ORIGINAL ARTICLE

Open Access

CrossMark

Evolving structure of the maritime trade network: evidence from the Lloyd's Shipping Index (1890–2000)

Zuzanna Kosowska-Stamirowska[1*] (iD), César Ducruet[1] and Nishant Rai[2]

*Correspondence:
zuzanna.stamirowska@iscpif.fr
[1] Centre National de la Recherche Scientifique (CNRS), UMR 8504 Géographie-cités, Université Paris 1 Panthéon-Sorbonne, 13 rue du Four, F-75006 Paris, France
Full list of author information is available at the end of the article

Abstract

Over 90 % of the world trade volumes is being carried by sea nowadays. This figure shows the massive importance of the maritime trade routes for the world economy. However, the evolution of their structure over time is a white spot in the modern literature. In this paper we characterise and study topological changes of the maritime trade network and how they translate into navigability properties of this network. In order to do so we use tools from Graph Theory and Computer Science to describe the maritime trade network at different points in time between 1890 and 2000, based on the data on daily movements of ships. We also propose two new measures of network navigability based on a random walk procedure: *random walk discovery* and *escape difficulty*. By studying the maritime network evolution we find that it optimizes over time, increasing its navigability while doubling the number of active ports. Our findings 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.

Keywords: Maritime trade, Network evolution, Network navigability, Graph theory

Introduction

The last decade had witnessed a surge in maritime flow visualization and maritime network analysis, especially at the global scale. This stands in sharp contrast with the very few works on such themes produced along the previous century. In the 1940s already, world maps showed the precise geographic distribution of British vessels (Siegfried 1940) and of US maritime trade (Ullman 1949). But it is only in the late 1960s that geographers, claiming the need to include maritime linkages in the analysis of ports, port systems, and port hinterlands (Rimmer 2012), pioneered the application of Graph Theory to maritime transport (Robinson 1968), but on a more local scale. Graph Theory which had been so popular for the analysis of other transport systems (e.g. road, rail, river, air, and telecommunications), lost ground in the discipline. Maritime transport remained the focus of broad cartographies of volume and distribution of main routes until the late 1990s, when other geographers proposed to measure the topological structure of the global container shipping network (Joly 1999) and to analyze the global strategies of ocean carriers such as Maersk and CMA-CGM (Frémont 2015).

Springer Open

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 2 of 17

The explosion of computer capacity, the revival of the "science of networks", and the growing availability of maritime traffic data soon gave birth to numerous analyses of global maritime flows, which greatly varied in objectives and outcomes. Physicists for instance found rather natural to investigate the topological properties of the global maritime network, responsible for no less than 90 % of world trade volumes, but focusing primarily on container shipping (Deng et al. (2009); Doshi et al. (2012); Hu and Zhu (2009)). They stressed the belonging to the classes of scale-free and small-world networks using standard measures from the then buoyant research field of complex networks. Other contributions of the kind consisted in comparing the networks of different fleet types (Kaluza et al. 2010) to better understand marine bioinvasions, analyzing the inter-similarity of the container shipping and airline networks (Parshani et al. (2010); Woolley-Meza et al. (2012)), and constructing global port-to-port matrices to estimate the impact of various scenarios on flow distribution (Tavasszy et al. (2011); Wang et al. (2012)). Geographers also contributed to this dynamics by mapping the nodal regions and centrality of pivotal hub ports for container shipping (Ducruet and Notteboom (2012); Gonzalez-Laxe et al. (2012); Wang and Wang (2011)), general cargo (Pais Montes et al. 2012), and in the multiplex graph (Ducruet 2013). Li et al. ((2015)) as well as Xu et al. (2015) departed from the classical view of port nodes to analyze the evolution of a global container shipping network made of large regions. Most of the other contributions to the field consisted in analyses of local and regional maritime networks using similar methodologies (see Tovar et al. (2015) and Ducruet (2015) for a synthesis).

This rapid review of the field raises several questions that this research would like to tackle. First, most of the aforementioned contributions focused on container shipping, known to be the most valuable and modern segment of maritime transport, having gone through rapid growth and transformation of its network configurations since its emergence in the mid-1950s and especially, with the advent of mega-ships since the 2000s. If we exclude fully-fledged density maps done in recent years but only to address other issues without any reference to networks, such as environmental impacts (see for instance Halpern et al. (2008)), we find that other fleet types received much less attention from a network perspective, so that the global maritime network as a whole remains poorly studied. Second, and related to the first, the focus on container shipping motivated scholars to be the most up-to-date and therefore to analyze current topologies, namely the shape of the network from the late 1990s onward. The extent to which recent and current topologies differ from earlier ones thus cannot be discussed or demonstrated. This lacuna is surprising, given the efforts put on understanding, for instance, the impact of the container revolution on world trade between 1962 and 1990 (Bernhofen et al. 2016) and the numerous works on the impact of technological change on the port and shipping industry (see Guerrero and Rodrigue (2013); Kuby and Reid (1992); Mayer (1973)). Perhaps, the high costs related with data acquisition and encoding motivated scholars to offer a recent, static view of the network.

In this paper, a new and dynamic analysis is proposed to fill-in such gaps. The main question to be tackled in this paper is whether the global maritime network has topologically changed over time and how this changes translated into its navigability properties. Based on a largely untapped historical database on worldwide merchant vessel movements, we compare current and past states of network configuration, assuming that successive and major technological (and wider economic) changes affected the dimension

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 3 of 17

and architecture of the macro-system. This research contributes as well to the wider research field on spatial and complex networks where dynamical analyses remain rather rare, given the scarcity of accurate time series data, thus resulting in a dominance of simulation experiments over empirical analyses (see Barthelemy (2011); Boccaletti et al. (2006)). Especially in terms of the studies of real-world networks growth and densification, as to the best of the authors knowledge all the real-world networks studied so far (Leskovec et al. (2007), Strano (2012)) exhibit super-linear growth of the number of edges with respect to the number of nodes, which contradicts the widely-spread Preferential Attachment model proposed by Barabási and Albert ((1999)). Moreover, this paper enters the field of navigability studies of transportation networks (De Domenico et al. (2014), Gulyas 2015) by looking at network's efficiency from the point of view of ease of navigation.

Our methodology consists of analyses of an unweighted network created by the ports, standing for nodes, and passages of the ships, standing for the edges. Then we take snapshots of the network almost every 5 years between 1890 and 2000. By applying tools from Graph Theory and navigability algorithms, we find that the maritime network has doubled its size in terms of the number of active ports over the studied period, but the rate of growth of the number of edges and the declining clustering coefficient indicate that the maritime network doesn't necessarily become denser with time, contrary to the findings of Leskovec et al. (2007) for other real-world networks. Our findings indicate that we might observe a process of network optimization which is due to some processes specific to the maritime industry as well as to economic and technological development. The random walk measures which we construct and apply in this paper show that navigation in the maritime network becomes easier with time, that is, the network's structure starts to privilege more efficient movements and that with time it becomes easier to reach a given port starting from any other port of the network. Surprisingly, we find that the observed processes begin before the widespread of containerization.

