



HAL
open science

Maritime Network Analysis: Connectivity and Spatial Distribution

César Ducruet, Justin Berli, Giannis Spiliopoulos, Dimitris Zissis

► **To cite this version:**

César Ducruet, Justin Berli, Giannis Spiliopoulos, Dimitris Zissis. Maritime Network Analysis: Connectivity and Spatial Distribution. Guide to Maritime Informatics, Springer International Publishing, pp.299-317, 2021, 10.1007/978-3-030-61852-0_10 . hal-03293379

HAL Id: hal-03293379

<https://hal.science/hal-03293379>

Submitted on 20 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Maritime Network Analysis: Connectivity and Spatial Distribution

Draft version of the chapter published in Artikis A. & Zissis D. (2021) Guide to Maritime Informatics, Springer, pp 299-317

César DUCRUET¹, Justin BERLI¹,
Dimitris ZISSIS^{2,3}, Giannis SPILIOPOULOS³

Abstract In this chapter we apply conventional graph-theoretical and complex network methods to a sample of port and inter-port shipping flows at and amongst the top 50 European ports in 2017 to detect the main topological and geographic structures of this network. Main results confirm earlier works by physicists about liner shipping network but our approach based on dry cargo and liquid cargo goes further with a mix of novelty and confirmation on how maritime networks and the European backbone in particular is driven by which forces.

1 Introduction

In this chapter we apply conventional methods of graph theory / complex networks and geomatics to maritime traffic at and between the top 50 European ports. Our initial aim is to transform the positional data collected through the Automatic Identification System (AIS) into a “port to port” graph connecting the top 50 European ports in terms of their total number of vessel arrivals (or calls) during the year 2017. The AIS, is a collaborative, self-reporting system that allows vessels to broadcast their identification information, characteristics and destination, along with other information originating from on-board devices and sensors, such as location, speed and heading¹. AIS messages are broadcast periodically and can be received by other vessels equipped with AIS transceivers, as well as by on the ground or satellite-based sen-

¹Centre National de la Recherche Scientifique (CNRS)

²University of the Aegean

³MarineTraffic

cdu@parisgeo.cnrs.fr, justinberli@gmail.com,
dzissis@aegean.gr, giannis.spiliopoulos@marinetraffic.com

¹ “M.1371: Technical characteristics for an automatic identification system using time-division multiple access in the vhf maritime mobile band,” <https://www.itu.int/rec/R-REC-M.1371/en>

sors. Since becoming obligatory by the International Maritime Organisation (IMO) for vessels above 300 gross tonnage to carry AIS transponders, large datasets are gradually becoming available and are now being considered as a valid source of traffic patterns insights². The ITU 1371-4 standard defines 64 different types of AIS messages. Types 1, 2, 3, 18, and 19 are position reports, which include latitude, longitude, Speed Over Ground (SOG), Course Over Ground (COG), and other fields related to ship movement; type 5 messages contain static-and-voyage information, which include the IMO identifier, radio call sign, name, ship dimensions, ship and cargo types.

For this study, we make use of AIS data collected in EU waters during the same year, with a focus on cargo vessels (AIS Ship Type:70-79) and tankers (AIS Ship Type:80-89) travelling both to and from the top 50 ports in the EU. AIS messages include both departure and destination port information, but unfortunately it cannot be considered as a reliable source of information, as it is manually entered by the vessel's crew, without following a specific standard, making it thus prone to errors. For this purpose, we calculate port calls by making use of the World Port Index dataset³, which contains the location and physical characteristics of major ports and terminals worldwide. The aim is to assign to all positional data accurate port arrival information. As an additional step and so as to improve result accuracy, we exclude smaller vessels that only operate within the given port area from this study by only calculating arrivals of vessels that have a valid IMO number and gross tonnage greater than 5000 tons.

As for any other (transport) network, such maritime flows are considered as a graph with ports as nodes (or vertices) connected by inter-port voyages as links (or edges) so that

$$[G = (v, e)] \quad (1)$$

The weight of connections among ports of the EU-28 network, is the number of trips and distances for each pair of top 50 ports per ship type (dry or liquid cargo). The difference between network and graph being that the graph exists in a relatively abstract topological space regardless of the real pattern of infrastructure or circulation (Figure 1). Such a transformation is a simplification allowing to apply a variety of measures and algorithms at the global and local levels (see Ducruet and Lugo, 2013 for a synthesis). Global level refers to the whole network's properties while local level considers individual nodes within the group.

² "Solas paper v - regulation 19 - carriage requirements for shipborne navigational systems and equipment," <http://solasv.mcga.gov.uk/regulations/regulation19.htm>

³ https://msi.nga.mil/NGAPortal/MSI.portal?_nfpb=true&_pageLabel=msi_portal_page_62&pubCode=0015

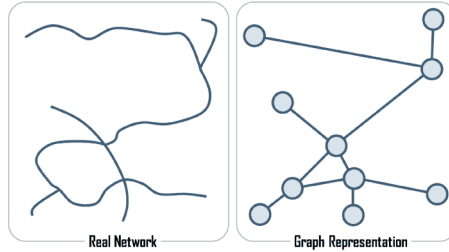


Fig. 1: Schematic representation of a network and a graph.

Source: Rodrigue and Ducruét (2018).

Maritime networks are typically *non-planar graphs* as they do not necessitate the existence of a vertex at each intersection between edges, like for airline networks, social networks, and multinational networks to name but a few, which are made of ephemeral or sometimes invisible interactions of flows of goods, people or information. This topology stands in sharp contrast with *planar graphs* made of physical infrastructure such as roads and railways but also power grids and the Internet. Cooperation between geographers and computer scientists pushed further the analysis and visualization of geographical networks of which maritime networks (Rozenblat and Melançon, 2013).

In this chapter, we explore both the topological and geographic properties of Europe's shipping network backbone to extract meaningful information about its structure and spatial distribution. This is not such a common approach in maritime network analysis, where topological approaches remain, unfortunately, separated from the spatial approach (Ducruét, 2015). These two approaches are actually complementary and provide readers a fully-fledged vision of how shipping flows exhibit specific features like other complex networks such as concentration, hierarchy, and density (see the recent work of Ansorena, 2018 for an example of a country-level empirical study). Ancient data extracted from historical records such as ship logs allowed for a precise analysis of past maritime networks for different purposes (see Siegfried, 1940; Garcia-Herrera et al., 2017; Mobasheri et al., 2017). More recently, AIS data became increasingly used for studying maritime networks to answer a wide variety of issues such as marine bioinvasions (Kaluza et al., 2010), the identification of abnormal vessel trajectories (Etienne et al., 2015), the delineation of fishing zones (Le Guyader et al., 2017), the foreland mix of European regions in relation to hinterlands (Guerrero et al., 2017), the estimation of delay and congestion risks (Stergiopoulos et al., 2018), to name but a few.

