www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Multiscale core-periphery structure in a global liner shipping network

Sadamori Kojaku[1,2], Mengqiao Xu[3], Haoxiang Xia[3] & Naoki Masuda[2,3]

Received: 16 August 2018
Accepted: 5 November 2018
Published online: 23 January 2019

Maritime transport accounts for a majority of trades in volume, of which 70% in value is carried by container ships that transit regular routes on fixed schedules in the ocean. In the present paper, we analyse a data set of global liner shipping as a network of ports. In particular, we construct the network of the ports as the one-mode projection of a bipartite network composed of ports and ship routes. Like other transportation networks, global liner shipping networks may have core-periphery structure, where a core and a periphery are groups of densely and sparsely interconnected nodes, respectively. Core-periphery structure may have practical implications for understanding the robustness, efficiency and uneven development of international transportation systems. We develop an algorithm to detect core-periphery pairs in a network, which allows one to find core and peripheral nodes on different scales and uses a configuration model that accounts for the fact that the network is obtained by the one-mode projection of a bipartite network. We also found that most ports are core (as opposed to peripheral) ports and that ports in some countries in Europe, America and Asia belong to a global core-periphery pair across different scales, whereas ports in other countries do not.

Transportation networks such as airways, railways and roadways underpin how the goods and people flow. An understanding of the structure of transportation networks is crucial in finding bottleneck of transportation and vulnerable parts, contributing one to improve its efficiency and resilience[1]. Maritime transport is by far the most cost-effective way to move goods and raw materials across the globe. More than 80% of global trade by volume is carried by ships and handled by seaports[2]. The most dominant type of global maritime transport in terms of seaborne trade value is the global liner shipping. To date, container ships carry over 70% value of the world trade[2], making the global liner shipping network (GLSN) indispensable to the development of international trade and the world economy.

Core-periphery (CP) structure is a meso-scale structure of networks that has been found in many networks including transportation networks such as airport networks[3–6], railway networks[7] and road networks[3,8]. With CP structure based on edge density, a network is decomposed into a set of core nodes and that of peripheral nodes[5–13]. The nodes within the core are densely interconnected, those in the periphery are sparsely interconnected, and a node in the core and one in the periphery are connected with some probability depending on the assumption. Previous studies suggested that transportation networks with CP structure would be robust against random failures (e.g., closure) of nodes[14] and realise a competitive trade-off between the cost and profit[15]. Moreover, the existence of a core may contribute to the functional stability of networks[11,16].

The portrait of core-periphery dichotomy was postulated as a means to explain the uneven trade development and economic growth of nations in the process of globalization[17]. Maritime shipping serves as the primary transportation mode for international trade. As such, investigating the CP structure of the GLSN may help us to understand heterogeneous international trade among world regions and countries[18,19]. Specifically, there are many practical questions one can address by uncovering CP structure in maritime networks. How can we plan shipping routes to improve the stability and economic efficiency of seaborne trade? Which are the ports playing key roles in regional trade and those in international trades? How are ports integrated to global trade markets? Therefore, we analyse the CP structure in the GLSN. Crucially, we use the extension of our previous algorithm, Kojaku-Masuda (KM) algorithm[5,6]. The algorithm generally detects multiple CP pairs in networks (Fig. 1), which many other algorithms do not. We use this algorithm for two reasons. First, individual CP pairs are expected to

[1]CREST, JST, Kawaguchi Center Building, 4-1-8, Honcho, Kawaguchi-shi, Saitama, 332-0012, Japan. [2]Department of Engineering Mathematics, Merchant Venturers Building, University of Bristol, Woodland Road, Clifton, Bristol, BS8 1UB, United Kingdom. [3]Faculty of Management and Economics, Dalian University of Technology, No. 2 Linggong Road, Ganjingzi District, Dalian City, Liaoning Province, 116024, China. Sadamori Kojaku and Mengqiao Xu contributed equally. Correspondence and requests for materials should be addressed to N.M. (email: naoki.masuda@bristol.ac.uk)

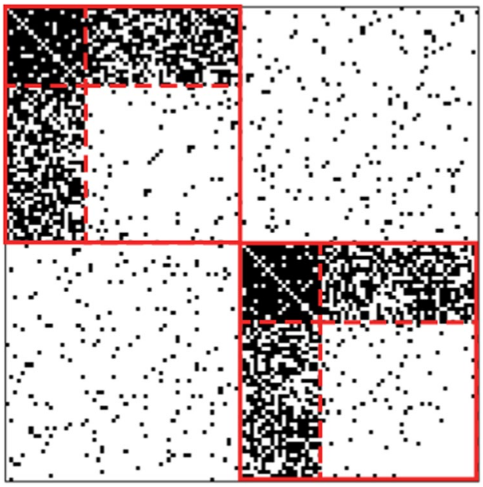


Figure 1. Adjacency matrix of a network with two CP pairs. The filled cell or empty cell indicates the presence or absence of an edge, respectively. The solid line indicates the partition of nodes into the two CP pairs. The dashed lines within each CP pair indicate the subpartition of nodes into the core and periphery. Each core block (top-left block in each CP pair) and periphery block (bottom right block in each CP pair) consist of 20 nodes and 40 nodes, respectively. The probability that each pair of nodes is adjacent by an edge is equal to 0.95 within each core block. The same probability is equal to 0.8 between the core and periphery blocks within each CP pair. The same probability is equal to 0.05 within each periphery block and between different CP pairs. We draw edges according to these probabilities, independently for the different node pairs.

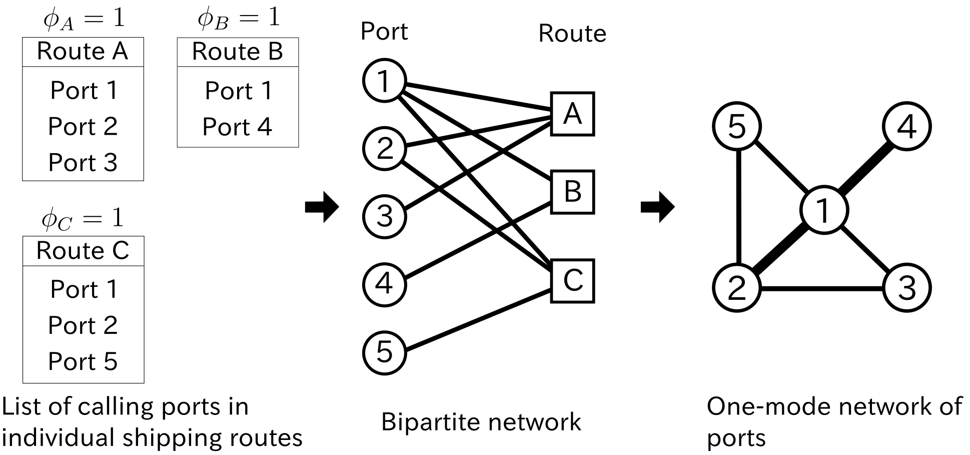


Figure 2. The construction of the GLSN. The width of edges in the one-mode network indicates the edge weight.

correspond to either regional or global (or intermediate) groups of ports in each of which some ports may serve as core ports whereas the others may play a role of peripheral ports. Second, in our previous studies[5,6], the algorithm found CP pairs more accurately than other algorithms did in artificial networks with planted CP pairs. We construct the GLSN from the empirical data on the liner shipping services operated by world's top 100 liner shipping companies in terms of fleet capacity (i.e., the twenty-foot equivalent unit capacity of the fleet). The data altogether account for over 92% of the total fleet capacity in the world.

To reveal the CP structure in the GLSN, we extend our previous algorithm in the following three manners. First, we adopt a null model that is compatible with the way we construct the GLSN from the data. Specifically, the original data set is regarded as a bipartite network composed of a layer of port nodes and a layer of shipping route nodes (Fig. 2). Edges represent which ports belong to which shipping routes. Our null model discounts the effects induced by the one-mode projection of an originally bipartite network. Second, our previous algorithms have a resolution limit, with which one can not find CP structure smaller than a threshold size[6,20]. To circumvent this problem, we use a multiresolution method for community detection[21,22] to extend the algorithm. Third, our previous algorithms provide different CP structures in the different runs of the same algorithm even if the initial condition is the same. In the present study, we run the algorithm 100 times and look at the consensus of the results obtained from the different runs.

