



HAL
open science

Maritime constellations: A complex network approach to shipping and ports

César Ducruet, Faraz Zaidi

► **To cite this version:**

César Ducruet, Faraz Zaidi. Maritime constellations: A complex network approach to shipping and ports. 2011. halshs-00551207v1

HAL Id: halshs-00551207

<https://shs.hal.science/halshs-00551207v1>

Preprint submitted on 2 Jan 2011 (v1), last revised 16 Jan 2012 (v3)

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 constellations: A complex network approach to shipping and ports

DUCRUET César

Centre National de la Recherche Scientifique (CNRS)
Paris I Sorbonne University / UMR 8504 Géographie-Cités
13 rue du Four F-75006 Paris (France)

Email: ducruet@parisgeo.cnrs.fr

Phone: +33-140464007

Fax: +33-140464191

ZAIDI Faraz

University of Bordeaux I / UMR 5800
Laboratory of Bordeaux for Research in Computer Science (LABRI)
351 Cours de la Libération F-33405 Talence (France)

Email: faraz.zaidi@labri.fr

This paper is currently submitted to Maritime Policy and Management Special Issue on “The Geography of Maritime Transport” Please do not quote or cite

Abstract

The analysis of community structures is one major research field in the science of networks. This exercise is often biased by strong hierarchical configurations as it is the case in container shipping. After reviewing the multiple definitions of port systems, this paper applies a topological decomposition method to worldwide inter-port maritime links. Isolating ports of comparable size reveals hidden substructures with the help of graph visualization. While geographic proximity is one main explanatory factor in the emergence of port systems, other logics also appear, such as specialized and long-distance trading links. This research provides interesting evidence about the role of geography, technology, and trade in the architecture of maritime networks.

Keywords: container shipping; port systems; scale-free networks; topological decomposition

1. Introduction

Half a century ago, the French geographer Perpillou (1959) considered that ports and maritime transport form a constellation¹. While such perspective advocated the existence of coherent substructures, little has been done identifying them in geography based on rigorous methods. Graph theory, social network analysis, and complex networks research principally describe the respective situation of individuals with regard to the overall structure of the network and the existence of communities (Boccaletti et al., 2006). The community² is a set of tightly connected nodes sharing denser relations with each other than with the rest of the network. Nodes through which such subgroups connect are labelled bridges or hubs. The seminal essay of Granovetter (1973) on the “strength of weak ties” insisted on the strategic role of such bridges in a social network. Bridge nodes have fewer connections but they are essential for the diffusion of information flows among groups, which depend on them to communicate with others. However, the presence of hubs and the density of communities depend on the wider topological properties of the network. For instance, scale-free networks contain few large degree³ nodes polarizing a majority of smaller nodes (Barabási and Albert, 1999). In comparison, small-world networks have a higher density of links due to their shorter diameter (i.e. longest shortest path) and higher transitivity (i.e. probability for a node to have its neighbours interconnected) (Watts and Strogatz, 1998). Many real-world networks are in fact both scale-free and small-world due to the intermingling of vertical (hierarchy) and horizontal (community) linkages.

The main questions addressed in this paper are as follows: are there bridges and communities in maritime networks, on what grounds, and how would their identification bring anything new to the geography of shipping and ports? Port and maritime geography have long identified a variety of similar concepts such as port systems and hubs but empirical evidence from a network perspective remains dramatically scarce (Ducruet et al., 2010a). Drawing

¹ In astronomy, the constellation is a group of stars forming a picture. Ptolémée identified 12 zodiacal figures as well as 36 other figures whose shape was inspired from animals and geometry. The delineation of current constellations is fixed by the International Astronomical Union since 1930 in Eugène Delporte’s official atlas of constellations (e.g. Orion). However nowadays, the contours of such figures do not have a scientific purpose due to the use of coordinates.

² A community is also called subgroup, subgraph, community, clique, or cluster. The clique is a complete maximal subgraph comprising at least three vertices (nodes).

³ The degree of a node (or degree centrality) is the number of adjacent nodes connected to it. Large degree nodes are also called stars or hubs.

from a dispersed literature, the emergence of port systems in maritime networks may be briefly explained by six main factors:

- a) Physical geography: port infrastructures are tied to waterside locations just like maritime transport is tied to ports and coastal morphology. This physical embedding is a specificity of spatial networks (Gastner and Newman, 2004), but its influence on shipping patterns is limited by the absence of track, the dredging of inter-oceanic canals and port entrance channels, and technological progress allowing flexible circulation of vessels. Yet, coasts, deltas, straits, basins, and seas constitute the main criteria behind the analysis of port systems.
- b) Spatial proximity: as a corollary of the previous factor, the criterion of spatial proximity is implicit in many studies selecting ports based on their adjacency on various levels (e.g. Pearl River Delta, Chinese ports, Asian ports). The *maritime façade* (or *seaboard*) also implies the notion of coastal continuity (e.g. North America's East Coast), while the *maritime region* may include two or more neighbouring coasts. The *port region* is, for a given port, a subpart of its continental hinterland (Ducruet, 2009) but it is often confounded with the maritime region. Finally, most theoretical models of port system evolution are based on spatial proximity between ports along a given coastline (Ducruet et al., 2009).
- c) Global and regional trade patterns: throughout history, various groups of ports have emerged based on frequency and density of trade linkages, such as the Hanseatic League from the North Sea to the Baltic in the 13th-17th centuries. According to Todd (1993, p. 4), "the crystallization of formerly disjointed ports into a ports system rests ultimately on the conditions of trade, conditions which wax and wane in correspondence with global business cycles". The concept of *port system* (Robinson, 1976) and *port* (or *maritime*) *range* (Vigarié, 1964) imply a certain level of regional integration, which results in interdependency among neighbouring ports. Lemarchand and Joly (2009) provided measures of this phenomenon highlighting higher integration levels at Northwest European and North American ranges than at Mediterranean and East Asian ranges. The North European range from Le Havre to Hamburg and the Asian maritime corridor from Singapore to Japan are classic examples of trade-based port systems.
- d) Carrier preferences and port selection: although carriers select ports according to trade demand, they also have their own requirements in terms of quality and cost of service

at and between ports (Notteboom, 2009), sometimes causing important deviations from trade routes. Especially for container ports, actors of the transport chain have gained considerable decision power in port selection while deploying their networks (Slack, 1993; Robinson, 2002). This makes ports (and transport nodes in general) different from individuals in social networks because their connectivity is dictated from external players to a large extent (e.g. shippers, forwarders, ocean carriers). Rapid changes in the organization of logistics led some scholars rethinking the meaning of the port concept itself (Olivier and Slack, 2006) and the role of proximity in port development (Hall and Jacobs, 2010), in a context of globalization and reduced transport costs. Yet, another view considers such changes having limited impact on the layout of port systems and hinterlands due to path-dependency and resiliency (Notteboom, 2009; Guerrero, 2010).

- e) Hierarchy: large ports generally handle a wide variety of cargoes, thereby concentrating and redistributing traffic flows towards smaller ports. Containerization and liner shipping have exacerbated such configurations with the establishment of regular hub-and-spoke systems. The spatial logics of transshipment hubs were conceptualized by Fleming and Hayuth (1994) through concepts of centrality and intermediacy. Rodrigue and Notteboom (2010) also underlined the fact that such intermediate hubs both connect global trunk lines and multiply intra-regional linkages towards secondary ports. The definition of hubs somewhat differs between maritime geography and social network analysis. A hub port generally has many connections, while a bridge node has few. Both have an intermediacy function between other connected ports. While the literature tends focusing primarily on large ports, a process of de-concentration has been confirmed since the work of Hayuth (1981) on the “challenge of the periphery”, stemming from diseconomies of scale in large ports (e.g. lack of space, congestion, delays, high handling costs) and new port development projects at more accessible sites.
- f) Functional specialization: ports can be differentiated according to the respective importance of hub functions and gateway (hinterland) functions. One traditional conception of ports is the *port triptych* (Vigarié, 1979) where hinterland, port, and foreland as essential and interdependent components. Some ports may share intense relations regardless of spatial proximity, such as London and Calcutta in colonial times (Smith, 1776, cited by Teboul et al., 2000). The distribution of forelands (i.e. overseas maritime linkages) is thus a combination of trade-related linkages and

transport-related linkages. North Korea's maritime forelands have spatially shrunk after the fall of the USSR, resulting in a process of hub-dependence upon South Korean ports reflecting humanitarian aid and North-South cooperation (Ducruet, 2008). Traffic shifts between competing ports tend to distort the direct line between hinterland and foreland: up to 40% of French exports passed through Antwerp (Belgium) instead of Le Havre or Marseilles in recent years (Gouvenal et al., 2010). However, measuring the hub function of seaports has always been difficult due to the lack of detailed data.

