a lower number of SCCs in average (cargo: 153; fishing: 203; tanker: 141; passenger: 194). 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 4). As expected, nodes are more evenly distributed among the SCCs for SRV networks, in which the largest connected components account for just around 40% of the nodes on average for the passenger networks and around 20% for the fishing networks. From a geographic perspective, the LRV giant component spans world-wide. Those ports that remain out of the giant component show a seasonal trend with a clear difference from winter and summer periods (see Figure 3).

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, 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 5). 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 [11]. By comparison, in SRV

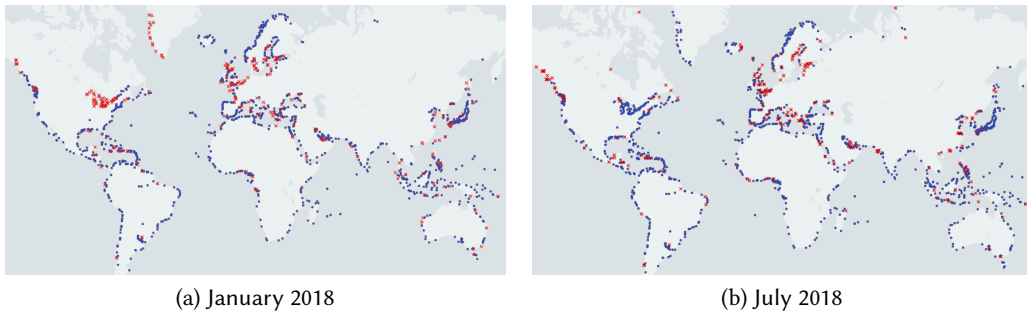


Figure 3: Ports in the giant connected components (blue circles) vs ports outside it (red crosses). During winter periods (left) several north-most areas are cut out from the giant component, such as in the Greenland or the Great Lakes of North America, whereas they are present during summer(right)

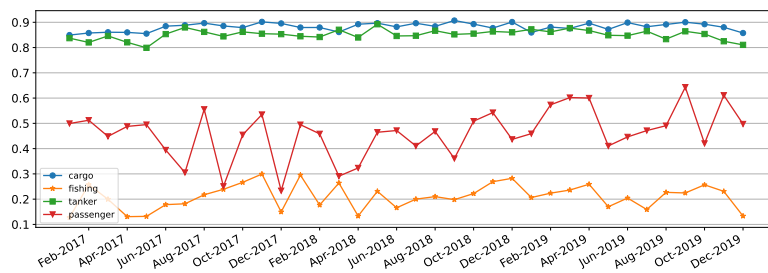


Figure 4: Fraction of nodes in the largest strongly connected component

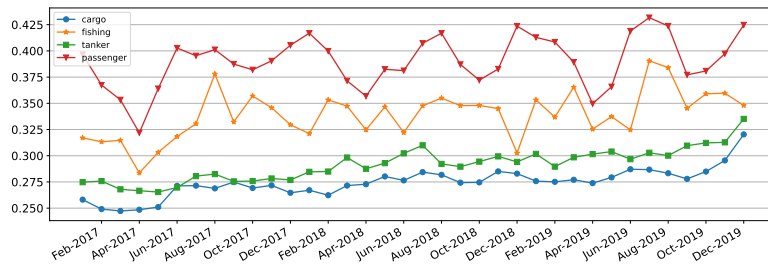


Figure 5: Fraction of bidirectional edges

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. In the vessel network, a lower average shortest path reveals more dense port connections. The average shortest path (computed on the giant connected component) of LRV networks is around 4 for tankers and cargo and is stable over time (see Figure 6). For SRV networks, the average shortest path is 4 for for fishing vessels, but with a much higher variability with respect to LRV networks. The average shortest path is relatively high and variable for passenger vessels, indicating a

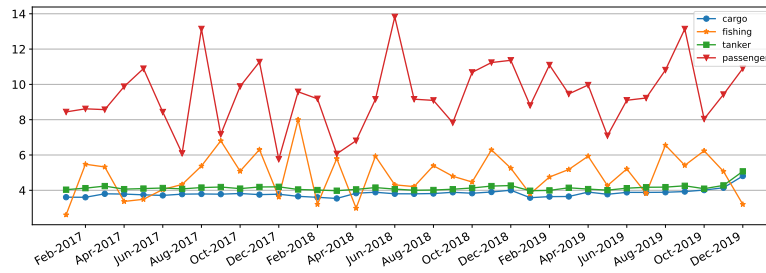


Figure 6: Average shortest path. As a matter of comparison, for similar size random networks, we measured the following average shortest path: 2.7 for cargo, 3.5 for fishing, 2.9 for tanker, and 3.9 for passenger.