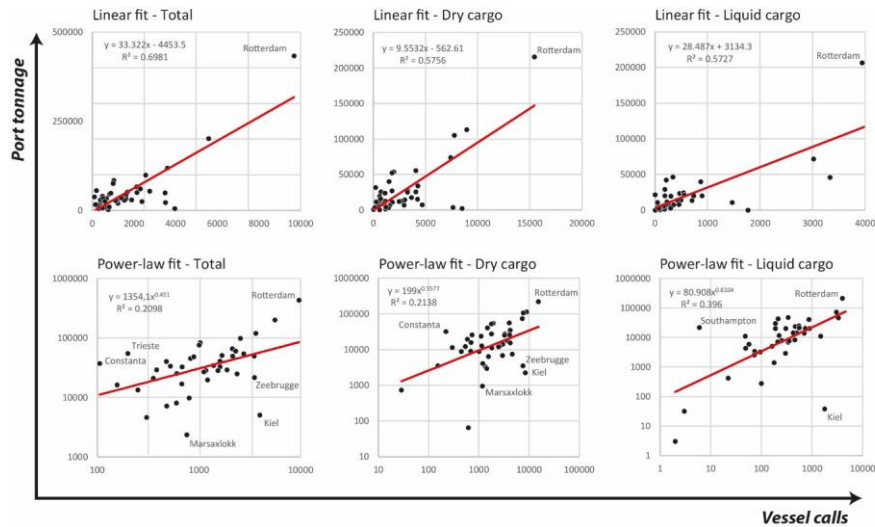


Fig. 2: Top European port hierarchy by type of traffic measurement, 2017.

Source: own elaboration based on MarineTraffic and Eurostat data.

The data sample under investigation here significantly correlates with official port throughput statistics like those provided by Eurostat (2018) as seen in Figure 2, with a determination coefficient of nearly 0.7 between port throughput and number of ship arrivals. The remaining 0.3 is mainly explained by the difference in unit, namely tons versus vessel calls. Yet, the log-log plot with a power-law function provides a much less significant coefficient (0.21). The figure points to a number of outliers which influence is blurred using a linear function. For instance, for total traffic, Kiel (Germany) and Marsaxlokk (Malta) have much more calls than their throughput would expect, and the inverse occurs for Trieste (Italy) and Constanta (Romania). Causes vary from case to case. Some ports handle large ships by fewer calls, such as in the latter case for those specialized in bulky commodities. Others have a high call frequency due to their role as transshipment hub (cf. hub-and-spokes pattern) but in the end handle lower weight but higher value goods (e.g. containers) like for Marsaxlokk.

However, another case appears that is due to the nature of traffic measurement itself. AIS data can be considered as surveillance data, tracking vessels as they perform their trips across the sea and between ports. A safe ship journey begins and ends at a sea port (or an anchorage within or close to the ports operational area), but during their voyages, vessels often pass through or in close proximity to other ports without making a port visit. As port calls/visits are measured by calculating when a vessel's trajectory intersects with a given ports "geographic boundaries" it is often the case that due to mistakes during the analysis or in the underlying port definition data, vessels are assumed to enter a port when only passing by. This issue

was clearly apparent in the work of Kaluza et al. (2010) where a number of minor ports appeared in their global ranking of the top 20 ports, such as Plaquemines and Terneuzen. This probably explains the overestimation of Kiel's traffic importance and the presence of tiny ports among the top 50 European ones such as Fischerhutte (Germany), a small river port along the Kiel Canal near Hamburg; Hochdown (?); Ijmuiden (outport of Amsterdam); Purfleet and Tilbury along the Thames River near London; and the river ports of Sas van Gent, Suiskil, and Zelzate (The Netherlands) lying between Terneuzen and Ghent near Antwerp.

Unfortunately, the throughput of these aforementioned ports is not reported by Eurostat although the latter source provides European-wide statistics from top ports to thousands of minor ports. We have decided, however, to keep these outliers in the following analyses despite the risk of providing a somewhat distorted view of the reality of Europe's maritime connectivity and port hierarchy (see Appendix 1 for a comparative table of Europe's 27 top ports). This will enable the comparison with other papers of the present book and test the role of these outliers in the analysis of the European maritime network. AIS data is often hard-to-handle for network-analytical purposes, as well explained by Alessandrini et al. (2017) when applying clustering techniques to Mediterranean ports. In addition, the directed inter-port graph was condensed into an undirected one to facilitate the analysis of the network.

The remainder of this chapter is organized as follows. The next section is a network-analytical approach to those flows by measuring global and local connectivity indices. It compares our results with those obtained in earlier works and conveys new findings about Europe's network structure in terms of centralization, hierarchy, and multiplexity. Thirdly, we map inter-port flows from the perspective of graph visualization on the one hand and flowmapping on the other to better understand the role of space and geography in shaping such flows. We discuss our results in the light of parent and broader studies at the crossroads of maritime transport and network analysis.

2 The topological approach

2.1 Multilayered structure

The structural and evolutionary features of multilayered maritime networks have been studied already by physicists (Parshani et al., 2010) and more specifically in the case of maritime networks by Ducruet (2017). The latter work demonstrated that while the most diversified ports (nodes) systematically exhibit higher centrality, traffic volume, and interaction range, the shortest inter-port linkages (edges) are more likely to carry the widest variety of commodities as an effect of corridor concentration, port complementarity, and market proximity. Such results also demonstrated that container flows were much space-dependent upon previously existing flow types such as bulks since the largest ports often developed a wider traffic variety (see also Ducruet et al., 2010). Throughout network science, many coupled communication

networks were studied such as Internet and power grids, airline and container networks, etc. The main rationale of this type of approach is that two different networks or layers share certain features such as common nodes and links so that their merger creates a new composite network having specific properties (Parshani et al., 2010), as in the case of node aggregation (Tsiotas and Serafeim, 2018) when merging nodes based on shared administrative units or functional-economical settings. The correlation coefficient can be expressed as follows:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

(2)

We transformed the network into an undirected graph by merging the directional links serving two same ports into one single link and added their respective number of calls (e.g. Amsterdam-Rotterdam + Rotterdam-Amsterdam) using an Excel macro, also as a means simplifying the graph by reducing the number of links and thus the statistical “noise”. Our results based on AIS data (Figure 3) exhibit a strong specialization of nodes more than links given their respective determination coefficient. Typical examples include Augusta (Italy), a Sardinian port specialized in liquid cargo, and on the other side container transshipment hubs such as Felixstowe, Piraeus, and Tilbury handling mainly dry cargo. The fact that node correlation remains lower than link correlation is revelatory of a significant gap with usual spatial networks where largest nodes embed within multiple layers. Here it is more the link level that exhibits such a trend, with 0.271 compared with 0.043 of determination coefficient, meaning that on average, and despite a low probability, links (edges) handling dry bulk cargo handle a comparable proportion of liquid cargo. But this is not the case for the majority of linkages, as domestic and heaviest links (e.g. Bremerhaven-Hamburg; Barcelona-Valencia) handle more cargo than international links such as Kiel-Rotterdam and Antwerp-Rotterdam. The determination coefficient is expressed as follows:

