Measuring the effect of distance on the network topology of the Global Container Shipping Network

Dimitrios Tsiotas, César Ducruet

Measuring the effect of distance on the network topology of the Global Container Shipping Network

Draft version of the paper published in Nature Scientific Reports (2021), 11: 21250

Dimitrios Tsiotas[1*] and César Ducruet[2]
1. Department of Regional and Economic Development, Agricultural University of Athens, Greece, Nea Poli, Amfissa, 33100, Greece
2. French National Centre for Scientific Research (CNRS), UMR 7235 EconomiX, Nanterre, France
E-mails: tsiotas@aua.gr; cesar.ducruet@economix.fr
* Corresponding author

Abstract
This paper examines how spatial distance affects network topology, on empirical data concerning the Global Container Shipping Network (GCSN). The GCSN decomposes into 32 multiplex layers defined at several spatial levels by successively removing connections of smaller distances. This multilayer decomposition approach allows studying the topological properties of each layer as a function of distance. The analysis provides insights into the hierarchical structure and (importing and exporting) trade functionality of the GCSN, hub connectivity, several topological aspects, and the distinct role of China in the network's structure. It also shows that bidirectional links decrease with distance, highlighting the importance of asymmetric functionality in carriers' operations. It further configures six novel clusters of ports concerning their spatial coverage. Finally, it reveals three levels of geographical scale in the structure of GCSN, where the network topology significantly changes; the neighborhood (local connectivity), the scale of international connectivity (mesoscale or middle connectivity), and the intercontinental market (large scale connectivity). The overall approach provides a methodological framework for analyzing network topology as a function of distance, highlights the spatial dimension in complex and multilayer networks, and provides insights into the spatial structure of the GCSN, which is the most important market of the global maritime economy.

Keywords: multilayer networks; maritime networks; global trade market; complex network analysis.

Introduction

Spatial networks, and their underlying socioeconomic structures, are described by a symbiotic relation: on the one hand, networks facilitate trade and other socioeconomic interactions supporting regional and economic development (1,2); on the other hand, the derived demand in the associated regional markets supports the development process of spatial and transportation networks, which are structures of considerable sunk costs affecting the future developmental dynamics of the spatial units participating in these networks (1). Geography plays a crucial role in the development and evolution of spatial and transport networks. It is related to the friction of movements between places and can be interpreted in terms of transportation costs (1,3). Due to spatial impedance, the connection probability between nodes usually decreases as distance increases (1-3). In addition, new nodes entering a spatial network preferentially connect nearby hubs, which already possess a high level of transport infrastructures (or amenities) (1,4,5).

The spatial property is therefore determinative for the structure and functionality of networks and affects their organization, evolution and growth, topology, and traffic (2-4,6). For instance, spatial networks are more likely to develop lattice-like than random, hub-and-spoke, or small-world topologies because spatial constraints usually impose planarity to these networks (3). Among transportation networks, road, rail, and other "technical" networks are considered as planar networks, whereas maritime and airline networks can

exceed planarity and develop more complex hierarchical structures (3,7). Also, in spatial networks, hubs are more likely to appear in central geographical places than in peripheral locations and to undertake the major traffic load of distant connectivity. At larger degree values, this property drives the correlation between node strength and degree linearly (3,6). In terms of functionality, spatial networks are usually quite different than those not embedded in space (3,6,8). While in social networks, nodes are actors that are directly operating and ruled by their behavioral and cognitive forces, spatial units – such as transport terminals or cities – function as the collective result of several actors (2). Spatiality also favors the development of "regional" rather than "global" hubs, with smaller degree and much larger traffic than if global hubs were available (3). In terms of community detection, spatial constraints reinforce the geographical properties of community configuration, which is mainly driven by adjacency and neighborhood forces

The spatial constraints are applicable at all levels of aggregation (neighborhood, local, global) in the network structure by introducing a distance cost in the development of connections (3). However, their effect on network topology is not the same for all real-world networks as it depends on their functionality. For instance, empirical research (3,9-13) has shown that road networks usually develop homogeneous and broadly invariant structures embedded in planar spaces, where severe constraints are applied to the node degree (that usually ranges around the average value $\langle k \rangle \approx 2.5$), degree distribution (that is generally peaked around its average value), average path length (that is expected to be large), and clustering coefficient (that is larger than that of random counterpart networks). Depending on the representation (where nodes usually represent road intersections or spatial units) and geographical scale, road networks are usually connective graphs described by mesh-like and lattice-like topologies relevant to two-dimensional lattices. Accordingly, railway networks (3,14-17) are usually embedded in planar spaces, with a similar topology than that of road networks, but vary in terms of nodes (stations, stops, or spatial units) and linear arrangements (bus-like structures). On the other hand, maritime networks (18) can display different properties than land transportation networks and can overcome planarity due to their attribute to conduct transportation on the sea surface instead of line infrastructure channels. This structural property is reflected on (3,7,18,19-23): higher average degree than land transportation networks, much larger average clustering coefficient than of counterpart lattices, and degree distributions described by power-law patterns that often are typical of standard non-spatial networks. Depending on the representation (where nodes usually represent ports) and geographical scale, maritime networks are not by default connective graphs (where components are sub-networks representing local markets) and are described either by composite ring-like or hub-and-spoke topologies. Amongst the various types of maritime transport, cargo ship networks (7,8,21,24-26) have received foremost attention in the literature due to their economic importance, as more than 80% of world trade volumes are carried by sea. In such networks, the weight configuration is very asymmetric ($w_{ij} \neq w_{ji}$), with almost 60% of all linked pairs existing only in one direction. This is mainly explained by the fact that container shipping is operated through asymmetric pendulum services with different legs back and forth within and between regions (24). Finally, airline networks are non-planar spatial networks that are privileged to conduct transportation in the 3d-space instead of line infrastructure channels. This property is reflected on (3,7,27-33): the shape of the degree distribution, which fits power-law curves and is heavy-tailed with a cutoff, due to the physical constraints on the maximum number of connections that a single airport can handle; the average path length, which is smaller than of counterpart lattices; the existence of important correlations between topology and geography, implying that larger airports have connections with larger traffic and also more distant connections; the large clustering, which is slightly decreasing at large degrees (k), illustrating the role of large airports that

provide non-stop connections to different regions which are not interconnected among them; and a clique-like configuration, because large weights are concentrated on links between large airports. Depending on the representation (where nodes usually represent airports) and geographical scale, airline networks are also not by default connective graphs and are described by hub-and-spoke topology with heterogeneous topological properties, where hubs have high connectivity, large traffic load, and long-distance connections. Airline networks emerge at two different geographical scales (3), global and domestic, where the first one defines two different groups of travel distances (intercontinental and regional).

Besides the complexity describing in scientific terms the effect of space on the topology of spatial networks, this relationship is further submitted to different disciplinary foci (2,6). For instance, economists and regional scientists tend to conceptualize space in econometric terms and to manage the spatial dimension either as a separate econometric variable or through a given cost or facilitation (e.g. trade cost, tariff, time); geographers, engineers, and spatial planners consider space as geospatial information expressed by geographical coordinates and territorial attributes (e.g. port city population, socio-economic features); and physicists and mathematicians analyze the topological differences captured before and after the embedding of networks in metric spaces (2,6). Despite this polyphony, it is possible to synthesize the literature on spatial networks and related constraints as follows (2,3,6,33): (a) through the examination of the shape of degree distribution, where a bell-shaped configuration peaked around the average degree implies the effect of planarity in the configuration of network topology; (b) in reference to theoretical (null) graph models, according to which topological aspects and measures of real-world or other empirical networks are compared with the results of counterpart null models of lattice-like, random-like, ring-like, small-world, hub-and-spoke, and other known topologies; (c) sometimes through spectral pattern recognition based on the sparsity (spy plot) patterns of the adjacency and connectivity matrices of spatial networks, and (d) by examining correlations between the measure of degree and the betweenness centrality (i.e. network intermediacy) and strength (i.e. traffic volume), to detect whether hubs (i.e. nodes of high connectivity) also intermediate to the majority of paths and undertake the major traffic load.

Although very informative and insightful, existing methods generally study the relationship between space and network topology within a holistic and structurally undisturbed context, thus conceptualizing networks as integrated and indivisible structures. This conceptualization is quite representative of such networks as systems of socioeconomic interaction (1-3), each system having a single balanced macroeconomic behavior. However, such networks also include elements of diverse importance and functionality (e.g. hubs, spokes, strong and weak ties, etc.) which are often hierarchically organized (3). Decomposition techniques in the research fields of community detection (33) and vulnerability analysis (21)are not yet applied to unravel the relationship between space and network topology.

In such respect, this paper decomposes a spatial network of prominent socioeconomic importance, the Global Container Shipping Network (GCSN), by successively removing edges (links) of shorter distances and by assigning the remaining links into layers. It thus configures a multilayer network model consisting of a family of layers, where each one includes links with lengths longer than a certain distance. This approach allows configuring networks of separate topologies for edge-groups of different spatial ranges and thus attempts to provide insights into the effect of space on the topological features of a spatial network, at different levels of spatial scale. To do so, the analysis builds on multilayer network modeling (21,34,35), which allows describing a system as a collection of layers instead of a single network. Some indicative approaches on multilayer analysis of spatial networks include the analysis of air-sea global networks (36), of the bi-layered Greek maritime

network (37), and of the intermingling of six maritime cargo layers (25). Generally, these and other similar works (21,38-39) build on a cross-layer conceptualization that is implemented through comparisons between layers of a multilayer spatial network model. Being inspired by such conceptualization, this paper generates a multilayer network from a source (original) network, where all layers consist of the same nodes (multiplex network model) compared with the original network but where each layer includes a subset of the original links exceeding a certain distance. This multiplex network model is a consistent collection of layers where the topological features of the original network can be expressed as a function of distance along the cross-layer direction (axis).

The analysis in this paper is applied to the Global Container Shipping Network (GCSN), first because it is a network of prominence economic importance, consisting of international trade flows of all over the globe (8,25). Maritime transportation is the dominant mode of transport in international trade, where around 80% of global trade volume and over 70% of global trade value are transported by sea and handled by ports (21). In the last half-century, the spatial characteristics of maritime networks changed tremendously (1). In particular, the GCSN has a complex structure driven by the counterbalance between spatial distance, market forces, and technology, which is related to (a) the transportation cost between places (2,41); (b) the availability of container transportation, which lowers the effects of distance due to massification (41); (c) the increase of trade demand due to urbanization and global manufacturing shift (25); (d) the changing composition of trade and the increasing containerization of all sorts of commodities of which bulks and even automobiles (42); (e) the growth in the size of container ships achieving economies of scale (25,41); and (f) the transformation and liberalization of the maritime market, which changed from a labor-intensive to a capital-intensive industry (18,26).