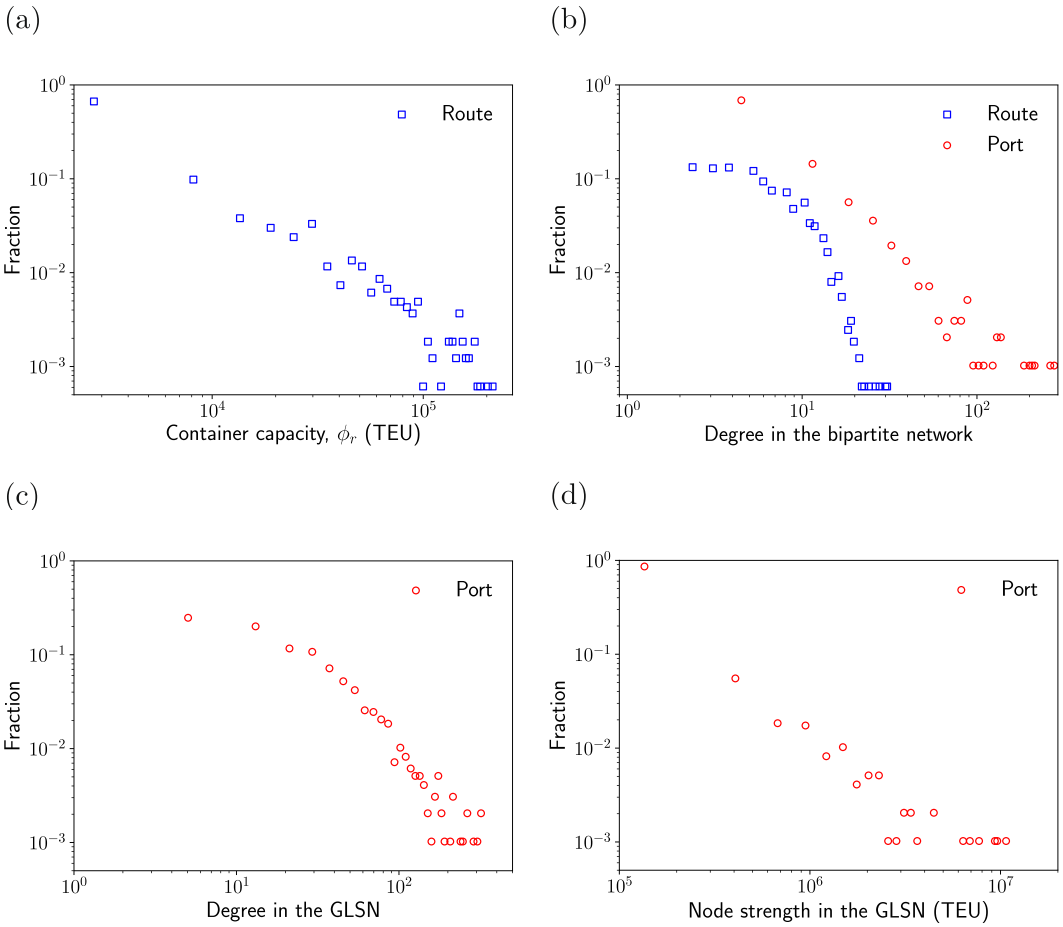


Figure 3. Distributions of (**a**) container capacity φ_r of each route, (**b**) the node's degree in the bipartite network, (**c**) the port's (unweighted) degree in the GLSN and (**d**) the port's weighted degree (i.e., node strength) in the GLSN.

The present algorithm is applicable to networks constructed from a one-mode projection of bipartite networks. Examples of such networks include human disease networks[23], metabolic networks[24] and mutualistic networks[25]. The Python code of the present algorithm is available on GitHub[26].

Results

Number of calling ports, number of serving routes, and node strength. The distribution of the container capacity of a route (i.e., the sum of the maximum volume of containers that shipping companies deploy on the shipping route) is shown in Fig. 3(a). The container capacity is heterogeneously distributed; a majority of the shipping routes has a capacity less than 10^2, while 2% of the routes has a capacity larger than 10^5. Degree d_i^{port} of ports in the bipartite network is also heterogeneously distributed (Fig. 3(b)). A majority (56%) of ports is shared by less than five routes, whereas 13 ports (1.3%) including Shanghai and Singapore are shared by more than 100 routes. Degree d_r^{route} of routes in the bipartite network is more homogeneously distributed than d_i^{port}. A majority (52%) of routes contains less than five calling ports. The largest number of calling ports in a route is 31, which covers only 3.2% of the $N = 977$ ports.

The degree of each port in the GLSN is shown in Fig. 3(c). A majority of ports (540 ports; 55%) has a degree less than 25 in the GLSN, while 60 (6%) ports have a degree larger than 100. We define node strength (i.e., weighted degree) of each port by the sum of the weight of edges attached to the port. As is the case for the container capacity, node strength is heterogeneously distributed (Fig. 3(d)). Most ports (813 ports; 83%) have a strength less than 2×10^5, while 51 ports (5%) have a strength larger than 10^6.

Multiscale CP structure. We identify consensus CP pairs (we call them CP pairs for short in the following text) using the algorithm presented in the Multiresolution algorithm section. The present algorithm is equipped with a resolution parameter γ, with which one can control the characteristic size of CP pairs to be detected. Different γ values may yield considerably different results. Therefore, we examine CP pairs across a range of γ, i.e., $\gamma \in \{0.01, 0.1, 0.2, 0.3, \ldots, 4\}$.

We show the CP pairs detected at some γ values in Figs. 4–6. There are at most five CP pairs. For $0.01 \le \gamma \le 1.9$, the algorithm identifies a unique CP pair containing ports in various geographical regions (Fig. 4). We refer to this CP pair as CP pair 1. The number of ports in CP pair 1 decreases from 951 ports at $\gamma = 0.01$ to 76 ports at

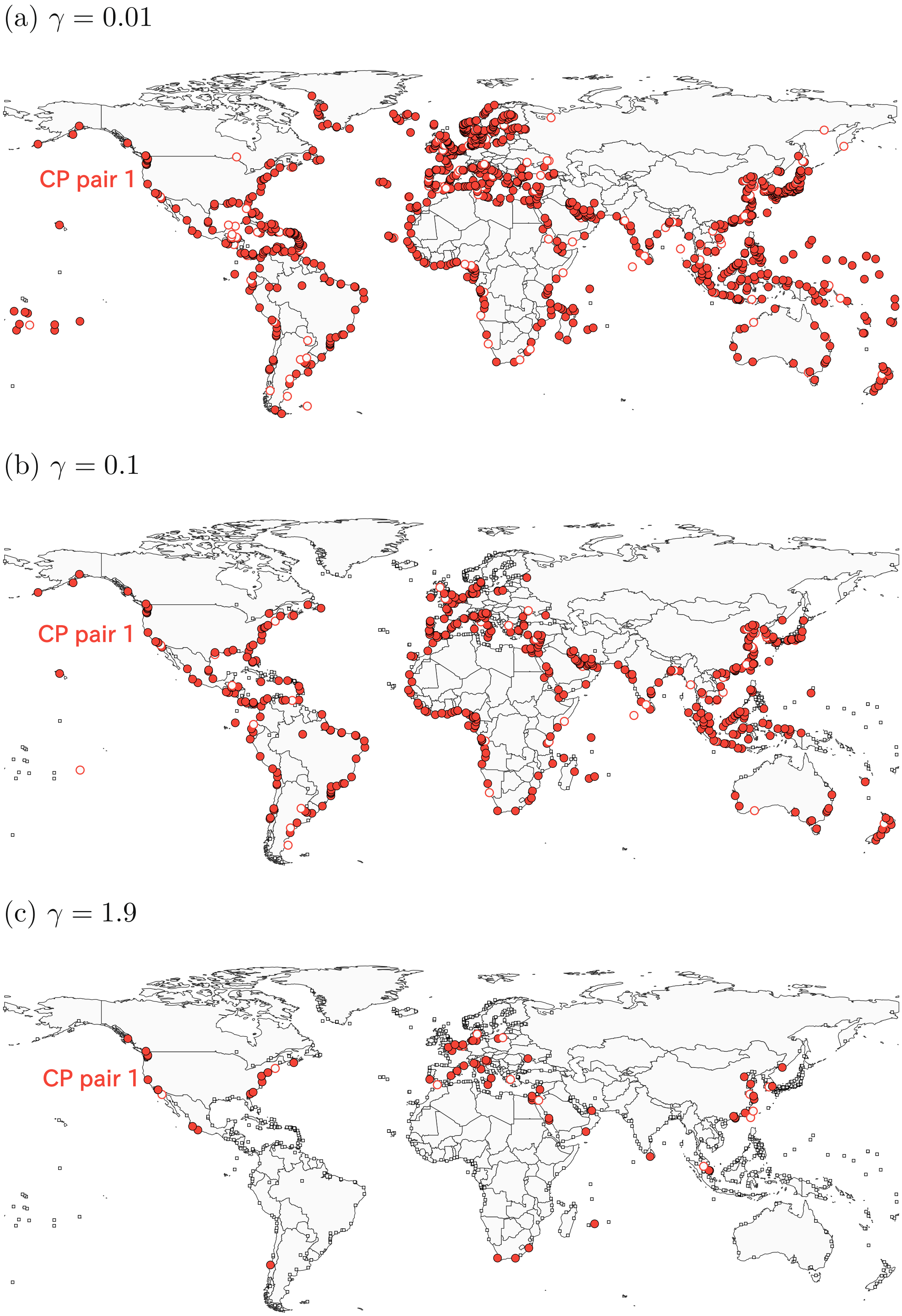
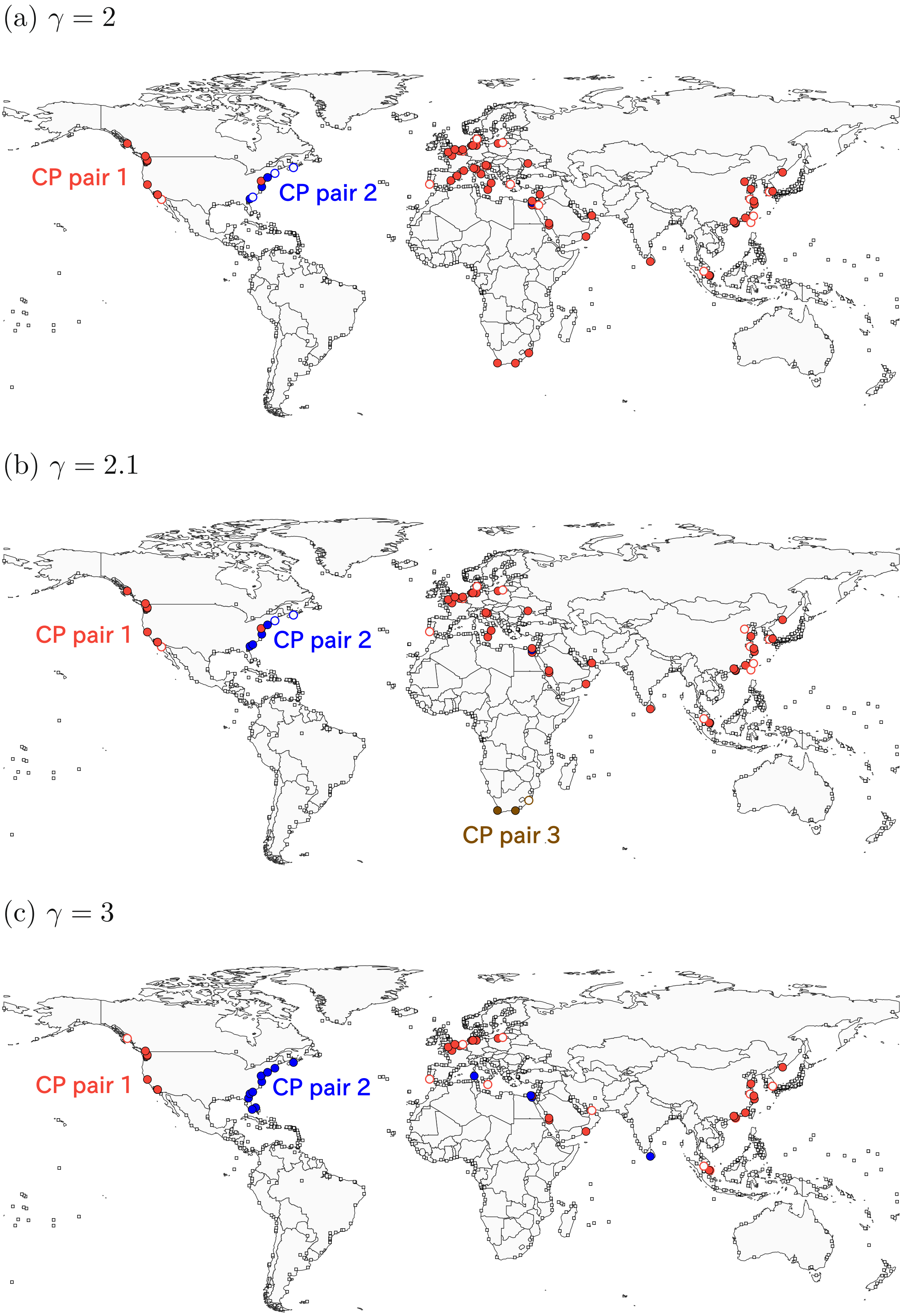