$$r^2 = 1 - \frac{\sum(y-y')^2}{\sum(y-\bar{y})^2}$$

$$= 1 \tag{3}$$

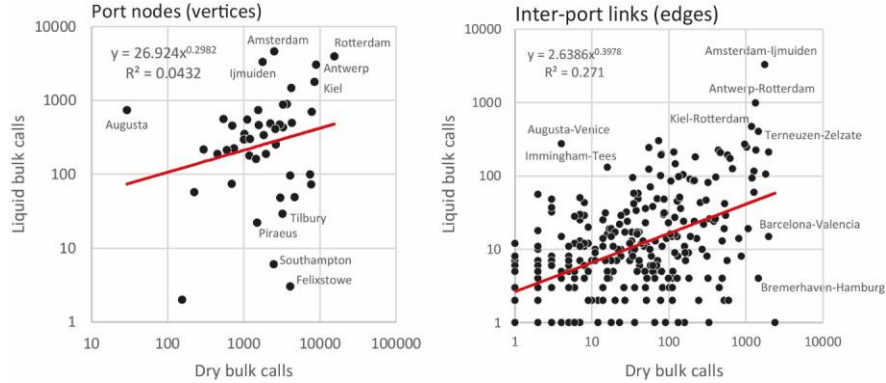


Fig. 3: Bilayered structure of the main European maritime network, 2017.

Source: own elaboration based on MarineTraffic data.

2.2 Interaction range

Based on a number of complementary features of complex spatial networks (Barthelemy, 2015), we investigate another dimension of European shipping flows to verify the influence of physical distance on the intensity of links (Table 1). As already demonstrated for airline networks (Guimera et al., 2005) and for liner shipping networks (Ducruet, 2017), there is a high statistical significance between node degree or weight and the distance to/from of its adjacent neighbors. Nautical distance was calculated using *GeoSeastems*, a Geographical Information System (GIS) dedicated to maritime flows (see Bunel et al., 2017).

Our results do not show, however, significant correlation between traffic size and average length of links, which stands in contrast with existing research. In the case of links, however, one may notice a decreasing average length of links opposite to the traffic ranking order. This corroborates the findings of Ducruet (2017) on the global shipping network. However, for nodes, previous findings do not apply for several possible reasons: sample size, intra-regional focus, etc. In theory, Rotterdam would be the largest and then the farthest-reaching European port, but the study focusing on Europe only blurs its global role. Yet, the fact that maximum average distance decays along with increasing traffic size in a very regular manner demonstrates that also in accordance with Ducruet (2017), stronger interactions occur over shorter distances (see Tobler, 1970 for a discussion about related fundamental issues and the *First Law of Geography* whereby near things are more closely related than distant things).

When it comes to the main traffic type (Table 1), different results were obtained according to the measurement unit. For nodes (ports), the specialization trend follows the global trend by which cargo types are not determined by heavier links, except for the maximal value for dry cargo for instance. This would mean that ports handling

more dry cargo are more likely to be larger on average, all commodities considered, while this is not the case for liquid bulks. On the contrary, heavier weight links always carry heavier traffic whatever its nature, but this is truer on average than according to maximal values. The top European maritime network is thus multifaceted with a mix of specialization and diversification forces at stake.

Table 1: Average and maximum nautical distance at nodes and links, 2017.

Network	Traffic class	Nodes		Links	
		Average	Maximum	Average	Maximum
Whole network	1	1,481	4,882	3,166	9,335
	2	2,851	3,493	3,029	7,381
	3	2,226	3,137	3,213	8,052
	4	1,761	2,625	2,118	6,914
	5	1,983	2,593	1,451	6,981
	All	2,062	4,882	2,272	9,335
Dry cargo network	1	1,713	4,946	3,449	9,335
	2	2,618	3,679	3,223	7,381
	3	2,157	3,303	3,170	8,052
	4	1,884	3,223	2,124	6,388
	5	2,024	2,580	1,481	6,981
	All	2,078	4,946	2,272	9,335
Liquid cargo network	1	1,608	7,381	2,847	7,381
	2	1,648	6,973	2,854	6,973
	3	2,506	6,423	3,617	5,666
	4	2,376	6,696	2,114	6,914
	5	1,772	6,947	1,394	6,947
	All	1,976	7,381	2,272	9,335

Source: own elaboration based on MarineTraffic data and Gephi software (Bastian et al., 2009).

Lastly, a “proximity effect” based on size is worth investigating here, namely the rich-club effect or “rich get richer” phenomenon in a network or other entity. The rich-club coefficient which corresponds to the density (*Gamma index*) of the “top network” divided by the density (*Gamma index*) of the whole networks is superior to “1”. The Gamma index refers to the share of observed links in the maximum possible number of links in the network, as refers to as follows for non-planar graphs:

$$\gamma = \frac{e}{\frac{u(u-1)}{2}} \quad (4)$$

Our results demonstrate that ports which average or mean degree (i.e. around 46) are more tightly connected with each other than the bottom of the (still upper) hierarchy, while the whole network of the top 50 ports has a rich-club coefficient around 1 (0.93). In comparison, ports with higher highest degree (i.e. port nodes) have a rich-club coefficient of 2.283, i.e. nearly three more time more connected than the whole network, and smaller but still large ports enjoying only 1.15 of rich-club

coefficient, nearly equal to the network's average. Such an imbalance between large and small ports even within the top 50 European ports motivate for a deepening of the analysis.

2.3 Graph connectivity

Other global-level analyses help to understand the network's structure. One of them is *assortativity* based on the average degree centrality of nodes' direct neighbors. The entire graph can be said *disassortative* as the correlation coefficient between node degree and average neighbor degree is negative (-0.178) but with low significance. It means that on average, larger nodes tend to connect smaller nodes and vice-versa, thereby suggesting a hierarchical structure. The same phenomenon applies to the two subnetworks of dry cargo (-0.154) and liquid cargo (-0.150) though with slightly less significance. Nevertheless, the fact that the significance is slightly higher for the whole network means that hierarchical tendencies are stronger in the composite network, which is in accordance with theory (Vespignani, 2010). The latter demonstrated that when its layers or subnetworks connect each other through mainly the same nodes, the whole network becomes both more centralized but more vulnerable to targeted attacks. Assortativity relates to the above formula:

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2} \quad (5)$$

Where j_i, k_i are the degrees of the vertices at the ends of the i th edge, with $i = 1 \dots M$ (M edges).

