

Indications of marine bioinvasion from network theory

An analysis of the global cargo ship network

A. Kölzsch^{1,2,a} and B. Blasius²

¹ Project Group Movement Ecology and Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6708PB Wageningen, the Netherlands

² Mathematical Modelling, ICBM, University of Oldenburg, 26111 Oldenburg, Germany

Received 24 March 2011

Published online 10 October 2011 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2011

Abstract. The transport of huge amounts of small aquatic organisms in the ballast tanks and at the hull of large cargo ships leads to ever increasing rates of marine bioinvasion. In this study, we apply a network theoretic approach to examine the introduction of invasive species into new ports by global shipping. This is the first stage of the invasion process where it is still possible to intervene with regulating measures. We compile a selection of widely used and newly developed network properties and apply these to analyse the structure and spread characteristics of the directed and weighted global cargo ship network (GCSN). Our results reveal that the GCSN is highly efficient, shows small world characteristics and is positive assortative, indicating that quick spread of invasive organisms between ports is likely. The GCSN shows strong community structure and contains two large communities, the Atlantic and Pacific trading groups. Ports that appear as connector hubs and are of high centralities are the Suez and Panama Canal, Singapore and Shanghai. Furthermore, from robustness analyses and the network's percolation behaviour, we evaluate differences of onboard and in-port ballast water treatment, set them into context with previous studies and advise bioinvasion management strategies.

1 Introduction

Biological invasions are geographical expansions of a species into areas not previously occupied by it. This is a natural process. However, lately bioinvasion events occur at extremely accelerated rates due to human actions, sometimes by deliberate introductions, but often accidentally [1–4]. Such increased rates of bioinvasion greatly threaten biodiversity and ecosystem functioning worldwide [5,6]. Additionally, they cause damages of human facilities, impact the economy, and pose unpredictable hazards to our health and livelihood [7–9]. Bioinvasion is a three stage process consisting of the introduction, establishment and proliferation of the invasive organisms [1,10]. Introduction is the stage where man can still intervene and the prevention of bioinvasion events may be possible. However, this phase of bioinvasion is least studied up to now [11], which motivated us to conduct the present study.

Bioinvasion of marine organisms is leading to an ever increasing level of homogenisation of the world oceans' ecosystems [12,13]. In some cases the introduction of exotic species has even caused habitat destruction and ecosystem degeneration, e.g. the introduction of *Caulerpa taxifolia* in the Mediterranean [14] and the pronounced fishing declines after invasion of *Mnemiopsis leydii* in the

Black Sea [15]. There are different vectors by which marine bioinvasive organisms can be spread. A large number of small organisms is transported by cargo ships between the ports of the world, mainly due to two different mechanisms. First, organisms can be transported over large distances by the exchange of ballast water [16]. It has been estimated that 3 to 4 billion tons of ballast water are released into the world's oceans per year [17]. Thereby, plankton, larval molluscs and small fishes travel within these huge amounts of ships' ballast water and are released again when the ballast water is discharged at destination ports. Plankton samples from several cargo ships from Japanese ports e.g. contained at least 367 different taxa [17,18]. For bioinvasion success one has to consider that most species are very sensitive to oxygen depletion in the ballast water tanks and high salinity of open seas water [3,18,19]. Second, besides the transport by ballast water exchange, a large number of invasive organisms can also be attached to the ships' hulls and thus be transported into new habitats [20]. Approximately 90% of the present world trade is being transported by ships and trade volumes increase greatly every year [21,22]. Therefore, a huge number of large cargo ships travel the world's oceans at ever increasing rates, thereby advancing levels of marine bioinvasion. For containment thereof the applicability of several ballast water management and anti-fouling options are presently

^a e-mail: a.koelzsch@nioo.knaw.nl

discussed more or less controversially [23,24]. Ballast water may e.g. be chemically cleaned in the ships' tanks, exchanged in the open sea or pumped safely through water exchange facilities of certain ports. A third, important way of marine bioinvasion is introduction by aquaculture. However, here we want to focus on and examine ocean shipping as the main vector of transport of marine bioinvasive organisms.

In this letter we apply techniques from complex network theory to study the potential for bioinvasive spread in the network of global shipping connections and to identify important donor and recipient ports. Network theory has become a widely applied and rather diverse field of study in physics, social sciences, transportation, ecology and epidemiology [25–27]. In several studies habitat patches were modelled as nodes that are linked by dispersal events. The spread of SARS during 2002/2003, for example, has been reasonably well reproduced simulating propagation dynamics on the worldwide airport network [28]. Many different measures have been developed to characterise real-world networks and applied to a variety of systems [29]. Often unexpected properties were revealed and similarities and differences to other networks became obvious. So, for example, in various large and very complex networks (i) pairs of nodes are connected by paths of only few consecutive links and (ii) nodes are locally densely clustered. These two properties have been summarised as “small world behaviour” [30], indicating that any pair of nodes is surprisingly well connected.

The global network of shipping connections, similar to other transportation networks, is a typical example for a geographically embedded network [31]. Recently, Drake and Lodge [32] have developed a global network of ship traffic from information on the inflow and outflow of goods in a number of ports worldwide based on gravity models, i.e. assuming spatially homogeneous trade flows [33]. A description in terms of gravity models however is necessarily crude since world trade and ship traffic are very heterogeneous, much influenced by cultural and political issues (see also [34]). Nevertheless, this network has been used in combination with the worldwide airport network [35] and climate information for studying disease dispersal by global traffic [9]. For improvement maritime shipping networks were reconstructed that take into account real shipping data [36–38]. In particular, we and others developed the global cargo ship network (GCSN) from real cargo ship trajectories based on Automatic Identification System (AIS) data [38]. Beside some overall structural similarities we find several differences between our network, the GCSN, and the shipping network developed from gravity models. Already the number and list of ports differs greatly. For example, there are many important ports that have not been included in [32], since the list of ship trade [39] that was used is not complete. On the other hand there are a number of ports that are not present in the GCSN, mainly African ports. However, when considering the small amount of trade that is transacted through these ports, they may not be very important for large scale consideration of bioinvasion by ships. Thus, we propose

that the GCSN is more appropriate for quantification and analysis of marine bioinvasion by ballast water transport and hull fouling.