Figure 4. Consensus CP pairs in the GLSN. The resolution is equal to (**a**) $\gamma = 0.01$, (**b**) $\gamma = 0.1$, (**c**) $\gamma = 1.9$. The filled circles indicate the ports with a coreness value larger than 0.5. The open circles indicate the ports with a coreness value less than or equal to 0.5. The open squares indicate homeless ports.

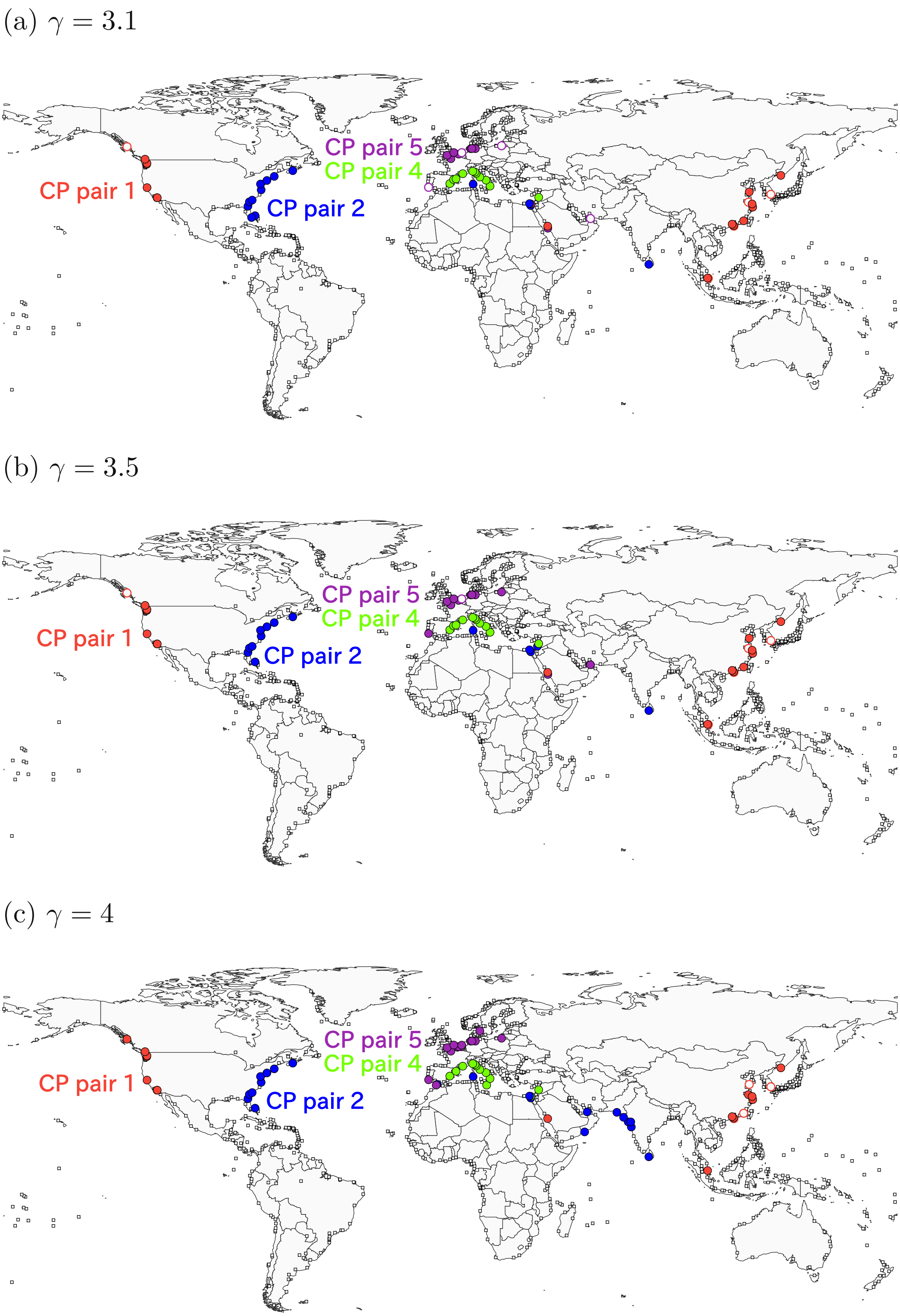


(a) $\gamma = 2$

(b) $\gamma = 2.1$

(c) $\gamma = 3$

Figure 5. Consensus CP pairs in the GLSN. The resolution is equal to (**a**) $\gamma = 2$, (**b**) $\gamma = 2.1$, (**c**) $\gamma = 3$.



(a) $\gamma = 3.1$

(b) $\gamma = 3.5$

(c) $\gamma = 4$

Figure 6. Consensus CP pairs in the GLSN. The resolution is equal to (**a**) $\gamma = 3.1$, (**b**) $\gamma = 3.5$ and (**c**) $\gamma = 4$.

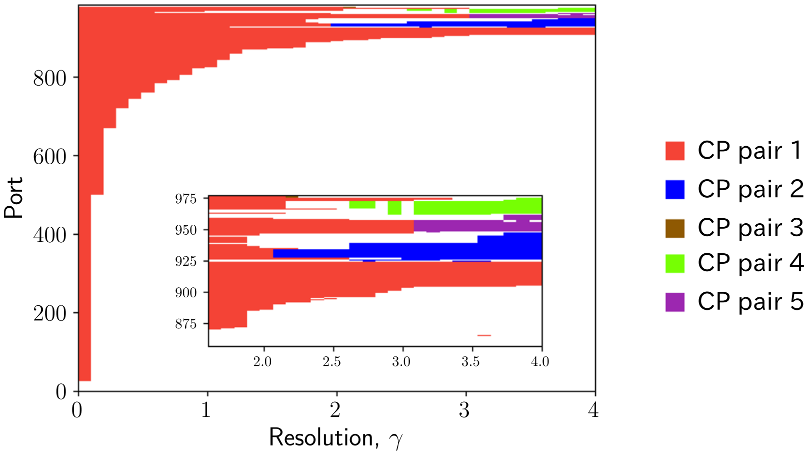


Figure 7. Membership of each port. The colour at (γ, i) indicates the index of the CP pair to which port i belongs at resolution γ. The colour code is the same as that used in Figs. 4–6.