And the average neighbor degree:

$$k_{nn,i} = \frac{1}{k_i} \sum_j a_{ij} k_j \quad (6)$$

Another possibility to test the hierarchical dimension of the network is to look at the latter's degree distribution (Figure 4a). We then plot the degree centrality of nodes (x axis) against the cumulated number of nodes (y axis) in a log-log diagram and insert a power-law function to obtain the slope exponent of the line. Interestingly, each of the networks (whole, dry, and liquid) exhibit scale-free properties as the exponent value ranges between 1 and 3 (Barabasi and Albert, 1999). This means that despite the absence of small and medium-sized ports in the studies sample, a few larger ports dominate smaller ones. In addition, we observe that although values are close among the three networks, the dry cargo network is the most hierarchical with a value of 1.3278 compared with 1.277 for the whole network and only 1.14 for the liquid cargo network. This result is in accordance with the nature of the flows, as dry cargo includes containers, i.e. the most hierarchical configuration among all types of maritime transport. Container shipping is highly selective and had become

since the mid-1990s strongly marked by economies of scale, route rationalization, and port concentration around large nodes performing transshipment activities of various forms such as hub-and-spokes and/or interlining (Cullinane and Khana, 2000; Notteboom, 2004). The slope exponent of power-law lines can be expressed as follows:

$$\alpha = 1 + N \left[\sum_i \ln \frac{k_i}{k_{\min} - \frac{1}{2}} \right]^{-1} \quad (7)$$

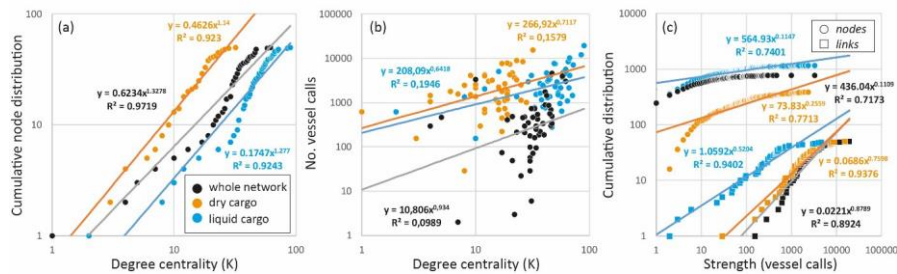


Fig. 4: Statistical distribution of the European maritime network, 2017.

Source: own elaboration based on MarineTraffic data and TULIP software.

We also tested the relationship between traffic size and degree centrality (Figure 4b) but it is not in accordance with already existing works such as Hu and Zhu (2009) in their study of the worldwide container shipping network. The whole European network exhibits zero correlation, certainly due to the aforementioned outliers in North Europe, while the relationship gets more significant when considering each layer independently. Regardless of degree centrality (K), the statistical distribution of links and nodes weights based on vessel calls (Figure 4c) exhibits power-law distribution only at node level for the whole network and the liquid cargo layer. All other slope exponents fall under 1 so that their topological structure remains hierarchically moderate (for an empirical validation of statistical laws obeying maritime networks, see Gastner and Ducret, 2015). Because nodes centralize flows and are fixed in space in non-planar spatial networks, hierarchical effects are higher than for links, which are more diversified in their morphology and distribution as they somewhat overcome pure physical and topological distances.

Last but not least, the extent to which the main European maritime network also exhibits small-world properties has been tested in Table 2. Small-world networks have defined by Watts and Strogatz (1998) as graphs with a relatively higher average

clustering coefficient and lower average shortest path length (ASPL) than a random network of identical size. Degree distribution in random graphs is known to follow a normal (i.e. Gaussian) distribution, contrary to scale-free and small-world networks. As a matter of fact, these conditions are fulfilled in our case study, because inter-port links are not randomly distributed but correspond, as said above for the scale-free dimension, to the patterns of a particular industry (shipping) spatial and economic features are most influential. The maritime network is thus easier to navigate (lower diameter and ASPL) and more tightly connected (higher density and average clustering coefficient). The average number of links per node (average degree centrality) is much higher than in the random network, also underlining the unevenness and hierarchical structure of the maritime network. The fact that the dry cargo layer exhibits much lower clustering coefficient and density is also in accordance with our previous results showing the stronger scale-free dimension of this particular layer. The clustering coefficient is calculated as follows:

$$C_v = \frac{N_v}{\frac{d_v(d_v-1)}{2}} \quad (8)$$

Table 2: Topological properties of the European maritime network vs. a random network.

Network	Nodes	Links	Average clustering coefficient	Diameter	Average shortest path length	Average degree centrality	Density (Gamma index)
Whole network	50	1165	0.8	3	1.366	46.6	0.951
Dry cargo network	50	387	0.401	4	1.781	15.48	0.316
Liquid cargo network	50	778	0.842	3	1.526	31.12	0.635
Random graph	50	62	0.039	11	4.029	1.24	0.05

Source: own elaboration based on MarineTraffic data and Gephi software (Bastian et al., 2009).

Lastly, European maritime connectivity may be envisaged with reference to Hu and Zhu (2009) by plotting node-level variables and check their compatibility. The outcome is often a good manner to conclude to specific features of networks in general. As shown in Figure 5a, the clustering coefficient is inversely proportional to the degree centrality of nodes, since large-degree nodes (many links) often act as hubs. The clustering coefficient is defined, for each node, by the average number of closed triangles (or cliques) it connects among its directly connected neighbors. It is called transitivity in sociology (Ducruet and Beauguitte, 2014). The lower its value, the more likely is the port to act as a hub among its neighbors; conversely, high clustering coefficients mean that such nodes belong to densely connected regions or communities with less hierarchy. While the inverse relationship is verified for the liquid cargo network (i.e. determination coefficient of 0.576 and Pearson correlation

coefficient of -0.759), the whole network and the dry cargo network exhibit nearly zero correlation. This is rather surprising given the stronger hierarchical structure of the dry cargo network. Results of Figure 5b based on nearest neighbors' degree do not exhibit specific trends, as in the work of Hu and Zhu (2009), because weights have not been taken into account. Otherwise, Hu and Zhu (2009) demonstrated that this correlation should be meaningful as a proof of the disassortative behavior of the network under study. Last but not least, and this time much more meaningfully than in the work of Hu and Zhu (2009), the power-law relationships between betweenness centrality and degree centrality (Figure 5c) is by far the strongest, with slope exponents over 0.2 for all networks, compared with 1.66 for Hu and Zhu (2009). This means that ports having a wider set of connections (K) are often the same nodes being at the crossroads of maritime routes within Europe.

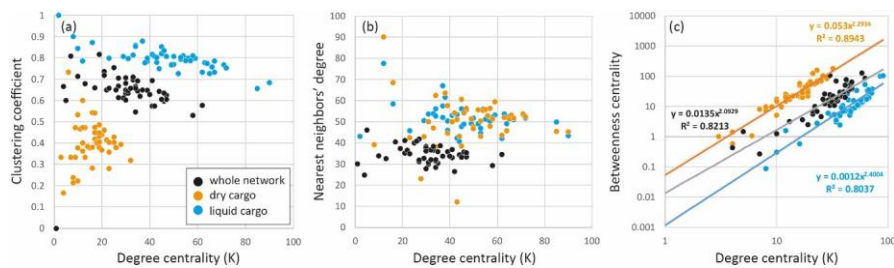


Fig. 5: Centrality correlations of the European maritime network, 2017.

Source: own elaboration based on MarineTraffic data and TULIP software.