In this study we examine structural and spread characteristics of the directed GCSN by applying a selection of network measures. These measures include characteristics such as small world properties, network efficiency and assortativity, but also consider measures that indicate community structure, port centrality and network robustness. We explain how these measures allow assertions about invasion spread and extract preliminary indications of the importance of different ports and trade structures for bioinvasion. Furthermore, we quantify how node deletions, i.e. application of ballast water treatment in ports, can affect bioinvasion spread on the GCSN and which transmission rates are of concern. Applying network measures for bioinvasion predictions and evaluation has the advantage that no complicated models have to be developed and no costly and time consuming field studies be conducted. Thus, already by simple examination of transportation vector networks, preliminary indications of spread characteristics can be derived for decision making authorities.

2 Methods

Because of the strong sensitivity of invasion success to travel conditions and propagule pressure it is very important to consider travel frequency and the duration of transport. Therefore, we consider it crucial to analyse the weighted cargo ship network for indications of the structural importance and dynamical properties of marine bioinvasive spread. In some cases we additionally analysed the unweighted and even the geographically embedded network for comparison. Many links in the GCSN are unidirectional, i.e. the network is highly asymmetric [38]. As this may be important for the patterns of spread, we retain the directionality of the GCSN for all analyses. Since most network measures have been developed for unweighted networks (but see e.g. [40]), we provide a summary of characteristics that are useful for spread analyses in directed, weighted networks (see Tab. 2).

2.1 Small world characteristics

One of the most important network properties for transportation is its connectivity. A quantitative concept of such a characterisation was introduced in [41]. It is based on directed topological or geographical distances between nodes and includes calculations of the local and global efficiencies, E_{loc} and E_{glob} , and the network cost c (see Tab. 1). High local and global efficiencies in combination with low cost indicate small world behaviour. This definition differs from the classical concept of small worlds (see above, [30]).

We calculated efficiencies and network cost for the topological as well as the geographically embedded GCSN,

Table 1. Global and local efficiencies and network cost for the topological and geographically embedded GCSN. In each formula n is the number of nodes in the network G . G_i is the subnetwork of neighbours of i and σ_{ij} are the shortest paths between nodes i and j . Different measures of pair distance are the length of the shortest topological path $|\sigma_{ij}|$, the length of the shortest geographically embedded path $|\sigma_{ij}^{geo}|$ and the waterway shortest distance d_{ij} . a_{ij} are the elements of the adjacency matrix, $a_{ij} = 1$ if there exists a link from i to j , $a_{ij} = 0$ else.

Network measure	Topological network	Geographically embedded network
Global efficiency	$E_{glob} = E(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{ \sigma_{ij} }$	$E_{glob}^{geo} = E^{geo}(G) = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{ \sigma_{ij}^{geo} }$
Local efficiency	$E_{loc} = \frac{1}{n} \sum_{i \in G} E(G_i)$	$E_{loc}^{geo} = \frac{1}{n} \sum_{i \in G} E^{geo}(G_i)$
Network cost	$c = \frac{\sum_{i \neq j \in G} a_{ij}}{n(n-1)}$	$c^{geo} = \frac{\sum_{i \neq j \in G} a_{ij} d_{ij}}{\sum_{i \neq j \in G} d_{ij}}$

where each node is characterised by its geographical position. This indicates how well connected the ports are, how long travels take and if it is possible for bioinvasive organisms to reach them. Geographical distances between ports were estimated as shortest sea travel routes, following the methods developed in [38].

2.2 Assortativity

Another characteristic of network structure and possible spread is the assortativity $k_{nn,i}$ of node linkages. It determines the correlation between the number of ingoing links (indegree) into node i and the average of its neighbours' mean number of outgoing links (outdegree) [42]. We calculated this relationship for the unweighted as well as the weighted GCSN. For the weighted assortativity, $k_{nn,i}^w$, links' weights were integrated (see Tab. 2a and [43]). A positive relationship, i.e. positive assortativity, indicates that nodes are mostly connected to neighbours with a similar number of links. For the large scale structure of the network this implies clustering into several strongly intertwined groups of nodes [44,45]. On the contrary, negative assortativity, also called disassortativity, indicates that strongly connected ports are connected mostly to less well-connected nodes, and the network has a relatively homogeneous, star-like structure [46]. Furthermore, the comparison between the unweighted and weighted nearest neighbours' degrees shows if nodes of high or low strength are clustered together and how degrees and strengths of neighbours correlate [40]. This indicates if bioinvasive spread concentrates on groups of strongly connected ports only or if organisms can quickly spread through the whole network.

2.3 Network community structure

A detailed analysis of the community structure [47] of the GCSN provides further insight into its transportation characteristics. The optimal network communities are obtained by minimising the network's modularity. This is

defined as the proportion of links that fall within compartments of the considered network compared to that expected in an equivalent random network (see Tab. 2b). In [48] an algorithm was developed that incorporates the directionality of links of directed networks in the partitioning method. This is appropriate and applied here for calculating the communities of the weighted GCSN. For comparison, note that the community structure of subnetworks of the unweighted GCSN has been presented in [38].