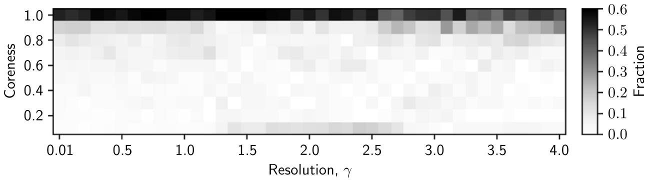


Figure 8. Distribution of coreness values of the ports in the consensus CP pairs.

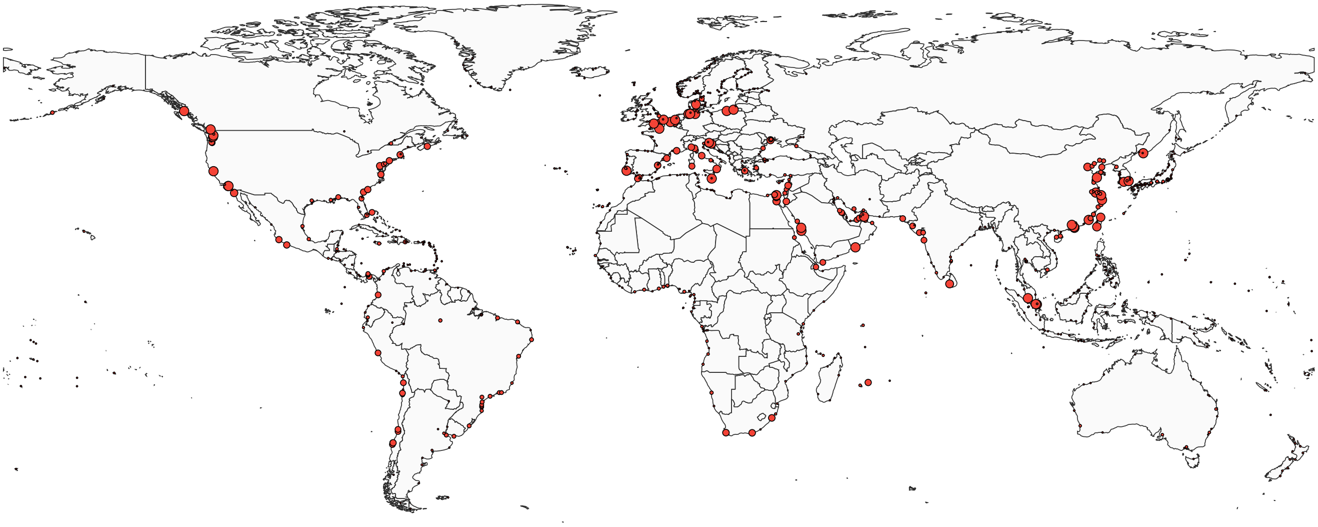


Figure 9. Persistence of each port, i.e., the largest resolution at which the port belongs to CP pair 1. The radius of the circle is proportional to the persistence of the port.

$\gamma = 1.9$. At $\gamma = 1.9$, the CP pair 1 contains many ports in China, the North Sea, the Mediterranean Sea and North America. Few ports in Oceania, the South America, the West Africa and the East Africa belong to CP pair 1.

For $2 \leq \gamma \leq 3$, the algorithm identifies three CP pairs (Fig. 5). As is the case for $0.01 \leq \gamma \leq 1.9$, CP pair 1 contains the ports across many regions. At $\gamma = 2.0$, the algorithm identifies CP pair 2 that branches from CP pair 1 (Fig. 5(a)). CP pair 2 contains most ports in the East Coast of the US, a Canadian port (Halifax) and an Egyptian port (Suez). At $\gamma = 2.1$, the algorithm identifies CP pair 3 located in the South Africa (Fig. 5(a)). CP pair 2 persists and enlarges in most cases as γ increases. In contrast, CP pair 3 is absent for $\gamma \geq 2.2$ (Fig. 5(b)).

For $3.1 \leq \gamma \leq 4$, the algorithm identifies four CP pairs. Each of CP pairs 1 and 2 spans different continents (Fig. 6). At $\gamma = 3.1$, CP pair 1 contains a majority of Chinese ports, the only port in Singapore and ports in the West Coast of the US. CP pair 2 contains most ports in the East Coast of the US, two ports in the Mediterranean

Name	Country	Strength	Persistence
Shanghai	China	10,831,283	4.0
Shenzhen	China	9,646,887	4.0
Ningbo-Zhoushan	China	9,415,002	4.0
Hong Kong	China	6,880,967	4.0
Busan	South Korea	6,309,888	4.0
Qingdao	China	4,369,577	4.0
Xiamen	China	3,416,973	4.0
Tanjung Pelepas	Malaysia	3,116,464	4.0
Guangzhou	China	2,361,131	4.0
Oakland	US	1,953,242	4.0
Los Angeles	US	1,360,462	4.0
Long Beach	US	1,230,191	4.0
Vostochny	Russia	931,440	4.0
Vancouver	Canada	902,164	4.0
King Abdullah	Saudi Arabia	848,920	4.0
Tacoma	US	634,167	4.0
Seattle	US	606,047	4.0
Prince Rupert	Canada	120,980	4.0
Jiangyin	China	30,156	4.0
Singapore	Singapore	7,732,268	3.8
Port Said	Egypt	1,798,944	3.3
Rotterdam	Netherlands	4,343,461	3.0
Hamburg	Germany	3,695,669	3.0
Port Kelang	Malaysia	3,512,490	3.0
Felixstowe	UK	2,378,804	3.0
Le Havre	France	2,142,498	3.0
Bremerhaven	Germany	1,905,953	3.0
Jeddah	Saudi Arabia	1,874,875	3.0
Salalah	Oman	1,673,569	3.0
Marsaxlokk	Malta	1,462,921	3.0
Southampton	UK	1,441,165	3.0
Khor Fakkan	UAE	921,245	3.0
Zeebrugge	Belgium	684,815	3.0
Sines	Portugal	441,525	3.0
Wilhelmshaven	Germany	415,918	3.0
Gdansk	Poland	225,241	3.0
Kaliningrad	Russia	182,586	3.0
Kwangyang	South Korea	1,587,926	2.9
Aarhus	Denmark	241,769	2.9
Trieste	Italy	348,319	2.8
Rijeka	Croatia	242,332	2.8

Table 1. Ports with the largest persistence values.

Sea, a port in Sri Lanka. The other CP pairs 4 and 5 also branch from CP pair 1 and are composed of geographically close ports. In fact, CP pairs 4 and 5 mostly consist of the Mediterranean ports and North European ports, respectively.

The membership of each port at each γ value is shown in Fig. 7. The number of ports in CP pair 1 decreases as γ increases. CP pairs 2, 4 and 5 detected for $2 \leq \gamma \leq 4$ are part of CP pair 1 detected for smaller γ values. CP pair 4 is absent for some γ values for $2.6 \leq \gamma \leq 3$ but persists for $3.1 \leq \gamma \leq 4$. As γ increases, CP pairs 2, 4 and 5 largely expand by absorbing ports that belong to CP pair 1 at small γ values.

The distribution of the coreness values of ports in any CP pair is shown in Fig. 8. For all γ values, most ports have a coreness value larger than 0.9. Therefore, the algorithm has classified most ports as core ports in most runs. If a CP pair only consists of core nodes, then the CP pair is a group of nodes that are densely interconnected with each other, which is equivalent to the usual notion of community. Therefore, the current result indicates that the detected CP pairs are close to communities. This property holds true for all γ values that we have examined.

Persistence of ports. CP pair 1 considered across different resolutions (i.e., γ) has a nested relation. In other words, CP pair 1 at resolution γ contains CP pair 1 at all larger γ values in a majority of cases. This is the case for all but two ports when one varies γ in the range $0.01 \le \gamma \le 4$. Based on this observation, we define the persistence of a port as the smallest γ value above which the port does not belong to CP pair 1 for the first time as one increases γ. In other words, the persistence is the largest value of γ such that the port belongs to CP pair 1 for all resolution values up to that γ value. We note that the persistence is independent of γ.

The persistence of each port is represented by the size of the circle in Fig. 9. In the figure, only the ports belonging to CP pair 1 at $\gamma = 0.01$ are shown. Highly persistent ports (e.g., persistence value larger than 3) are concentrated in China, the North Sea, the Mediterranean Sea, the Malay Peninsula, the Red Sea, and the West Coast of the US. The two highly persistent ports in the Malay Peninsula, Singapore and Tanjung Pelepas, face the Strait of Malacca, which is an important shipping lane in the world[27]. There are few highly persistent ports in the Caribbean Sea, Japan, Oceania, the East Coast of the South America, the East Africa and the West Africa. Therefore, these regions may be relatively segregated from the main international shipping trade networks.

We show the ports with the persistence value larger than 2.8 in Table 1. Highly persistent ports have a relatively large node strength (i.e., weighted degree). More precisely, the persistence and node strength are positively correlated with the Spearman correlation coefficient being equal to 0.83. We find that 497 ports (51%) have a persistence value less than or equal to 0.1, while 64 ports (7%) have a persistence value larger than 2.