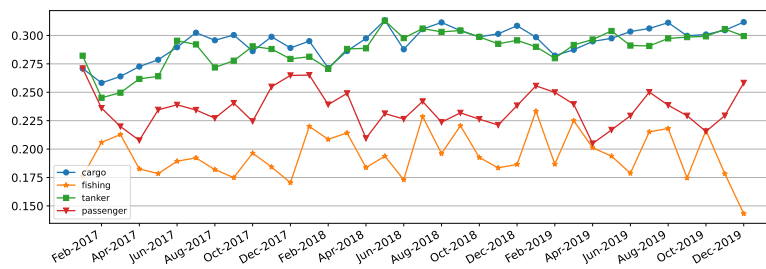


Figure 7: Average clustering coefficient. As a matter of comparison, for similar size random networks, we measured the following average clustering coefficient: 0.01 for cargo, 0.02 for fishing, 0.01 for tanker, and 0 for passenger.

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 a node 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 [17]. In other words, it can serve to evaluate how many voyages happen around the same set of ports. The results (Figure 7) show that cargo and tanker networks create networks of higher density with respect to passenger and fishing networks. The average clustering coefficient variability is high for all the type of vessels, 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. Conclusion

This paper presented an analysis of the evolution of networks of voyages of vessels between ports, based on several Time-series of topological features of the Port Network. The networks were built in a bottom-up and data-driven fashion, considering three years of worldwide AIS data. The empirical evaluation of the Time-series 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.

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] D. Zissis, K. Chatzikokolakis, G. Spiliopoulos, M. Vodas, A Distributed Spatial Method for Modeling Maritime Routes, *IEEE Access* 8 (2020) 47556–47568. URL: <https://ieeexplore.ieee.org/document/9028151/>. doi:10.1109/ACCESS.2020.2979612.
- [2] D. Filipiak, K. We, W. Abramowicz, Extracting Maritime Traffic Networks from AIS Data Using Evolutionary Algorithm, *Bus Inf Syst Eng* (2020) 17.
- [3] A. Soares, R. Dividino, F. Abreu, M. Brousseau, A. W. Isenor, S. Webb, S. Matwin, 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), IEEE, 2019, pp. 1–7.
- [4] Z. Wang, C. Claramunt, Y. Wang, Extracting global shipping networks from massive historical automatic identification system sensor data: a bottom-up approach, *Sensors* 19 (2019) 3363.
- [5] I. Varlamis, I. Kontopoulos, K. Tserpes, M. Etemad, A. Soares, S. Matwin, Building navigation networks from multi-vessel trajectory data, *GeoInformatica* (2020) 1–29.
- [6] F. G. Laxe, M. J. F. Seoane, C. P. Montes, Maritime degree, centrality and vulnerability: port hierarchies and emerging areas in containerized transport (2008–2010), *Journal of Transport Geography* 24 (2012) 33–44.
- [7] C. P. Montes, M. J. F. Seoane, F. G. Laxe, General cargo and containership emergent routes: A complex networks description, *Transport Policy* 24 (2012) 126–140.
- [8] C. Ducruet, Multilayer dynamics of complex spatial networks: The case of global maritime flows (1977–2008), *Journal of Transport Geography* 60 (2017) 47–58.
- [9] E. Carlini, V. M. de Lira, A. Soares, M. Etemad, B. Brandoli, S. Matwin, Understanding evolution of maritime networks from automatic identification system data, *GeoInformatica* (2021) 1–25.
- [10] E. Carlini, V. M. de Lira, A. Soares, M. Etemad, B. B. Machado, S. Matwin, Uncovering vessel movement patterns from ais data with graph evolution analysis, in: EDBT/ICDT Workshops, 2020.
- [11] P. Kaluza, A. Kölzsch, M. T. Gastner, B. Blasius, The complex network of global cargo ship movements, *Journal of the Royal Society Interface* 7 (2010) 1093–1103.
- [12] Z. Kosowska-Stamirowska, C. Ducruet, N. Rai, Evolving structure of the maritime trade network: evidence from the lloyd’s shipping index (1890–2000), *Journal of Shipping and Trade* 1 (2016) 10.
- [13] P. Coscia, P. Braca, L. M. Millefiori, F. A. N. Palmieri, P. Willett, Multiple Ornstein–Uhlenbeck Processes for Maritime Traffic Graph Representation, *IEEE Transactions on Aerospace and Electronic Systems* 54 (2018) 2158–2170. doi:10.1109/TAES.2018.2808098.
- [14] exactearth.com, ExactEarth, last accessed July 2020. URL: <https://www.exactearth.com/>.
- [15] msi.nga.mil, World Port Index, last accessed July 2020. URL: <https://msi.nga.mil/Publications/WPI>.
- [16] G. H. Blake, Maritime boundaries, in: *The Oceans: Key Issues in Marine Affairs*, Springer, 2004, pp. 63–76.
- [17] J. M. Hernández, P. Van Mieghem, Classification of graph metrics, Delft University of Technology: Mekelweg, The Netherlands (2011) 1–20.