Additionally, we characterised the members of each community of the weighted GCSN according to its role based on its intra- and intercommunity connections. On that account we calculated the within-community degree's z-score of each node and its participation coefficient (see Tab. 2b and [35,49]). The participation coefficient describes how many of a node's links are connected only within its community or also to nodes of other communities. The phase plot of the within-community degrees and the participation coefficients provides a simple approach to distinguish nodes into provincial hubs, connector hubs or non-hub connectors [50]. In the light of bioinvasion one can conclude from these measures between which ports there is regular, strong exchange of organisms, which ports function as hubs globally or within their community and which ports are the most important for spread of invasives outside of their community.

2.4 Network node centralities

Small world networks usually contain a giant component that is a strongly connected subnetwork of the majority of all ports, i.e. there is a directed path between any pair of nodes [51]. This is the case for the GCSN. It has 951 nodes, 935 of which comprise the giant component [38]. In the following section we consider properties of the nodes of the giant component only.

Centrality of a network addresses the issue of which nodes are most important for the structure and dynamical characteristics of the network [51]. According to the question of what is considered important there are a number of different centrality measures; the three most widely

Table 2. Network measures for the characterisation of the directed, weighted GCSN. (a) Formulae for the calculation of the nearest neighbours' degrees for the examination of assortativity in the unweighted and weighted network. w_{ij} is the weight of the link from i to j , k_i^{in} and k_i^{out} are the number of in- and outgoing links (in/outdegrees) of node i , and s_i^{in} and s_i^{out} the cumulative weights of in- and outgoing links (in/outstrength). (b) For modularity calculations m is the number of links in the network G , δ_{ij} the Kronecker delta and c_i the label of the community to which node i belongs. ς_i is the sum of the weights of all links that connect node i to nodes in its community c_i . $\bar{\varsigma}_{c_i}$ is the mean ς over all nodes belonging to community c_i and $SD(\varsigma_{c_i})$ is the corresponding standard deviation. ς_{ic} represents the summed weights of the links of node i to nodes in any community c . (c) Formulae for centrality measures further contain the largest eigenvalue λ of the weight matrix and shortest weighted paths σ_{jk}^w between j and k (link distance is set to $1/w_{ij}$). The shortest path length is $|\sigma_{jk}^w|$ and $\{\sigma_{jk}^w\}$ is the number of paths of shortest length between j and k in general and $\{\sigma_{jik}^w\}$ the number of paths passing node i on their way from j to k . α is the proportion of unidirectional connections between pairs of ports and \bar{w}_i the mean weight of outgoing links from i . (d) Variables for calculating cluster size growth R_0 are the transition rates r per year and link or unit weight (here per 16,232 GT, the mean value of all included ships), α , in- and outdegrees and -strengths, and the mean outweight from i , \bar{w}_i . For the critical transition to global spread, $R_0 = 1$, r is the percolation threshold.

Network measure	Definition
a	
Nearest neighbours' degree (unweighted)	$k_{nn,i} = \frac{1}{k_i^{in}} \sum_{j \in N_i} k_j^{out}$
Nearest neighbours' degree (weighted)	$k_{nn,i}^w = \frac{1}{s_i^{in}} \sum_{j \in N_i} w_{ij} s_j^{out}$
b	
Modularity	$Q = \frac{1}{\sum_{i \in G} s_i^{in}} \sum_{i,j \in G} \left(w_{ij} - \frac{s_i^{out} s_j^{in}}{\sum_{i \in G} s_i^{in}} \right) \delta_{c_i, c_j}$
Within-community degree z-score	$z_i = \frac{\varsigma_i - \bar{\varsigma}_{c_i}}{SD(\varsigma_{c_i})}$
Participation coefficient	$P_i = 1 - \sum_{s=1}^{N_c} \left(\frac{\varsigma_{ic}}{s_i} \right)^2$
c	
Eigenvector centrality	$x_i = \frac{1}{\lambda} \sum_{j \in G} w_{ij} x_j$
Closeness centrality	$cc_i = \sum_{j \in G} \sigma_{ij}^w $
Betweenness centrality	$bc_i = \sum_{j,k \in G} \frac{\#\{\sigma_{jik}^w\}}{\#\{\sigma_{jk}^w\}}$
R_0 centrality	$rc_i = \alpha s_i^{in} s_i^{out} + (1 - \alpha) s_i^{in} (s_i^{out} - \bar{w}_i)$
d	
Cluster size growth	$R_0 = r \frac{\alpha \sum_{i \in G} s_i^{in} s_i^{out} + (1 - \alpha) \sum_{i \in G} s_i^{in} (s_i^{out} - \bar{w}_i)}{\sum_{i \in G} s_i^{in}}$

used ones are presented here (see Tab. 2c). Additionally, we develop a new measure of centrality, the R_0 -centrality, that is especially suited to characterise the general possibilities of spread on a network (see below). For each of the four centralities we provide a distribution of the proportion of ports, weighted by centrality, situated in each of the six continents. Finally, we examine the correlations of the four centrality measures, using rank correlation by Kendall.

One centrality that has been widely used for characterising spread is the eigenvector centrality x [52,53]. Going beyond simple measures of node degree and strength [25], that count the number of links of each node or sum its

weights, respectively, the eigenvector centrality additionally considers that not all links of a node are equally important. It incorporates a recursive dependency of a node's centrality on the average centrality of its neighbours (see Tab. 2c). Applied to invasive spread, ports with many well connected neighbouring ports are considered to be most relevant.