The multiple and complex combinations among these factors make it difficult to delineate port systems and verify their coherence. Based on the case of liner shipping, this paper wishes to estimate the respective influence of these factors in the emergence and interconnection of coherent port systems. A global database on containership movements in 1996 and 2006 allows for an empirical verification by means of recently developed methods of network analysis. The remainder of this paper is as follows. Section 2 describes the data, structure of the global liner shipping network, and introduces the topological decomposition method in order to reveal communities and bridges. Section 3 proposes an interpretation of the results in the light of the aforementioned factors. Finally, directions for improvement and further research are proposed in conclusion.

2. Data and methodology

2.1 Structure of the global liner shipping network

By means of vessel movement data obtained from Lloyd's Marine Intelligence Unit (LMIU)⁴, we build a graph where nodes are ports and links are inter-port connections realized by the circulation of vessels. For more simplicity, we do not take into account the weight of links nor do we consider directionality between ports. Instead, the analysis compares two dimensions of the network: the graph of direct links (GDL) only includes direct successive calls between ports, whereas the graph of all links (GAL) includes direct and indirect calls. Ducruet and Notteboom (2010) highlighted important differences between the respective structures of these two dimensions. The GDL is a typical scale-free network with a strong hierarchical

⁴ <http://www.seasearcher.com/lmiu/index.htm> (Accessed December 2010)

structure, whereas the GAL is more a small-world network. Figure 1 confirms the higher coefficients of the power-law line in the GDL (1.6) than in the GAL (1.08) as well as their stability between 1996 and 2006. Being built upon the overlap of all complete graphs formed by vessels, the GAL exhibits a higher transitivity (average clustering coefficient) and lower average shortest path length than the GDL. Tightly connected communities are thus more likely to appear in the GAL than in the GDL, which is more dependent upon large hubs and thus more vulnerable than the GAL due to the strong influence of large degree ports.

[Insert Figure 1 about here]

Another dimension of the port hierarchy is the relation between degree, length of links, and traffic. Based on orthodromic distances between connected ports, we calculated the average length of the longest links for each degree value. We find that large degree ports connect longer distance links than small degree ports on average (Figure 2). The correlation remains high at both years, but it has slightly dropped because of some secondary ports expanding their links. The trend is less linear in the GAL due to the inclusion of indirect links. Overall, large hubs dominate the network by the number and spatial extension of their maritime connections. However, while such long-distance links are important for globalization, their traffic weight is relatively small with regard to short-distance links. Ducruet and Notteboom (2010) calculated on a world level that links of 0-500 kilometres concentrate about 80% of worldwide container traffic in the GDL and 50-60% in the GAL, due to the importance of intra-regional linkages. This demonstrates the importance of smaller ports in ensuring network connectivity and continuity.

[Insert Figure 2 about here]

The relationship between degree and traffic is also presented in Figure 2. Traffic shares in world total were calculated for each degree value based on the amount of total vessel capacity passing through each port after one year of movements. While correlations are all very high regardless of the year and type of network considered, they also have slightly increased between 1996 and 2006, thus indicating a reinforcement of traffic concentration at the largest ports. This also confirms the analysis of Gini coefficients by Ducruet and Notteboom (2010) on world port throughputs measured at port level from 1970 to 2009 highlighting a constant increase of concentration among ports. In other words, traffic is more polarized at large

degree ports, while lower degree ports have lost traffic in relative terms. The high correlation between degree and traffic is one common feature of complex networks, as empirically demonstrated by Deng et al. (2009) for maritime networks.

While such evidence confirms the relevance of degree centrality for comparing ports, the strong hierarchical structure of liner shipping networks necessitates specific methods of graph partitioning (or clustering) for revealing communities and bridges in such scale-free / small-world networks.

2.2 The topological decomposition method

There are several ways to analyze networks. Among numerous statistical measures, the degree centrality of a node (i.e. number of adjacent neighbours, or number of connections to other nodes in the network) is the least controversial. At network level, the degree distribution provides clues about the overall structure of the network (Brandes and Erlebach, 2005). Another approach is the visualization or graphical representation of the network, which is a useful method to discover hidden knowledge and extract interesting patterns in data (Herman et al., 2000). However, the size of most real-world (complex) networks presents new challenges necessitating more advanced methods of analysis, such as the decomposition of the network into subcomponents. In addition, centrality and structural measures describing the organization of the network are limited in describing the role of groups and subsets as well as the complex interactivity among them.

The discovering of clusters in complex networks is a wide field of analysis using several methods. For example, Girvan and Newman (2002) used edge (link) betweenness to produce a divisive hierarchical clustering algorithm. The basic idea is to distinguish among intra-cluster edges and inter-cluster edges. Edges lying between clusters will have a higher betweenness centrality than edges within a cluster. This algorithm removes high betweenness centrality edges to identify clusters and recalculates the betweenness centrality. Wu et al. (2004) introduced a multilevel mesh structure to cluster large networks through an agglomerative process. A set of representative nodes is defined based on nodes' betweenness centrality and degree centrality, while all other nodes are assigned the nearest representative nodes to obtain clusters. At each level, a branching factor determines the number of clusters, but this number might not represent the actual number of clusters in the dataset as they are

determined by the user without the use of any heuristic or statistical measure. Boccaletti et al. (2007) proposed a clustering method based on the cluster de-synchronization properties of phase oscillators. Starting from a fully synchronized state of the network, a dynamical change in the weights of the interactions that retain information on the original betweenness distribution yields a progressive hierarchical clustering that fully detects dense communities. Newman (2004) introduced a faster agglomerative hierarchical clustering algorithm based on a quality function called modularity Q . The algorithm repeatedly joins communities together in pairs, choosing at each step the join that results in the greatest increase in Q . Boutin et al. (2006) used a focus based filtering and clustering technique for complex networks. This technique extracts a tree-like graph so that the resulting structure can be drawn using any force directed algorithm leaving the final drawing easily readable. One of the drawbacks of this system is that the user has to choose an initial entry point to filter the graph, as in the Nystuen-Dacey algorithm applied to the worldwide container shipping network by Ducruet and Notteboom (2010) revealing nodal regions in which one independent port dominates a number of subordinate ports. Moreover since edges are removed to simplify the structure of the network, important information can be lost. Efficient algorithms to cluster networks with only small-world properties have been proposed (Auber et al., 2003; van Ham and van Wijk, 2004). These systems perform well if the topology of the network follows small-world properties but fail to perform in the presence of scale-free properties. This is due to the fact that in a scale-free network, few nodes dominate the entire network, what makes it difficult identifying the clusters. Similarly methods to cluster networks with only scale free properties have been proposed. For example Päivinen (2005) proposed a method based on the Minimum Spanning Tree (MST). While constructing the MST, they take into account the nodes having high degree and increase the importance of the edges connected to these high degree nodes which are called hubs. This makes the edges connected to the hubs more important and thus are retained during the construction of the spanning tree. This approach works well if the underlying network structure has only scale-free properties, but it faces some limitations in the presence of small-world phenomena. This is because the approach assumes that the high degree nodes are the probable cluster centres, while in reality, it is possible to have a clique in the network in which two nodes connect to two different hubs, the MST method will force the clique structure to break and assign the nodes to two different clusters resulting in a loss of information. Finally, an important class of clustering algorithms are called Spectral Clustering algorithms have attracted considerable interest (Spielman and Teng, 1996). They are based on computing the eigenvectors of the adjacency matrix, or some other matrix representing the

graph structure. The biggest advantage of these algorithms is that they are able to detect clusters without a specific form as compared to classical algorithms such as k-means. They are well suited for large size networks as well. But these algorithms are suited only for data sets where the similarity graphs are sparse (Luxburg, 2007). Further more, choosing a good similarity graph is not trivial, and spectral clustering can be quite unstable under different choices of the parameters for the neighbourhood graphs.