Discussion

We developed a multiscale algorithm to identify CP structure in a one-mode projection of bipartite networks, which intends to reveal multiscale CP pairs across different scales. We applied the algorithm to a GLSN and revealed the inequality of regions in terms of the extent to which they are integrated into the global maritime transportation system. Specifically, our algorithm uncovered the following properties of the CP structure in the GLSN.

First, at a coarse resolution, we detected a unique CP pair (CP pair 1) that mainly consists of ports in Asia, Europe and North America (Fig. 4(c)). As major production and consumption centres on a global scale, these three regions have long been seen as dominating poles in global trade and container shipping activities[28]. Container shipping services that connect Asia and Europe, Asia and North America, and Europe and North America constitute the world's main East-West trading lanes, well-known as "East-West Corridor" in the maritime shipping industry[29]. Our result also provides some information on the integration of the economy in different regions into the global markets. For instance, the ports in CP pair 1 are located in leading countries in trades (e.g., China, France, Germany, the United Kingdom and the United States) but not in Japan. The absence of Japanese ports indicates that the integration of Japan into the global maritime transportation system may be insufficient, despite its status as the world's fourth-largest export economy in value. This situation might have a negative influence on the country's international trade development in the long run.

Second, for finer resolutions, the algorithm identified four small CP pairs that branch from CP pair 1 (Fig. 6). These CP pairs involve main regional liner shipping markets of North Europe, Mediterranean, East Asia and North America, respectively. Two out of the four CP pairs, which are composed of major container ports in Northern Europe (CP pair 4) and the Mediterranean (CP pair 5), respectively, are geographically concentrated. In the liner shipping industry, they are highly developed conventional markets of intra-regional seaborne trade in Europe. In contrast, the other two CP pairs extend across distinct geographical regions, corresponding to two inter-regional shipping routes in the West-East direction: North American East Coast-Mediterranean Sea-Indian Subcontinent shipping route via Suez Canal (CP pair 2) and North American West Coast-East Asia shipping route across the Pacific Ocean (CP pair 1). In particular, the dominance of China and the US in CP pair 1 is consistent with the high intensity of the bilateral trade between China and the US, the world's two largest countries in commodity trades[30].

Third, the present algorithm classified a majority of ports in the GLSN as core ports as opposed to peripheral nodes (Fig. 8), as indicated by their high coreness values. Although we do not know why this is the case, the result underlines the specificity of the GLSN. In fact, in worldwide airport networks, more than half of the airports were classified as peripheral nodes[5,6]. This comparison indicates that the GLSN may be better regarded as a collection of communities, which is in agreement with the previous work reporting the community structure of global maritime shipping networks[31]. It should be noted that we found that CP pairs in the GLSN were similar to communities because we actually ran CP analysis.

Fourth, the persistence that we calculated for each port might be useful in evaluating the extent to which a port is integrated into the main international seaborne trade markets. The majority of the most persistent ports are regional load centres in the container shipping markets, i.e., world's leading container ports in terms of the yearly container throughput volume[32]. Examples include East Asian ports of Busan, Guangzhou, Hong Kong, Ningbo-Zhoushan, Qingdao, Shanghai, and Shenzhen, Southeast Asian ports of Singapore and Tanjung Pelepas, North American West Coastal ports of Long Beach and Los Angeles, and European ports of Antwerp, Hamburg and Rotterdam.

Our study has the following limitations. First, we did not inform the edge weight by the actual container traffic between ports due to the commercial confidentiality. Instead, we used traffic capacity deployment data provided by shipping companies to approximate the actual traffic, assuming that the traffic capacity between any port pair on a same shipping service route was equal and bidirectional. Second, one-mode projection discards much information about the original bipartite network composed of the ports and routes. To mitigate this problem, one can use other one-mode projection methods that reflect some properties of the bipartite network to the projected networks[33–35]. Another approach is to study the original bipartite network without one-mode projection. Third, we did not analyse another family of CP structure, i.e., transportation-based CP structure[3,7,8,11]. Transportation-based CP structure dictates that a core is a group of nodes that are frequently used in paths connecting nodes, e.g., nodes with high betweenness centrality. Because GLSNs underlie maritime transportation, analysis of transportation-based CP structure may yield useful knowledge of the flow of cargo across the world.

Methods

Data set. We use an empirical data set provided by Alphaliner[36], which reports the statistics of $R = 1{,}631$ major liner shipping service routes in the world for the year 2015. On each liner shipping service route (hereafter, shortened as service route), container ships call at a sequence of ports with a fixed service schedule. Cargo ships may call at ports for bunkering and maintenance, which are not directly associated with trade. The present data set contains only the calling ports for cargo loading and unloading, ensuring a high relevance to world seaborne trade. There are $N = 977$ ports in total. We denote by d_r^{route} the number of calling ports for route r. Additionally, we denote by d_i^{port} the number of routes that port i serves. The container capacity of route r, denoted by ϕ_r, is given by the sum of the maximum volume of containers (counted in Twenty Equivalent Unit; TEU) deployed on shipping route r by world shipping companies.

The data set does not contain the amount of containers transported between ports owing to the commercial confidentiality. Therefore, we assume that the same amount of containers is transported between any pair of ports belonging to the same route. This procedure is equivalent to the following one-mode projection of the bipartite network.

We represent the data as a bipartite network composed of ports and routes, where a port i and a shipping route r are adjacent if and only if port i is a calling port of route r (Fig. 2). We denote by $B = (B_{ir})$ the $N \times R$ adjacency matrix of the bipartite network, where $B_{ir} = 1$ or $B_{ir} = 0$ indicates that port i and route r are adjacent or not adjacent, respectively.

We construct the GLSN composed of ports by projecting the bipartite network to a one-mode network (Fig. 2). For example, in collaboration networks between academic authors, one connects all pairs of authors of a paper by an edge, resulting in a clique. Because a larger clique (i.e., a paper involving more authors) implies that the pairwise relationships between each pair of authors would be weaker, one often normalises the edge weight by dividing it by $d-1$[37,38], where d is the number of authors of the paper. We apply the same method to the GLSN because the pairwise relationship between ports on a route would be relatively weak if the route involves many ports. We assume that a route is worth a summed edge weight of unity for each port. Then, we obtain

$$W_{ij} \equiv [1 - \delta(i, j)] \sum_{r=1}^{R} \frac{\phi_r}{d_r^{\text{route}} - 1} B_{ir} B_{jr}, \tag{1}$$

where $\delta(\cdot,\cdot)$ is Kronecker delta. The sum of the weight of edges incident to each port (i.e., node strength) is equal to the sum of the container capacity deployed in all the individual service routes in which the port is involved. This quantity is used for calculating the well-known country-level liner shipping connectivity index (LSCI)[39]. We note that the GLSN is a weighted network and does not contain self-loops (i.e., edges whose endpoints are the same node).

Multiresolution algorithm. We regard a network as a collection of C non-overlapping CP pairs (Fig. 1). Each CP pair consists of one core block (i.e., group of nodes) and one periphery block. By construction, there are many edges within each core block, whereas there are relatively few edges within each periphery block. One may assume that there are many edges between the core and periphery blocks[9,13] or few edges[40,41]. We assume that there are many edges between the core and periphery blocks because we need to pair each periphery block with a particular core block.

The present algorithm is an extension of our previous algorithm, which we call the KM algorithm[5,6]. Therefore, we start by explaining the KM algorithm. The algorithm identifies multiple CP pairs in networks, which many previous algorithms do not. In the KM algorithm, we quantify the intensity of CP structure of a network by

$$S \equiv \frac{1}{2\Omega} \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} x_i x_j \delta(c_i, c_j) + \frac{1}{2\Omega} \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} [(1 - x_i) x_j + x_i (x_j - 1)] \delta(c_i, c_j), \tag{2}$$

where c_i is the index of the CP pair to which node i belongs, and $x_i = 1$ or $x_i = 0$ indicates that node i is a core node or a peripheral node, respectively. The first and second terms on the right-hand side of Eq. (2) are the fraction of the weight of edges confined within the core blocks and that connecting the core and periphery blocks within a CP pair, respectively. Quantity $\Omega = \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij}/2$ is the sum of the edge weight in the entire network, which normalises the value of S between 0 and 1. The KM algorithm seeks CP pairs by maximising

$$Q^{\text{CP}} \equiv S - \mathbb{E}[\tilde{S}] = \frac{1}{2\Omega} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(W_{ij} - \mathbb{E}[\widetilde{W}_{ij}] \right) (x_i + x_j - x_i x_j) \delta(c_i, c_j), \tag{3}$$

where $\tilde{S}$ is the value of S in a sample network generated from a null model. The adjacency matrix of the sampled network is denoted by $\widetilde{\mathbf{W}} = (\widetilde{W}_{ij})$. The expectation with respect to the null model is denoted by $\mathbb{E}[\cdot]$. We note that Q^{CP} is equivalent to the modularity[21,42] when all nodes are core nodes, i.e., $x_i = 1$ $(1 \le i \le N)$.