Also, the global shipping network is an assemblage of different types of liner shipping services, each of them being characterized by a varying degree of shipping distance (43). Deep-sea services occur over longer edges, including round-the-world services and pendulum services, the latter being deployed on both inter and intra-continental legs. Short-sea services occur over shorter edges and they include feeder services between hub and spokes, coastal shipping, and the short-sea service itself usually bound to a closed sea or within a maritime range (44). The interlining service is a specific type of transshipment activity whereby two deep-sea services exchange containers at a given hub. The emergence of transshipment hubs in the mid-1990s led to the increasing use of feeder services on short distances, while the progress of regional integration favored the multiplication of short-sea shipping (see 45). Concrete examples of transshipment hubs include ports with a high ratio of sea-sea transshipment, such as Singapore, Salalah (Oman), Gioia Tauro (Italy), and Kingston (Jamaica), often located within a maritime basin with the least deviation distance from the trunk line (46). Intra-continental pendulum services are deployed through multiple calls among ports in proximity, such as at the Northeastern seaboard of America, the Le Havre-Hamburg range in North Europe, and the Japanese megalopolis from Fukuoka to Tokyo. The increasing ship size, up to the present era of mega-ships, motivated longer voyages and fewer port calls. Lastly, the GCSN is specific as vessels need to avoid coastlines.

Such elements motive this paper to question whether rapid technological improvements, increased globalization, and lowering costs contributed to modify – or even to nullify – the effect of distance on the topological structure of the GCSN. On the one hand, recent research demonstrated the persistence of gravitational properties affecting the GCSN (18), as larger (port) cities connect more with each other but less over longer distances. Such a result is in line with international trade studies examining the “puzzling effect of distance” on bilateral exchanges (66), one of the most important findings in economics. On the other

hand, the introduction of containers is believed to have lowered distance impacts since the 1980s (67), an effect which had, however, been challenged by China's trade growth (68). While the search for distance effects in networks now has a long tradition in natural (69,70) and social (71,72) sciences, the existing literature usually considers the network as one single, aggregated entity. We propose to go one step further, by decomposing the network into layers of varying link distance. Such an angle of attack has the potential to highlight the ability of ocean carriers to overcome distance, and to better understand the centrality and functionality of hubs, from the local to the global.

The remainder of the paper is organized as follows: the methods section describes the modeling and the methods used for the analysis of the GCSN; the next section shows the results of the network and empirical analysis applied to the multilayer GCSN; and finally, in the last section, conclusions are given.

Methods

■ Conceptual framework, modeling, and data

The Global Container Shipping Network (GCSN) is modeled to a multilayer graph $M(\Gamma, X=\emptyset)$ (34,35), where $\Gamma=\{G_0, G_1, G_2, \ldots, G_{31}\}$ is a family of 32 layers generated from the original layer G_0, and $X=\emptyset$ is the null set of interlayer connections (implying that all GCSN connections are within-layer links, and no connection between layers exist). The original layer $G_0(V,E)$ was constructed on the annual direct routes between the GCSN ports and particularly based on 282'785 movements of fully cellular container vessels in 2016. Ports are nodes (belonging to the set $V \mid |V| = n$) and links (belonging to the set $E \mid |E| = m$) inter-port voyages (18,43). The GCSN has a dual weighted representation: the first one is distance-weighted ($w_{ij}^{(a)} = d_{ij}$), where links are weighted by their orthodromic distance, namely by the geographical length of a straight line considering the rotundity of the Earth, and are measured in nautical miles (1nm = 1.852km); the second one is freight-mass-weighted ($w_{ij}^{(b)} = w_{ij}$), where edge weights represent annual carried freight mass expressing the cumulated vessel capacity measured in deadweight tons (DWT). The total DWT of a link corresponds to the product between the number of voyages and the vessel capacities (25). The data used in this study was obtained from Lloyd's List Intelligence, a world leader in maritime insurance and information, which tracks the daily movements of more than 80% of the world fleet and 98% of the world container fleet (18,25).

The original layer (G_0) of the multilayer GCSN (Γ) is a directed graph composed of n=1'109 nodes (ports of worldwide container connectivity) and m=12'069 links (direct container connections between ports). The other 31 layers are also directed and are induced from $G_0 \in \Gamma$ by sequentially removing edges of shortest distances. In particular, the first layer (G_1) is composed of the same number of nodes as the original layer ($n_1=n_0$), but includes only those edges having a greater distance than 100nm, namely $G_1(V_1, E_1) = G(V, E \geq 100\text{nm})$. Similarly, the other layers are generated by applying the respective restrictions $E \geq$ 200, 300, 400, 500, 600, 700, 800, 900, 1'000, 1'250, 1'500, 1'750, 2'000, 2'500, 3'000, 3'500, 4'000, 4'500, 5'000, 5'500, 6'000, 6'500, 7'000, 7'500, 8'000, 8'500, 9'000, 9'500, 10'000, and 10'500nm. Let us denote this set of spatial distances as Σ. This method of layer construction builds on an unevenly stratified distance sampling applying: (a) a step of 100nm to the interval [100-1'000]; (b) a step of 250nm to the interval [1'000-2'000]; and (c) a step of 500nm step to the interval [2'000-10'500]. The concept behind this unevenly stratified sampling is to empirically counterbalance data resolution, computational complexity, sample adequacy for statistical inference analysis, and empirical intuition, in the available

family of layers. In particular, while it is desirable to reduce the resolution of the layers' dataset and to ease computations, the available number of layers should be also enough (>30) to apply statistical inference analysis based on the normality assumption (to construct normal distribution confidence intervals) (47), and the level of resolution should be satisfactory enough at the shorter distances defining the scale of the neighborhood, where the major volume of cargo activity (18,21,25) is by definition applicable. The property $V_0 = V_1 = V_2 = \ldots = V_{31} = V$ makes Γ a multiplex network, where all layers have the same number of nodes and the edge-set (E_i) in each layer is produced by applying a distance filter ($E \geq d(i)$ nm | $d(i) \in \Sigma$, i=0,1,2,…,31). Let ΔE_i denote the edge difference ($\Delta E_i = E_i - E_{i-1} \mid i$=1,2,…,31) between layers G_{i-1} and G_i (see Fig.1). Thereby, we can define each layer $G_i \in \Gamma$ as a function of the original one $G_i = f(G_o)$, as follows:

$$G_i\left(V_i, E_i\right) = G_o\left(V, E(i)\right) = G_o\left(V, E - \sum_{j=1}^{i} \Delta E_j\right) \quad (1).$$

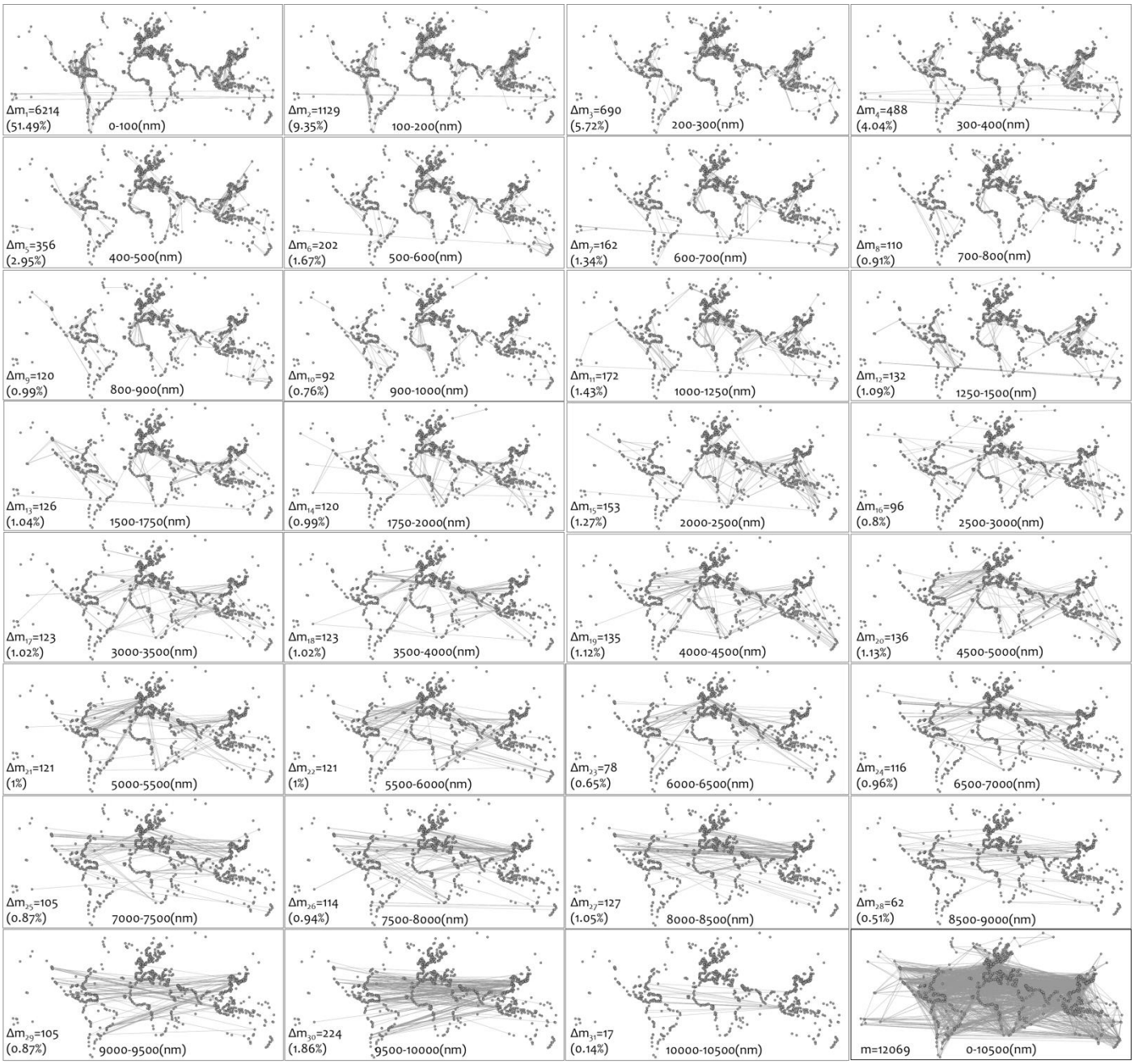

Fig.1. Edges (Δm_i) added in each layer of the multilayer GCSN (n=1'109 nodes, m=12'069 edges). The edges m_i included in layer G_i are those with distance shorter than (≥) i nm, where: i=100, 200, 300, 400, 500, 600, 700, 800, 900, 1'000, 1'250, 1'500, 1'750, 2'000, 2'500, 3'000, 3'500, 4'000, 4'500, 5'000, 5'500, 6'000, 6'500, 7'000, 7'500, 8'000, 8'500, 9'000, 9'500, 10'000, 10'500 (data from the year 2016); and are computed by the formula $m_i = m_{i-1} + \Delta m_i$ (m_o=0).