The remainder of the paper is organized as follows. The second section presents the elaboration of the historical database after its extraction from archival documents and the network analytical tools to be applied to the resulting graph to best unravel dynamics of change. Main results are offered in the third section, ranging from the most common methods of complex network analysis to more advanced ones in relation to the evolving navigability of the network. The last section discusses the results and concludes about their usefulness to further understand the specificity of current port and maritime transport challenges.

Elaboration of a global historical database using the Lloyd's Shipping Index

Maritime flows of merchant vessels among ports of the world have been recorded by the maritime insurance company Lloyd's List since the late sixteenth century, focusing primarily on the British fleet but since 1890, on any other. An in-depth review of all the research works having used such a unique source of information concluded that it still remains unknown to most shipping specialists (including historians, geographers, and economists). Only a dozen references to Lloyd's List could be identified in the entire academic literature to date, mostly to retrieve the port calls of a given ship for genealogical purposes, to identify the location of shipwrecks for underwater archaeology, to count the vessel calls at a given port, and to measure the time gap between call date and publication

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 4 of 17

date to analyze the evolution of telecommunications (Ducruet et al. 2015). Given their main focus on container shipping, most studies of maritime networks rather use carrier schedule data provided by Containerisation International, Barry Rogliano Salles (Alphaliner database), or company websites, while others compile information on real-time positioning of ships such as Automated Identification System (AIS). Data from Lloyd's List appear to be the world's only possible source to map and analyze global maritime flows back in time, i.e. prior to containerization.

Since its origin, the Lloyd's Shipping Index reports on a daily or weekly basis the latest movement of each vessel between two or more ports, including dates of departure and arrival, tonnage capacity, operating company, flag, date of build, and additional comments in the case of damage, loss, or war event. The somewhat difficult readability of the older publications and the limitations of existing Optical Character Recognition (OCR) software forced us to concentrate our efforts on the extraction of vessel calls by port and inter-port link. The choice was made to extract one entire publication every five years or so between 1890 and 2000, a couple of years before the paper version ceased to exist. From 2009 onwards, such data is only available in expensive digital format. We believe that this period is a relevant time frame to cover the most important transitions from sail to steam, combustion, containerization, and mega-carriers, with a good balance between the periods pre- and post-containerization. Nowadays, the Lloyd's company insures about 80 % of the world fleet and therefore is historically a leader on the market with a monopoly power, centralizing most of the information on maritime transport flows.

The stability of the document structure and contents, notwithstanding a huge growth in the number of movements, makes the 5-year snapshots comparable over time. But given the fact that this publication was daily or weekly, extracting only one item in the entire year inevitably created a potential bias in the representativeness of the data sample, difficult to estimate in comparison to the yearly figure. One solution has been to target the same period for every item, namely around April, to strengthen the robustness of our database to seasonal effects. However, the fact that we have a sample of data only from one month in a specific season (Spring) can potentially bias the results, as traffic can exhibit different patterns along the year, for example when goods need to be delivered before Christmas. Moreover, the global historical database did not come out ready. Immense efforts were put on data verification and cleaning: 10,253 place names were checked with scrutiny taking into account regular changes in port names (e.g. Port Swettenham in Malaysia becoming Klang or Port Klang) and exclude passage points such as straits and channels in order to keep only commercial ports in the database. Some test of the accuracy of the Lloyd's data were conducted in Ducruet et al. (2015), where, for example, the authors confirmed that over the entire period 1890–2000, the correlation with Chinese port tonnage was over 0.8, showing that regardless the data extraction technique and partial time coverage of the data base, the extracted data are sufficiently representative for maritime network flows. The resulting flow matrix or network is an undirected graph encompassing 22 different years of observation and constructed to allow the application of various tools originating from Graph Theory and Complexity Science.

Methodology

Graph Theory provides us with a powerful toolkit for the modelling and treatment of data which exhibit pairwise relations, such as a transportation network. In the case of the data

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 5 of 17

derived from the Lloyd's Shipping Index, we treat each port as a node of the network, and the edges (the connections between the ports) are added if we observe at least one movement of a ship from port A to B in a given time period. We obtain in this way an undirected and unweighted graph $G = (N, E)$, with the set of nodes (N) and edges (E). In this work we do not take into account the intensity of movements among the ports; that is, each edge in the network has a weight of 1 and contributes equally to the scores that we obtain while performing calculations on the network. This is certainly a simplification, which overlooks the intensity of flows, by looking only at the existing connections which we can observe in the maritime network. As a result, busy links, such as Singapore-Shanghai contribute to the the network in the same way as the links at which we observed only one movement during the studied period. This simplification is not harmful if one wants to study just the topological properties of the network, as it is the case in the present paper, and not necessarily the intensity of flows or congestion effects. For the studies of flows per se, weighting by tonnage or number of calls seems to be a necessity. As previously discussed, we only have a portion of data every 5 years, covering only a part of the yearly movements of the world fleet. However, they do provide a reasonable overview of the most important connections and ports, and they do keep track of the networks evolution over time.

Topological measures used to analyze the network's evolution

In the first part of our analysis we use the classical network measures in order to describe topological properties of the maritime trade network derived from the Lloyd's Shipping Index publications. The different network measures, which will be further discussed, allow us to draw conclusions about the structure of connections between the ports of the world maritime network and about its evolution over time. In this work our goal is to investigate the structure of the maritime trade network and to see how it relates to its efficiency. Understanding the underlying structural properties of the network is the first step to future research and modelling of the maritime network evolution, which in turn can be useful for simulations of its future developments. Thanks to the availability of data from different moments in time, we can constructs snapshots of the maritime trade network every 5 years, and therefore follow some global measures to see what evolutionary processes can be observed in this network. For a comprehensive overview of different network measures consult Newman (2010).

The first and the most classic network measure which is largely used to characterize both node's centrality and global network evolution is the node degree. This measure corresponds to the number of neighbors each individual node has in the network and can be explained intuitively as the number of unique "trading partners" of a given port in a given time period. In this paper we will focus on the average degree, as it is a measure of the density of the network and of the proportional increase of the number of sea connections (edges) with the number of ports (nodes).