This algorithm has a resolution limit[6]. In other words, CP pairs whose size is smaller than a threshold cannot be detected. The modularity maximisation for finding communities in networks also shares this shortcoming[43]. To discuss the CP structure at different resolutions, here we extend the algorithm[6,20] using multiresolution methods[21,22]. In the new algorithm presented in this study, we seek CP pairs by maximising

$$Q_{\gamma}^{\text{CP}} \equiv \frac{1}{2\Omega} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(W_{ij} - \gamma \mathbb{E}[\widetilde{W}_{ij}] \right) (x_i + x_j - x_i x_j) \delta(c_i, c_j), \tag{4}$$

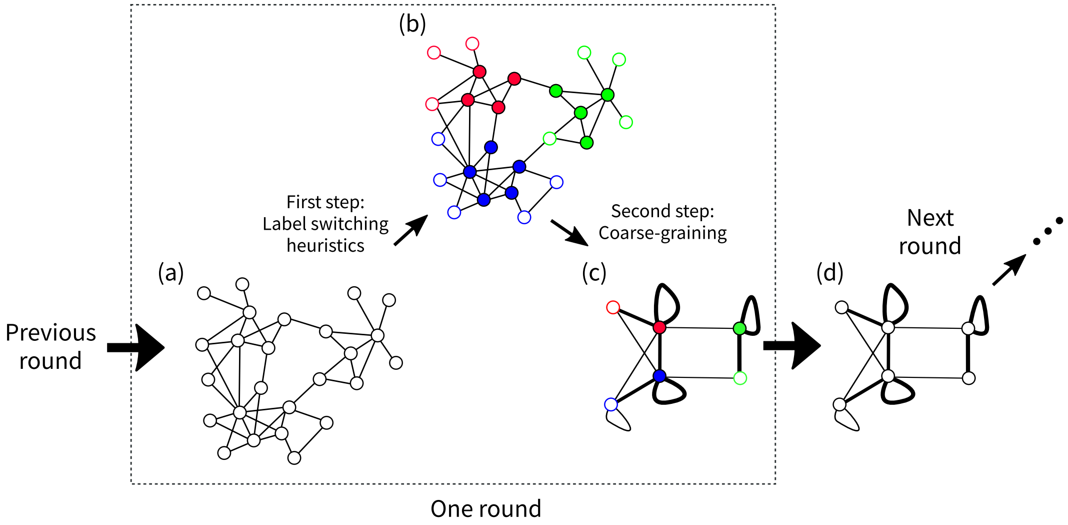


Figure 10. Schematic illustration of the variant of the Louvain algorithm. At the beginning of the current round, we have an input network of nodes (i.e., (**a**)). In the first step, we detect CP pairs in the input network using a label switching heuristic (i.e., (**b**)). In the second step, we construct a coarse-grained network by contracting the nodes in the input network having the same label into a super-node (i.e., (**c**)). Then, we perform the next round of which the input network is the coarse-grained network of the current round (i.e., (**d**)). We iterate the rounds until the value of Q_{γ}^{CP} stops increasing. (**a**) Input network for the current round. (**b**) CP pairs detected in the first step. The colour of each node indicates the CP pair to which the node belongs, i.e., c_i $(1 \leq i \leq N)$. The filled and blank circles indicate core and peripheral nodes, respectively, i.e., x_i. (**c**) Coarse-grained network constructed in the second step. The colour and openness of circles indicate the label (c_i, x_i) of super-node i. The thickness of the edge between super-nodes indicates the weight of the edge, i.e., the sum of the weight of the edges between a node in the input network belonging to one super-node and a node in the input network belonging to the other super-node. (**d**) The input network for the next round.

where γ $(\gamma \geq 0)$ is a resolution parameter that controls the effect of the null model term (i.e, $\mathbb{E}[\widetilde{W}_{ij}]$). The value of γ affects the size of the CP pairs. A detected CP pair is typically large if γ is small. It should be noted that Q_{γ}^{CP} is equivalent to Q^{CP} when $\gamma = 1$.

The KM algorithm accepts various null models. We exploit this property to mitigate the artificial effect induced by the one-mode projection of bipartite networks such as the abundance of large cliques in the projected network. In our previous algorithms[5,6], we have adopted the Erdős-Rényi random graph[44] or the configuration model[45] as the null model. With the configuration model, we rewire the edges by preserving the degree of each node; the Erdős-Rényi random graph does not preserve the degree of each node. Here we use the configuration model as the null model because it is a standard null model in community detection[42], rich-club detection[46] and motif analysis[47]. However, applying the configuration model directly to the GLSN is problematic because the GLSN is obtained as the one-mode projection of a bipartite network (i.e., Eq. (1)). To circumvent this problem, we incorporate the effect of the one-mode projection into the configuration model, similar to a previous study on community detection[37], as follows.

We generate a randomised bipartite network, whose adjacency matrix is denoted by $\widetilde{\mathbf{B}} = (\widetilde{B}_{ir})$, using the configuration model. In other words, the randomised network preserves the degree of each node and the bipartiteness; otherwise, the network is uniformly randomly generated. We allow multi-edges (i.e., multiple edges between the same pair of nodes) in the randomised bipartite networks for computational ease. We carry out the one-mode projection of $\widetilde{\boldsymbol{B}}$ to obtain a randomised unipartite network. The expected edge weight of the randomised unipartite network, $\mathbb{E}\ [\widetilde{W}_{ij}]$, is given by

$$\mathbb{E}[\widetilde{W}_{ij}] = [1 - \delta(i, j)]\mathbb{E}\left[\sum_{r=1}^{R}\frac{\phi_r}{d_r^{\mathrm{route}} - 1}\widetilde{B}_{ir}\widetilde{B}_{jr}\right]. \tag{5}$$

The randomised bipartite network (whose adjacency matrix is **B**) preserves the degree d_r^{route} of each route r. Therefore, Eq. (5) simplifies to

$$\mathbb{E}[\widetilde{W}_{ij}] = [1 - \delta(i, j)]\sum_{r=1}^{R}\frac{\phi_r}{d_r^{\mathrm{route}} - 1}\mathbb{E}\left[\widetilde{B}_{ir}\widetilde{B}_{jr}\right]. \tag{6}$$

The term $\mathbb{E}\ [\widetilde{B}_{ir}\widetilde{B}_{jr}]$ represents the probability that ports i and j are adjacent to route r in the randomised bipartite network. With the configuration model, the probability that ports i and j are adjacent to route r is equal to[37]

$$\mathbb{E}\left[\widetilde{B}_{ir}\widetilde{B}_{jr}\right] = d_i^{\text{port}} d_j^{\text{port}} \frac{d_r^{\text{route}}(d_r^{\text{route}} - 1)}{M(M - 1)}, \tag{7}$$

where $M = \sum_{r'=1}^{R} d_{r'}^{\text{route}}$ is the number of edges in the randomised bipartite network. Substitution of Eq. (7) into Eq. (6) yields

$$\mathbb{E}[\widetilde{W}_{ij}] = [1 - \delta(i, j)] d_i^{\text{port}} d_j^{\text{port}} \sum_{r=1}^{R} \frac{\phi_r d_r^{\text{route}}}{M(M - 1)}. \tag{8}$$

By substituting Eq. (8) into Eq. (4), we obtain the quality function

$$Q_\gamma^{\text{CP}} = \frac{1}{2\Omega} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(W_{ij} - \gamma d_i^{\text{port}} d_j^{\text{port}} \sum_{r=1}^{R} \frac{\phi_r d_r^{\text{route}}}{M(M - 1)} \right) (x_i + x_j - x_i x_j) \delta(c_i, c_j). \tag{9}$$

Maximisation of Q_γ^{CP}. We used a label switching heuristic to maximise Q_γ^{CP} in our previous algorithms[6,20]. In our preliminary analysis, we found that the label switching heuristic in the present case detected multiple CP pairs in the GLSN for $\gamma = 0$, whereas a single CP pair is natural anticipation in this case. This result suggests that the label switching heuristic may return notably suboptimal results for various γ values. Therefore, we implemented the following Louvain algorithm[48] to maximise the Q_γ^{CP}, which in fact yielded larger values of Q_γ^{CP} than the label switching heuristic for all γ values that we investigated.

We iterate rounds, each of which consists of two steps (Fig. 10). In the first step, we identify CP pairs in a network using a label switching heuristic. In the second step, we coarse-grain the network by contracting the nodes belonging to the same CP pair detected in the first step into a super-node. (To avoid the confusion with the nodes in the original GLSN, here we use the term super-node to refer to a node in the coarse-grained network.) Then, we apply another round of the two steps to the coarse-grained network. We iterate the rounds of the two steps until the value of Q_γ^{CP} stops increasing. Then, we set the label of each node in the original network (i.e., $\mathbf{W}$) to the label of the super-node to which it belongs in the final coarse-grained network.