The definition context of the GCSN's layers illustrates that this study does not conceptualize spatial distance as a direct variable to measure its effect on network topology. Instead, it conceives spatial distance indirectly, through the criterion applied to generate the network layers of Γ. Taking into account that each layer is a separate network with characteristic topological properties, this spatial-distance criterion (that is applied to the network layers' production) is projected through an unspecified mechanism to the topological features of each layer. Therefore, this paper conceives spatial distance through its transformation to the topological properties of layer G_i that is produced under certain

spatial constraints $d(i) \in \Sigma$. So, if we denote as Ξ the topological space where the original layer G_o is embedded, we can define the topological space $\Xi_i \equiv \Xi(i) \equiv \Xi(d(i))$ of layer G_i as a function of Ξ and the spatial threshold $d(i) \in \Sigma$, implying more broadly that $\Xi(i)$ is a function of distance. Further, we loosely define the topological space $\Xi(i)$ as what in literature is called "network topology". Many academics could use a single characterization to describe this property, such as random-like, lattice-like, small-world, and hub-and-spoke topology (2-4). In this paper, we define network topology $\Xi(i)$ in vector terms, as the composition of fundamental topological measures and attributes (x_1, x_2, …, x_p) describing the topological features of a single layer G_i, as follows:

$$\Xi(i) = \Xi(x_1, x_2, \ldots, x_p) = \Xi(x_1(i), x_2(i), \ldots, x_p(i)) \quad (2).$$

This approach allows studying network topology in a multivariable context, expressed by the collection of various measurable attributes and not by a single topological characterization. Besides, the GCSN has a known hub-and-spoke network topology with scale-free characteristics (18,20,24,25), and thus a mono-variable consideration of its network topology (e.g. based on the power-law distribution exponent or a categorical variable describing network topology) would not be much insightful. On the other hand, the multivariable context developed in this paper incorporates in more detail topological information by decomposing topological space $\Xi(i)$ into a set of measurable topological attributes $\Xi(i)=\Xi(x_1(i), x_2(i), \ldots, x_p(i))$. Thereby, this paper conceives spatial distance not as a single geographical variable but as the effect of the generative criterion of the spatial constraints ($E \geq d(i)$), which are applicable for the GCSN layers' construction $G_i(V, E \geq d(i))$ on the topological attributes $\Xi(x_1(i), x_2(i), \ldots, x_p(i))$ of the GCSN layer G_i. The overall approach goes beyond the typical comprehension of network topology in the literature and provides insights into a multivariable conceptualization of this concept.

■ Network and empirical analysis

This paper studies a set of topological properties $\Xi(i) =\Xi(x_1(i), x_2(i), \ldots, x_p(i) \mid i=0,1,2,\ldots,32)$ on the available family $\Gamma=\{G_o, G_1, G_2, \ldots, G_{31}\}$ of the 32 GCSN layers. The rationale behind this approach is based on the research hypothesis that important connections are determinant to the configuration of network topology and thus that network topology should considerably change when important connections are omitted from the network. Taking into account that each layer is generated under a spatial constraint $E \geq d(i)$, with $d(i) \in \Sigma$, the scores' collection for a topological measure $x_j \in \Xi$ across the GCSN layers $\{x_j(0), x_j(1), x_j(2), \ldots, x_j(32)\}$ configures a series expressing the behavior of x_j for different distances $\{d(0), d(1), d(2), \ldots, d(32)\}$. This allows considering each topological measure $x_j \in \Xi$ as a function of distance $x_j = x_j(d(i))$ and thus interpret its behavior across the different distance constraints as the effect of distance on the topological measure $x_j \in \Xi$. Further, the collections $\{x_j(0), x_j(1), x_j(2), \ldots, x_j(32)\}$ for all topological measures $j=1,2,\ldots,p$ can provide an approximation of the effect of space on network topology. However, although this approach succeeds in describing the topological properties of the GCSN layers as a function of distance, it does not suggest a typical case of "distance attenuation law" (1,3), due to the way conceptualizing distance (through the transformation of spatial constraints to the topological properties of GCSN layer and not as a direct variable). Within this context, a set of network measures are included in the analysis, which is extracted from the literature (4,48-50) and is briefly described in Table 1.

Table 1

Network measures (*) included in the analysis of the GCSN.

Measure	Symbol	Description	Math Formula
Graph density	ρ	The fraction of the existing connections of the graph (m) to the number of the possible connections (that are equal to $\binom{n}{2}$, where n is the number of nodes). It expresses the probability to meet in the GMN a connected pair of nodes.	$\rho = m \Big/ \binom{n}{2} = \frac{2m}{n \cdot (n-1)}$
Node Degree	k	The number of edges $k(i)$ being adjacent to a given node i belonging to a graph $G(V,E)$, where V is the node-set and E is the edge-set. Node-degree expresses the node's communication potential.	$k_i = k(i) = \sum_{j \in V} \delta_{ij}$, where $\delta_{ij} = \begin{cases} 1, \text{ if } e_{ij} \in E \\ 0, \text{ otherwise} \end{cases}$
Node strength	s	For a network edge $e_{ij} \in E$, where E is the edge-set, the node-strength $s(i)$ is defined by the sum of edge weights w_{ij} being adjacent to a given node i.	$s_i = s(i) = \sum_{j \in V} \delta_{ij} \cdot w_{ij}$
Average Path Length	$\langle l \rangle$	The average length of the network shortest-paths $d(i,j)$, where n is the number of nodes in the network.	$\langle l \rangle = \frac{\sum_{i \in V} d(i,j)}{n \cdot (n-1)}$
Network diameter	$d(G)$	The maximum length of the shortest-paths $d(i,j)$ in the network.	$d(G) = \max \{ d(i,j) \}_{i,j \in V}$
Clustering Coefficient (local)	$C(i)$	The probability of meeting linked neighbors around node i, which is equivalent to the number of the node's connected neighbors $E(i)$ (i.e., the number of triangles that are configured in the neighborhood), divided by the number of the total triplets shaped by this node, which equals to $k_i(k_i-1)$, where k_i is the degree of node i.	$C(i) = \frac{E(i)}{k_i \cdot (k_i - 1)}$
Modularity	Q	An objective function expressing the potential of a network to be subdivided into communities. In its mathematical formula, g_i is the community of node $i \in V$ (where V is the node-set), $[A_{ij} - P_{ij}]$ is the difference of the actual (A_{ij}) minus the expected (P_{ij}) number of edges falling between a particular pair of vertices $i,j \in V$, and $\delta(g_i,g_j)$ is an indicator (the Kronecker's) function returning 1 when $g_i=g_j$.	$Q = \frac{\sum_{i,j} [A_{ij} - P_{ij}] \cdot \delta(g_i, g_j)}{2m}$

* Sources: (4),(48)-(50).

Within this context, the methodological approach in this paper is applied toward a double direction. The first builds on the conceptualization of the current methods examining the relationship between space and network topology within a holistic and structurally undisturbed context. This direction performs an analysis on the original layer (G_o) of the GCSN to detect topological properties related to the effect of space. The second direction builds on the decomposition rationale developed in this paper and examines the collections of the available topological measures across the 32 layers of the multilayer GCSN as a function of distance (as previously were described in detail).

In technical terms, the methods used in the analysis are first the computation of the degree distribution $p_i(k)$ of each layer G_i, expressed by the frequency distribution (k_j, $n(k_j)$) of the unique node degree values k_j in the network, as follows (3,4,51):

$$p(k) = \left(k_i, n\left(k_i\right)\right) \quad (3),$$

where $n(k_i)$ is the node frequency (number of nodes) of degree k. When node frequencies are divided by the total number of nodes ($n(k_i)/n$), the degree distribution interprets the probability to meet a node of degree k_i. Further, both distance and deadweight tonnage (cumulated vessel capacity) edge distributions are computed by constructing histograms (47,52), expressing the edge frequency within classes of certain ranges of edge weights.

Next, parametric curve fittings are applied to degree distributions and the series of topological measures as a function of distance. This process estimates the parameters of a curve $y=f(x)$ that optimally fit to the observed data y_i by minimizing the square differences $e = y_i - \hat{y}_i$, according to the relation (47,52):

$$\min\left\{e = \sum_{i=1}^{n}\left[y_i - \hat{y}_i\right]^2\right\} = \min\left\{\sum_{i=1}^{n}\left[y_i - f(k_i)\right]^2\right\} \quad (4).$$

The optimization method used for the estimations is the Least-Squares Linear Regression (LSLR), which is based on the assumption that differences e follow the normal distribution $N(0, \sigma_e^2)$.

Next, on the series of (p in number) topological measures $\{x_j(0), x_j(1), x_j(2), \ldots, x_j(32) \mid j=1,2,\ldots,p\}$ we compute the first-order differences, according to the formula (53,54):

$$\Delta^{(1)}x_j(i) = x_j(i) - x_j(i-1),\ i=1,\ldots,32 \quad (5)$$

where $x_j(i)$ is the value of the attribute-series x_j at a place (distance) i, and $x_j(i–1)$ at the previous position i-1. This approach expresses the discrete analogy of the first derivative function and allows capturing the changes in a network attribute without removing the variable's scale. After computing the first-order differences, we statistically test whether their values are significantly greater than the average (mean value) of the first-order series. To do so, we construct 95% confidence intervals (CIs) of the mean, according to the formula:

$$\mu_{lb,ub} = \hat{\mu} \pm z_{a/2} \cdot se \quad (6),$$

where $\hat{\mu}$ is the sample estimator of the mean value, $z_{a/2}$ is the z-score computed for 1–a% confidence level, and $se = s/\sqrt{n}$ is the sample's standard error (47). Cases exceeding the (zone) interval [μ_{lb}, μ_{ub}] are considered statistically different from the mean value and thus imply that the attribute-series exhibit significant changes.

To detect whether connectivity in hubs is an effect of distance, and at what level the hubs undertake the distant connections of the GCSN, we examine the attribute-series of the measures of (a) degree (k); (b) in-degree (k+); and (c) out-degree (k–); for the Top 10 highly connected ports (hubs) of the original layer, along with some other chosen ports showed considerable constancy to distance. The ports included in the hubs' connectivity analysis are shown in Table 2.

Table 2

Top 10 ports with highest connectivity (hubs) and other selected ports that are included in the analysis of the GCSN.

Degree			In-Degree			Out-Degree		
Rank	Port	Measure	Rank	Port	Measure	Rank	Port	Measure
1	Singapore	345	1	Singapore	170	1	Singapore	175
2	Busan	300	2	Busan	148	2	Busan	152
3	Shanghai	247	3	Shanghai	135	3	Hong Kong	120
4	Hong Kong	246	4	Rotterdam	128	4	Rotterdam	116
5	Rotterdam	244	5	Hong Kong	126	5	Port Klang	115
6	Port Klang	233	6	Port Klang	118	6	Shanghai	112
7	Algeciras	206	7	Algeciras	106	7	Algeciras	100
8	Tanjung Pelepas	180	7	Qianwan	106	8	Tanjung Pelepas	93
9	Qianwan	174	8	Antwerp	98	9	Jebel Ali	82
10	Antwerp	167	9	Beilun	87	10	Beilun	78
11	Beilun	165	9	Kaohsiung	87	13	Kaohsiung	74
12	Kaohsiung	161	9	Tanjung Pelepas	87	17	Antwerp	69
12	Jebel Ali	161	12	Jebel Ali	79	17	Yangshan	69
16	Yangshan	147	13	Yangshan	78	18	Qianwan	68
28	New York	115	19	New York	66	28	New York	49

Finally, we apply a factor analysis on edge distribution (DWT) variables configured by percentile of link distance, to detect commonalities that may provide insights into the hierarchy of GCSN due to the effect of space. This method describes the variability among possibly correlated variables by this reflected on a potentially lower number of unobserved variables, called factors (underlying or latent variables or components). Factor analysis models the observed variables as linear combinations of the potential factors, including stochastic terms, and generally contributes to the dimension reduction of the observed data. This is done by applying an orthogonal transformation to the original variables, which can be considered as fitting a *p*-dimensional ellipsoid to the data, where each axis of the ellipsoid corresponds to a factor. The dimension is reduced when some axes of the ellipsoid are relatively small, implying that the variance along that axes is also small so that these axes can be omitted from the dataset (52).