3 The flowmapping approach

3.1 Substructures within the European maritime network

The variety of algorithms which serve to reveal hidden information from a network is extremely wide and numerous. This paper picked up two of the most common ones with the goal to compare two distinct approaches based on the same dataset. One first algorithm relates to the well-known *traveling salesman problem* in mathematics (and hence, graph theory) whereby a given person should optimize its route in terms of cost and/or time to best serve its customers for pick-up and delivery. The Kruskal algorithm creates a so-called *minimum spanning tree* (MST), i.e. the aforementioned least cost/time route; the most efficient “line” connecting all nodes according to the desired criterion. Our reasoning had been to use the inverse quantity of flows (i.e. vessel calls) per link as a base for the application of this algorithm, to reveal which subnetwork or optimal route actually carries the largest traffic among all possible paths (Figure 6, left). As we obtain a tree, i.e. a graph without loops (or

cliques) it was preferred to use the Strahler number (often noted U) to measure the branching complexity of the obtained subnetwork and identify its main junctions. Such a measure was first used in hydrology in the 1940s to define stream size based on a hierarchy of tributaries from the source to the sinks. We added betweenness centrality to better reflect the situation of nodes in this new network, as defined by the number of occurrences of each node on all possible shortest paths throughout the graph. While the diameter is the longest of all shortest paths, ASPL refers to the following formula:

$$l_G = \frac{1}{n(n-1)} \sum_{i,j} d(v_i, v_j) \quad (9)$$

Betweenness centrality for the following formula:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (10)$$

And degree centrality the following one:

$$\sum_{k=0}^{\infty} A_R^k \beta^k \quad (11)$$

The second method is a partitioning algorithm (Figure 6, right) called single linkage analysis (SLA). Its philosophy is rather simple as it retains for each node its heaviest traffic link to another node. This also results into a tree-like structure made often split into connected components (or *nodal regions*), each of them being more or less centered upon one main node, as initially applied in geography by Nystuen and Dacey (1961). Degree centrality is then used to show the connectivity of port nodes. As in the next section, the underlying hypothesis of such analyses is that space matters, should it be Euclidian distance, coastal morphology, nautical distance, or other geographic feature, in the distribution of maritime networks.

When it comes to the MST⁴, a very interesting result is the higher betweenness centrality of an Atlantic arc between Rotterdam (the most central node) and Bilbao, including Le Havre and Sines. This segment is crucial in terms of centrality as it connects North and South Europe. This strategic importance is confirmed by the highest Strahler number of Bilbao, which lies in between those two regions. Bilbao acts a source of all traffic that irrigates both regions. Of course, this is not the case in reality as North-South maritime traffic is actually part of a wider circulation system connecting Europe with Asia and North America, the round-the-world trunk line, which does not influence this intra-European analysis. Despite the merits of Atlantic ports such as Le Havre (the first and last port of call of the Northern range) and Sines (where the global stevedore Port of Singapore invested to turn it into a new

⁴ There is no specific formula as the route equals the minimum value of all links connecting a single path.

transshipment hub and gateway in 2004 and recently became one of the leading South European ports), those two nodes are often bypassed by the giant containerships for instance on the Europe-Asia route, as well as Bilbao. Although the geographic dimension of these results are somewhat fuzzy outside this main segment, one can observe the specific importance of certain hubs such as Hamburg and Marsaxlokk but also large gateways such as Antwerp, Amsterdam, and Bremerhaven. South European ports are thus clearly disadvantaged in terms of centrality and seem rather peripheral in this optimal system of flows.

Results obtained from SLA provide complementary evidence on the role of the geographic factors influencing network and node connectivity. As such, there is an overwhelming importance of spatial proximity in the emergence of nodal regions. Except from the one dominated by Rotterdam, Europe's largest port, which extends to the Iberian Peninsula, all nodal regions are confined in the vicinity of the connected ports, a typical example being the nexus (or *dyad*) Las Palmas-Tenerife, but are not necessarily bound to the national level, as seen with Piraeus-Marsaxlokk, Genoa-Fos, Venice-Constanta, Dublin-Liverpool, Kiel-Klaipeda, and Zeebrugge-Goteborg. Nevertheless, such results confirm that even in the case of supposedly footloose flows and networks, distance plays a very important role in the configuration of maritime networks. This also corresponds to an economic and operational reality because hub-and-spokes patterns are based on concentration but also high sailing frequency between the hub and the spokes, which is not possible everywhere the same. For less hierarchical substructures, physical proximity is the main explanatory factor and also given the nature of data, as AIS signals trace "imaginary routes" and corridors with strong spatial friction, contrary to other maritime traffic data such as origin-destination flows regardless of the true morphology of maritime routes.

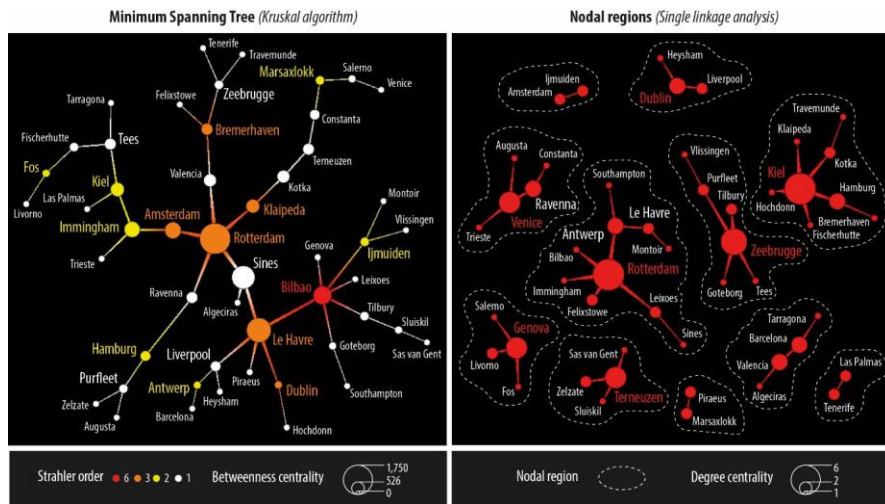


Fig. 6: Substructures within the main European maritime network, 2017.

Source: own elaboration based on MarineTraffic data and TULIP software.