The second and the third network measures which we use are the average shortest path length and the effective diameter. The first can be defined as the average of all topological distances between all pairs of nodes present in G along the shortest paths, while the diameter is the longest shortest path of the network. Formally, the average shortest path can be expressed as

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 6 of 17

$$S_G = \frac{1}{n(n-1)} \sum_{i \neq j} d(n_i, n_j)$$

Where $d(n_i, n_j)$ is the topological distance between a pair of two nodes and it means the number of "hops" between two nodes of the network. Both the average shortest path length and the diameter rely on topological distance between the pairs of ports, which can be a proxy for the speed of delivery of goods that are being shipped around the network. The shorter the average shortest path is, the faster (at least, in topological sense) the goods can arrive to their final destination. Following the steps of Leskovec et al. (2007), we compute the *effective diameter*, taking into account the 0.9th percentile distance in the network in order to avoid noise which often appears in the measurement of the diameter, and to to ensure comparability between our studies and those of the evolution of other real world networks. In order to calculate the effective diameter it was necessary to compute the shortest paths between all pairs of nodes and to plot a cumulative density function of the distances. We then took the 0.9th percentile to define the effective diameter. Both measures are calculated at the global level.

The last global classic measure borrowed directly from Graph Theory is the *clustering coefficient*, which tells us how dense the network is and captures the probability with which the neighbors of n_i are also connected to one another. The clustering coefficient can be defined as the ratio of the number of edges present in the node's direct neighborhood over the number of all potential edges in this neighborhood. Formally,

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

where k stands for the number of nodes in the neighborhood, e_i stand for the edges present in the neighborhood of n_i Just like discussed above, the nominator stands for the number of edges present in the neighborhood (multiplied by two, because this is an undirected graph), and the denominator captures the number of all possible edges which could exist in this neighborhood. We use *average clustering coefficient* in order to get a global measure which enables us to describe the network as a whole in one number per time period. This measure, combined with others, enables us to draw conclusions about the density of the network and its organization. It allows us to see if the network tends to evolve towards a more hub-and-spoke model, where shipping companies rely on transhipment rather than on direct links between all ports.

Turning towards local measures, we calculate the *closeness centrality* which captures the number of hops from a node to any other node in a network. In other words, it measures the topological distance from n_i to all other nodes in a network along the shortest paths. The higher the closeness centrality, the easier it is to get to other nodes of the network, as closeness centrality is the inverse of the sum of topological distances, formally

$$C_c(n_i) = \frac{1}{\sum_j d(n_i, n_j)}$$

Where d stands for topological distance.

Closeness centrality is really important from the point of view of individual ports, as it delivers us information about the time of delivery of goods from the port of interest to any other port in the network. The more central the port is, the shorter the route taken by

good shipped from this port will be, which can be a proxy for costs. However, closeness does not take into account the overall efficiency of the network, as it favors direct links rather than transshipment.

Another well-known centrality measure which we use is the betweenness centrality, which tells us if the node lays on a crossroads of many routes in the network, therefore occupying a priviledged position of a so-called "middle man", or a hub, where the goods are transshiped. Formally, the betweenness centrality corresponds to the proportion of the shortest paths passing through n_i to the number of all the shortest paths in the network between all pairs of nodes. Formally,

$$C_b = \sum_{s,t \in N} \frac{\sigma_{st}(n_i)}{\sigma_{st}}$$

where $\sigma_{st}(n_i)$ is the number of shortest paths passing through the node and σ_{st} is the number of shortest paths between all pairs of nodes.

Random walk measures - locally computable centrality metrics

In the second part of our analyses we run algorithms on the network to measure its navigability. In a way, we leave the concept of a static network in order to analyse the potential flows on the underlying structure, metaphorically treating the network as a sort of preexisting infrastructure, like rail or pipelines. Navigability is a crucial concept in any transportation network (De Domenico et al. (2014), Gulyás (2015)). Intuitively it captures the ease with which one can travel from any point A to B in a network, which is of huge importance when the delivery times and efficient route planning are of essence, as it is in the maritime network. The algorithmic measures, which we propose in this paper, tell us how easy it is to move around the network, and also which nodes are the privileged ones, meaning that if we start a walk at that node, we will be able to visit many other nodes. Our aim is to see how the navigability of the maritime network changed over time together with the underlying topological structure. We would like to know if globalization, new technologies (like containerization) have pushed the network towards an optimal, more navigable organization. According to de Domenico et al. (2014) random walks are a good proxy to determine networks navigability, as they capture the dynamic functionality of the network. However, the classical random walk measures which attract substantial attention in the literatue (Lovasz 1996) are usually global and require time consuming computations, which depend on the size of the network. The best known examples of such measures are the *cover time*, which gives us the number of steps necessary to visit all the nodes in the network, *mixing time*, which gives us the number of steps that the random walker needs to perform to get lost in the graph, or the *hitting time*, which tells us how many steps will be needed before the random walker reaches the node of interest. Short random walks present an interesting, yet not that popular alternative to these global measures, especially provided that they can give us interesting information about the neighbourhood structure of the network and are computable locally, which reduces time complexity of such operations.

In the present paper we propose a *random walk discovery* measure, which tells us how many unique nodes can be visited, if a walker (say, a ship) moving randomly around the network, starts at node n_i and performs 100 steps. The question we ask here is: how many

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 8 of 17

nodes are discovered in T steps of a random walk? By knowing the properties of short random walks on graphs, we know that some specific topological structures enable better scores than others. In a good case, when the graph looks locally like a tree with nodes of degree equal to at least 3, the number of nodes discovered will be close to T. However, in a bad case, which is a line, the number of nodes discovered will be roughly equal to $\sqrt{T}$. The theoretical lower bound is $\sqrt[3]{T}$ nodes discovered in T steps in any network (Barnes and Feige 1993). In practice, this means that a higher degree of the node or of nodes in its neighborhood leads to better discovery rates. Secondly, in clustered networks, better discovery rate is observed when communities are strongly interconnected. The more nodes are visited by the random walker, the better the position of the node in the network from the point of view navigability. We iterate the random walk discovery algorithm 100 times for each node of the network and take the average as final score in order to avoid statistical biases. The choice of a proper leght of the random walk is important, however, in order to be able to study the local neighborhood of nodes it is necessary and enough to make sure that the chosen number of steps is not too large and that it is smaller than the mixing time in the graph in question, because once the limit of the mixing time is exceeded, our measure will be insensitive to the starting node. The mixing time in for example expander graphs is of the order $\log(N)$, where N is the total number of nodes in the graph. The algorithm written in pseudocode can be found in the Appendix.