Results

■ Degree and edge distributions

The GCSN has extensively been studied in the literature as a complex network (7,8,21,24-26), due to its prominence economic (and particularly trade) importance. In all different available aspects of this cargo shipping network in the literature, scale-free characteristics are evident in the power-law (PL) fittings of its degree distribution, although they do not suggest a typical case (4) of a scale-free network. This analysis, applied to the degree and edge distributions of the GCSN according to the first methodological dimension structurally undisturbed context, supports the existing literature findings. In particular, in Fig.2, we can observe that all undirected (k), in-degree (k–), and out-degree (k+) distributions of the original GCSN layer (G_o) strongly fit PL patterns, with high scores of coefficient of determination $R^2(k)$=0.994, $R^2(k)$=0.9987, and $R^2(k)$=0.9991, respectively. Further, although they do not fall within the typical interval of scale-freeness ($2 < \gamma < 3$), the slopes of in-degree and out-degree distributions are significantly (at 95% confidence level) greater than one (>1). This result illustrates a tendency of the GCSN to configure structures of hierarchy (see 3,4,51) in its incoming (related to import trade activity) and outgoing (related to import trade activity) connectivity. On the other hand, the PL exponent of the undirected degree distribution is significantly (95%) lower than one (<1), describing a smoother slope of PL decay in comparison with the directed cases. These results indicate that hubs in the

undirected representation of the GCSN are more in number (frequency) in comparison with the directed cases, a fact that reveals the asymmetric functionality in carriers' operations of the GCSN.

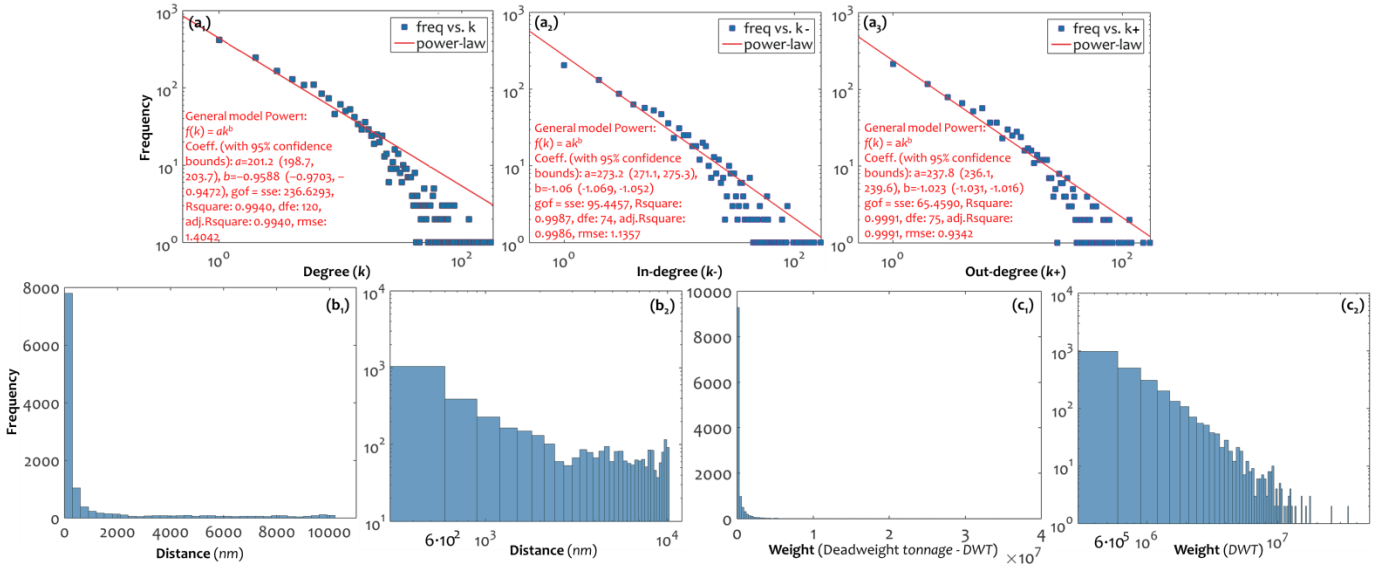

Fig.2. (a_1) Degree (k, $n(k)$), (a_2) In-degree (k–, $n(k$–)), and (a_3) In-degree (k+, $n(k$+)) distribution, showing the edge (distance-weighted, measured in nautical miles - nm) distribution at (b_1) metric and (b_2) log scale, and histograms showing the edge (mass-weighted, measured in deadweight tonnage - DWT) distribution at (c_1) metric and (c_2) log scale, of the original layer (G_o) of the GCSN.

A similar PL-like picture is also observed at the edge distributions in Fig.2. In particular, both the distance and deadweight tonnage (DWT) edge distributions of the GCSN configure distinguishable PL patterns. On the one hand, the edge distribution of distance is long-tailed, with the majority of connections being shorter than 1000nm. In the log-log scale, we can observe a change in the edge distribution slope for distances greater than 3000nm, where, in the metric scale, this distribution change appears almost horizontal. This status describes a uniform structure of the GCSN at medium-range and long-range distances, which may relate to a more linear than hierarchical structure and provide an aspect of the linear relationships empirically observed between global maritime trade and the economic size of its hinterland (7). Further, taking into account that in spatial networks hubs undertake the majority of distance connectivity (3,6), this value can be related to the existence of hubs and loosely provide a characteristic spatial range of the hubs activity in the GCSN, for distances over 3000nm. On the other hand, the edge distribution of deadweight tonnage is short-tailed, with the majority of connections having annual transport capacity up to $6 \cdot 10^6$tn. In the log-log scale, we can similarly observe a change in the edge distribution slope for annual transport capacity greater than $9 \cdot 10^6$tn, where, in the metric scale, this distribution change also appears almost horizontal. According to a similar hub-based interpretation, this value of $9 \cdot 10^6$tn can be seen as a characteristic value of the carrying capacity of the activity of hubs. Between these two cases of edge distribution, the slope of the distance-weighted edge distribution is smoother than this of tonnage-weighted. This implies that the hierarchical structure in the freight (DWT) distribution throughout the GCSN is more intense in comparison with the distance distribution and, more intuitively, that hubs proportionally undertake greater freight transport load than serve distant transportation. This observation is in line with the integration of the world economy and the lowering of border effects in maritime transportation (18).

■ Hub connectivity as a function of distance

This part of the analysis is coordinated to the second dimension of the methodological framework, which is based on a decomposition rationale. It examines the connectivity performance of hubs, as a function of geographical distance, for the cases of highly connected ports shown in Table 2. The results of the analysis are shown in Fig.3, where it can be seen that the top ports can be ranked by their ability to deploy their connectivity over longer distances. Interestingly, ports with a dominance of gateway functions (i.e. import, export, sea-land transshipment) have longer-range connections, namely Yangshan, Antwerp, Rotterdam, and New York. The port of Yangshan is Shanghai's new port, the largest of the country, serving the 26.3 million inhabitants global city and acting as a key point of entry for the Yangtze River port system (55). New York is another global city on its own; and the second-largest container port in the U.S. after Los Angeles-Long Beach, with massive land transport connections reaching a vast hinterland (56). Antwerp and Rotterdam have a much smaller city size in comparison but serve wide European hinterlands through multimodal connections. While Rotterdam has important transshipment functions as a hub for the British Isles, this activity only represents about 35% of its total activity. These four ports thus have in common a low level of transshipment incidence (below 50%), which is the share of containers shifted between vessels via the port terminals, with or without storage in between.

At the bottom of the figure are ports with a high level of transshipment incidence, exerting hub functions through hub-and-spokes and/or interlining configurations. This level reaches 80% for Singapore, meaning that only 20% of Singapore port's activity is devoted to Singapore's trade. A strong transshipment hub function implies a multiplication of high-frequency links between the hub and the feeder ports within a given region, thereby reinforcing short-range connectivity (57). The geographic distribution of transshipment activity across the globe (57) confirms that container hubs with a high transshipment incidence locate along the round-the-world trunk line and at strategic passages such as the Gibraltar Straits (Algeciras), the Taiwan Straits (Kaohsiung) or are centrally located within a given region, such as Southeast Asia (Singapore) and the Persian Gulf (Jebel Ali).

These two groups echo the fundamental distinction between centrality and intermediacy proposed by the authors of (58) about the spatial characteristics of transportation hubs. Ports like Rotterdam and New York have a high centrality, i.e. a strong traffic self-generation power derived from the accessibility to a vast market on the land side (hinterland). On the contrary, Singapore and Algeciras have a limited landward market and accessibility, relying mostly on intermediacy, i.e. the ability to embed the networks of transport operators for sea-sea transshipment. In fine, trade ports have longer-range connections due to the majority of direct deep-sea callings of containerships, while transshipment hubs, despite their global importance, have stronger connectivity locally as pivots through which many secondary ports consolidate their cargo to access the rest of the network.

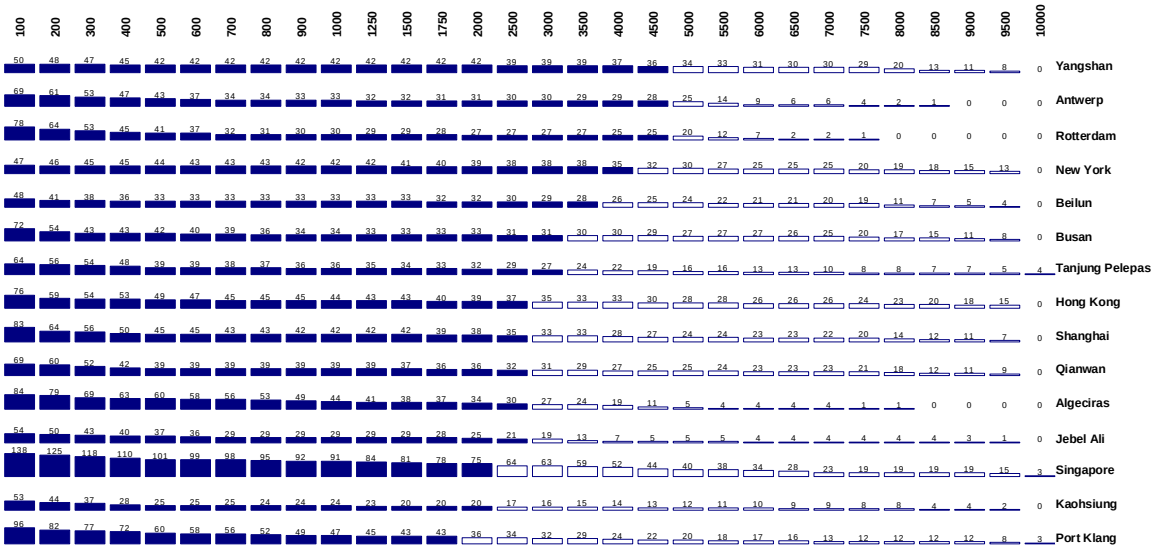

Fig.3. Bar charts of node degree across the available 32 GCSN layers (expressing node degree as a function of distance), for the selected ports shown in Table 2 (bar heights are proportional to the degree; dark bars represent values higher than row's average).