3.2 Cartography of the network

To render sailing vessels across the ocean, we constructed a maritime grid representing every potential routes one vessel may use (see Bunel et al., 2017 for further details). This grid had to be a wise compromise between lightweight and precision to depict maritime paths as close to reality as possible whilst keeping a small amount of links and nodes, thus allowing quick network-specific calculations. Achieving

that goal allows us to map and analyse shipping flows extracted from any sources whatsoever. We had to develop an automated method that creates a worldwide maritime grid connecting every ports from one another. To begin with, the world was divided into eight squares of equal dimension; then, a process subdivided each mesh iteratively and kept on dividing remaining squares as long as they were intersecting both the ocean and continents. It resulted in higher mesh density along the coastline, thus easing the future connection of ports. We decided to limit the algorithm to seven iterations as it seemed to bear the best amount of details and the least expected amount of links. In the end, we obtained 23,000 squares covering the world ocean. Each square's centroid, — i.e. the centre of gravity — was calculated and represents every nodes of the maritime grid. Those centroids were linked from one another using the Moore neighbourhood and weighted according to their length. As ports needed to be connected to the grid, we decided to group them by clusters based on their geographic proximity; the centroid of every cluster of points is calculated and linked to the maritime grid; every ports, except isolated ones, are linked to their respective cluster, consequently offering us a fully functional network upon which we can calculate shortest paths between pairs of ports using Dijkstra's algorithm. Finally, one last enhancement of the grid was to include main rivers to link fluvial ports to the maritime network and thus ensure the network's completeness. The following illustration (Figure 7) depicts the European maritime grid which is used to project the AIS port-to-port matrix; it includes every potential path a vessel can use to sail across the sea and reach its port of arrival. Clusters as well as links between ports and the main grid are not represented as it would conceal the main network and impact overall readability.

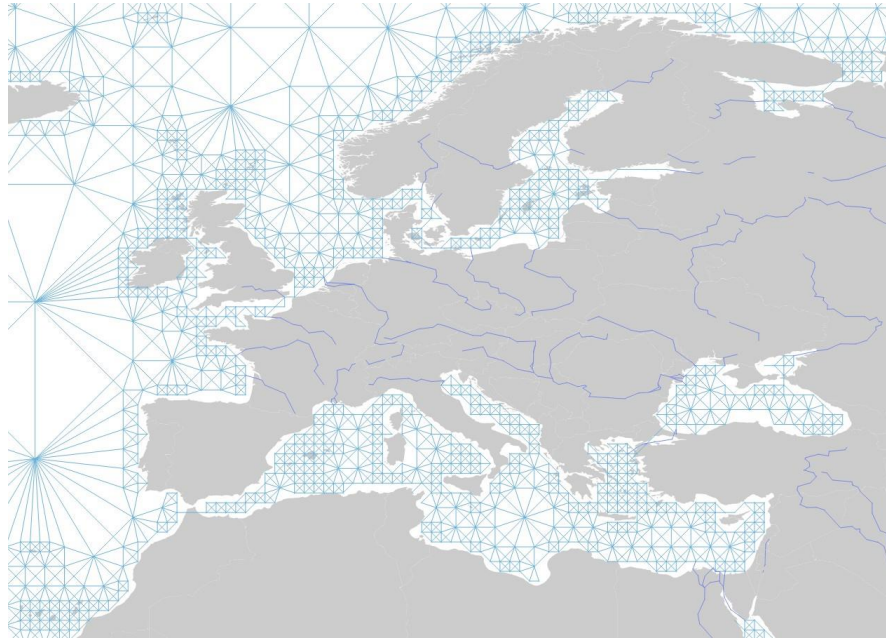


Fig. 7: The European maritime grid.

Source: own elaboration based on GeoSeastems.

Visualizing the network for a given date or period of time requires to calculate shortest paths between each pair of ports supplied by the analyzed matrix. To facilitate this workflow, we developed a web platform that enable automatic calculations of routes throughout the maritime grid at a user-specified date or period; furthermore, we can also calculate topological graph indexes such as betweenness and closeness centrality, eccentricity, etc. Data are stored in PostgreSQL/PostGIS database to access

powerful spatial indexes and significantly enhancing time efficiency.

To create the following flow map (Figure 8), several scripts were involved to obtain the number of ship per arcs and per nodes, as well as the percentage of containers. The results were calculated through the use of SQL queries wrapped inside a Python script, deployed mainly for automation purposes. The workflow is divided in two parts: the first calculates shortest paths between each pair of ports of the matrix and effectively represents every segment of the grid actually used by vessels; the second part aggregates the amount of vessels as well as the amount of containers for every segment of the grid by summing those values. Finally, we can calculate the percentage of containers and tankers for each arc.

The generated map displays two information: arcs width and points size are directly proportional to ship traffic — i.e. the number of vessel using this network section and the number of vessel arriving or departing from a port — as a classic flow map; the color gradation indicates the fraction of containers, and consequently the fraction of tankers as well, composing the ship traffic. The importance of containers traffic (69,402) in comparison with tankers traffic (15,620) which is approximatively 4.5 times higher, has been taken into consideration to construct classes. A map insert is necessary here to have a better overview of Dutch ports even though the grid has been designed for global analysis as shown by the coarse resolution of links.

The cartography of main European maritime traffic flows mainly concentrates in North Europe, especially along the well-known “Le Havre – Hamburg” range, as the “hot spot” of a longer corridor between Kiel and the English Channel, including main Benelux ports such as Rotterdam, Antwerp, Amsterdam, and German ports such as Bremerhaven, Hamburg, and Kiel. To corroborate the previous assertion, proximity between ports seems to notably influence maritime transportation. That phenomenon is obvious when we consider some pair of ports such as Dublin-Liverpool, Valencia-Barcelona, Genoa-Livorno, Las Palmas-Tenerife, that indicates substantial connections between nearby ports and confirms geographic proximity as an important factor in the maritime shipping network. Containers appears to be the dominant shipping method deployed between those local/regional transport routes while tankers appear to gain importance on long-distance transportation. It is revealed by the main corridor north/south sailing along the western coasts of Europe which indicates a more balanced fraction of tankers than in nearby-ports transportations. Only two Portuguese ports, Sines and Leixoes, seems to go against this statement as they appear to be focusing on tankers rather than containers. Despite its recent growth as container hub and gateway, Sines remains mostly a petrochemical complex. From the Gibraltar Straits and Algeciras in particular, traffic volume on the main route regularly decreases up to Marsaxlokk, Piraeus, and Constanta. Yet, one can also observe the South European range between Valencia and Livorno (Leghorn) as another important port system made of “multi-port gateway regions” (Notteboom, 2010), especially for containers.

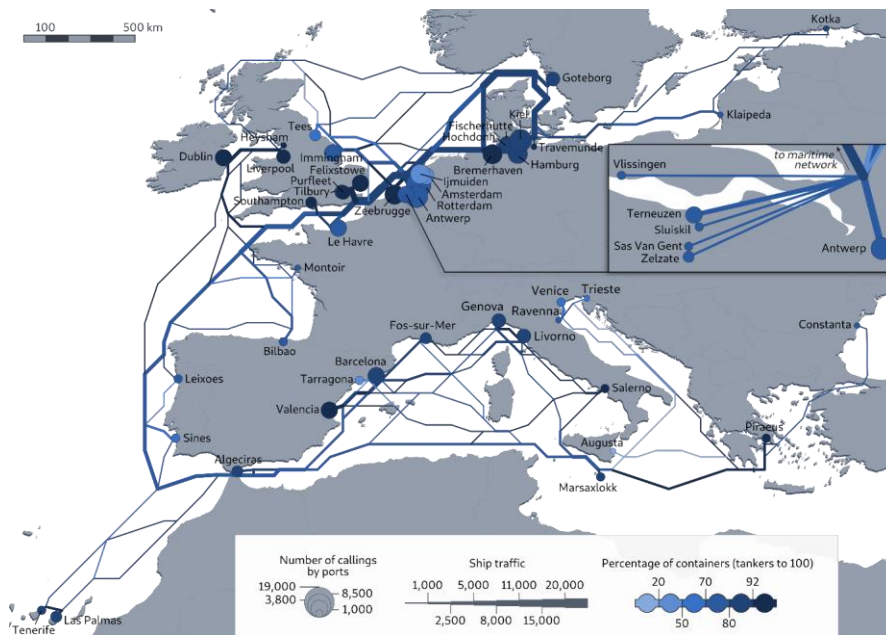


Fig. 8: Cartography of the main European maritime network, 2017.