The second algorithmic measure which we use is the *escape difficulty*. In this algorithm we ask the walker to start her random walk at node n_i and we count how many steps she had to make in order to be at least 4 hops away from n_i. More formally speaking, we want to see how many steps need to be performed by the mobile agent moving randomly around the network to escape the 3-neighborhood of n_i. The considered measure corresponds precisely to the hitting time of the node outside of the 3-neighborhood of the starting node in the considered graph. It may be represented as the hitting time from the staring node n_i to the special node v in the graph obtained from our network by merging all nodes outside of the 3-neighborhood of n_i into a single node. The hitting time is a basic and well studied random walk parameter on graphs (Lovasz 1996). We remark that a symmetrized version of the hitting time between a pair of nodes, known as the commute time, describes the electrical resistance between this pair of nodes (Tetali 1991). Thus our measure reflects the electrical flow between the node n_i and the outside of its 3-neighborhood. Electrical flows are in turn related to the maximal flow problem (Christiano et al. (2011)), under an appropriare weighting of links. Same as in the case of the random walk discovery, we iterate the algorithm 100 times and then take the average in order to obtain the final score. The escape difficulty measure will provide us with information about the direct neighborhood of the node. If the score is high we can conclude that it is plausible that the node lies in a small, highly connected cluster with few links to the outside world. If the score is low, it means that it is easy for the mobile agent to escape the 3-neighorhood and that node must lie on bridge between clusters. The algorithm written in pseudocode can be found in the Appendix.

Results

By analyzing the results obtained for the database covering a portion of movements of ships during the period 1890–2000 we can say that the maritime network has changed its structure over time by quite a bit.

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 9 of 17

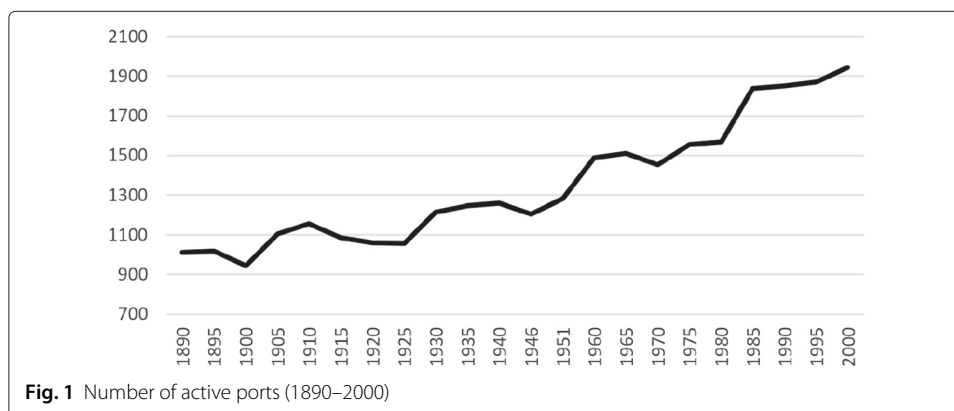**The size of the maritime network**

First of all, the network has substantially increased its size between 1890 and 2000. Note that we take into account only the active ports, that is the ones which either received or sent at least one ship in a given time period. In Fig. 1 we have reported the fluctuations in the number of active ports over the years. We observe a growth of the network which almost doubles its number of nodes between 1890 and 2000. We also see that this growth is rather steady over time. In the year 1890 we have 1011 ports that were active during the period of data collection, while in 2000 this number goes up to 1944 active ports. This constant increase in the number of active ports can be explained with progressive globalization and strengthen international trade, as well as the development of global supply chains.
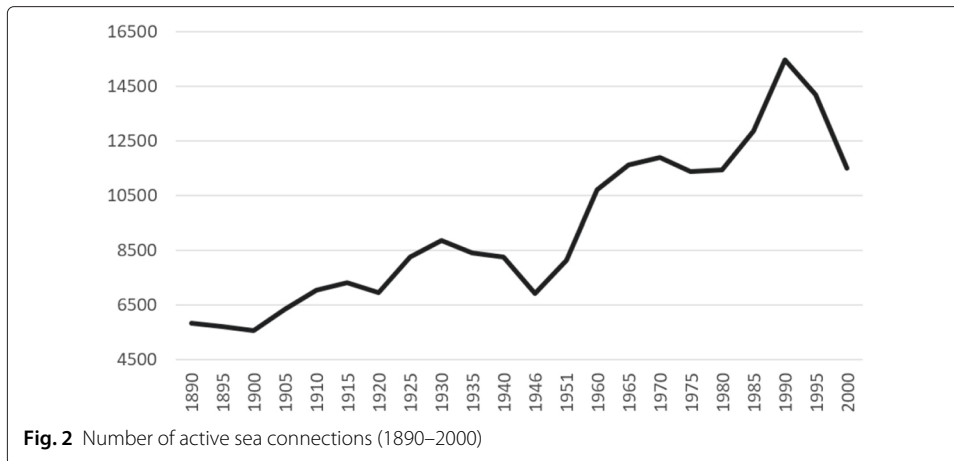
In terms of the number of existing connections, we see that their number grows as well, and that the increase of the number of edges is not much faster than the growth of the number of nodes (Fig. 2). This indicates that while the network grows, it evolves towards its specific structure.

There exist models, especially the well-known Preferential Attachment, which predict that the increase in the number of edges should be linear in the number of nodes, which means that the average degree should not change over time. By observing the average degree computed for each time period in the studied maritime network, we find it to be equal to 11.6 in 1890 and then to grow slowly with some fluctuations until 1930 when we observe a drop, and then in 1951 it goes up again to reach its peak in 1990 (16.9), only to drop dramatically in the 2000 to a value similar to that from 1890. In the year 2000 the average degree is equal to 11.9, only by 0.3 higher than in 1890, with twice as many active ports (Fig. 3). Provided that the scores for average degree are the same at the beginning and the end of the studied period, we cannot exclude the possibility that the network evolved accordingly to some version of the Preferential Attachment model, which could potentially take into account the question of geographical distance between ports.

Evolving topology

The considerations of the average network degree from the previous subsection are especially interesting when compared to the existing literature on network densification. Leskovec at al. (2007) study the evolution of a number of real world networks, such as the scientific paper citation network, network of actors, email network etc. and find that

Fig. 1 Number of active ports (1890–2000)