Ports in between those two categories have a mixed profile, being both hinterland ports and transshipment hubs, as seen with the cases of Busan and Hong Kong. Their interaction range is lower than gateway ports due to the importance of hub-and-spokes activities but higher than transshipment hubs as they also serve as gateways for a noticeable part of their activity. Busan for instance is South Korea's main port (80% of total container throughput) and second-largest city, serving the Seoul capital region at distance by land transport, while it also acts as a hub for numerous Japanese and Northern Chinese secondary ports (59). Its transshipment share does not exceed 40%. Before its partial retrocession to China, Hong Kong had long been a transshipment hub like Singapore, with limited if no landside connectivity across the border (60). Its transshipment activity had been a mix of Chinese re-exports and international transshipment (notably for Japan trade), combined with strong connections with Taiwan and in particular Kaohsiung until China-Taiwan direct trade had been relaxed across the Taiwan straits (61,62). Afterward, Hong Kong gradually shifted its cargo handling operations to South China, evolving towards a global financial center with fewer transshipment activities (63), and a growing gateway function serving the mainland Chinese hinterland.

This analysis shows that while larger ports tend to connect farther than smaller ports, trade (gateway) ports have a wider interaction range than transshipment ports. In some cases, this can be accentuated by geography, as seen with New York, which is relatively remote from the trunk shipping line connecting the Europe-Mediterranean area with the Caribbean basin. The latter region has become one important transshipment zone due to its proximity to the Panama Canal and its central position between North and South America. Most transatlantic flows are now rerouted through the Caribbean for consolidation. Similar observations could be made for other ports, such as in South Africa, which connections are longer on average due to peripherality (64). Our results confirm a more generic process by which transshipment volumes (and shares) are higher as the deviation distance from the trunk line is lower (65). Other factors come into play such as urban size, hinterland connectivity, and port performance as a whole.

■ Classification of links and ports

This approach builds on redistributing node strength (weighted degree) by percentiles of distance. One first result is striking, as the number and share of bidirectional links decrease with distance (Table 3), showing the importance of asymmetric pendulum services whereby ocean carriers deploy their vessels along with different loops between Europe-Asia and Asia-Europe (24), for instance. Namely, the closer the ports the higher the probability for mutual interaction is, such as through short-sea, coastal, or feeder services. It is a corollary of the fact that trade decreases with distance, from an operational perspective. Such a result has no equivalent in network theory and its applications, and can pave the way towards further research in all types of communication networks.

Table 3

Distribution of bidirectional links by distance percentiles

Distance percentile	No. Links	Bi-directional links	% in total	% in percentile
1 (shortest)	1149	406	12.6	35.3
2	1150	393	12.2	34.2
3	1150	398	12.4	34.6
4	1149	354	11.0	30.8
5	1152	343	10.7	29.8
6	1149	328	10.2	28.5
7	1151	310	9.6	26.9
8	1150	279	8.7	24.3
9	1150	213	6.6	18.5
10 (longest)	1151	191	5.9	16.6
Total	11501	3215	100.0	28.0

At next, factor analysis was launched based on ten variables, each of them representing the natural logarithm of traffic (DWT) by percentile of link distance (Fig.4), from the shortest (distance percentile 1) to the longest (distance percentile 10). For all ports, their edges were classified into 10 classes based on their kilometric length. For each port, the variables thus correspond to total traffic per class. Namely, each variable represents the amount of traffic per port (in deadweight tons, DWT) along 10 classes of link distance. Those classes were defined based on the percentiles of distance, from the shortest to the longest links. Therefore, world ports will have more or less distant traffic based on this decomposition. Interestingly, factor 1 (horizontal) depicts a size effect by which all variables are projected on positive values. Factor 2 illustrates more a “scale effect” opposing shorter-range traffic (positive values) and longer-range traffic (negative values), with a gradual order from the shortest to the longest. A hierarchical clustering analysis considered the four main factors that concentrate 72% of the total variance with respective eigenvalue > 1. In this analysis, six clusters were found, each of them containing a relatively well-balanced number of ports. The nature of clusters was then revealed by visualizing their traffic distribution by percentile. Despite certain resemblances, each cluster has a specific distribution of traffic weight throughout the different distance ranges. Cluster#6 (including “trans-scalar” ports) handles the largest amount of traffic, Cluster#3 (short-range ports) manages traffic within ports’ vicinity, Cluster#4 (local ports) has the lowest traffic in the longest-range category, Cluster#2 (medium-, long-range ports) and #5 (medium-range ports) have nearly no local traffic, while Cluster#1 (short/medium-range ports) is an intermediary class.

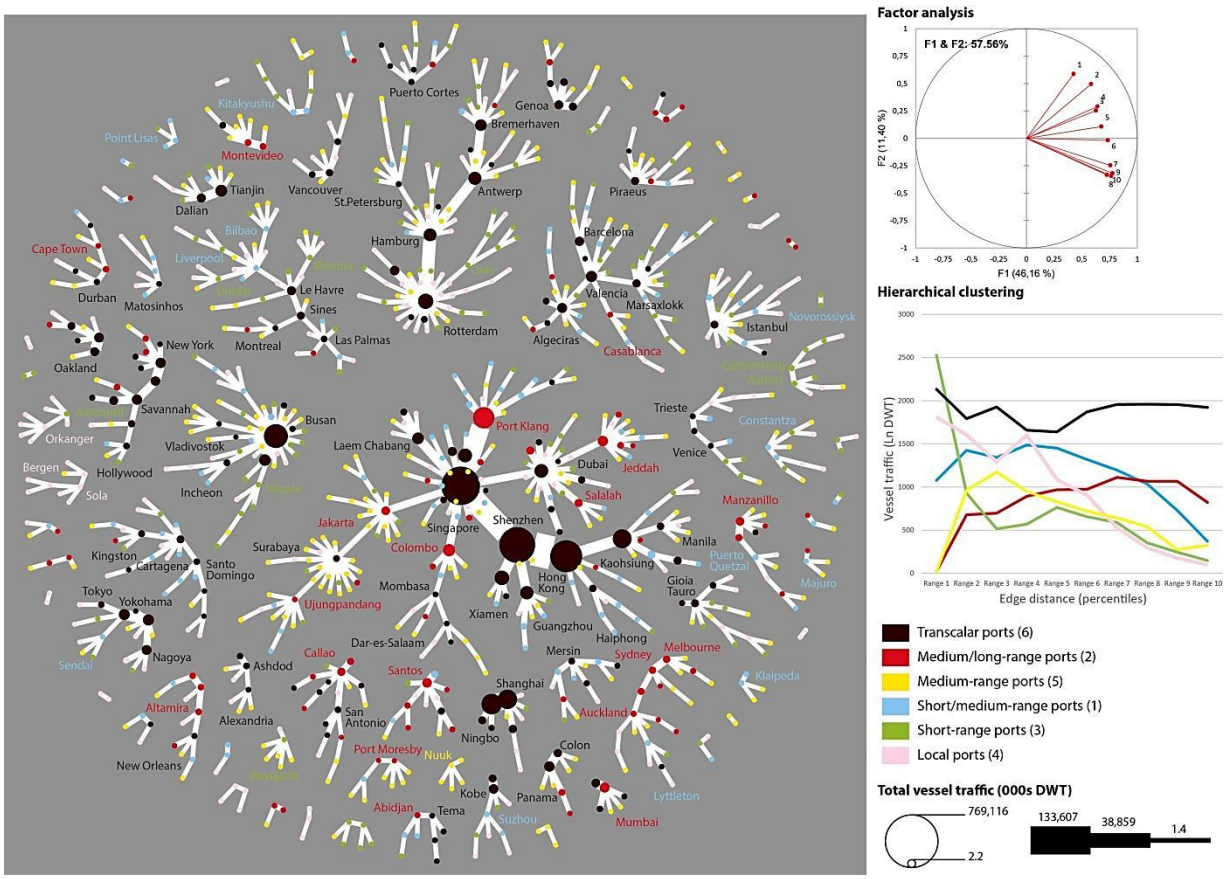

Fig.4. Multivariate analysis and visualization (space-L) of the GCSN (own elaboration based on TULIP 3.0.0; https://tulip.labri.fr/site/?q=node/110)

A single linkage analysis algorithm was employed to bisect the graph, transform it into a tree, and reveal its main hubs, based on the hypothesis that hubs centralize all local, regional, and global level flows as distribution platforms (1). Expectedly, Fig.4 shows that hub-like structures mostly appear around black cluster ports (transcalar), among which are Singapore, Rotterdam, Valencia, Gioia Tauro, Istanbul, Piraeus, and New York, to name but a few. In addition, while such ports dominate their belonged subgraph, it is possible to observe a "range effect" whereby those subgraphs are organized as multi-hub structures, along a given maritime route. This is particularly the case of the largest component centered upon Singapore, which includes several other Asian hubs and gateways and even extends towards East Africa. Other trans-scalar Asian ports having a hub position appear as independent substructures, among which Busan (East Sea), Dalian-Tianjin (Yellow Sea), while main Japanese ports appear in relative isolation from the rest (Tokyo-Yokohama-Nagoya, Kobe-Osaka). The hub function is less developed for gateways as they are more specialized in sea-land transshipment ensuring direct trade, resulting in a lesser number of nodes in the star-like structures (e.g. Shanghai, Barcelona). The red cluster is the one with another noticeable number of hub-and-spokes configurations, differing from the black cluster by its limited local traffic. Ports in this category are often located in low-density port systems, such as Manzanillo (Central America West Coast), Callao (Peru), Australian main ports, Jeddah (Red Sea), Salalah (Arabian Peninsula), Makassar (Sulawesi), Casablanca, and Cape Town (South Africa), while their (relative) remoteness or their function as transshipment hub exacerbates their long-distance traffic. Santos and Montevideo are main national ports with strong transatlantic linkages, like Jakarta, Colombo, and Port Klang along the Europe-Asia route. Yet, Singapore's dominance remains overwhelming within this

corridor. Other clusters are thus mostly composed of ports dominated by the aforementioned hubs. For the medium-range cluster (yellow), the only exception is Nuuk in Greenland as a national hub. For local ports (light red), the three minor hubs are located in Norway, with a centrality also limited to the national (or even regional) scale. Short-range ports (green) as well are confined to North Europe, while short/medium-range ports (blue) are geographically diverse.