Due to the limitations of the aforementioned clustering methods, a look at decomposition techniques is necessary. Decomposition means that the nodes of the network can be grouped into subsets based on some criteria such that the structural relations between nodes are preserved, without any preconditions. Decomposition is motivated by two important features of complex networks. First, such networks have non-uniform degree distribution compared with random and regular networks. Second, visualization of these networks produces highly entangled and hard to read drawings, despite advanced visualization tools and interactive techniques (Herman et al., 2000). Notably, a number of tools have been proposed that combine decomposition methods with visualization to analyze these networks (Auber, 2003; Heer and Boyd, 2005). The idea to decompose and study complex networks is not new. A decomposition based on the connectivity of vertices (nodes) was proposed by Batagelj and Zaversnik (2002). The method consists of identifying subsets of the network called k-cores. These subsets are obtained by recursively removing all the vertices of degree smaller than k, until the degree of all remaining vertices is larger than or equal to k. For example, to obtain a 2-core of a network, we remove all the vertices with degree less than 2 in the network, i.e. nodes of degree 1 (with only one link). After this removal, certain vertices that previously had high degree might now have degree 1: the removal process is repeated until there are no vertices of degree 1 left in the network. All the vertices left after this removal become part of 2-core and the process can be restarted for higher values of k. Cores with larger values of k correspond to sets of vertices with high degree and are connected to high degree vertices only. This gives cores with larger values of k a more central position in the network's structure (Alvarez-Hamelin et al., 2006). This method has been used in several domains to analyze networks and the connectivity of vertices such as, in the analysis of protein interaction networks (Bader and Hogue, 2003; Wuchty and Almaas, 2005) and networking to filter out peripheral Autonomous Systems (Gaertler and Patrignani, 2004). Apart from its utilization in analysis, it has also been used to visualize large-scale networks as it decomposes a network

into subsets of vertices of increasing centrality. It can also help focus on certain regions of interest in a network (Baur et al., 2004).

The method used in this paper is significantly different from k-cores, although both k-core and the topological decomposition method proposed by Zaidi (2010) are both based on the degree of vertices and create subsets. Topological decomposition focuses on studying how edges are distributed in high and low degree nodes, while k-cores focus on recursively identifying central nodes and has clearly different objectives. This method isolates either high or low degree nodes from the rest of the network, resulting in the formation of Degree Induced Subgraphs (DIS), thus focusing on a minimum value of k (MinDIS) and a maximum value of k (MaxDIS). Figure 3 presents the result of the MaxDIS method at k=4, showing the existence of one interesting subgroup when nodes over degree 4 are removed. The MinDIS method is the opposite: for instance in the same figure, removing nodes having less than five connections would remain only two connected nodes (5 and 6), thus highlighting which nodes dominate the network.

The choice of the k value is difficult because there is no recognized threshold valid for all networks. Due to the inherent hierarchical structure of scale-free networks, the average degree is biased by few extreme values. The minimum or maximum value of k (also called cutoff limit, percolation threshold or fragmentation threshold) may thus be chosen depending on the case under study based on specific criteria, such as traffic share or average distance measures presented in Figure 2 for container shipping. Another possibility is to look at the density of the resulting DIS using either method based on the following equation:

$$CD_q = (e_q * 2) / (n_q * (n_q - 1))$$

Where CD_q represents the Component Density (CD) of connected component q, e_q represents the number of edges in q and n_q represents the number of nodes in q. The equation represents the ratio of the number of edges in component q to the maximum number of edges possible in the component; it is equivalent to the Gamma index defined by Kansky (1963) as a measure of connectivity. This measure is also the same than the rich-club coefficient proposed by Colizza et al. (2006) for whom the rich-club phenomenon “refers to the tendency of nodes with high centrality, the dominant elements of the system, to form tightly interconnected

communities". The measure is applied to all connected components having more than four nodes, and excludes isolated nodes (zero degree) to avoid biases. The density level of connected components is a useful indicator for determining which large degree nodes should be removed from the network in order to obtain tight subgroups.

[Insert Figure 3 about here]

3. Communities and bridges in maritime networks

3.1 Defining fragmentation thresholds

We first consider the density of all resulting connected subgroups at every removal operation for both MaxDIS and MinDIS methods (Figure 4). With the MinDIS method, the density of subgroups regularly increases overall, but in a faster way after removing ports having less than 10 connections on average. More precisely, the curves become stiffer at degree 15 in the GDL and at degree 40 in the GAL, thus providing interesting cutoff thresholds. With the MaxDIS method, the density curve exhibits a distinct behaviour. Following successive removal iterations of larger degree ports, a peak of density appears as disconnected ports and components start to reconnect. The density then decreases along with the emergence of a large subgroup. In mathematics, this phenomenon is called the emergence of a giant component (Janson et al., 1993). Physicists call it percolation and refer to this phenomenon as phase transition (Essam, 1980). The network changes drastically as certain links are introduced, and become a single connected component (Barabási, 2002). In this case these links are introduced in the network by higher degree nodes that are responsible of connecting all smaller components.

Higher density values appear for the GAL due to its small-world property and larger size than the GDL. In all cases, the relevant fragmentation threshold can be the degree value of highest density (11 in the GDL and 28-35 in the GAL), with the risk to produce too many subgroups. Another possibility is to select a higher value while avoiding the absorption by the giant component of too many ports: 18-23 in the GDL and 53-63 in the GAL. For MaxDIS methods, a low threshold will make the interpretation of numerous subgroups difficult, while a high threshold will bring out a large single subgroup with less readability. One possible solution is to use an intermediate value so as to obtain a reasonable of subgroups with good readability:

15-17 in the GDL and 40-49 in the GAL. Such threshold values are, interestingly, the same than for the MinDIS method. For every graph, the retained values have in common to situate the threshold at about 8-9% of the maximum degree. The proportion of ports contained in the resulting subgroups is 76-73% in the GDL and 56-53% in the GAL.

Another important dimension of the search for a relevant k value is the proportion of ports and links included in the largest component. For each graph at both years, low thresholds produce many subgroups of smallest ports, while higher thresholds produce one large subgroup (or giant component) including more ports. The share of the largest component at each value of k indicates whether this entity dominates the results. In Figure 4, the number and size of connected components is thus implicit.

[Insert Figure 4 about here]

3.2 The “rich-club” oligopoly of largest ports (MinDIS)

Largest degree ports are isolated from the rest of the network in order to verify the latter's connectivity and pattern (Figure 5). Results are presented for each network at both years retaining the top 5, 10, and 15 degree ports only for better clarity. In all cases, one interesting outcome is the connexity of the network, because removing many ports and links does not split the network into disconnected components. Another result is the clear dominance of European and Asian ports, with the North European range and the East Asian corridor interconnected. This provides a rather different picture than the map of the top hundred direct traffic links, where trans-Pacific flows dominate by their volume (Ducruet and Notteboom, 2010).

In the GDL, the role of Singapore as a bridge between Europe and Asia is made evident. European ports are more numerous and have stronger connectivity than Asian ports in 1996. Le Havre and Felixstowe have in 1996 a forefront position with direct links to Kaohsiung, Hong Kong, and Yokohama. In 2006, this pattern is still visible but some changes occurred, with more transversal linkages between the two regions. Intermediate hubs and gateways have emerged in each region, superseding Le Havre, Piraeus, Felixstowe, Yokohama, and Kobe at the top of the hierarchy. The bridge role of Mediterranean hubs (e.g. Marsaxlokk, Gioia Tauro, Algeciras, and Valencia) in facilitating East-West flows is clearly underlined in 2006. This is

explained by carrier choices and hub strategies because such ports have limited cargo base and act as intermediary pivots between trading regions (Rodrigue and Notteboom, 2010). In Asia, this phenomenon is better explained by the rise of Chinese ports (Ningbo, Shanghai). Only Houston is included outside of Europe and Asia in 1996. Thus, the MinDIS method not only highlights which are the largest ports but also how they are interconnected. In the GAL, a similar evolution occurs, with a drop in the number of European ports and the disappearance of New York between 1996 and 2006. While the visualization of the GAL does not show explicit patterns due to complete graphs (all ports interconnected), we observe that large gateways rather than large hubs rank at the top of the hierarchy (e.g. Shenzhen, Barcelona).