The details of each step are as follows. Let $\overline{\mathbf{W}}$ be an $N' \times N'$ weighted adjacency matrix of the network in the beginning of the rth round, where N' is the number of super-nodes in the beginning of the rth round. We note that $\overline{\mathbf{W}} = \mathbf{W}$ and $N' = N$ in $r = 1$. In the first step of each round, we initialise the label of each super-node i by $(c_i, x_i) = (i, 1)$, where $1 \le i \le N'$. Then, we inspect each super-node in a random order. For each inspected super-node i, we propose a new label $(c_i, x_i) = (c_j, 0)$, where super-node j is a neighbour of super-node i in the network specified by $\overline{\mathbf{W}}$. We also propose new label $(c_i, x_i) = (c_j, 1)$. After carrying out this procedure for all neighbours of super-node i, we adopt the proposed label that yields the largest increment in Q_γ^{CP}. If the largest increment in Q_γ^{CP} is negative, then we do not change the label of super-node i. The increment in Q_γ^{CP} caused by changing the label of super-node i from (c, x) to (c', x') is given by

$$\begin{aligned} \frac{1}{\Omega}\Bigg[& \overline{W}_{i,(c',1)} + x'\overline{W}_{i,(c',0)} - \overline{W}_{i,(c,1)} - x\overline{W}_{i,(c,0)} + (x' - x)\overline{W}_{ii} \\ & - \gamma \overline{d}_i (\overline{D}_{(c',1)} + x'\overline{D}_{(c',0)} - \overline{D}_{(c,1)} - x\overline{D}_{(c,0)}) \left(\sum_{r=1}^{R} \frac{\phi_r d_r^{\text{route}}}{M(M - 1)} \right) \Bigg], \end{aligned} \tag{10}$$

where $\overline{d}_i$ is the sum of d_j^{port} values of the nodes belonging to super-node i, $\overline{W}_{i,(c,x)} = \sum_{j=1, j\neq i}^{N'} \overline{W}_{ij} \delta(c, c_j) \delta(x, x_j)$ is the sum of the weight of the edges between super-node i and other super-nodes with label (c, x), and $\overline{D}_{(c,x)} = \sum_{j=1}^{N'} \overline{d}_j \delta(c, c_j) \delta(x, x_j)$ is the sum of $\overline{d}_i$ over the super-nodes with label (c, x). We note that $\overline{W}_{ii}$ is the edge weight of the self-loop of super-node i. If no label has changed in the process of inspecting the N' super-nodes, then we proceed to the second step. Otherwise, we repeat to draw a new random order of the N' super-nodes and inspect the N' super-nodes for possible label switching, until no further increase in Q_γ^{CP} occurs.

In the second step, we coarse-grain the network by contracting the super-nodes having the same label as a result of the first step into one super-node. In the new network, the edge weight between two super-nodes representing labels (c, x) and (c', x') is given by the sum of the weight of the edges between a super-node with label (c, x) before the coarse-graining and a super-node with label (c', x') before the coarse graining. We note that the super-nodes may have self-loops (Fig. 10).

Statistical test. We examine the statistical significance of individual CP pairs using the so-called (q, s)-test[6,20] that we previously proposed. The (q, s)-test evaluates the significance of individual CP pairs. For a CP pair in question, the (q, s)-test computes the quality of a CP pair composed of the same number of nodes in randomised networks. Then, the (q, s)-test judges the CP pair in question as significant if its quality value is statistically larger than that of the CP pair of the same number of nodes in randomised networks. The (q, s)-test requires a quality function q for individual CP pairs. We compute the quality of the CP pair c, denoted by q_c, by the contribution of the cth CP pair to Q_γ^{CP}, i.e.,

$$q_c \equiv \frac{1}{2\Omega}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(W_{ij} - \gamma d_i^{\rm port} d_j^{\rm port}\sum_{r=1}^{R}\frac{\phi_r d_r^{\rm route}}{M(M-1)}\right)(x_i + x_j - x_i x_j)\delta(c_i, c_j)\delta(c_i, c). \tag{11}$$

We note that the sum of q_c over all CP pairs is equal to $Q_\gamma^{\rm CP}$.

The value of q_c would be positively correlated with the number n_c of nodes in the cth CP pair[6]. In other words, a large q_c value may be caused by a large number of nodes in the CP pair. To discount the effect of the correlation, the (q, s)-test assesses the significance of the cth CP pair using the conditional probability $P(\tilde{q} \geq q_c|n_c)$ that the quality $\tilde{q}$ of a CP pair of the same size n_c detected in a randomised network is larger than q_c. If $P(\tilde{q} \geq q_c|n_c)$ is smaller than a significance level α $(0 < \alpha \leq 1)$, then one judges the CP pair in question to be significant. Otherwise, the CP pair is insignificant.

In the (q, s)-test, one infers $P(\tilde{q} \geq q_c|n_c)$ as follows. First, we generate 500 randomised networks using the null model discussed in the Multiresolution algorithm section. Second, we detect the CP pairs in the randomised networks using the present algorithm with the same resolution parameter used for finding the CP pair in question. For each $\bar{c}$th detected CP pair in the 500 randomised networks, we compute the quality $\tilde{q}^{(\bar{c})}$ and the number $\tilde{n}^{(\bar{c})}$ of nodes in the CP pair. Third, we infer a joint probability $P(\tilde{q}, \tilde{n})$ using the Gaussian kernel density estimator[49], i.e.,

$$P(\tilde{q}, \tilde{n}) = \sum_{\bar{c}=1}^{\bar{C}} f\left(\frac{\tilde{q} - \tilde{q}^{(\bar{c})}}{h\sigma_{\tilde{q}}}, \frac{\tilde{n} - \tilde{n}^{(\bar{c})}}{h\sigma_{\tilde{n}}}\right)\Big/ \bar{C}, \tag{12}$$

where $\overline{C}$ is the sum of the number of CP pairs detected in the 500 randomised networks, and $\sigma_{\tilde{q}}$ and $\sigma_{\tilde{n}}$ are the unbiased estimation of the standard deviation for $\{\tilde{q}^{(\bar{c})}\}$ and $\{\tilde{n}^{(\bar{c})}\}$ $(1 \leq \bar{c} \leq \overline{C})$, respectively. Function $f(\cdot,\cdot)$ is the bivariate standard normal distribution given by

$$f(y_1, y_2) \equiv \frac{1}{2\pi\sqrt{1 - \rho^2}}\exp\left(-\frac{y_1^2 - 2\rho y_1 y_2 + y_2^2}{2(1 - \rho^2)}\right), \tag{13}$$

where ρ is the Pearson correlation coefficient between $\{\tilde{q}^{(\bar{c})}\}$ and $\{\tilde{n}^{(\bar{c})}\}$ $(1 \leq \bar{c} \leq \overline{C})$. Using Eq. (12), we obtain

$$\begin{aligned} P(\tilde{q} \geq q_c|n_c) &= \frac{\int_{q_c}^{\infty} P(\tilde{q}, n_c)\mathrm{d}\tilde{q}}{\int_{\infty}^{\infty} P(\tilde{q}, n_c)\mathrm{d}\tilde{q}} \\ &= 1 - \frac{\sum_{\bar{c}=1}^{\bar{C}}\exp\left(-\frac{(n_c - \tilde{n}^{(\bar{c})})^2}{2\sigma_{\tilde{n}}^2 h^2}\right)\Phi\left(\frac{\sigma_{\tilde{n}}(q_c - \tilde{q}^{(\bar{c})}) - \rho\sigma_{\tilde{q}}(n_c - \tilde{n}^{(\bar{c})})}{\sigma_{\tilde{n}}\sigma_{\tilde{q}}h\sqrt{1 - \rho^2}}\right)}{\sum_{\bar{c}=1}^{\bar{C}}\exp\left(-\frac{(n_c - \tilde{n}^{(\bar{c})})^2}{2\sigma_{\tilde{n}}^2 h^2}\right)}, \end{aligned} \tag{14}$$

where $\Phi(y) = (2\pi)^{-1/2}\int_{-\infty}^{y}\exp(-u^2/2)\mathrm{d}u$ is the cumulative function of the standard normal distribution.

We note that the Gaussian kernel estimator converges to any form of the probability distribution as the number of samples, $\overline{C}$, increases[50]. Parameter h is a free parameter that affects the speed of the convergence. We use Scott's rule of thumb[51], i.e., $h = \overline{C}^{-1/6}$. We adopt the Šidák correction[52] to evade the multiple comparisons problem. In other words, we test each CP pair in the original network at a significance level of $\alpha = 1 - (1 - \alpha')^{1/C}$, where α' is the targeted significance. We set $\alpha' = 0.05$.

Consensus CP pairs. Even one starts with the same initial condition, the present algorithm yields different significant CP structures in different runs due to the stochasticity of the algorithm. We address this issue by gathering the consensus of the results of different runs, which is regarded as a type of consensus clustering of data points[53–55].

To this end, we first run the present algorithm 100 times for a given value of γ. (We show the results for 6 runs at each γ value in the Supplementary Figures S1–S9). Second, for each pair of ports i and j, we compute the fraction of runs in which ports i and j belong to the same CP pair, which we denote by P_{ij}. Third, we construct an undirected and unweighted network composed of the $N = 977$ ports, where two ports i and j are adjacent if and only if $P_{ij} \geq \theta$. We set $\theta = 0.9$. Finally, we regard each connected component of the network as a consensus CP pair. We refer to the ports that do not belong to any consensus CP pair as homeless ports. We define the coreness of each port i in the consensus CP pair as the fraction of runs in which port i is classified as core port.

Matching CP pairs across resolutions. Given consensus CP pairs calculated at different resolutions, we match consensus CP pairs detected at two consecutive resolutions γ and γ' as follows. For each consensus CP pair c at resolution γ and each consensus CP pair c' at resolution γ', we compute the similarity $\tau_{c,c'}$ between them using the Jaccard index, i.e.,

$$\tau_{c,c'} \equiv \frac{|V_c \cap V_{c'}|}{|V_c \cup V_{c'}|}, \tag{15}$$

where V_c and $V_{c'}$ are the sets of ports in consensus CP pairs c and c', respectively. We match c and c' if $\tau_{c,c'} > \max_{\bar{c}\neq c}\tau_{\bar{c},c'}$ and $\tau_{c,c'} > \max_{\bar{c}\neq c'}\tau_{c,\bar{c}}$. We note that some consensus CP pairs at resolution γ may not be matched with any consensus CP pair at γ' or vice versa. We did not find ties in the $\tau_{c,c'}$ value during the matching procedure.