Two other measures of centrality that are widely used in social studies are closeness centrality and betweenness centrality. Both are based on the set of shortest directed paths between all pairs of ports. For the weighted network we use shortest paths with respect to the link distances proportional to the inverse weight, $1/w_{ij}$, with w_{ij} being

the cumulated cargo capacity of ships travelling through. Shortest paths then minimise the sum of these distances. Thus, links that are heavily frequented are related to a “shorter” distance, and bioinvasion along those routes is more likely. Closeness centrality cc [54] for each port i is the reciprocal of the average of shortest path lengths from i to all ports $j \neq i$. It points out from which, and to which, ports invasive organisms can spread quickly. Betweenness centrality bc [55] is the number of shortest paths between all pairs of nodes that pass through node i (Tab. 2c). Ports that are most “between” are especially important for guaranteeing short paths and high connectedness. If they are deleted, shortest path lengths increases and bioinvasion spread will slow down.

Finally, we developed a new centrality measure of spread from percolation theory [56] that we denote as R_0 centrality rc , with $rc_i = \alpha s_i^{in} s_i^{out} + (1 - \alpha) s_i^{in} (s_i^{out} - \bar{w}_i)$. It is split into two terms according to α , the proportion of unidirectional, as opposed to bidirectional, links in the network. In the bidirectional case the subtraction of \bar{w}_i , the mean strength of outgoing links of i , is necessary to account for ineffective back-infections to the donor node(s). In networks with only unidirectional links $rc_i = s_i^{in} s_i^{out}$, thus R_0 centrality is fully characterised by the in- and outstrengths of a node. The R_0 centrality is derived from the cluster size growth R_0 (Tab. 2 (c, d)), being proportional to each nodes’ contribution to it. It is negatively related to the percolation threshold r^* (see below and [57]) that is the transmission rate above which spread through the whole network is likely within a certain time interval. In contrast to the other centralities, the R_0 centrality is a local measure, depending on a node’s in-strength and outstrength only. It is a general characteristic and measures how much each node and its neighbours contribute to the possibility of global spread, for example of bioinvasive organisms, through the network.

2.5 Network robustness and percolation

In the following, we use network theory to examine how different ballast water management and anti-fouling options affect bioinvasion. First, we analyse error and attack tolerances [58] of the GCSN. This means that we delete nodes and all their links one by one and examine the size of the giant component of the remaining network. In analogy to error tolerance we select the nodes to be deleted in random order, while for attack tolerance analyses we consecutively delete nodes of maximum degree or strength. If the networks are more sensitive to selective deletions than to random ones, bioinvasion could be slowed down by equipping a certain proportion of the largest ports with ballast water management devices.

Inspired by the results of Drake and Lodge [32] we examined how invasive spread is determined by per ship transmission intensity and how this is affected by node deletions. For sensible quantitative predictions and comparability, in this simulation analysis we adjusted link strengths of the GCSN multiplying them with a factor of 1.72, because only ships comprising 58% of the global

cargo capacity had been available for network development [38]. Average ship numbers for each link were calculated by dividing the link’s weight in gross tonnage (GT) by the mean of all ships’ GTs. We calculated the average fraction of ports infected from one random, initially infected port for the full GCSN (1000 simulations each) and varied over a set of different per year transmission probabilities r . Furthermore, we determined the critical transmission probability, the percolation threshold r^* , for which the cluster size growth $R_0 = 1$. For transmission probabilities above this value, $r > r^*$, the number of invaded ports grows exponentially as ships travel the oceans, and soon all will be invaded. However, if $R_0 < 1$ and thus transition rates $r < r^*$, bioinvasion spread will decrease and diminish during the considered year. R_0 is well-known as the basic reproductive number in epidemiology [51,59].

To put the issues of transmission intensity decrease and node deletion into perspective we calculated r^* for the GCSN under random and selective node deletion. This provides indications of the magnitude in the reduction of the transition probability per year and the number of ports with ballast water treatment required for controlling bioinvasion. These results were compared to a specific per ship transmission probability that has been estimated from empirical data [32].

3 Results

3.1 Small world properties

Global and local efficiencies and network cost for the plainly topological, but directed GCSN are $E_{glob} = 0.43$, $E_{loc} = 0.75$ and $c = 0.04$. This reveals that network topology very cost efficiently connects the ports of the world [41]. The high value of E_{loc} points out that the ship network is highly fault tolerant, i.e. failure of one node will not much affect the efficiency of its neighbouring nodes. E_{glob} indicates that the topological GCSN is approximately half as efficient as a fully connected network of the same size at a cost of only 4%.

Efficiencies and cost for the geographically embedded GCSN are somewhat different, $E_{glob}^{geo} = 0.99$, $E_{loc}^{geo} = 0.70$ and $c^{geo} = 0.22$. It becomes obvious that, when considering the geographical locations of each port, the GCSN is only 1% less efficient than an ideal ship traffic network with a direct connection between each pair of ports. This is very remarkable. The local efficiency E_{loc}^{geo} does not change a lot and, in comparison to the topological network, error tolerance is still given. The cost c^{geo} , in accordance with the increased E_{glob}^{geo} , increased somewhat, but is still far from the 100% for the fully connected network. The geographically embedded GCSN is practically as efficient as the fully connected one at a cost of only 22%. Concluding, one might state that the GCSN is a small world network.

3.2 Assortativity

The relationship between the indegree k_i^{in} of each port i and the average degree of its neighbours $k_{nn,i}$ and $k_{nn,i}^w$

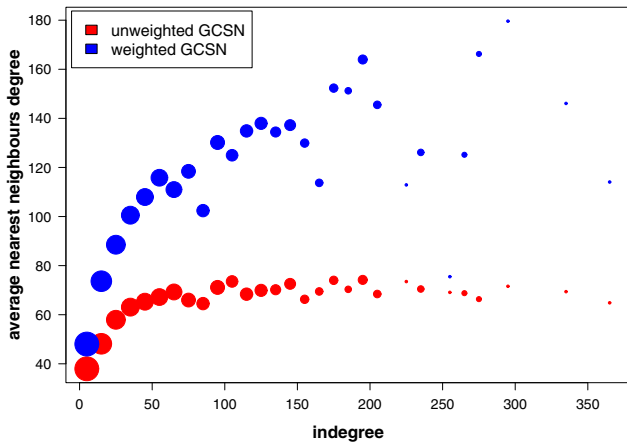


Fig. 1. (Color online) Degree correlations as signs of network assortativity. Average nearest neighbours' degrees of each node of the unweighted (red dots) and weighted (blue dots) GCSN (Tab. 2a) averaged over bins of size 10 for k^{in} . Radii of the dots indicate the sample size, i.e. number of different ports within the respective bin.