[Insert Figure 5 about here]

3.3 Small worlds of smaller ports (MaxDIS)

Each connected component can be interpreted as one port system, as showed in Figure 6 through the example obtained from the application of different thresholds. Numerous cliques of ports for low thresholds ($k = 11$) as well as series of complete graphs, thereby illustrating preferential links among smaller ports. The higher the threshold ($k = 15$ or 18), the giant component begins to emerge by absorbing all connected components into one entity. While it is impossible to map and analyze in detail each subgroup due to their number and size, a closer look at exemplary cases chosen from both graph types and at both years is proposed. Results can be confronted to the aforementioned factors of port system formation as follows:

- The majority of subgroups are explained internally by geographic proximity between ports, regardless of the year and type of graph. It is a constant behaviour for smaller ports to cluster together forming densely connected port systems and maritime regions, because they connect over shorter distances on average (Figure 2). Subgroups appear in every area of the world, implying the absence of region-specific connectedness, although the Scandinavia-Baltic area often stands out with tightly connected communities of small ports;
- Many exceptions exist based on long-distance trading links between Western European ports and non-European ports (e.g. Portugal/Brazil/Angola, Spain/Mexico, Spain/Brazil, France/West Africa/Antilles, etc.), thus reflecting other types of proximities based on former colonial ties, shared language and culture. Western European ports often act as bridges between their former colonies through North-South relations, such as the French ports of

Rouen (Normandy), Montoir (Nantes), and Le Verdon (Bordeaux) keeping important overseas links across the Atlantic with Africa and the Caribbean. These results are important because regardless of the type of graph and over time, the most salient hidden structures recall a centre-periphery organization of the world system with very few transversal and resilient linkages. European ports are the only notable exceptions of small ports keeping long-distance connections despite their limited size, as a reflection of their specific function and history. Surprisingly, there are very few examples of subgroups including North American, Asian, and European ports together, except for the case of Canadian ports (Argentia, Shelburne) being included with North European ports (e.g. Aberdeen, Cuxhaven).

- Another case is the emergence of specialized circuits, as seen with the “oil cluster” composed of Nigerian, Italian, Venezuelan, and U.S. ports having only in common to be oil ports. Geographically, this cluster forms a triangle across the Atlantic and has the peculiarity to emerge outside the giant component even at high degree thresholds, meaning that it is very distinct from the general shipping pattern where most other ports are included. Other examples of this kind include a cluster including Algiers, Port Cartier (Canada), Velsen (Netherlands), Bandirma (Turkey), together with some Swedish, Finnish and Spanish ports, as well as the cluster including ports from Great Lakes, Canada, U.S. Gulf coast together with Warri (Nigeria), Odense (Denmark), and Dordrecht (Netherlands), but their internal logic remains hardly interpretable. We can only hypothesize that the two latter clusters emerged accidentally from the combination of different logics that were not necessarily planned or interdependent.

[Insert Figure 6 about here]

The analysis of giant components wishes to verify which bridge ports allow the connectivity of the global network without the inclusion of large hubs, and the logics behind the emergence of maritime regions? We first remove all ports which cumulated traffic is below 10% of world totals based on Figure 2 in order to reduce the noise and to shift the analysis towards important nodes, thus providing the following degree thresholds: 18 and 30 for the GDL, 105 and 160 for the GAL. The choice of the degree value for visualization is the one at which the giant component emerges, i.e. when the global network becomes connected so as to allow the identification of bridge ports. Such bridge ports are defined by two criteria: a low

clustering coefficient⁵ and at least one connection with a port located in a distinct maritime region.

Results in Figure 7 are different according to the type of graph and present interesting evolutions. Firstly, the giant component in the GDL is bigger than in the GAL in terms of the number of ports because the GDL is more constrained by distance than the GAL, which includes all intermediate calls. Thus, links are primarily “local” and it needs including more ports before one single component emerges. In the GDL, a circular or ring-like pattern is visible that broadly corresponds to a round-the-world trunk line. The evolution of the network makes this pattern less readable in 2006 because of more many transversal links between different regions, resulting in a larger number of bridge ports. We interpret such changes as the effect of technological improvement in the shipping industry that “made the world smaller” (Levinson, 2006) in a context of increased globalisation and shrinking transport costs. Distances between regions become less constraining, as the maritime network as a whole increased in size, length, density, traffic, and connectivity (Ducruet and Notteboom, 2010).

[Insert Figure 7 about here]

In the 1996 GDL, the ports of Rouen, Montoir (outport of Nantes), and Amsterdam have a very interesting bridge role between Western Europe, West Africa (Abidjan, Dakar) and Latin America (Pointe-à-Pitre, Fort de France, Oranjestad), illustrating the heritage of historical linkages. Other examples include Vigo and Tenerife connecting Europe with Peru (Callao) and Brazil (Vitoria) respectively. Long-term trading links between Germany and Brazil complement this Latin proximity by the position of Bremen. The internal composition of maritime regions is also interesting, as we observe for Western Europe a strong connectivity among Scandinavia-Baltic ports recalling the ancient pattern of the Hanseatic League.

Several bridge ports are identical in the 2006 GDL, such as Tenerife, Liverpool, Zeebrugge, Vigo, Callao (Lima), Jakarta, and Penang, while some former ones still appear inside maritime regions, such as Thamesport (London), Venice, and Amsterdam. This shows the goodness of the chosen thresholds for comparing different years. Such ports have maintained

⁵ The clustering coefficient calculates for each node the ratio between the number of observed triplets and the maximum number of triplets based on the links to adjacent nodes. In other words, it is the probability for the

neighbours of a given node to be connected. Its formula is as follows:
$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i - 1)}$$

their long-distance connectivity outside of large hubs. Most of them are also large adjacent coastal cities that lost their central role in shipping networks due to port competition and inner pressures at the port-city interface (Ducruet and Lee, 2006). This analysis shows that such port cities keep their roles at lower levels of the hierarchy. Other examples of bridge ports include Sines (Portugal) that was the focus of a container hub strategy by Port of Singapore Authority (PSA) since 2004, and Guangzhou in China having a long history of maritime trade and a diversified foreland (Wang and Ng, 2010). Conversely, some ports already present in 1996 have become bridge ports, such as Baltimore, Brisbane, and Montevideo. North American ports that did not appear very much in 1996 stand out as important bridges between the Americas and Africa (e.g. Jacksonville, Philadelphia). The central position of Zeebrugge in the giant component clearly corresponds to the map of its overseas connections provided by official sources⁶.

Results for the graph of all linkages (GAL) provide a complementary picture. Overall, the giant components form a dominant North Europe - East Asia link. Caribbean and South European ports act as intermediaries in 1996. In 2006 relations are more direct between the two main regions, but some Southern European and Mediterranean ports are now included in the European pole (Ceuta, Algiers, Poti, Ravenna). One common aspect with the GDL is the dominance of Scandinavia-Baltic ports in Europe, together with Dublin and Liverpool. In the Asian region, there is balance between North and South, with Incheon and Surabaya appearing at both years: they have in common to act as regional hubs thus complementing the domination of nearby Busan and Singapore (Ducruet et al., 2010b). We can also interpret such results by the influence of regional integration in the formation of maritime regions. As Europe becomes more integrated, the distinction between North and South has lost grounds; on a world level, some ports get closer due to continued diffusion of containerization and international trade growth. Yet, one may wonder whether the evolution of the network followed similar paths since the global financial crisis that started in 2007. We can hypothesize that trade decrease and transport increase would result in a weaker network where distance and proximity play the major role in the formation of maritime regions.

4. Conclusion

⁶ <http://www.zeebruggeport.be/en/node/491> (Accessed December 2010)

Looking at port systems and maritime regions through methods of network science brings interesting results for geography. Based on the assumption that degree centrality (i.e. number of links to other ports) is a key indicator of the situation of ports in a maritime network, topological decomposition appears as a relevant tool to verify the existence of coherent subgroups. Relations among largest centrality ports confirm the crucial importance of the Europe-Asia link, with Singapore as a vital bridge, while this simple scheme was strengthened in recent years by the development of other intermediate hubs (e.g. Mediterranean). A comparison between different years (1996 and 2006) also underlined important shifts of centrality from Europe to Asia. The study of links among smaller ports offers a somewhat different perspective about the geography of the network. While most subgroups of smaller ports are explained by spatial proximity, we find several exceptions revealing the permanency of specialized long-distance links between Western European ports and their former colonies, and some other subgroups simply based on a mix of other logics such as bunkering-related circulations. As smaller ports agglomerate with each other, they form a giant component with a global coverage. Therefore, the connectivity of the global network is maintained even without large hubs, due to the permanency of longstanding links between maritime regions. Bridge ports are often large coastal metropolises having lost their central function in shipping networks to global hubs, but which maintain long-distance trading links. The global network of regional hubs is thus a specific entity corresponding to other logics than sole hierarchy and proximity.

The outcomes of this research are important for network studies in general because they confirm the belonging of container shipping to the wider category of complex networks having both scale-free and small-world properties. The emergence of a giant connected component also makes container shipping networks comparable with other networks in biology and sociology. For port and maritime geography, the detour via network analytical methods may contribute to rethink the traditional definition of port systems and maritime regions as simple alignments of ports along a coastline. The reality is indeed more complex, as port systems emerge at various levels from the local to the global, sometimes making Euclidian distance irrelevant. Perhaps, further analysis based on the frequency of links would provide different results than the topological approach. Researches on systems of cities in urban geography using similar methods would also prove helpful for the study of ports and maritime networks. In terms of policy implications, the identification of tightly connected ports can be a useful guide to address inter-port cooperation guidelines. The identification of

bridge ports has also strong implications for ports willing to evaluate their position and specialization in transport and logistics chains.