■ Network measures as a function of distance

This part of the analysis builds on the decomposition rationale (second dimension) of the methodological framework, expressing the topological features of the GCSN as a function of distance, according to the relation [2]. The results of the analysis are shown in Fig.5, where the measures of (number of) edges (Fig.5a); average degree (Fig.5b); clustering and average clustering (Fig.5h) coefficient; and total edge weight (Fig.5k), are all described by a decaying (PL-like) pattern, expressing a decline of these measures as distance increases. In terms of clustering, this pattern illustrates that circular (triangular) connections become less in number as distance decreases and thus that more linear structures in the connectivity of GCSN emerge. Next, the patterns of network diameter (Fig.5c) and average path length (Fig.5d) appear "noisy constant", described by a horizontal slope equipped with several local fluctuations that generally imply a relative indifference of the path-defined GCSN measures to distance. However, these local fluctuations may reveal levels of geographical scale where the path-determined network topology is either consistent or asymmetric. For instance, at the range of 6'500-8'500, we can observe a concave area in the directed aspects of both measures, implying more intense asymmetric forces in the topology of GCSN.

On the other hand, the measures of average edge length (Fig.5i) and average edge weight (Fig.5l) are described by ascending PL patterns, which impressively illustrate that the average trade volume carried throughout the GCSN appears indifferent to distance, for distances greater than 100nm. The number of connected components (Fig.5f) also follows an ascending pattern, implying that the connectedness of the GCSN decomposes as distance increases, but this effect is more intense at smaller distances (<1'000nm) and becomes smoother at greater ones. Next, the patterns of graph density (Fig.5g), degree distribution exponent (Fig.5m), and coefficient of determination R^2 (Fig.5n) are decaying exponential ones, illustrating a decline through distance. For the degree distribution exponent, this decaying pattern illustrates that the topology of the GCSN has more distinctive hub-and-spoke characteristics as distance increases. Next, modularity (Fig.5e) configures a "U"-shaped pattern, illustrating a general increase of the measure through distance, but not at the neighborhood (<300nm) and large (<8'500nm) scale. This shape implies a tendency of the GCSN to separate into communities due to local and distant connectivity. Finally, the total edge length (Fig.5j) configures a linear declining pattern, implying that large distances play an important role in the cohesion of the network, as imprinted on the uniform decay of the total edge length.

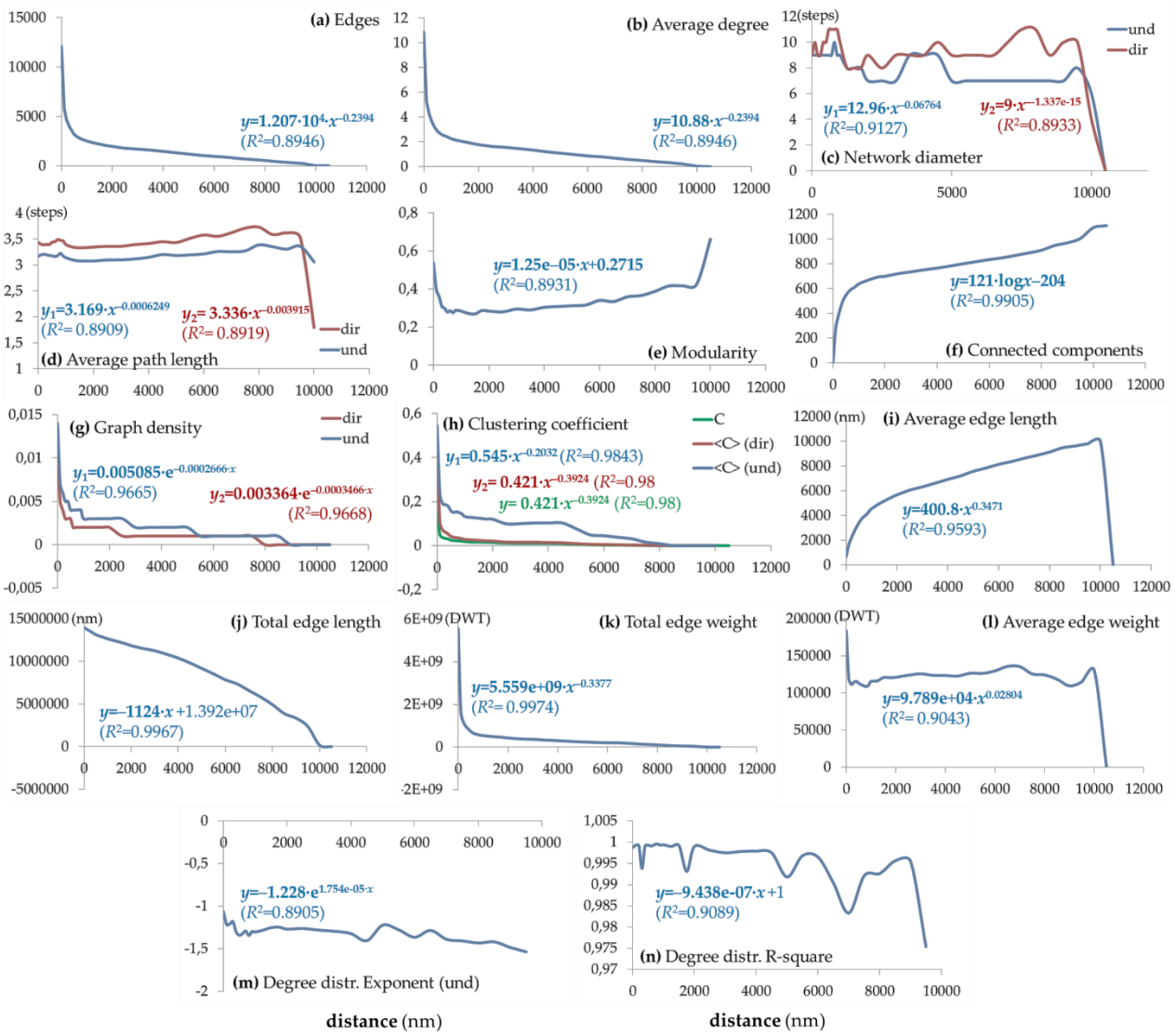

Fig.5. Line plots (metric scale) showing the distribution of major network measures of (a) network edges, (b) average degree (und), (c) network diameter (dir/und, measured in steps), (d) average path length (dir/und, measured in steps), (e) modularity, (f) number of connected components, (g) graph density (dir/und), (h) clustering and average clustering coefficient (dir/und), (i) average edge length (nm), (j) total edge length (nm), (k) total edge weight (DWT), (l) average edge weight (DWT), (m) PL degree distribution exponent (und), and (n) PL degree distribution determination (R^2), which are expressed as a function of distance and are computed on the series of layers composing the multilayer model of GCSN (as shown in Fig.1.). Equations of the best possible (those with max R^2) fitting curves are shown per diagram.

To examine topological changes between successive layers of GCSN and thus across different levels of geographical distance, we apply a first-order differences analysis, according to the relation [1]. The results of the analysis are shown in Fig.6, which illustrates significant changes in measures of network topology across the available distance levels. As it can be observed, a significant amount of the GCSN edges (Fig.6i) is distributed at the neighborhood scale (<200nm). For average edge length (Fig.6ii), up to a range of 400nm, and at the distances of 2'500nm and 5'500nm, changes are significantly greater than the mean. However, the total edge length (Fig.6iii) shows significantly great changes at distances greater than 5'000nm. For average edge weight (Fig.6iv), we can observe three zones, a neighborhood (0-200nm), a middle ("mesoscale") geographical scale (1'000-

1’500nm), which perhaps describes international market regions, and a large (mega) geographical scale (≥5’000nm). This allows defining three different types of markets in the GCSN, according to the average freight distribution and to the geographical scale: the *neighborhood*, the *international*, and the *intercontinental* market. For the total edge weight (Fig.6v), significant changes appear at the level of the neighborhood, highlighting the importance of the neighborhood market in the GCSN. For average degree (Fig.6vi), a decaying pattern is evident implying that a significant number of routes in the GCSN are distributed within the scale of the neighborhood. Next, for max in-degree (Fig.6vii), we can observe significantly great changes at distances <500nm and locally at 2’500nm, 4’500nm (mesoscale), and 10’000nm (large or mega-scale), which can apply another zoning in terms of import maritime market: the neighborhood (<500nm), the international (2’500nm-4’500nm), and the intercontinental market (10’000nm).

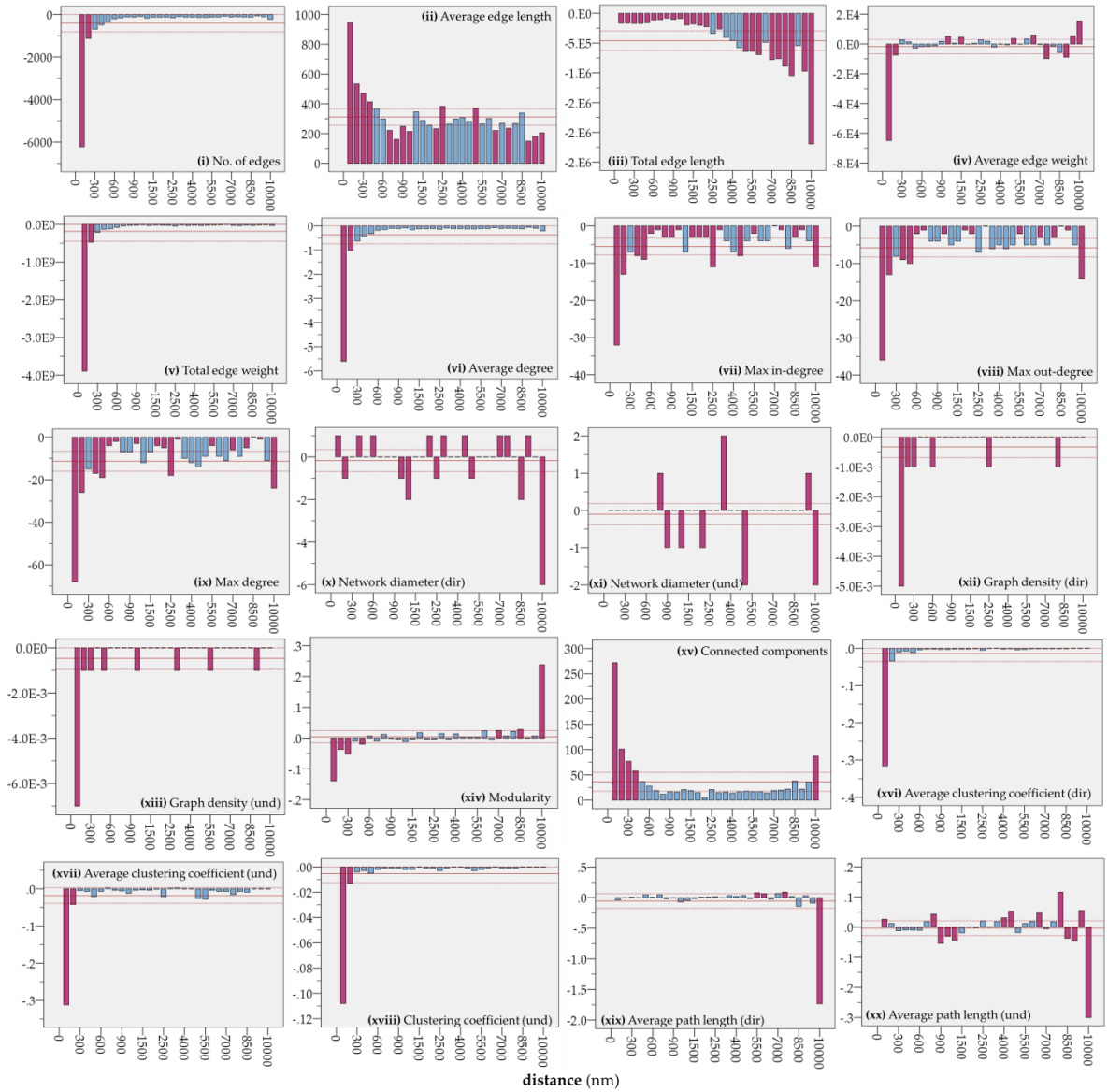

Fig.6. Bar plots (categorical axis) showing the first-order differences of the GCSN’s attribute-series of (i) network edges, (ii) average edge length (nm), (iii) total edge length (nm), (iv) average edge weight (DWT), (v) total edge weight (DWT), (vi) average degree, (vii) max in-degree, (viii) max out-degree, (ix) max degree, (x) network diameter (und, steps), (xi) network diameter (dir, steps), (xii) graph density (dir), (xiii) graph density (und), (xiv) modularity, (xv) connected components, (xvi)