(see Fig. 1) reveals that the global cargo ship network is positively assortative. Especially up to port sizes of $k^{in} = 70$ there is a strong increase of $k_{nn,i}$ and $k_{nn,i}^w$. Above that average neighbours' degrees of the unweighted GCSN seem to asymptotically approach a limit at around $k_{nn} = 70$. This means that at this point larger ports cannot obtain more links to other ports. Then, as can be observed from the further growth of the weighted k_{nn}^w with k^{in} , ports with extremely many connections are preferentially linked with larger ports. The positive assortativity, and this special kind of it in particular, indicates that the global cargo ship network contains several communities, i.e. groups of strongly within-group connected ports. The fact that k_{nn}^w is always above k_{nn} underlines that links with large weights are directed towards neighbours with large degrees. This is emphasised by their averages, $\langle k_{nn,i} \rangle = 52.21 < \langle k_{nn,i}^w \rangle = 81.68$. Whether k_{nn}^w approaches a limit is not clear, because sample sizes for ports with large indegrees are rather small.

In the parameter dependence of both measures (Fig. 1) one can observe slight periodic behaviour, k_{nn} and k_{nn}^w showing a number of equidistant, local minima. This may be another indication of the network's structure and sizes of its compartments.

3.3 Network community structure

The community structure of the weighted GCSN (Fig. 2a) reveals that there are nine smaller groups and two large trading communities. The large communities are the Atlantic group including European and American countries and the Pacific group of Asia and Australia. Small communities are specialised, often local trade routes (e.g. West Africa-Argentina/Brazil) and ferry connections (e.g. Dover-Calais). The phase plot of participation coefficients indicates which ports are especially important

for global and local network connectivity (Fig. 2b). Oil ports at the coast of Louisiana as well as some ferry ports in Europe fall into the group of provincial hubs. Many large, well known and globally significant ports are connector hubs, e.g. the Panama and Suez Canal, Shanghai, New York & New Jersey, Singapore and Antwerp. Note that the Panama Canal has a very high within-module degree whereas the Suez Canal has an extremely high participation coefficient, thus being very important for overall network connectedness. Some ports, like Santos and Le Havre are nonhub connectors. They are not of a very high within-module degree, but important to keep the GCSN globally connected. Note that there are no kinless hubs or nonhubs, i.e. ports the links of which are homogeneously distributed among all communities.

3.4 Network node centralities

The rankings of the ports of the weighted GCSN (Fig. 3) highlight the 100 most central ports, according to each of the four centrality measures. Furthermore, pie charts of the weighted proportions of central ports in each continent indicate centres of high shipping intensity. The Panama and Suez Canal frequently appear in the top ten central ports (see Tab. 3). This underlines their high importance in topologically holding together the network. Furthermore, Shanghai and Singapore are always among the first ranks for whichever type of centrality. Singapore is even first for three of the four kinds of centrality. One can observe that there are always very few ports of Africa, Australia or South America among the most central.

There are several differences between the most important ports in respect to the four kinds of centrality. Ports of highest eigenvector centrality that are important for spread in the long term, are mostly situated at the Gulf of Mexico. Closeness centrality, a more short term indicator for invasive spread, reveals a more homogeneous distribution of the most central ports. Among them are many European, North American and Asian ports, with a slight preference of ports in the Suez region. The high closeness of this region to any other port underlines its importance for keeping the two major components of the weighted GCSN connected (see previous section). In terms of betweenness centrality again Singapore, Shanghai and the two big canals are most important besides a large number of ports in northern Europe. The R_0 centrality provides somewhat intermediate results. Ports of highest R_0 centrality are Shanghai, Singapore, the Suez and Panama Canal, some ports at the Gulf of Mexico and several ones in northern Europe (Tab. 3).

Rank correlations between all pairs of the four centralities are positive, $0.43 < \tau < 0.82$ (all $p < 0.001$). The two largest ones are (i) closeness centrality and R_0 centrality and (ii) eigenvector centrality and R_0 centrality, emphasising that the R_0 centrality is a measure that incorporates several different notions of spread. Correlations that involved betweenness centrality are lowest, the mainly global character of this measure demarcating it from the others.