References

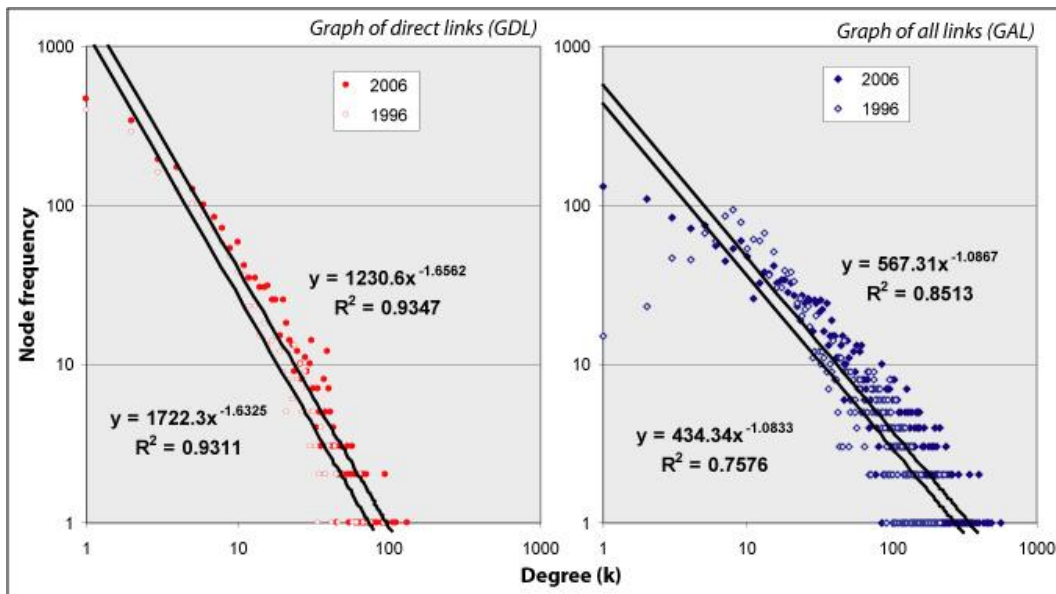
- Alvarez-Hamelin, J.I., Dall'Asta, L., Barrat, A., Vespignani, A. (2006) K-core decomposition: a tool for the visualization of large scale networks. *Advances in Neural Information Processing Systems*, 18: 41-50.
- Auber, D. (2003) Tulip - a huge graph visualization framework. In: P. Mutzel and M. Junger (eds.) *Graph Drawing Software, Mathematics and Visualization Series*. Springer Verlag.
- Auber, D., Chiricota, Y., Jourdan, F., Mélançon, G. (2003) Multiscale visualization of small world networks. In: *INFOVIS '03: Proceedings of the IEEE Symposium on Information Visualization*, pp. 75-81.
- Bader, G., Hogue, C. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4(1): 2.
- Barabási, A.L. (2002) *Linked: The New Science of Networks*. Basic Books, 1st Ed., May.
- Barabási, A.L., Albert, R. (1999) Emergence of scaling in random networks, *Science*, 286, 509-512.
- Batagelj, V., Zaversnik, M. (2002) Generalized cores. arxiv cs.DS/0202039.
- Baur, M., Brandes, U., Gaertler, M., Wagner, D. (2004) Drawing the as graph in 2.5 dimensions, *Proc. Graph Drawing 2004*, 43-48.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.U. (2006) Complex networks: Structure and dynamics, *Physics Reports*, 424, 175-308.
- Boccaletti, S., Ivanchenko, M., Latora, V., Pluchino, A., Rapisarda, A. (2007) Detection of complex networks modularity by dynamical clustering. *Physical Review E*, 75.
- Boutin, F., Thievre, J., Hasco, M. (2006) Focus-based filtering + clustering technique for power-law networks with small world phenomenon. In: *VDA'06: Visual Data Analysis - SPIE-IS&T Electronic Imaging*, 6060(001-012). SPIE P., U.S.
- Brandes, U., Erlebach, T. (2005) *Network Analysis: Methodological Foundations*. Lecture Notes in Computer Science, Springer.
- Colizza, V., Flammini, A., Serrano, M.A., Vespignani, A. (2006) Detecting rich-club ordering in complex networks, *Nature Physics*, 2: 110-115.
- Deng, W.B., Long, G., Wei, L., Xu, C. (2009) Worldwide marine transportation network: Efficiency and container throughput, *Chinese Physics Letters*, 26(11): 118901.
- Ducruet, C. (2008) Hub dependence in constrained economies: the case of North Korea, *Maritime Policy and Management*, 38(4): 377-394.

- Ducruet, C. (2009) Port regions and globalization, in: Notteboom, T.E., Ducruet, C., Langen de, P.W. (eds.), *Ports in Proximity: Competition and Cooperation among Adjacent Seaports*, Aldershot, Ashgate, pp. 43-51.
- Ducruet, C., Lee, S.W. (2006) Frontline soldiers of globalisation: Port-city evolution and regional competition, *Geojournal*, 67(2): 107-122.
- Ducruet, C., Lee, S.W., Ng, K.Y.A. (2010b) Centrality and vulnerability in liner shipping networks: Revisiting the Northeast Asian port hierarchy, *Maritime Policy and Management*, 37(1): 17-36.
- Ducruet, C., Notteboom, T.E., De Langen, P.W. (2009) Revisiting inter-port relationships under the New Economic Geography research framework, in: T.E. Notteboom, C. Ducruet and P.W. De Langen (eds), *Ports in proximity: Competition and coordination among adjacent seaports*, Aldershot: Ashgate, pp. 11-28.
- Ducruet, C., Notteboom, T.E. (2010) The worldwide maritime network of container shipping: spatial structure and regional dynamics, *GaWC Research Bulletin*, 364: <http://www.lboro.ac.uk/gawc/rb/rb364.html>
- Ducruet, C., Rozenblat, C., Zaidi, F. (2010a) Ports in multi-level maritime networks: Evidence from the Atlantic (1996-2006), *Journal of Transport Geography*, 18(4): 508-518.
- Essam, J.W. (1980) Percolation theory. *Reports on Progress in Physics*, 43(7): 833.
- Ferrari, C., Parola, F., Morchio, E. (2006) Southern European ports and the spatial distribution of EDCs, *Maritime Economics and Logistics*, 8(1): 60-81.
- Fleming, D.K., Hayuth, Y. (1994) Spatial characteristics of transportation hubs: Centrality and intermediacy, *Journal of Transport Geography*, 2(1): 3-18.
- Gaertler, M., Patrignani, M. (2004) Dynamic analysis of the autonomous system graph. In: *IPS 2004, International Workshop on Inter-domain Performance and Simulation*, pp. 13-24.
- Gastner, M.T., Newman, M.E.J. (2004) The spatial structure of networks, *The European Physical Journal B*, 49: 247-252.
- Girvan, M., Newman, M.E.J. (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99: 8271-8276.
- Gouvernal, E., Slack, B., Franc, P. (2010) Short sea and deep sea shipping markets in France, *Journal of Transport Geography*, 18(1): 97-103.
- Granovetter, M. (1973) The strength of weak ties. *American Journal of Sociology*, 78(6): 1360-1380.
- Guerrero, D. (2010) *Les aires d'influence des ports de la France entre réseau et gravitation*, Unpublished PhD Dissertation in Geography, University of Paris 7 Denis Diderot.
- Hall, P.V., Jacobs, W. (2010) Shifting proximities: The maritime ports sector in an era of global supply chains, *Regional Studies*, 44(9), 1103-1115.