Source: own elaboration based on MarineTraffic data and GeoSeastems.

Another interesting learning from this cartography is the importance of smaller but yet top European port regions such as North Adriatic, Black Sea, East Baltic, and most of all, the British Isles. The latter faced important challenges keeping their position vis-à-vis mainland competitors but recent policies indicate proactive efforts (e.g. the Liverpool mega-port project) although the relative importance of British and Irish ports is artificially due to shortsea shipping flows and ro-ro (which is counted for the cargo part in the “dry bulk” category), which also applies to Zeebrugge. Last but not least, we observe that port size and port connectivity are not necessarily in direct relationship with traffic specialization, as Northern Adriatic port handle more liquid bulks, such as Amsterdam and Sines (a former petro-chemical complex), just like Tees and Immingham, being the main ports of industrially declining economic bases (Ducruet et al., 2010). This is also the case for large dry cargo ports with advanced actions in attracting transshipment (Valencia, Marsaxlokk), environmental progress (Goteborg), acting as gateways of a global city like London (Tilbury, Felixstowe), or specializing in particular commodities such as automobiles (e.g. Bremerhaven), being the focus of global terminal handling operators like COSCO (c.f. Antwerp) or CMG-CGM (c.f. Marsaxlokk), or multipurpose-ports doing both dry and bulk cargoes such as Le Havre, Rotterdam, and Antwerp? Is commodity and/or functional specialization or diversity a desired policy and vision for European port actors?

The share of containers is also an interesting feature as it reveals some regional specializations such as the Italian ports located in the northern Adriatic Sea and north-eastern England which seems to focus on tankers rather than containers. It brings to light a specialized corridor between Tees/Immingham and Amsterdam/Ijmuiden. On the other hand, Venice, Trieste and Ravenna appears to focus on longer tankers transportation towards northern Europe.

4 Conclusion

This paper investigated the topological, statistical, and spatial structure of the main European maritime network. Based on AIS data transformed into an origin-destination matrix, conventional and advanced graph-theoretical and complex network methods applied to such flows revealed a number of stylized facts and confirmed at least partly earlier research on maritime network analysis. As such, the current European maritime network shares many similarities with other networks in terms of scale-free / small-world dimension, centralization, multilayered structure, subnetworks, and so on. Combined with geomatics and cartography (GIS), this paper also revealed the importance of physical distance on the distribution of flows and their specialization according to the two main traffic categories retained, dry cargo and liquid cargo.

Further research shall investigate in depth the applicability of the homophily concept, borrowed from sociology, to transport networks such as the maritime network, to verify how nodes of similar profile (here commodity specialization) connect more density with each other than with nodes of different profile. But this is only one of the possible further research pathways this paper would like to open. Expanding from the European scale to the global scale and compare regional connectivity differences (see Arvis et al., 2018 for a recent focus on the Mediterranean basin); refining algorithms and network indices to better adjust them to the reality of shipping and make academic research closer to industry needs; expand the maritime analysis to hinterlands and territorial/inland areas that depend directly or indirectly on maritime transport for connecting global markets; better convey such findings towards the related European Commission experts in charge of a more balanced and durable continent. Other research pathways are underway such as the integration between maritime networks and land-based networks (Berli et al., 2018) to better understand the intermodal importance of certain nodes as well as the maritime (albeit indirect) centrality of non-port, inland cities in Europe and the rest of the world.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. [313847] "World Seastems".

Appendix

Table 3: Top 27 European ports in 2017 according to MarineTraffic and Eurostat data. *N.B. shaded cells correspond to ports appearing in both sources*

Rank	MarineTraffic (incoming calls)	Eurostat (000s tons)
1	ROTTERDAM 9674	ROTTERDAM 431944
2	ANTWERP 5577	ANTWERP 198691
3	KIEL 3965	HAMBURG 120323
4	HAMBURG 3612	AMSTERDAM 96343
5	ZEEBRUGGE 3511	ALGECIRAS 83422
6	BREMERHAVEN 3493	FOS SUR MER 76427
7	IMMINGHAM 2756	LE HAVRE 60033
8	AMSTERDAM 2553	VALENCIA 58321
9	DUBLIN 2376	IMMINGHAM 54403
10	VALENCIA 2304	BREMERHAVEN 52294
11	BARCELONA 2146	LONDON 50380
12	LE HAVRE 2128	TRIESTE 49311
13	FELIXSTOWE 1887	SINES 48056
14	GENOVA 1682	GENOVA 45049
15	GOTEBORG 1614	BERGEN 44654
16	LIVERPOOL 1613	PIRAEUS 41021
17	LIVORNO 1540	GOTEBORG 40974
18	SOUTHAMPTON 1391	BARCELONA 39142
19	LAS PALMAS 1219	CONSTANTA 37521
20	TEES 1172	DUNKERQUE 36864
21	TERNEUZEN 1117	SOUTHAMPTON 36046
22	ALGECIRAS 1029	RIGA 35822
23	FOS SUR MER 1009	MILFORD HAVEN 34768
24	SINES 890	TEES 26873
25	PIRAEUS 825	WILHELMSHAVEN 24474
26	TENERIFE 802	TARANTO 20982
27	MARSAXLOKK 757	TALLINN 19937

References

1. Alessandrini A., Fernandez-Arguedas V., Vespe M. (2017) Vessel tracking data usage to map Mediterranean flows. In: Ducruet (Ed.), *Advances in Shipping Data Analysis and Modeling*.