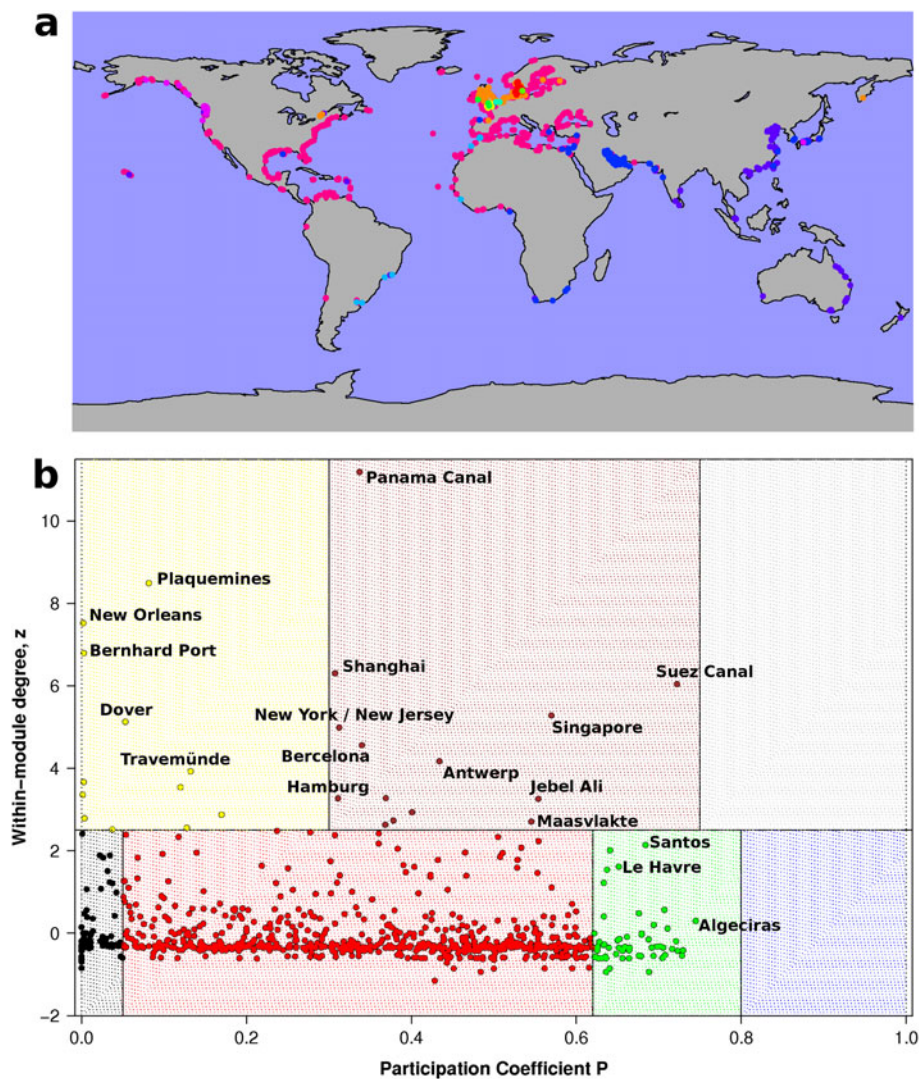


Fig. 2. (Color online) Community structure and representation of the roles of each port. (a) The 11 communities of the weighted GCSN are shown in different colors. It becomes apparent that trading preferences and geographical distances basically determine these communities. (b) Locations of ports in the phase-space of the z -score of within-module degree vs. the participation coefficient P . The positions of the ports indicate their different roles for network structure and connectivity. Hubs always have a high within-module degree and the larger the participation coefficient of a port the greater its influence on connecting the different communities of the networks. Regions in this phase space are coloured according to [35] into provincial hubs (yellow), connector hubs (brown), kinless hubs (grey), ultraperipheral (black), peripheral (red), nonhub connectors (green) and kinless nonhubs (blue). The names of the largest hubs and some connecting ports are given.

3.5 Network robustness and percolation

The global cargo ship network is very robust to random node deletions (Fig. 4a). Up to a fraction of 80–90% of ports deleted, the size of the largest component decreases almost linearly in steps of size one. Thus, if randomly deleting any proportion less than 80% of all nodes the leftover ports are still strongly connected to each other, and spread of bioinvasive organisms through the remaining network prevails highly likely. Against selective, ordered deletion of preferentially high degree (blue lines in Fig. 4a) and high strength (red lines) nodes, respectively, the GCSN is less robust. Interestingly, the effect of node deletion, selected either by degree or strength, on network

connectivity is very similar. Figure 4a shows that the connectivity of the giant component is only retained up to a deletion of 35–40% of the most strongly connected ports; else the whole network falls apart. Additionally, the slopes of the red and blue lines are relatively steep. This means that the network becomes disconnected quickly, losing its strong connectivity already after removing a very small proportion of the best connected nodes.

Results of network percolation (Fig. 4b) show how the probability of global spread per year changes with different transmission rates. With increasing transmission probability per ship and year the fraction of infected ports grows from 0 to 0.99, steeply increasing at transmission probabilities slightly above the critical percolation threshold