- Hayuth, Y. (1981) Containerization and the load center concept, *Economic Geography*, 57: 161-176.
- Heer, J., Boyd, D. (2005) Vizster: Visualizing online social networks. In: *INFOVIS '05: Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization*, 5.
- Herman, I., Mélançon, G., Scott Marshall, M. (2000) Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1): 24-43.
- Janson, S., Knuth, D.E., Luczak, T., Pittel, B. (1993) The birth of the giant component. *ArXiv Mathematics e-prints*, September.
- Kansky, K. (1963) The structure of transportation networks, Research Paper 84, Chicago: University of Chicago.
- Lemarchand, A., Joly, O. (2009) Regional integration and maritime range, in Notteboom, T.E., Ducruet, C., De Langen, P.W. (Eds) *Ports in Proximity: Competition and Coordination Among Adjacent Seaports*, Aldershot: Ashgate, pp. 87-98.
- Levinson, M. (2006) *The box: How the shipping container made the world smaller and the world economy bigger*, Princeton & Oxford: Princeton University Press.
- Luxburg, U. (2007) A tutorial on spectral clustering. *Statistics and Computing*, 17(4): 395-416.
- Newman, M.E.J. (2004) Fast algorithm for detecting community structure in networks. *Physical Review E*, 69: 066133.
- Notteboom, T.E. (2009) Complementarity and substitutability among adjacent gateway ports, *Environment and Planning A*, 41(3): 743-762.
- O'Kelly, M.E. (1998) A geographer's analysis of hub-and-spoke networks, *Journal of Transport Geography*, 6(3), 171-186.
- Olivier, D., Slack, B. (2006) Rethinking the port, *Environment and Planning A*, 38(8): 1409-1427.
- Päivinen, N. (2005) Clustering with a minimum spanning tree of scale-free-like structure. *Pattern Recogn. Lett.*, 26(7): 921-930.
- Perpillou, A. (1959) *Géographie de la circulation*. Paris: Centre de Documentation Universitaire.
- Robinson, R. (1976) Modelling the port as an operational system: A perspective for research, *Economic Geography*, 52: 71-86.
- Robinson, R. (2002) Ports as elements in value-driven chain systems: The new paradigm, *Maritime Policy and Management*, 29(3): 241-255.
- Rodrigue, J.P., Notteboom, T.E. (2010) Foreland-based regionalization: Integrating intermediate hubs with port hinterlands, *Research in Transportation Economics*, 27(1): 19-29.

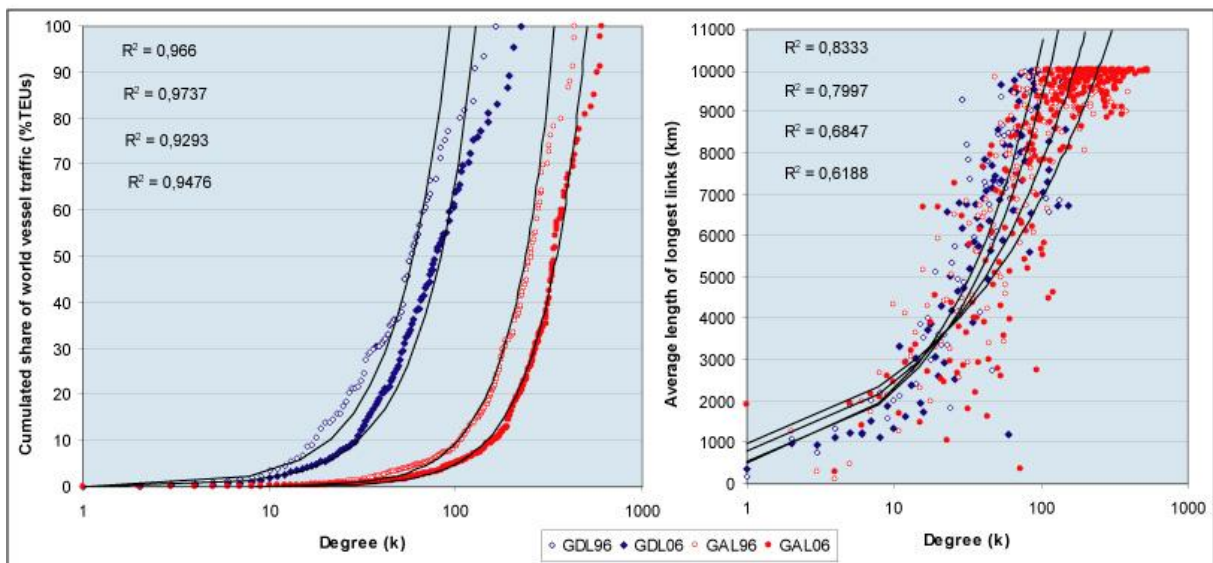
- Slack, B. (1993) Pawns in the game: Ports in a global transportation system, *Growth and Change*, 24(4): 579-588.
- Spielman, D.A., Teng, S.H. (1996) Spectral partitioning works: Planar graphs and finite element meshes. In: *IEEE Symposium on Foundations of Computer Science*, pp. 96-105.
- Teboul, R., Cuenca, C., Richaud, A. (2000) *La question urbaine dans l'histoire de la pensée économique*, Paris: L'Harmattan.
- Todd, D. (1993) The interplay of trade, regional and technical factors in the evolution of a port system: The case of Taiwan, *Geografiska Annaler B*, 75(1): 3-18.
- van Ham, F., van Wijk, J.J. (2004) Interactive visualization of small world graphs. In: *INFOVIS 2004. IEEE Symposium on Information Visualization*, pp. 199-206.
- Vigarié, A. (1964) *Les grands ports de commerce de la Seine au Rhin*, Paris: SABRI.
- Vigarié, A. (1979) *Ports de Commerce et Vie Littorale*, Paris: Hachette.
- Wang, J.J., Ng, K.Y.A. (2010) The geographical connectedness of Chinese seaports with foreland markets: a new trend? *Tijdschrift voor Economische en Sociale Geografie*, DOI: 10.1111/j.1467-9663.2010.00600.x
- Watts, D.J., Strogatz, S.H. (1998) Collective dynamics of small-world networks, *Nature*, 393, 440-442.
- Wu, A.Y., Garland, M., Han, J. (2004) Mining scale-free networks using geodesic clustering. In: *KDD '04: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 719-724.
- Wuchty, S., Almaas, E. (2005) Peeling the yeast protein network. *Proteomics*, 5(2): 444-449.
- Zaidi, F. (2010) *Analysis, structure and organization of complex networks*, Unpublished PhD Dissertation in Computer Sciences, University of Bordeaux I.

Figure 1: Degree distribution by graph type, 1996-2006



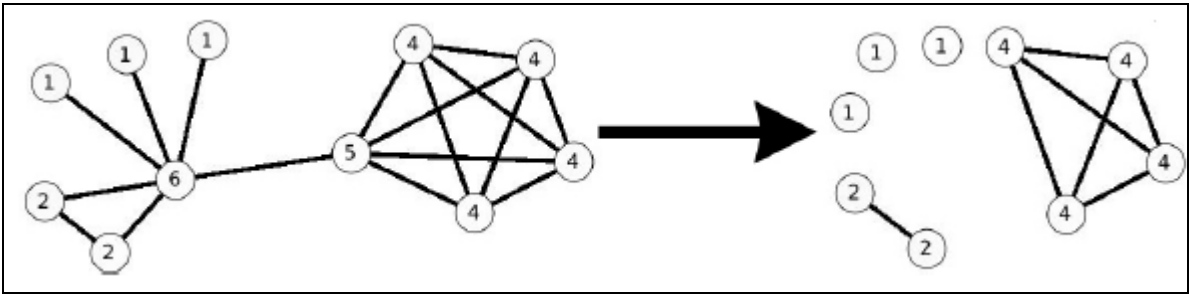
Source: realized by authors based on LMIU data and TULIP software