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 10 of 17

Fig. 2 Number of active sea connections (1890–2000)

the number of edges always grows super-linearly with respect to the number of nodes. This effect of growth of the average degree in real world networks is puzzling, because it contradicts Preferential Attachment model (Barabási and Albert 1999) which predicts a perfectly linear increase of the number of edges with the number of nodes. One example of a real world network which is close to the linearity of growth is the road network in the Milan region (Strano et al. 2012). The network which they study exhibits a rather constant average degree over time, with only a very slight increase of the order of 0.2, this however can be due to the nature of the road network, which is planar, which means that it can be drawn on a sphere in such a way that no links will overlap or cross. This property leads to important consequences for the network structure, because the maximal degree of nodes in constrained by space. In the case of road network nodes are defined as road junctions, therefore it is hard to expect many nodes of degree more than 10 and large variations in the average degree over time. The maritime network, like the road network, is a spatial and transportation network, but with one major difference — it is not planar. Therefore, the maritime network does not suffer from the "natural" limitation of the maximal number of neighbors, as each port can develop as many connections with as many ports as it wishes to. In the case of the maritime network we seem to observe an unusual effect where first the number of edges grows super-linearly, but at the end of the sample goes back to its initial level.
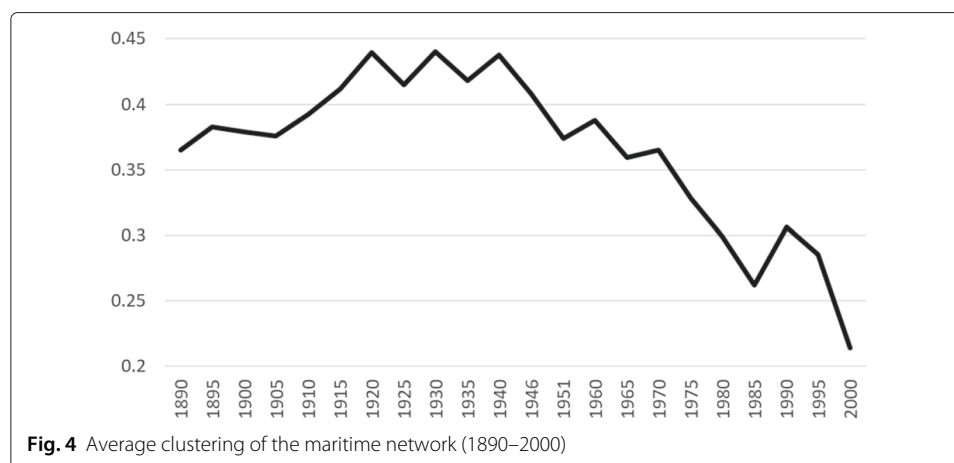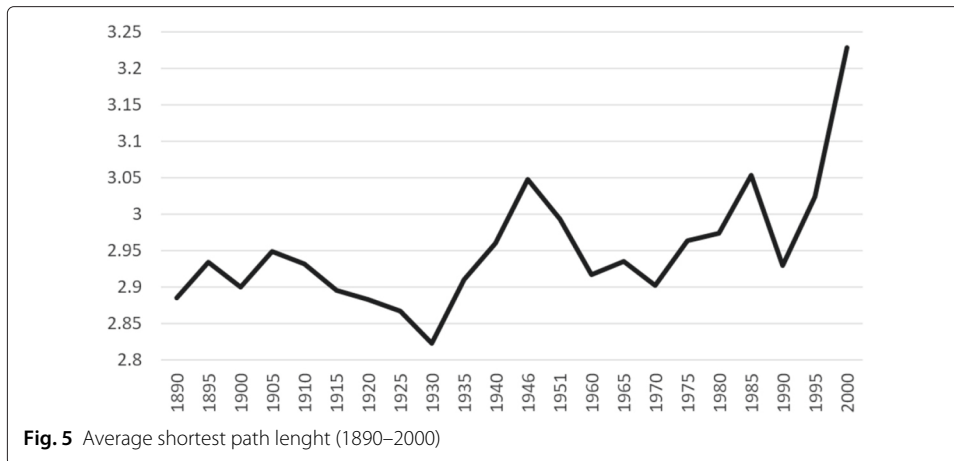
Fig. 3 Average node degree (1890–2000)
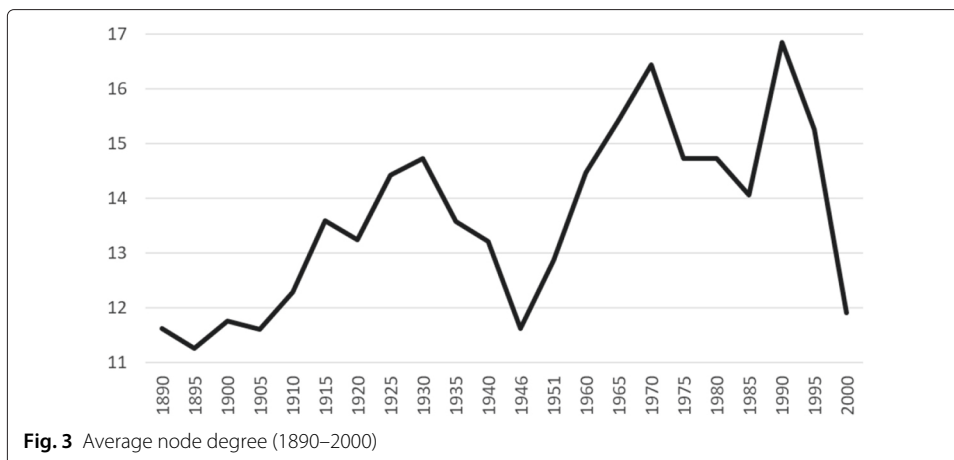
Another major difference between the maritime network and the findings of Leskovec et al. (2007), is that the maritime network exhibits a constant effective diameter equal to 4 over the entire period, so we do not observe the phenomenon of shrinking diameter which Leskovec et al. (2007) find for all the networks studied by them. It seems that the maritime trade network is a network of unique evolutionary properties, which have not been yet observed in other real world networks, that, surprisingly, shared many common evolutionary traits. These findings place the maritime network at a hot spot for studies of the evolution of the real-world networks in complexity science, because it creates a need for better understanding of its evolution and a need for a potentially completely new model of network growth.

In order to deepen our understanding of the evolution and densification of the maritime network we have calculated the average clustering coefficient for each time period (the results are presented in Fig. 4). We find that the average clustering coefficient decreases steadily over the period between 1940 and 2000, which indicates a change in the network structure. Perhaps we cannot go as far as to say that we observe network sparsification, but, especially by looking at the clustering coefficient, we can say that we observe a reorganization of the network, and that it becomes less clustered with time, which supports the hypothesis that the network develops into a more hub-and-spoke structure. We also find that the clustering starts to fall in 1940, that is long before the widespread of containerization, which would be the usual suspect for the cause of network optimization, understood as a tradeoff between network's navigability and maintenance cost. This network optimization process can be noticed also in the behavior of the average shortest path length. The results of the average shortest path are reported in Fig. 5, where we can see that the average topological distance between each pair of nodes is small, around 3 for the entire period under study. It increases over the time period, but only very slightly, passing from 2.88 to 3.23, even though the size of network increases tremendously.
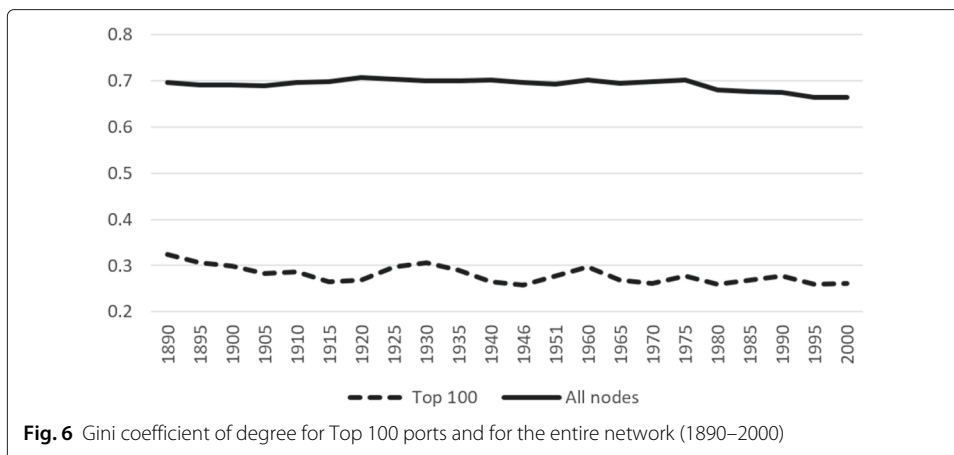
Network centralization