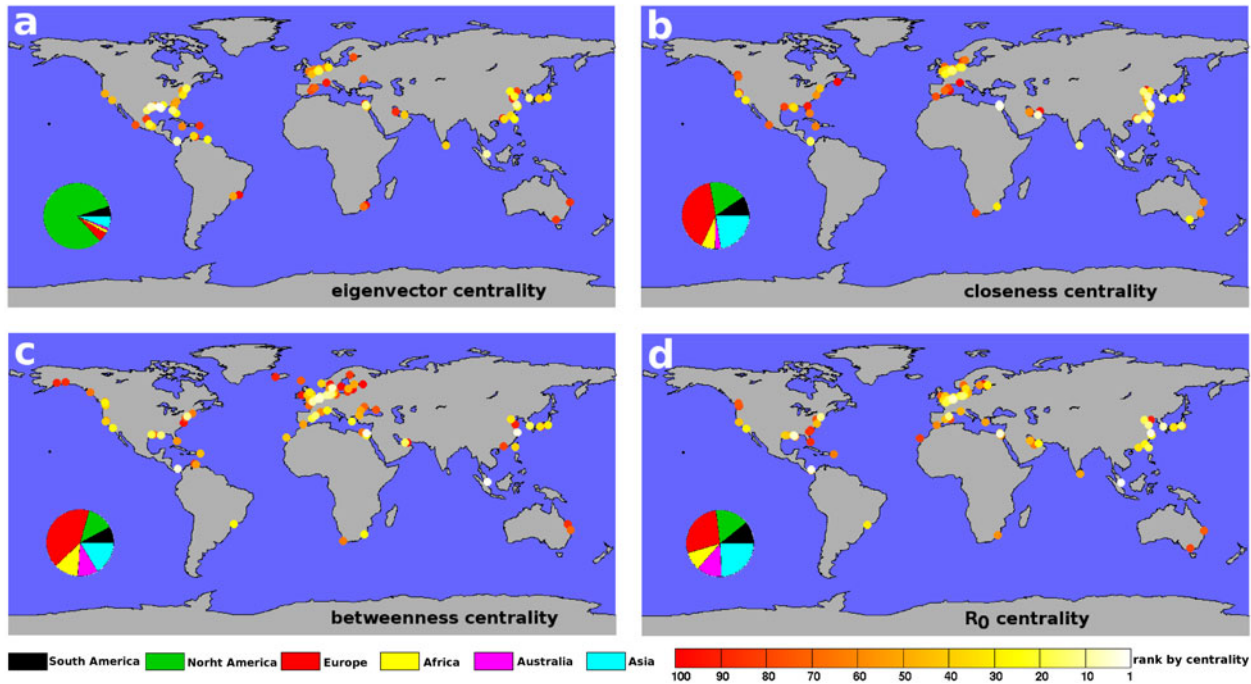


Fig. 3. (Color online) Top ports of different centrality measures of the weighted global cargo ship network. Each map shows the positions and colour coded the ranks of the 100 highest ranked ports according to different centrality measures. They are (a) eigenvector centrality, (b) closeness centrality, (c) betweenness centrality and (d) R_0 centrality. The most central ports are coloured light yellow and white. Insets depict the respective distributions of ports in the six continents weighted by the different centrality measures.

Table 3. Top ten most central ports in terms of eigenvector, closeness and betweenness centrality and the novel R_0 centrality.

rank	Eigenvector centrality	Closeness centrality	Betweenness centrality	R_0 centrality
1	St Bernhard Port	Singapore	Singapore	Singapore
2	New Orleans	Suez Canal	Suez Canal	Panama Canal
3	Plaquemines	Shanghai	Shanghai	Shanghai
4	Port of South Louisiana	Port Said	Panama Canal	Suez Canal
5	Mississippi River	Suez	Maasvlakte	Plaquemines
6	Houston	Jebel Ali	Gothenburg	Antwerp
7	Panama Canal	Busan	Antwerp	Busan
8	Barbours Cut	East Port Said	Zeebrugge	New Orleans
9	Singapore	Kaohsiung	Le Havre	St Bernhard Port
10	Shanghai	Qingdao	Bremerhaven	Maasvlakte

$r^* = 9.14 \times 10^{-5}$. The proportion of infected ports never reaches 1, because due to edge effects of data sampling a subset of 10 ports has no ingoing links and cannot be infected. They are Papenburg, Leer, Stralsund and Wolgast in Germany, Marugame and Hashihama in Japan, Haikou and Jinshan in China, Cagliari in Italy and Cekisan in Turkey. The value of the percolation threshold is very small, even below the transmission probability estimated from empirical data [17,32], $r^{DL} = 4.4 \times 10^{-4}$. The cluster size growth for r^{DL} in our network is $R_0 = 4.81$. To push r^{DL} below r^* for the full GCSN it would have to be reduced by at least $1 - r^*/r^{DL} = 79.2\%$. When additionally nodes are deleted one by one, the percolation threshold r^* increases, in the beginning a little faster for node deletion selected according to strength (Fig. 4c). The increase is, however, very slow, and above deleting 40% of all nodes there is no difference in r^* between random and

selective node removal. This is a sign of high clustering and local efficiency throughout the GCSN. To push the threshold value r^* above the estimated transmission rate r^{DL} by port deletions only, about 70% of all ports would have to be deleted, even if by then any strongly connected component of the network has fallen apart (see above). An optimal combinations of node deletion and decrease of transmission probability can be read off Figure 4c in several ways. For example, when deleting 40% of the strongest connected nodes r^* would have to be reduced by more than 53.3% to avert global spread.

4 Discussion

In this study we extracted novel information about ship traffic structures and its indications for marine bioinvasion