Figure 2: Traffic and distance over degree by graph type, 1996-2006



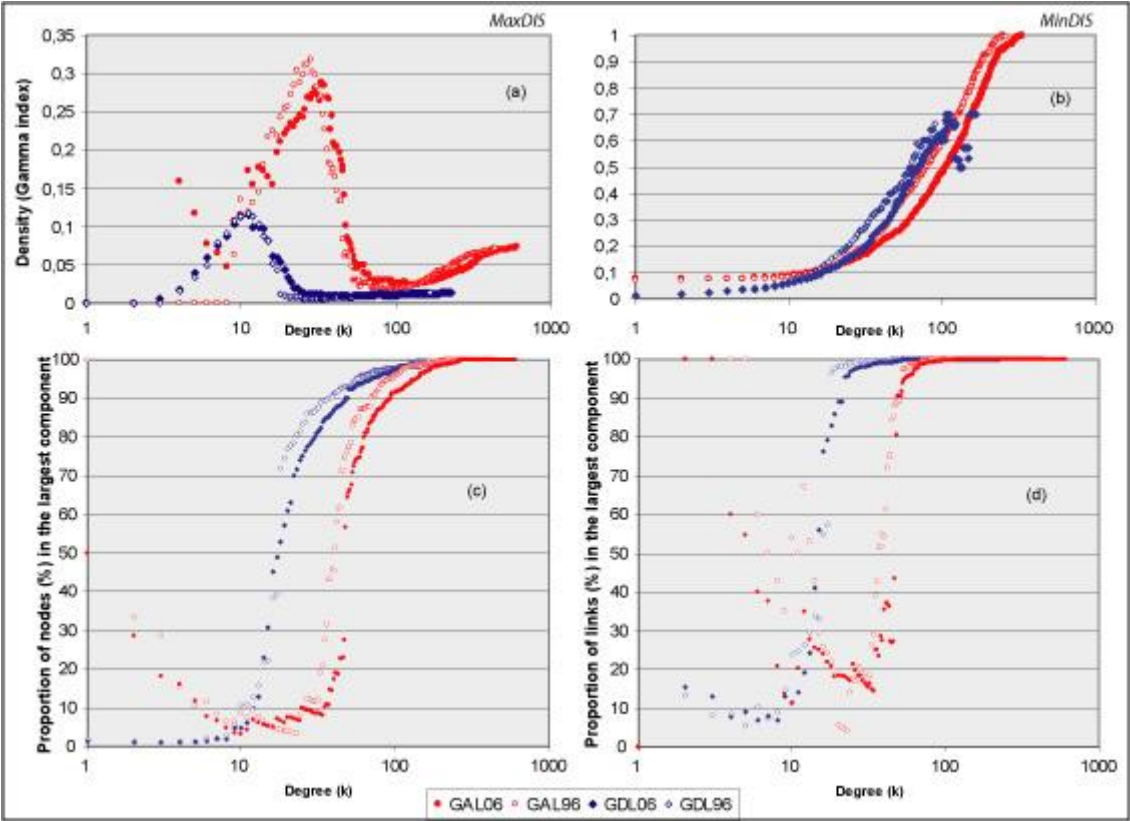
Source: realized by authors based on LMIU data

Figure 3: The topological decomposition method



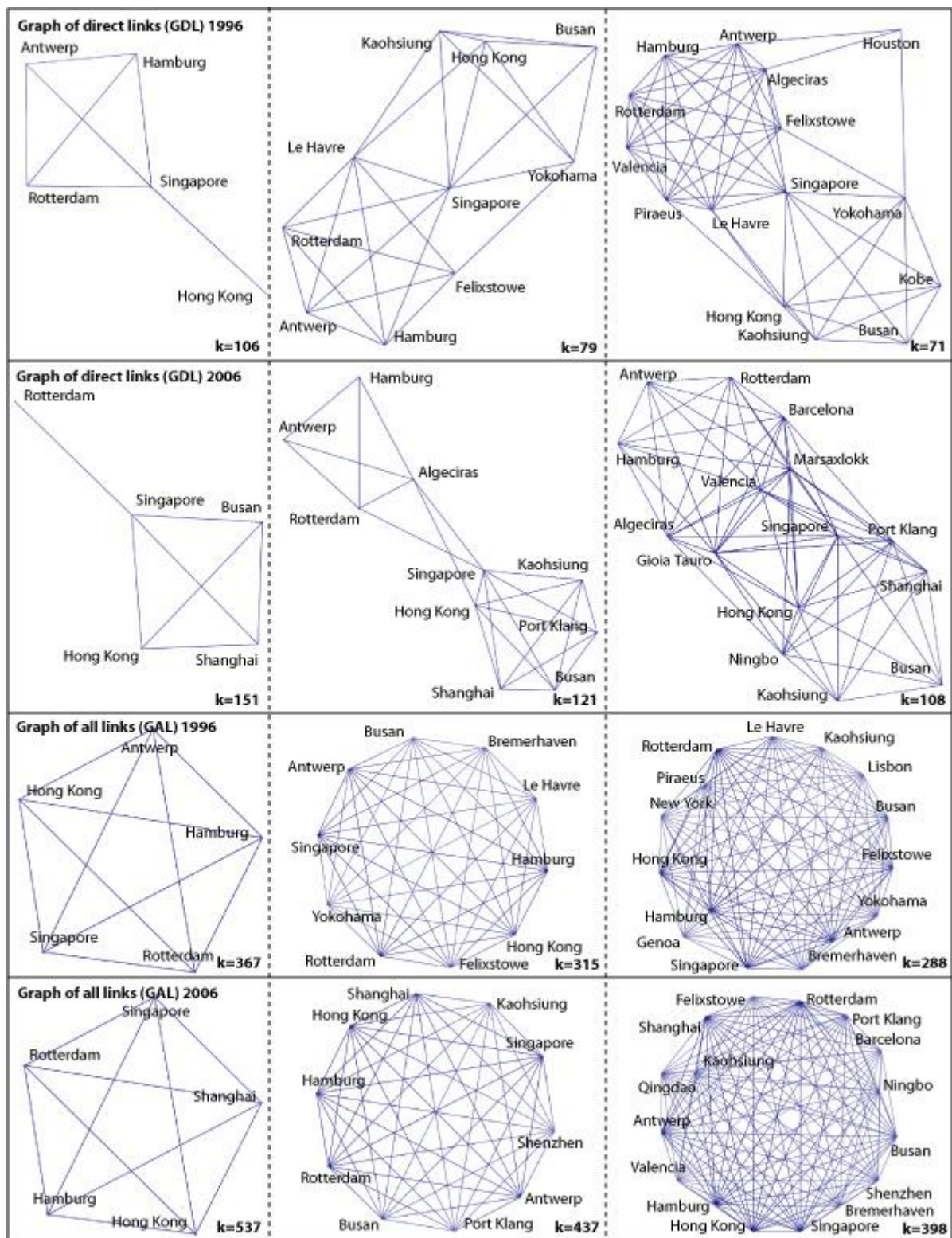
Source: Zaidi (2010)

Figure 4: Description of MaxDIS and MinDIS results by graph type, 1996-2006



Source: realized by authors based on LMIU data and TULIP software

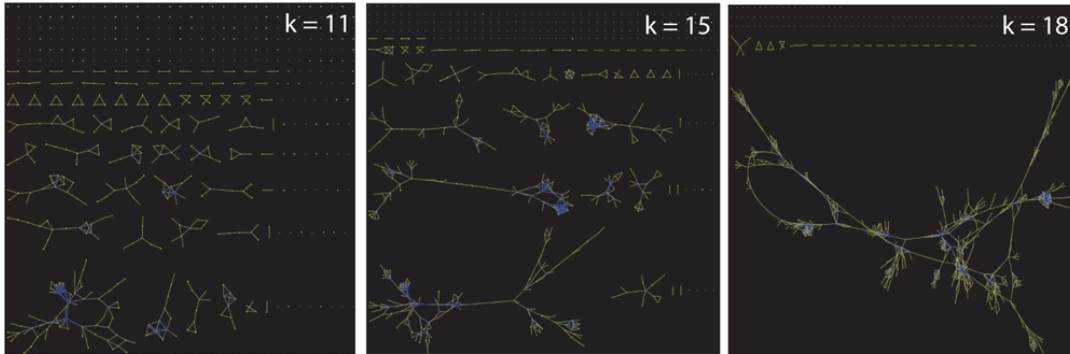
Figure 5: Illustration of selected MinDIS, 1996-2006



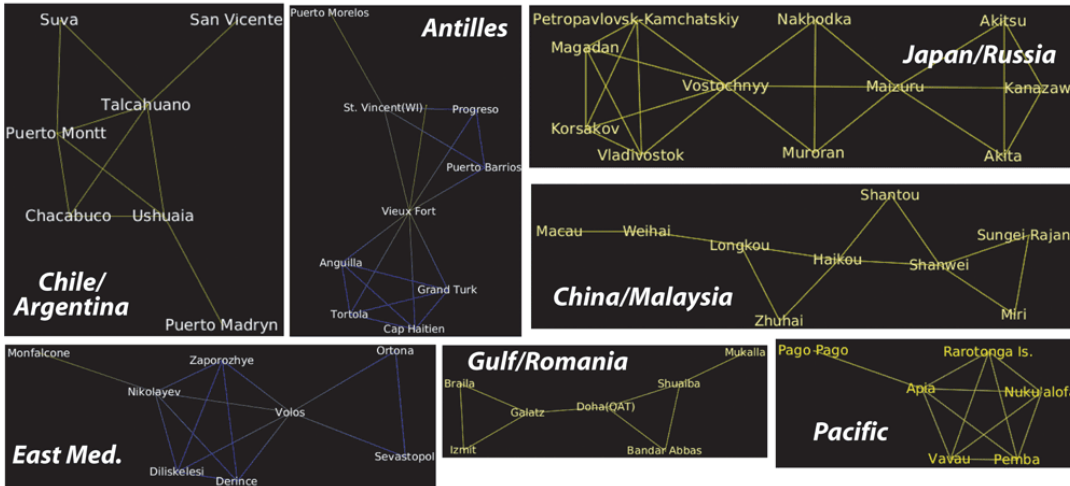
Source: realized by authors based on LMIU data and TULIP software