average clustering coefficient (dir), (xvii) average clustering coefficient (und), (xviii) clustering coefficient (und), (xix) average path length (und, steps), (xix) average path length (und, steps), which are expressed as a function of distance and are computed on the series of layers composing the multilayer model of GCSN (as shown in Fig.1.).

For the max out-degree (Fig.6viii), the neighborhood (<500nm) and intercontinental (10’000nm) zones are also evident, but not the middle (mesoscale) zone. The difference in the patterns between max in- and out-degree implies that export trade occurs at larger distances than import trade, a fact that can be verified by applying an independent samples *t*-test for the equality of means between the upper (out-degree) and lower triangular (in-degree) distance weights matrix of the GCSN. This observation provides evidence about the asymmetric structure of the GCSN functionality at the mesoscale zone. Next, network diameter shapes a rather variable picture, both for its directed (Fig.6x) and undirected (Fig.6xi) expressions. Its pattern implies that this measure is more dependent on mesoscale and large-scale connections. A variable decaying pattern also describes graph density (Fig.6xii, Fig.6xiii). Next, the pattern of modularity (Fig.6xiv) also supports that at the neighborhood and large scales significant changes occur. The patterns of clustering (Fig.6xviii) and average clustering coefficient (Fig.6xvi, Fig.6xvii) illustrate that clustering in the GCSN is mainly a matter of neighborhood connections and thus is related to the dynamics of the local markets. Finally, the average path length (Fig.6xix, Fig.6xx)) produces zoning at the neighborhood scale (100nm), the very beginning (800-1250nm), and the end of the middle scale (4’500-5’500nm), and the large scale (≥6’500nm).

An attempt to combine all the previous observations to a single outcome results in three levels of geographical scale in the structure and functionality of GCSN, which are: (a) the neighborhood (local) connectivity scale, defined by distances shorter than 600 nautical miles (< 600nm); the international (mesoscale or middle) connectivity scale, defined by distances at the range of 1’250-4’500nm; and the markets of intercontinental (large) connectivity scale, defined by distances greater than 5’000 nautical miles (> 5’000nm). These geographical scales are visualized into buffers at the map of Fig.7.

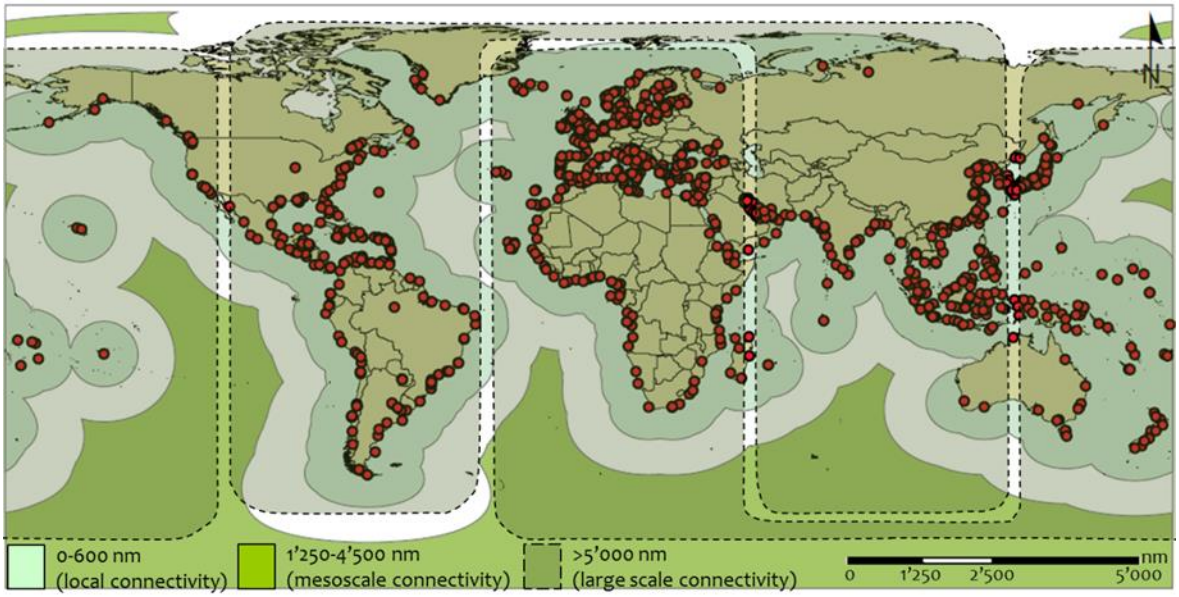

Fig.7. Zone buffering of the geographical scales (local, mesoscale, and large connectivity) of the GCSN functionality, as revealed from the analysis (own elaboration based on ESRI ArcGIS 10.50; https://www.arcgis.com).

Conclusions

This paper examined how spatial distance affects the topology of the Global Container Shipping Network (GCSN) by using a multilayer network model consisting of 32 multiplex layers generated by successively removing connections of smaller distance. The degree and edge distribution analysis showed that the structures of hierarchy in the GCSN are more intense and distinguishable per trade role (import and export), whereas when the GCSN is interpreted under the trade balance these directed structures of hierarchy appeared more counterbalanced. Also, it showed that the hierarchical structure in the freight distribution of the GCSN is more intense and homogenized than the distance distribution. The hubs connectivity showed a significant loss at shorter distances (≤ 800nm) and a smoother decay at longer distances. Trade/gateway hubs like Antwerp, Rotterdam, and Yangshan have a longer interaction range than transshipment hubs such as Singapore, Algeciras and Kaohsiung, which connectivity is more local due to the high frequency and multiplication of local linkages with feeder ports. A balanced profile was detected for Busan and Hong Kong, which are both gateway (hinterland) ports and hub (transhipment) ports. Also, the analysis revealed that bidirectional links in GCSN decrease with distance, highlighting the importance of asymmetric functionality in the carriers' operations. Ports were classified into six clusters (local, short-range, short/medium-range, medium-range, medium/long-range, and trans-scalar ports) according to the distribution of their traffic over classes of distance. Transcalar ports dominate the network when the latter is transformed into a tree-like structure revealing hubs and their belonged subgraphs. At last, the analysis on network measures verified the scale-free-like topology of the GCSN and illustrated a linear configuration of the GCSN in higher distances. The path-defined measures and the average trade load showed a relative indifference to distance, supporting the empirical finding of the integration of the world economy and maritime transportation. The connectedness of the GCSN was also found to decompose as distance increases, but mainly at smaller distances (<1'000nm). Finally, the analysis revealed three levels of geographical scale in the structure of GCSN, the neighborhood (local connectivity, <600nm), international (mesoscale or middle connectivity, 1'250-4'500nm), and intercontinental markets (large scale connectivity, >5'000nm).

The overall approach contributed to the study of the spatial dimension in complex and multilayer networks and provided insights into the spatial structure of the GCSN maritime market. In particular, the identification of transcalar nodes is an addition to the literature on hub theory (73), as it demonstrates that such nodes dominate through their ability connecting all traffic scales, compared with recent studies of the GCSN only focusing on topology (74). The directional decay over distance is another new finding, which can serve as a benchmark to study any other domain subject to asymmetrical relationships, from trade imbalance to empty container repositioning and power structures in social and firms' networks.

Author Contributions Statement

D.T. and C.D. designed research; performed research; contributed new reagents/analytic tools; analyzed data, and wrote the paper.

References

1. Rodrigue, J. P., Comtois, C., Slack, B., (2013) *The Geography of Transport Systems*, New York, Routledge Publications.
2. Tsiotas, D., Polyzos, S., (2018) "The complexity in the study of spatial networks: an epistemological approach", *Networks and Spatial Economics*, 18(1), pp.1–32.

3. Barthelemy, M., (2011) “Spatial networks”, *Physics Reports*, 499, 1–101.
4. Albert, R., Barabasi, A.-L, (2002) “Statistical mechanics of complex networks”, *Reviews of Modern Physics*, 74(1), pp.1–47.
5. Polyzos, S., Tsiotas, D., (2020) “The contribution of transport infrastructures to the economic and regional development: a review of the conceptual framework”, *Theoretical and Empirical Researches in Urban Management*, 15(1), pp.5-23
6. Ducruet, C., Beauguitte, L., (2014) “Spatial science and network science: Review and outcomes of a complex relationship”, *Networks and Spatial Economics,* 14, pp.297–316.
7. Liu, C., Wang, J., Zhang, H., Yin, M., (2018) “Mapping the hierarchical structure of the global shipping network by weighted ego network analysis”, Int. J. Shipping and Transport Logistics, 10(1), pp.63-86.
8. Ducruet, C., (2013) “Network diversity and maritime flows”, *Journal of Transport Geography*, 30, pp.77–88.
9. Buhl, J., Gautrais, J., Reeves, N., Sole, R. V., Valverde, S., Kuntz, P., Theraulaz, G., (2006) “Topological patterns in street networks of self-organized urban settlements”, *European Physical Journal B,* 49, pp.513-522.
10. Cardillo, A., Gomez-Gardenes, J., Zanin, M., Romance, M., Papo, D., del Pozo, F., Boccaletti, S., (2013a) “Emergence of network features from multiplexity, *Scientific Reports*, 3, 1344;
11. Crucitti, P., Latora, V., Porta, S., (2006) “Centrality in networks of urban streets”, *Chaos,* 16, pp.(015113)1-9.
12. Lammer, S., Gehlsen, B., Helbing, D., (2006) “Scaling laws in the spatial structure of urban road networks”, *Physica A,* 363, pp.89-95.
13. Tsiotas, D., (2020) “Drawing indicators of economic performance from network topology: the case of the interregional road transportation in Greece”, *Research in Transportation Economics* (10.1016/j.retrec.2020.101004).
14. Kurant, M., Thiran, P., (2006) “Extraction and analysis of traffic and topologies of transportation networks”, *Physical Review E*, 74, 036114.
15. Kurant, M., Thiran, P., (2006) “Layered Complex Networks”, *Physical Review Letters*, 96, 138701.
16. Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P.A., Mukherjee, G., Manna, S.S., (2003) “Small-world properties of the Indian railway network”, *Physical Review E*, 67, 036106.
17. Tsiotas, D., (2017) “Links between network topology and socioeconomic framework of railway transport: evidence from Greece”, *Journal of Engineering Science and Technology Review*, 10(3), pp.175-187.
18. Ducruet, C., (2020) “Port specialization and connectivity in the global maritime network”, *Maritime Policy & Management*, pp.1-17.
19. Hu, Y., Zhu, D., (2009) “Empirical analysis of the worldwide maritime transportation network”, *Physica A*, 388, pp.2061-2071.
20. Kaluza, P., Koelzsch, A., Gastner, M. T., Blasius, B., (2010) “The complex network of global cargo ship movements”, *Journal of the Royal Society Interface,* 7, pp.1093–1103.
21. Calatayud, A., Mangan, J., Palacin, R., (2017) “Vulnerability of international freight flows to shipping network disruptions: A multiplex network perspective”, Transportation Research Part E, 108, pp.195-208.
22. Tsiotas, D., (2017) “The imprint of tourism on the topology of maritime networks: evidence from Greece”, *Anatolia: An International Journal of Tourism and Hospitality Research*, 28(1), pp.52–68