Let us turn towards the measures on the local level. First we look at the degree distribution in each time period, then we compute the gini coefficient in order to see the level of inequalities in our network. We have calculated the gini coefficient for all the nodes

Fig. 4 Average clustering of the maritime network (1890–2000)

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 12 of 17

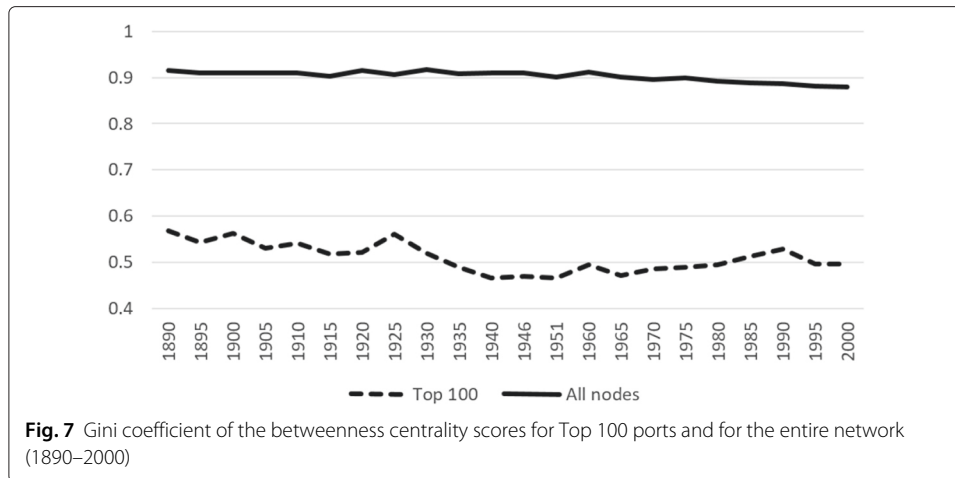**Fig. 5** Average shortest path lenght (1890–2000)

in the network and the top 100 nodes in order to check for potential hierarchical structure. The results are reported in Fig. 6. We find that the inequalities in degree are much smaller among the top 100 nodes than for the entire network, as it oscillates around 0.3 for the top 100 and around 0.7 for the whole network. A similar pattern can be found by looking at the gini coefficient of betweenness, where we find that the distribution for the entire network is really unequal, whereas the scores for top 100 nodes are much more equal (Fig. 7). These findings indicate that the network has some well-connected nodes which span the network and which are rather equal among each other, while we observe significant inequalities in the network as a whole, indicating that apart from the top 100 nodes, there must be numerous not so well-connected ports which create links to those of higher degree. These findings would go in favor of the hypothesis that the network develops towards a more hub and spoke structure, favouring efficient transshipment, which is already indicated by the results of the average clustering coefficient.
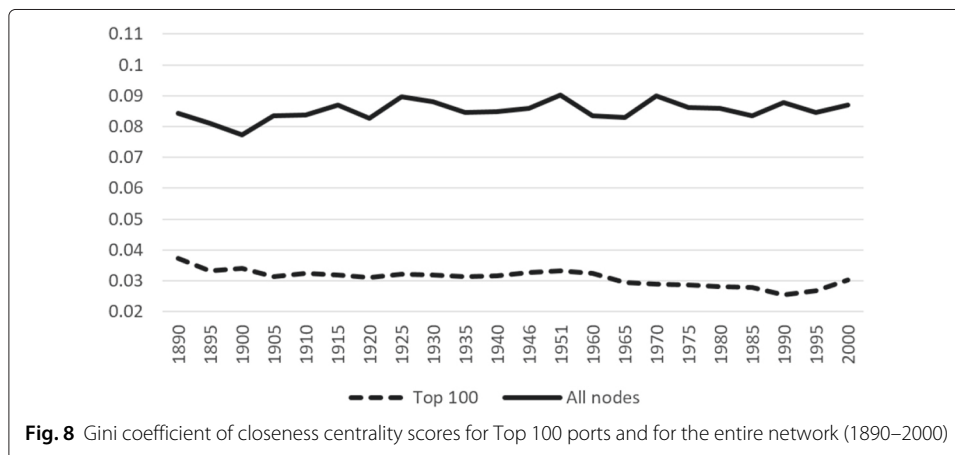
We have applied the same methodology to analyze the individual scores of closeness centrality and we report the results in Fig. 8. We find that the distribution of closeness scores is very equal, as it is close to 0. We still find that the closeness distribution for the top 100 ports is almost perfectly equal, while the inequalities in the entire network are relatively larger, but these differences are really small.
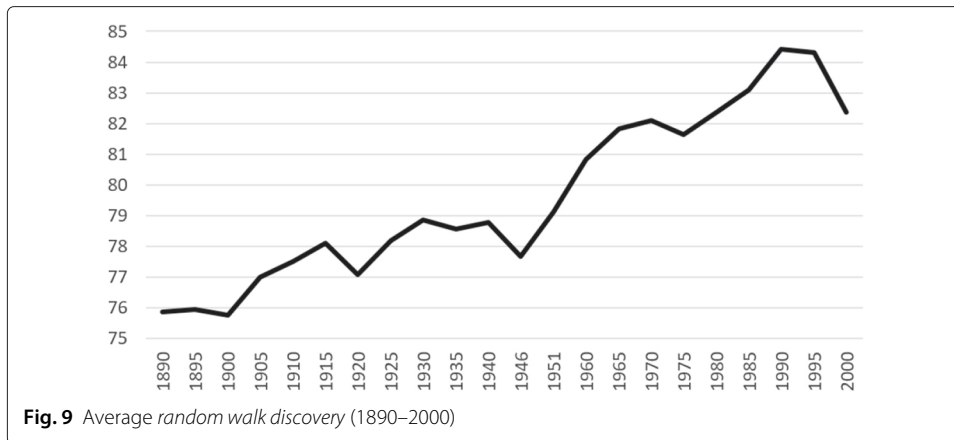