Figure 6: Illustration of selected MaxDIS

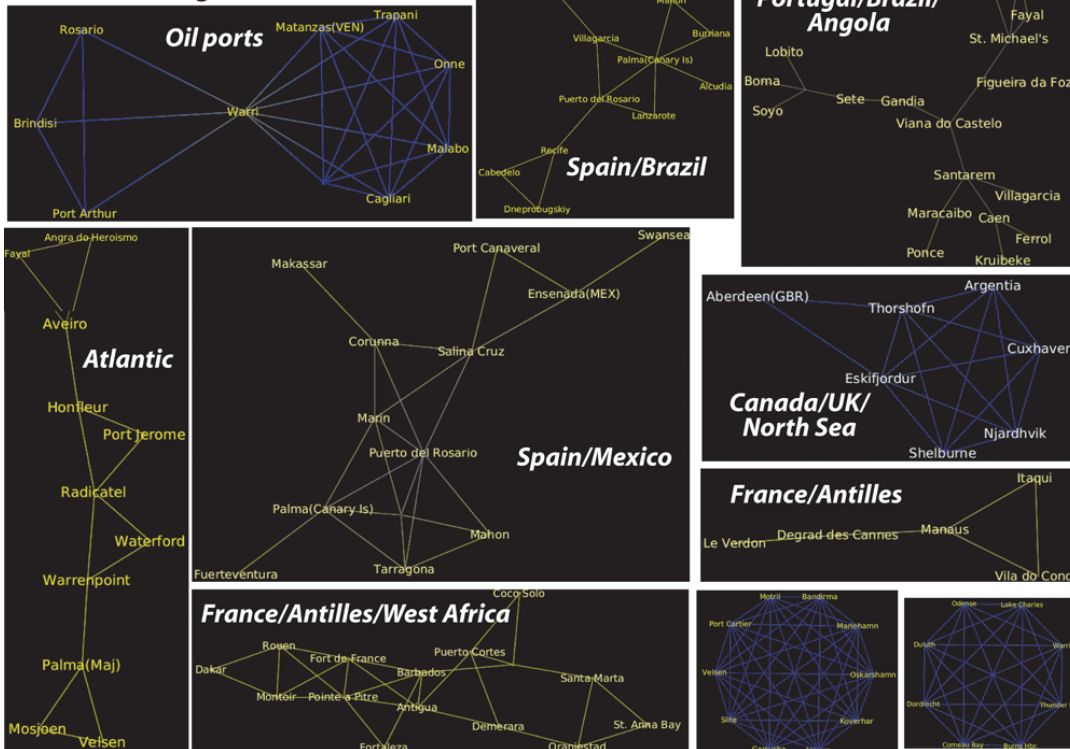
Example of results at different thresholds (GDL, 1996)



Spatial proximity



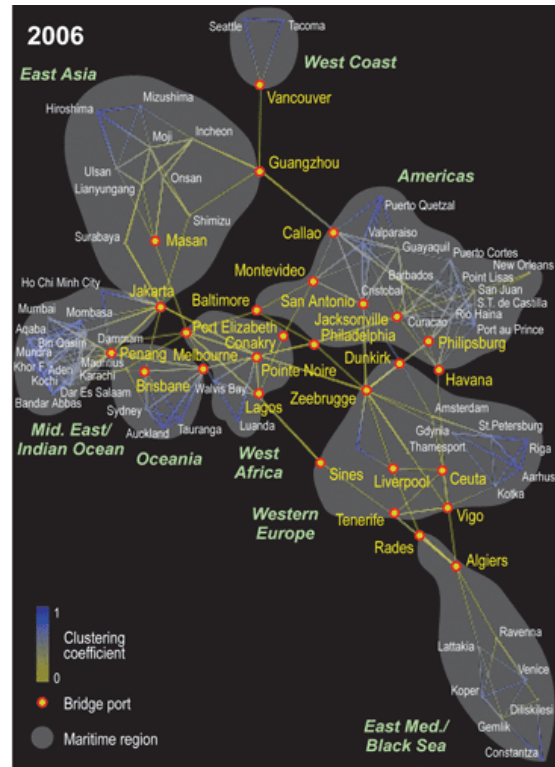
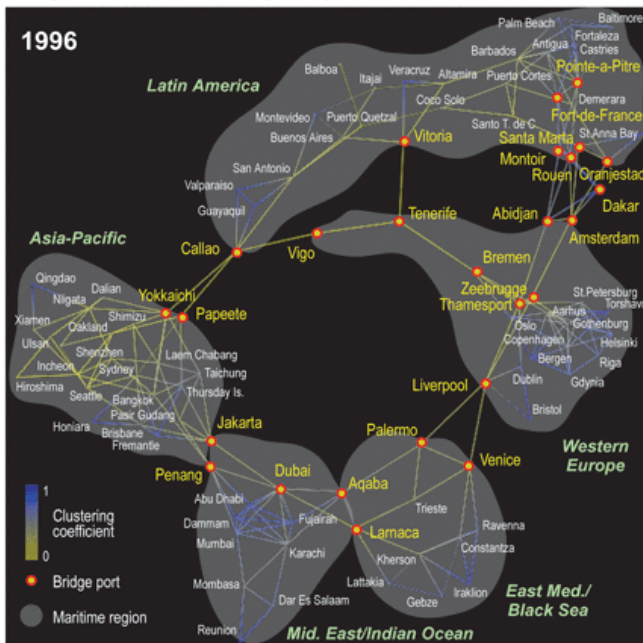
Functional linkages



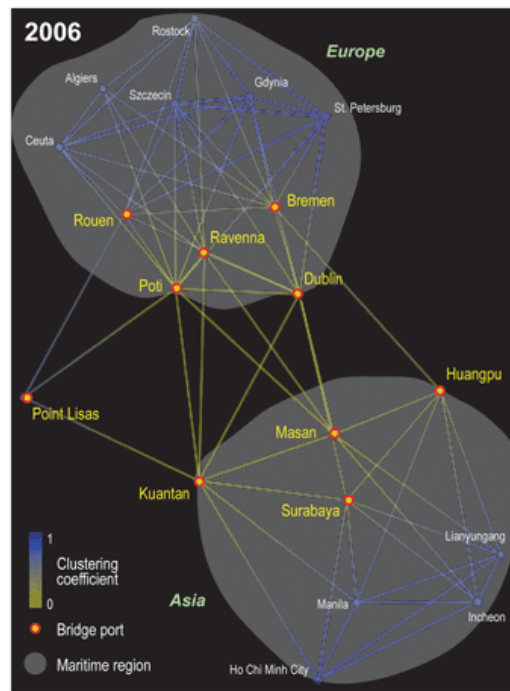
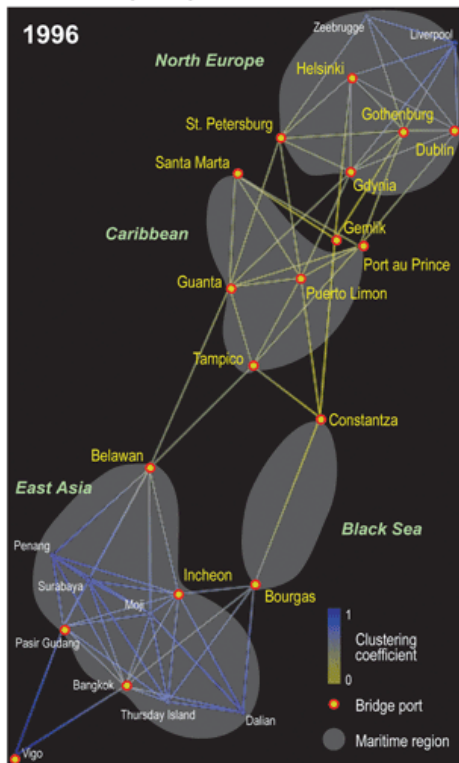
Source: realized by authors based on LMIU data and TULIP software

Figure 7: Bridges and communities within giant components

Graph of direct links (GDL)



Graph of all links (GAL)



Source: realized by authors based on LMIU data and TULIP software