As a result of the matching, we found seven consensus CP pairs across the resolution values. In fact, three of them (shown in green in Figs. 4–6 and 7) are composed of almost the same set of nodes and reside in different ranges of γ separated by gaps (therefore not contiguous in terms of the γ value). Therefore, we regard these three consensus CP pairs as a single consensus CP pair.

References

1. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).
2. Hoffmann, J. *et al.* Review of maritime transport (United Nations Publications, 2017).
3. Holme, P. Core-periphery organization of complex networks. *Phys. Rev. E* **72**, 046111 (2005).
4. Rossa, F. D., Dercole, F. & Piccardi, C. Profiling core-periphery network structure by random walkers. *Sci. Rep.* **3**, 1467 (2013).
5. Kojaku, S. & Masuda, N. Finding multiple core-periphery pairs in networks. *Phys. Rev. E* **96**, 052313 (2017).
6. Kojaku, S. & Masuda, N. Core-periphery structure requires something else in the network. *New J. Phys.* **20**, 43012 (2018).
7. Rombach, M. P., Porter, M. A., Fowler, J. H. & Mucha, P. J. Core-periphery structure in networks (revisited). *SIAM Rev.* **59**, 619–646 (2017).
8. Lee, S. H., Cucuringu, M. & Porter, M. A. Density-based and transport-based core-periphery structures in networks. *Phys. Rev. E* **89**, 032810 (2014).
9. Borgatti, S. P. & Everett, M. G. Models of core/periphery structures. *Soc. Netw.* **21**, 375–395 (2000).
10. Boyd, J. P., Fitzgerald, W. J. & Beck, R. J. Computing core/periphery structures and permutation tests for social relations data. *Soc. Netw.* **28**, 165–178 (2006).
11. Csermely, P., London, A., Wu, L.-Y. & Uzzi, B. Structure and dynamics of core/periphery networks. *J. Comp. Netw.* **1**, 93 (2013).
12. Tunç, B. & Verma, R. Unifying inference of meso-scale structures in networks. *PLOS ONE* **10**, e0143133 (2015).
13. Cucuringu, M., Rombach, P., Lee, S. H. & Porter, M. A. Detection of core-periphery structure in networks using spectral methods and geodesic paths. *Eur. J. Appl. Math.* **27**, 846–887 (2016).
14. Peixoto, T. P. & Bornholdt, S. Evolution of robust network topologies: Emergence of central backbones. *Phys. Rev. Lett.* **109**, 118703 (2012).
15. Verma, T., Russmann, F., Araújo, N. A. M., Nagler, J. & Herrmann, H. J. Emergence of core-peripheries in networks. *Nat. Commun.* **7**, 10441 (2016).
16. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167 (2011).
17. Krugman, P. & Venables, A. J. Globalization and the inequality of nations. *The Quarterly J. Econ.* **110**, 857–880 (1995).
18. Mahutga, M. C. The persistence of structural inequality? A network analysis of international trade, 1965–2000. *Soc. Forces* **84**, 1863–1889 (2006).
19. García-Pérez, G., Boguñá, M., Allard, A. & Serrano, M. Á. The hidden hyperbolic geometry of international trade: World Trade Atlas 1870–2013. *Sci. Rep.* **6**, 33441 (2016).
20. Kojaku, S. & Masuda, N. A generalised significance test for individual communities in networks. *Sci. Rep.* **8**, 7351 (2018).
21. Reichardt, J. & Bornholdt, S. Statistical mechanics of community detection. *Phys. Rev. E* **74**, 016110 (2006).
22. Heimo, T., Kumpula, J. M., Kaski, K. & Saramäki, J. Detecting modules in dense weighted networks with the Potts method. *J. Stat. Mechanics: Theory and Experiment* **2008**, P08007 (2008).
23. Goh, K.-I. *et al.* The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685–8690 (2007).
24. Guimerà, R. & Amaral, L. A. N. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
25. Padrón, B., Nogales, M. & Traveset, A. Alternative approaches of transforming bimodal into unimodal mutualistic networks. The usefulness of preserving weighted information. *Basic and Appl. Ecol.* **12**, 713–721 (2011).
26. Kojaku, S. & Masuda, N. *Python code of our algorithm*. Available at https://github.com/skojaku/multiresolcp.
27. Qu, X. & Meng, Q. The economic importance of the Straits of Malacca and Singapore: An extreme-scenario analysis. *Transp. Res. Part E: Logis. Transp. Rev.* **48**, 258–265 (2012).
28. César, D. & Theo, N. The worldwide maritime network of container shipping: spatial structure and regional dynamics. *Global Netw.* **12**, 395–423 (2012).
29. Notteboom, T. & Rodrigue, J.-P. Containerisation, box logistics and global supply chains: The integration of ports and liner shipping networks. *Mari. Econom. & Logis.* **10**, 152–174 (2008).
30. United Nations Comtrade Database. UN Comtrade. Available at https://comtrade.un.org/data/ Accessed: 22 Jul 2018.
31. Kaluza, P., Kölzsch, A., Gastner, M. T. & Blasius, B. The complex network of global cargo ship movements. *J. R. Soc. Interface* **7**, 1093–1103 (2010).
32. Lloyd's L. Available at https://lloydslist.maritimeintelligence.informa.com/ Accessed: 30 Jul 2018.
33. Zhou, T., Ren, J., Medo, M. & Zhang, Y.-C. Bipartite network projection and personal recommendation. *Phys. Rev. E* **76**, 046115 (2007).
34. Gualdi, S., Cimini, G., Primicerio, K., Di Clemente, R. & Challet, D. Statistically validated network of portfolio overlaps and systemic risk. *Sci. Rep.* **6**, 39467 (2016).
35. Saracco, F. *et al.* Inferring monopartite projections of bipartite networks: an entropy-based approach. *New J. Phys.* **19**, 53022 (2017).
36. Alphaliner. Available at https://www.alphaliner.com/ Accessed: April 2015.
37. Guimerà, R., Sales-Pardo, M. & Amaral, L. A. N. Module identification in bipartite and directed networks. *Phys. Rev. E* **76**, 036102 (2007).
38. Newman, M. E. J. *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
39. United Nations Conference on Trade and Development. Available at http://unctadstat.unctad.org/wds/ReportFolders/reportFolders.aspx Accessed: 10 July 2018.
40. Boyd, J. P., Fitzgerald, W. J., Mahutga, M. C. & Smith, D. A. Computing continuous core/periphery structures for social relations data with MINRES/SVD. *Soc. Netw.* **32**, 125–137 (2010).
41. Craig, B. & von Peter, G. Interbank tiering and money center banks. *J. Financ. Intermed.* **23**, 322–347 (2014).
42. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
43. Fortunato, S. & Barthélemy, M. Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA* **104**, 36–41 (2006).
44. Erdős, P. & Rényi, A. On random graphs I. *Publ. Math.* **6**, 290–297 (1959).
45. Fosdick, B., Larremore, D., Nishimura, J. & Ugander, J. Configuring random graph models with fixed degree sequences. *SIAM Rev.* **60**, 315–355 (2018).
46. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 110–115 (2006).
47. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).

48. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008).
49. Wand, M. P. & Jones, M. C. Comparison of smoothing parameterizations in bivariate kernel density estimation. *J. American Stat. Assoc.* **88**, 520–528 (1993).
50. Parzen, E. On estimation of a probability density function and mode. *Annal. Math. Stat.* **33**, 1065–1076 (1962).
51. Scott, D. W. Multivariate density estimation and visualization. In *Handbook of Computational Statistics*, 549–569 (Springer, Berlin, 2012).
52. Šidák, Z. Rectangular confidence regions for the means of multivariate normal distributions. *J. Am. Stat. Assoc.* **62**, 626–633 (1967).
53. Strehl, A. & Ghosh, J. Cluster ensembles – A knowledge reuse framework for combining multiple partitions. *J. Machi. Learning Res.* **3**, 583–617 (2002).
54. Topchy, A., Jain, A. K. & Punch, W. Clustering ensembles: models of consensus and weak partitions. *IEEE Trans. Patt. Anal. and Machi. Intel.* **27**, 1866–1881 (2005).
55. Goder, A. & Filkov, V. Consensus clustering algorithms: Comparison and refinement. In Proc. Meeting on Alg. Eng. & Experiments, 109–117 (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008).

Acknowledgements

M.X. acknowledges the support provided through China Postdoctoral Science Foundation, Grant Number 2017M621141. H.X. acknowledges the support provided through National Natural Science Foundation of China under Grant Numbers 71371040 and 71871042. N.M. acknowledges the support provided through JST, CREST, Grant Number JPMJCR1304.

Author Contributions

M.X. and N.M. conceived and designed the research; M.X. preprocessed the empirical data; S.K. and N.M. proposed the algorithm; S.K. performed the computational experiment; S.K., M.X., H.X. and N.M. wrote the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-35922-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.