Fig. 6 Gini coefficient of degree for Top 100 ports and for the entire network (1890–2000)

Fig. 7 Gini coefficient of the betweenness centrality scores for Top 100 ports and for the entire network (1890–2000)

Network's navigability

In order to measure the navigability of the maritime network we have created the random walk discovery measure, which tells us how many unique nodes are visited by a walker on the network in a random walk starting from a node n_i. Each time the walker performed 100 steps and the procedure has been repeated for each node 100 times. We constructed the individual scores by taking the average of all iterations. We then took the average of all individual scores and reported them in Fig. 9. We find that the average number of unique nodes visited in a single walk increases over time, indicating an increase in the network's navigability. In general we observe a clear upward trend in the average random walk discovery starting from 1900, which becomes even more visible starting from 1946 - long before containerization has even begun. However, we do observe some drops in 1920, 1946, and, most surprisingly in the year 2000, when the drop is the largest in the whole period under study. The last drop is especially puzzling, because the beginning of the 21st century is known to be the period of increased globalization and increased optimization performed by maritime operators. It is possible that the effect which we observe in the year 2000 is due to the limited amount of data which has been used for this study, or is linked to the fact that since the year 2000 we started to observe an important trend in ship

Fig. 8 Gini coefficient of closeness centrality scores for Top 100 ports and for the entire network (1890–2000)

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 14 of 17

Fig. 9 Average *random walk discovery* (1890–2000)

upscaling, which has led to exclusion of smaller ports that were unable to handle such large vessels. If the studies of the fuller dataset confirm that the average random walk discovery started to deteriorate in the 21st century, we would have a really interesting phenomenon to explain.

Another measure of navigability which we propose in this paper is the escape difficulty, where we check how many steps need to be performed by a random walker to leave the 3-neighborhood of n_i. In theory, it is most difficult for the walker to leave the 3-neighborhood if n_i is located in a small clique or dense subnetwork with few links to the rest of the network. Such network structure would correspond to a very regionalized world, where ports tend to develop connections mostly with their neighbors. If the score of the escape difficulty is low, we can suspect that the node lies in a sparse and highly connected neighborhood (formally, in a part of the graph with good expansion), such as a tree, or a forest, rather than a collection of weakly connected clusters. This would also be the case for a network with hub-and-spoke structure. Indeed, this is what we find by launching the escape difficulty procedure for each node of the maritime network and by taking the average of all the scores (just as in the case of random walk discovery, we iterate the algorithm 100 times for each node). The results are reported in Fig. 10, where we observe a downward-sloping trend in the average escape difficulty, whose values become smaller

Fig. 10 Average number of steps needed to escape the 3-neigborhood (1890–2000)

with time, passing from 8.37 to 5.93 between 1890 and 2000. However, it goes through rather large variations, especially the peak between 1925 and 1930, when the value of the average escape difficulty was over 11, so we cannot claim that the trend is very clear.

Conclusion

In the present paper we have constructed a network of maritime connections thanks to the data extracted from the Lloyd's Shipping Index, a database containing information on daily movements of ships of almost the entire world fleet. Our data cover the years between 1890 until 2000, where we have information about the movements of ships during at least 2 weeks in regular intervals of 5 years. This data enabled us to construct snapshots of the network of sea connections at different moments in time and to follow its evolution. Most of the existing studies of real world networks focus on the static networks due to scarcity of quality data. One of the few examples of such studies is the work by Leskovec et al. (2007) who find that real world networks tend to follow two laws, that is densification and shrinking diameters. Our work proposes a dynamic view on a real world and truly global network. In particular, we find that the maritime network doesn't necessarily densify with time and that its effective diameter remains constant over the period of a century, even though during this period the size (number of nodes) of the network doubles. In the case of the maritime network we seem to observe a strange phenomenon of network optimization, which begins long before the widespread of containerization, and exhibits itself in the decreasing clustering coefficient and increasing navigability. The maritime network tends to be also quite unequal, having the top ports creating a sort of a "rich club", which again, together with global network measures, suggests that the network structure tends to evolve towards a hub-and-spoke structure. Moreover, we construct two new algorithmic measures of network's navigability which are based on the random walk procedure. The *random walk discovery* measures the ease of exploration of a network in a given number of steps, while the *escape difficulty* tells us how hard is it to leave a 3-neighborhood of a given node, and therefore provides us with valuable insights about the global network structure. Similar studies need to be conducted on a fuller data sample in order to confirm the observed trends and check for possible seasonal effects. It would be certainly interesting to study the peaks and the falls of the network measures which seem to align with some well-known events from the world history, such as the Great Depression and the 2nd World War. At this stage of research we are unable to isolate the effects of precise events on the network structure in such a way that we could establish a causal relationship. Such studies would require data of much finner density than just 5 years and potentially external control variables to isolate the precise effects. All of which we leave for future research.

Appendix

Algorithm 1 Random Walk Discovery

---

1. Starting at node n_i, **repeat** 100 times:

    - Pick a neighbor of the current node uniformly at random.
    - Move to that neighbor.

2. **Return** the number of distinct nodes visited during the 100 steps of the walk.

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 16 of 17

Algorithm 2 Escape difficulty

---

1. Starting at node n_i, **repeat**:

    - Pick a neighbor of the current node uniformly at random.
    - Move to that neighbor.

    until the distance of the current node from n_i is more than 3.

2. **Return** the number of performed steps of the walk.

Authors' contributions

ZKS: Study conception and design. CD: Acquisition of data. ZKS and NR: Analysis and interpretation of data. ZKS and CD: Drafting of manuscript. ZKS and NR: Critical revision. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Centre National de la Recherche Scientifique (CNRS), UMR 8504 Géographie-cités, Université Paris 1 Panthéon-Sorbonne, 13 rue du Four, F-75006 Paris, France. [2]IIT Kanpur, G102 Kanpur Uttar Pradesh, Kanpur, India.