23. Tsiotas, D., Niavis, S., Sdrolias, L., (2018) “Operational and geographical dynamics of ports in the topology of cruise networks: the case of Mediterranean”, *Journal of Transport Geography*, 72, pp.23–35
24. Ducruet, C., Notteboom, T.E., (2012) “The worldwide maritime network of container shipping: spatial structure and regional dynamics”, *Global Networks*, 12(3), pp.395-423.
25. Ducruet, C., (2017) “Multilayer dynamics of complex spatial flows: The case of global maritime flows (1977-2008)”, *Journal of Transport Geography*, 60, pp.47–58.
26. Ducruet, C., Itoh, H., (2021) “Introduction to global container shipping market”, In *Global Logistics Network Modelling and Policy*, Elsevier, pp.3- 30.
27. Bianconi, G., Pin, P., Marsili, M., (2009) “Assessing the relevance of node features for network structure”, *Proceedings of the National Academy of Sciences* (USA), 106, 11433.
28. Guimera, R., Amaral, L. A. N., (2004) “Modeling the world-wide airport network”, *European Physical Journal B*, 38, pp.381-385.
29. Guimera, R., Mossa, S., Turtschi, A., (2005) “The worldwide air transportation network: anomalous centrality, community structure, and cities’ global roles”, *Proceedings of the National Academy of Science*, 102(22), pp.7794-7799.
30. Jia, T., Jiang, B., (2012) “Building and analyzing the US airport network based on en-route location information”, *Physica A*, 391, pp.4031–4042.
31. Wang, H., Mo, H., Wang, F., Jin, F., (2011a) “Exploring the network structure and nodal centrality of China’s air transport network: A complex network approach”, *Journal of Transport Geography*, 19, pp.712–721.
32. Tsiotas, D., Polyzos, S., (2015) “Decomposing multilayer transportation networks using complex network analysis: A case study for the Greek aviation network”, *Journal of Complex Networks*, 3(4), pp.642-670
33. Fortunato, S., (2010) “Community detection in graphs”, *Physics Reports*, 486, pp.75-174.
34. Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gomez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., Zanin, M. (2014) The structure and dynamics of multilayer networks. *Phys. Rep.*, 544, 1–122.
35. Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J., Moreno, Y., Porter, M. A., (2014) “Multilayer networks”, *J. Complex Netw.*, 2, pp.203–271.
36. Ducruet, C., Ietri, D., Rozenblat, C., (2011) “Cities in worldwide air and sea flows: A multiple networks analysis”, *Eur. J. Geogr*. 528, 23603.
37. Tsiotas, D., Polyzos, S., (2018) “Effects in the network’s topology due to node aggregation: empirical evidence from the maritime transportation network in Greece”, *Physica A*, 491C, pp.71-88.
38. Alderson, D. L., Funk, D., Gera, R., (2020) “Analysis of the global maritime transportation system as a layered network”, *Journal of Transportation Security*, 13(3), pp.291-325.
39. Aleta, A., Meloni, S., Moreno, Y., (2016) “A Multilayer perspective for the analysis of urban transportation systems”, *Scientific Reports*, 7, 44359.
41. Rodrigue, J.P., Notteboom, T.E., (2009) “The geography of containerization: half a century of revolution, adaptation and diffusion”, *Geojournal*, 74, pp.1-5.
42. Brun, J.F., Carrere, C., Guillaumont, P., de Melo, J., (2005) “Has distance died? Evidence from a panel gravity model”, World Bank Economic Review, 19(1), pp.99–120.

43. Ducruet, C., Notteboom, T. E., (2021) “Developing liner service networks in container shipping”, In: Song D.W., Panayides P. (Eds.), *Maritime Logistics. A Guide to Contemporary Shipping and Port Management*, Kogan Page.
44. Notteboom T.E., Ducruet C., de Langen P.W. (2009) *Ports in Proximity. Competition and Coordination among Adjacent Seaports*. Routledge
45. Mohamed-Cherif, F.Z., Ducruet, C., (2016) “Regional integration and maritime connectivity across the Maghreb seaport system”, *Journal of Transport Geography*, 51, pp.280-293.
46. Zohil J., Prijon M. (1999) The MED rule: the interdependence of container throughput and transhipment volumes in the Mediterranean ports. Maritiime Policy and Management, 26(2): 175-193.
47. Walpole, R.E., Myers, R.H., Myers, S.L., Ye, K., (2012) *Probability & Statistics for Engineers & Scientists*, ninth ed., New York, USA, Prentice Hall Publications.
48. Barabasi, A-L., (2013) “Network science”, *Philosophical Transactions of the Royal Society of London A: Mathematical Phys Eng Sci*, 371(1987), 20120375.
49. Koschutzki, D., Lehmann, K., Peeters, L., Richter, S., (2005) “Centrality indices”, In: Brandes, U., Erlebach, T., (eds) *Network analysis*. Springer-Verlag Publications, Berlin, pp 16–61.
50. Newman, M. E. J., (2010) *Networks: an introduction*, Oxford, Oxford University Press.
51. Tsiotas, D., (2019) “Detecting different topologies immanent in scale-free networks with the same degree distribution”, *Proceedings of the National Academy of Sciences*, 116(14), pp.6701-6706.
52. Norusis, M., (2008) *SPSS 16.0 advanced statistical procedures companion*, Prentice Hall Press.
53. Box, G., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., (2015) *Time series analysis: forecasting and control*, New Jersey, John Wiley & Sons.
54. Das, S., (1994) *Time series analysis*, New Jersey, Princeton University Press.
55. Wang, C., Ducruet, C., (2012) “New port development and global city making: Emergence of the Shanghai-Yangshan multilayered gateway hub”, *Journal of Transport Geography*, 25, pp.58-69.
56. Rodrigue, J. P., (2004) “Appropriate models of port governance: Lessons from the Port Authority of New York and New Jersey”, In: D. Pinder and B. Slack (eds) *Shipping and Ports in the 21st Century*, London: Routledge.
57. Rodrigue, J. P., (2015) “Transshipment hubs: connecting global and regional maritime shipping networks”, *Port Economics*, September 17, available at the URL: https://www.porteconomics.eu/transshipment-hubs-connecting-global-and-regional-maritime-shipping-networks/ [accessed: 25/8/211].
58. Fleming, D. K., Hayuth, Y., (1994) “Spatial characteristics of transportation hubs: centrality and intermediacy”, *Journal of Transport Geography*, 2(1), pp.3-18.
59. Fremont, A., Ducruet, C., (2005) “The emergence of a mega-port: The case of Busan, from the local to the global”, *Tijdschrift voor Economische en Sociale Geografie*, 96(4), pp.421-432.
60. Lee, S. W., Ducruet, C., (2009) “Spatial glocalization in Asia-Pacific hub port cities: A comparison of Hong Kong and Singapore”, *Urban Geography*, 30(2), pp.162-184.
61. Wang, J. J., (1998) “A container load center with a developing hinterland: a case study of Hong Kong”, *Journal of Transport Geography*, 6(3), pp.187-201.
62. Wang, L., Lin, Y. S., Ducruet, C., (2019) “Study on the changes of ports’ connection across the Taiwan Strait in the global maritime network”, *Acta Geographica Sinica*, 73(12), pp.2282-2296

63. Wang, J. J., Chen, M. C., (2010) “From a hub port city to a global supply chain management center: a case study of Hong Kong”, *Journal of Transport Geography*, 18(1), pp.104-115.
64. Fraser, D. R., Notteboom, T. E., Ducruet, C., (2016) “Peripherality in the global container shipping network: The case of the Southern African container port system”, *Geojournal*, 81(1), pp.139-151.
65. Zohil, J., Prijon, M., (1999) “The MED rule: the interdependence of container throughput and transhipment volumes in the Mediterranean ports”, Maritime Policy and Management, 26(2), pp.175-193.
66. Disdier, A.C., Head, K., (2008) “The puzzling persistence of the distance effect on bilateral trade”. *Review of Economics and Statistics*, 90(1), pp.37-48.
67. Hummels, D., (2001) *Time as a Trade Barrier*. Center for Global Trade Analysis, Purdue University, GTAP Working Paper, No. 1152.
68. Brun, J.F., Carrère, C., Guillaumont, P., de Melo, J., (2005) “Has distance died? Evidence from a panel gravity model”, *The World Bank Economic Review*, 19(1), pp.99-120.
69. Gastner, M.T., Newman, M.E.J., (2006) “The spatial structure of networks”, *The European Physical Journal B*, 49, pp.247-252.
70. Gastner, M.T., Newman, M.E.J., (2006) “Optimal design of spatial distribution networks”, *Physical Review E*, 74, 016117.
71. Rietveld, P., Vickerman, R., (2004) “Transport in regional science: The “death of distance” is premature”, *Papers in Regional Science*, 83(1), pp.229-248.
72. Markusen, A., (1996) “Sticky places in slippery space: A typology of industrial districts”, *Economic Geography*, 72(3), pp.293-313.
73. Alumura, S.A., Campbell, J.F., Contreras, I., Karad, B.Y., Marianov, V., O’Kelly, M.E., (2021) “Perspectives on modeling hub location problems”, *European Journal of Operational Research*, 291(1), pp.1-17.
74. Xu, M., Pan, Q., Muscolini, A., Xia, H., Cannistraci, C.V., (2020) “Modular gateway-ness connectivity and structural core organization in maritime network science”, *Nature Communications*, 11, 2849, https://doi.org/10.1038/s41467-020-16619-5