- Tracking and Mapping Maritime Flows in the Age of Big Data. London and New York: Routledge, pp. 173-187.
2. Ansorena I.L. (2018) Bilateral connectivity in the liner shipping network: An overview. *World Review of Intermodal Transportation Research*, 7(4): 295-309. Arvis J.F., Vesin V., Caruthers R., Ducruet C., De Langen P.W. (2018) Maritime networks, port efficiency and hinterland connectivity in the Mediterranean. World Bank Group, International Development in Focus, <http://documents.worldbank.org/curated/en/508771540319329808/Maritime-Networks-Port-Efficiency-and-Hinterland-Connectivity-in-the-Mediterranean>
 3. Auber D. (2004) Tulip: A huge graph visualisation framework. In: Mutzel P., Junger M. (Eds.), *Graph Drawing Software*, Berlin & Heidelberg: Springer, Mathematics and Visualization, pp.105-126.
 4. Barabasi A.L., Albert R. (1999) Emergence of scaling in random networks. *Science*, 286(5439): 509-512. Barthelemy M. (2015) Spatial networks: Tools and perspectives. In: Ducruet C. (Ed.), *Maritime Networks: Spatial Structures and Time Dynamics*, London and New York: Routledge, pp. 50-60.
 5. Bastian M., Heymann S., Jacomy M. (2009) Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*. Berli J., Bunel, M., Ducruet C. (2018) Sea-land interdependence in the global maritime network: The case of Australian port cities. *Networks and Spatial Economics*, Online First, <https://link.springer.com/article/10.1007/s11067-018-9403-4>
 6. Bunel M., Bahoken F., Ducruet C., Lagesse C., Marnot B., Mermet E., Petit S. (2017) Geovisualizing the sail-to-steam transition through vessel movement data. In: Ducruet (Ed.), *Advances in Shipping Data Analysis and Modeling. Tracking and Mapping Maritime Flows in the Age of Big Data*. London and New York: Routledge, pp. 189-205.
 7. Cullinane K.P.B., Khanna M. (2000) Economies of scale in large containerhips: Optimal size and geographical implications. *Journal of Transport Geography*, 8(3): 181-195.
 8. Ducruet C. (2015) Maritime flows and networks in a multidisciplinary perspective. In: Ducruet C. (Ed.), *Maritime Networks: Spatial Structures and Time Dynamics*, London and New York: Routledge, pp. 3-26.
 9. Ducruet C. (2017) Multilayer dynamics of complex spatial networks: the case of global maritime flows (1977-2008). *Journal of Transport Geography*, 60: 47-58.
 10. Ducruet C., Beauguitte L. (2014) Network science and spatial science: Review and outcomes of a complex relationship. *Networks and Spatial Economics*, 14(3-4): 297-316.
 11. Ducruet C., Koster H.R.A., Van der Beek D.J. (2010) Commodity variety and seaport performance. *Regional Studies*, 44(9): 1221-1240.
 12. Ducruet C., Lugo I. (2013) Structure and dynamics of transportation networks: Models, concepts, and applications. In: Rodrigue J.P., Notteboom T.E., Shaw J. (Eds.), *The SAGE Handbook of Transport Studies*, SAGE Publications, pp. 347-364.
 13. Etienne L., Alincourt E., Devogele T. (2015) Maritime network monitoring. From position sensors to shipping patterns. In: Ducruet (Ed.), *Maritime Networks. Spatial Structures and Time Dynamics*. London and New York: Routledge, pp. 190-209.
 14. Eurostat (2018) Database. Brussels: European Commission, <https://ec.europa.eu/eurostat/data/database> [Accessed October 2018]
 15. Garcia-Herrera R., Gallego D., Barriopedro D., Mellado J. (2017) Ship logbooks help to understand climate vulnerability. In: Ducruet (Ed.), *Advances in Shipping Data Analysis and Modeling. Tracking and Mapping Maritime Flows in the Age of Big Data*. London and New York: Routledge, pp. 37-51.
 16. Gastner M.T., Ducruet C. (2015) The distribution functions of vessel calls and port connectivity in the global cargo ship network. In: Ducruet C. (Ed.), *Maritime Networks: Spatial Structures and Time Dynamics*, London and New York: Routledge, pp. 242-261.
 17. Guerrero D., Gonzalez-Laxe F.I., Freire-Seoane M.J., Pais Montes C. (2017) Foreland mix and inland accessibility of European NUTS-3 regions. In: Ducruet (Ed.), *Advances in Shipping Data Analysis and Modeling. Tracking and Mapping Maritime Flows in the Age of Big Data*. London and New York: Routledge, pp. 207-229.

18. Guimera R., Mossa S., Turtschi A., Amaral L.A. (2005) The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences USA*, 102(22): 7794-7799.
19. Hu Y., Zhu D. (2009) Empirical analysis of the worldwide maritime transportation network. *Physica A*, 388(10): 2061-2071.
20. Kaluza P., Kölsch A., Gastner M.T., Blasius B. (2010) The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48): 1093-1103.
21. Le Guyader D., Ray C., Brosset D. (2017) Identifying small-scale fishing zones in France using AIS data. In: Ducruet (Ed.), *Advances in Shipping Data Analysis and Modeling. Tracking and Mapping Maritime Flows in the Age of Big Data*. London and New York: Routledge, pp. 251-262.
22. Mobasher A., Zipf A., Ducruet C. (2017) Spatio-temporal modeling of ship-to-ship interactions in mid-19th century high seas. In: Ducruet (Ed.), *Advances in Shipping Data Analysis and Modeling. Tracking and Mapping Maritime Flows in the Age of Big Data*. London and New York: Routledge, pp. 263-272.
23. Notteboom T.E. (2004) Container shipping and ports: An overview. *Review of Network Economics*, 3(2): 86-106.
24. Notteboom T.E. (2010) Concentration and the formation of multi-port gateway regions in the European container port system: an update. *Journal of Transport Geography*, 18(4): 567-583.
25. Nystuen J.D., Dacey M.F. (1961) A graph theory interpretation of nodal regions. *Papers in Regional Science*, 7(1): 29-42.
26. Parshani R., Rozenblat C., Ietri D., Ducruet C., Havlin S. (2010) Inter-similarity between coupled networks. *Europhysics Letters (EPL)*, 92: 68002.
27. Rodrigue J.P., Ducruet C. (2018) Graph theory: definition and properties. *The Geography of Transport Systems*, Paper 10, https://transportgeography.org/?page_id=5976 [accessed October 2018].
28. Rozenblat C., Mélançon G. (Eds.), *Methods for Multilevel Analysis and Visualization of Geographical Networks*, Springer Geography.
29. Siegfried A. (1940) *Suez, Panama et les routes maritimes mondiales*. Paris: Armand Colin.
30. Stergiopoulos G., Valvis E., Mitrodimas D., Lekkas D., Gritzalis D. (2018) Analyzing congestion interdependencies of ports and container ship routes in the maritime network infrastructure. *IEEE ACCESS*, doi: 10.1109/ACCESS.2018
31. Tobler W. (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46: 234-240.
32. Tsiotas T., Serafeim P. (2018) Effects in the network topology due to node aggregation: Empirical evidence from the domestic maritime transportation in Greece. *Physica A*: 491(1): 71-88.
33. Vespignani A. (2010) Complex networks: The fragility of interdependency. *Nature*, 464(7291): 984-985.
34. Watts D.J., Strogatz S.H. (1998) Collective dynamics of "small-world" networks. *Nature*, 393: 440-442.
35. Wessa P. (2018) *Free Statistics Software*, Office for Research Development and Education, version 1.2.1, <https://www.wessa.net/> [Accessed October 2018]