References

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Barnes G, Feige U (1993) Short random walks on graphs. In: Proceedings of the Twenty-fifth Annual ACM Symposium on Theory of Computing. STOC '93. ACM, New York, NY, USA. pp 728–737. doi:10.1145/167088.167275. http://doi.acm.org/10.1145/167088.167275

Barthelemy M (2011) Spatial networks. Physics Reports 499(1–3):1–101

Bernhofen DM, El-Sahli Z, Kneller R (2016) Estimating the effects of the container revolution on world trade. J Int Econ 98:36–50

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU (2006) Complex networks: Structure and dynamics. Phys Rep 424(4–5):175–308

Christiano P, Kelner JA, Madry A, Spielman DA, Teng S (2011) Electrical flows, laplacian systems, and faster approximation of maximum flow in undirected graphs. In: Fortnow L, Vadhan S. P (eds). Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6–8 June 2011. ACM. pp 273–282. doi:10.1145/1993636.1993674. http://doi.acm.org/10.1145/1993636.1993674

De Domenico M, Solé-Ribalta A, Gómez S, Arenas A (2014) Navigability of interconnected networks under random failures. Phys Sci Appl Phys Sci 111(23):8351–8356

Deng WB, Long G, Wei L, Xu C (2009) Worldwide marine transportation network: efficiency and container throughput. Chin Phys Lett 26(11):118901

Doshi D, Malhotra B, Bressan S, Lam JSL (2012) Mining maritime schedules for analyzing global shipping networks. Bus Intell Data Min 7(3):186–202

Ducruet C (2013) Network diversity and maritime flows. J Transp Geograph 30:77–88

Ducruet, C (2015) Maritime flows and networks in a multidisciplinary perspective. In: Ducruet C (ed). Maritime Networks. Spatial Structures and Time Dynamics. Routledge Studies in Transport Analysis, Abingdon, UK, pp 3–26

Ducruet C, Haule S, Ait-Mohand K, Marnot B, Kosowska-Stamirowska Z, Didier L, Coche MA (2015) Maritime shifts in the world economy: Evidence from the Lloyd's List corpus, eighteenth to twenty-first centuries. In: Ducruet C (ed). Maritime Networks. Spatial Structures and Time Dynamics. Routledge Studies in Transport Analysis, Abingdon, UK, pp 134–160

Ducruet C, Notteboom TE (2012) The worldwide maritime network of container shipping: Spatial structure and regional dynamics. Global Netw 12(3):395–423

Frémont A (2015) A geo-history of maritime networks since 1945: The case of the Compagnie Générale Transatlantique's transformation into CMA-CGM. In: Ducruet C (ed). Maritime Networks. Spatial Structures and Time Dynamics. Routledge Studies in Transport Analysis, Abingdon, UK, pp 37–49

Gonzalez-Laxe F, Freire-Seoane MJ, Pais Montes C (2012) Maritime degree, centrality and vulnerability: Port hierarchies and emerging areas in containerized transport (2008–2010). J Transp Geograph 24:33–44

Guerrero D, Rodrigue JP (2013) The waves of containerization: Shifts in global maritime transportation. J Transp Geograph 34:151–164

Kosowska-Stamirowska *et al. Journal of Shipping and Trade* (2016) 1:10

Page 17 of 17

Gulyás A, Bíró J. J, Kőrősi A, Rıetvári G, Krioukov D (2015) Navigable networks as Nash equilibria of navigation games. Nat Commun 6(7651). doi:10.1038/ncomms8651

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C, Bruno JF, Casey KS, Ebert C, Fox HE, Fujita R, Heinemann D, Lenihan HS, Madin EMP, Perry MT, Selig ER, Spalding M, Steneck R, Watson R (2008) A global map of human impact on marine ecosystems. Science 319(5865):948–952

Hu Y, Zhu D (2009) Empirical analysis of the worldwide maritime transportation network. Physica A 388(10):2061–2071

Joly O (1999) La structuration des réseaux de circulation maritime. PhD thesis, Le Havre University, CIRTAI

Kaluza P, Koelzsch A, Gastner MT, Blasius B (2010) The complex network of global cargo ship movements. J R Soc Interface 7:1093–1103

Kuby MJ, Reid N (1992) Technological change and the concentration of the U.S, general cargo port system: 1970–88. Econ Geograph 68(3):272–289

Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrinking diameters. ACM Transp Knowl Discov Data 1(2):1

Li Z, Xu M, Shi Y (2015) Centrality in global shipping network basing on worldwide shipping areas. Geojournal 80(1):47–60

Lovasz L (1996) Random walks on graphs: A survey. Combinatoris, Paul Erdos is Eighty 2:353–398

Marnot B (2005) Interconnexion et reclassements: l'insertion des ports français dans la chaîne multimodale au XIXème siècle. Flux 59(1):10–21

Mayer HM (1973) Geographical aspects of technological change in maritime transportation. Econ Geograph 49:145–155

Newman MEJ (2010) Chapters 7 and 8. In: Networks: an Introduction. Oxford University Press, New York, USA

Pais Montes C, Freire Seoane MJ, Gonzalez-Laxe F (2012) General cargo and containership emergent routes: A complex networks description. Transp Policy 24:126–140

Parshani R, Rozenblat C, Ietri D, Ducruet C, Havlin S (2010) Inter-similarity between coupled networks. Europhys Lett 91:68002

Rimmer PJ (2012) The changing status of New Zealand seaports, 1853–1960. Ann Assoc Am Geograph 57(1):88–100

Robinson R (1968) Spatial structuring of port-linked flows: The port of Vancouver, Canada, 1965. PhD thesis, Vancouver: University of British Columbia, Geography Department

Siegfried A (1940) Suez, Panama et les routes maritimes mondiales, Paris: Armand Colin

Strano E, Nicosia V, Latora V, Porta S, Barthélemy M (2012) Elementary processes governing the evolution of road networks. Nat Sci Rep 2(296). doi:10.1038/srep00296

Tavasszy L, Minderhoud M, Perrin JF, Notteboom TE (2011) A strategic network choice model for global container flows: Specification, estimation and application. J Transp Geograph 19(6):1163–1172

Tetali P (1991) Random walks and the effective resistance of networks. J Theor Probab 4(1):101–109. doi:10.1007/BF01046996

Tovar B, Hernandez R, Rodriguez-Deniz H (2015) Container port competitiveness and connectivity: The Canary Islands main ports case. Transp Policy 38:40–51

Ullman EL (1949) Mapping the world's ocean trade: a research proposal. Prof Geograph 1(2):19–22

Wang C, Wang J (2011) Spatial pattern of the global shipping network and its hub-and-spoke system. Res Trans Econ 32(1):54–63

Wang J, Pulat PS, Shen G (2012) Data mining for the development of a global port-to-port freight movement database. Int J Shipping Transport Logistics 4(2):137–156

Woolley-Meza O, Thiemann C, Grady D, Lee JJ, Seebens H, Blasius B, Brockmann D (2012) Complexity in human transportation networks: A comparative analysis of worldwide air transportation and global cargo-ship movements. Eur Phys J B 84:589–600

Xu M, Li Z, Shi Y, Zhang X, Jiang S (2015) Evolution of regional inequality in the global shipping network. J Transp Geograph 44:1–12