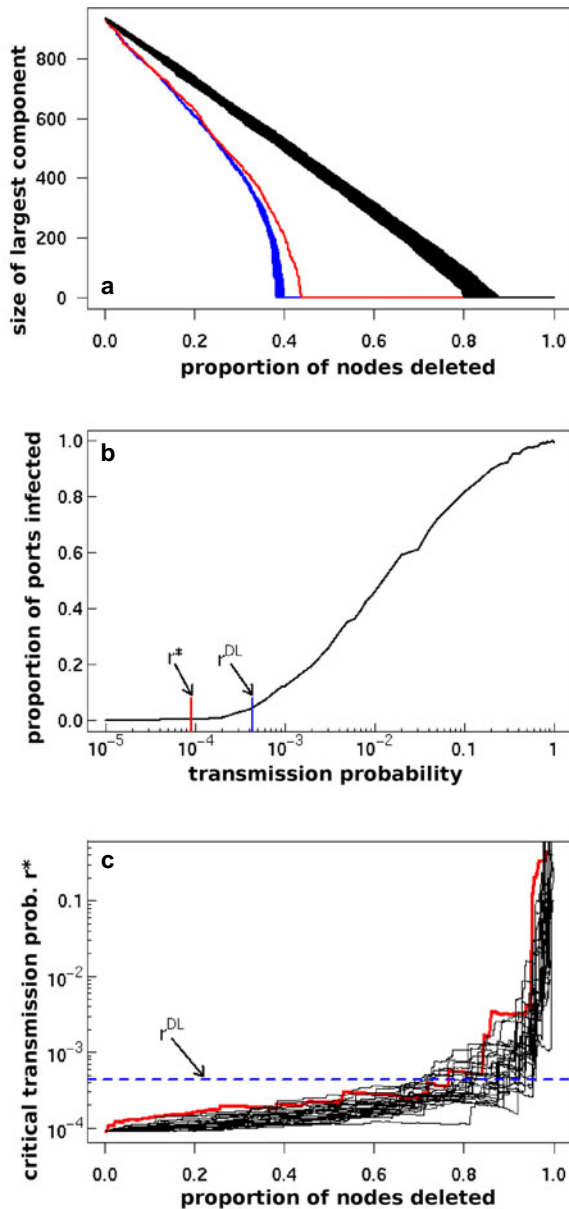


Fig. 4. (Color online) Robustness to node removal and critical transmission probabilities. (a) Decrease of network connectivity with random and selective node removal. Presented are sizes of the largest component after random node removal (black lines), successive removal of nodes of highest degree (blue lines) and successive deletion of nodes of highest strength (red lines). For each scenario 100 samples were calculated. (b) Simulated mean proportion of infected ports in the GCSN for different transmission probabilities r per mean ship, i.e. 16,232 GT. For each r we averaged over 1000 samples, each time infection starting in one random port. The critical transmission probability, the percolation threshold r^* , (red line) was calculated for $R_0 = 1$ and the by 1.72 adjusted node strengths (see Methods and Tab. 2d). r^{DL} (indicated by the blue line) is a probability for transmission by ship, which has been estimated from empirical data [32]. (c) Critical transmission probabilities r^* for the cargo ship network with successively deleted nodes; randomly (black lines) and selective by strength (red lines). We marked the transmission probability r^{DL} estimated from data with a blue dashed line.

by calculating several network characteristics of the global cargo ship network. Our results point out that cargo ship traffic is globally very efficient and the ports are closely connected. This agrees well with other transportation networks [60,61] and previous results concerning the GCSN, i.e. high clustering $C = 0.49$ and small topologically shortest paths $\langle l \rangle = 4.4$ [38]. Concerning bioinvasion this means that the structure of the GCSN strongly supports quick spread of invasive organisms, so that if species can establish in new habitats sooner or later a global homogenisation of marine ecosystems is likely. Such is even accelerating as global trade of goods is increasing at unprecedented rates [21].

The efficiency and network cost measures that have been applied here for the quantification of network connectivity differ from the classical concept of small world determination [30], however the results nicely coincide [38]. The efficiency-cost concept allows for a clear physical interpretation of how well pairs of nodes are connected. They do not regard the network links' weights, but are calculated for the purely topological, directed network and the geographically embedded one. For several invasion events it has been suggested that not propagule pressure, but mainly conditions in the recipient regions determine invasion success [62]. Therefore, a simple examination of pure network structure may reveal new insights of general bioinvasion. Especially the optimized transport efficiency on the geographical map of ports is impressive, as the GCSN seems to be almost as efficient as a fully connected network of the same size.

The positive assortativity of global shipping underlines that large and well connected ports are strongly linked to other large ones between which bioinvasive organisms quickly spread. In this sense, port size is a good indicator for invasibility and invasiveness. Positive assortativity also indicates a distinct community structure of the GCSN, resulting in more strongly pronounced spread within certain groups of ports [63].

Community structure, as calculated by modularity minimisation, underlines differences in the ports spread behaviour. It is strongly influenced by intensive local ship movement, as can be seen by the identified small, local port communities. Often they are highly frequented by ferries. Apart from that the GCSN contains only two very large communities, an Atlantic group of ports and a Pacific one. Two similar regions where large numbers of harmful invasive marine species were present have recently been determined in a metastudy [63]. Furthermore, the importance of different types of hubs is pointed out. Provincial hubs that are mainly European ferry ports and US oil ports at the Gulf of Mexico keep local passenger traffic going and receive oil from off-shore. Bioinvasive organisms that originate from them can be quickly spread throughout their local community. Such local spread and its importance for the establishment of bioinvasive organisms has been examined e.g. at the US West coast [64]. Global spread over large distances is basically propagated by connector hubs. They are disproportionately important for bioinvasion spread and coincide with the highly connected

ports of Shanghai, Singapore, Antwerp, the Panama Canal and Suez Canal. The importance of hubs for bioinvasion spread has been proposed already [65]. Consequently, global as well as local connectivity is very high and invasion spread likely to quickly advance.

The fact that the GCSN contains a giant component [38] underlines that almost all ports are connected with each other and thus are prone to bioinvasion. The more specific centrality measures focus on various aspects of spread through the network. They coincide well with the results of the community structure analysis, but additionally reveal that some ports are only important for certain issues of transport, whereas others seem to be central in any. The latter ones coincide well with the connector hubs. Singapore, appearing first in the centrality rankings three times, seems to be of special importance. This conforms with the fact that Singapore is currently, i.e. for the year that the GCSN was constructed for, the world's busiest port in terms of total shipping tonnage and container and crude oil transshipment [66]. Considering its position at the Strait of Malacca it also provides the functions of a canal, connecting the Pacific with the Indian Ocean. The eigenvalue centrality ranking, revealing a concentration of most central ports around the Gulf of Mexico, deviates from these general findings. The accumulation of highly connected ports at the South coast of Northern America may be surprising, but can be explained by the intensive oil transportation and trade between close ports within this region. The large proportion of highly "between" and "close" ports in Northern Europe indicates that they are very well connected to a wide range of distant ports worldwide, and that European ports are important as intermediate transshipment ports on shortest routes. European ports also contain among the highest numbers of harmful invasive species [63].

The intermediate results by the novel R_0 centrality point out that this measure includes contributions of short and long term global spread characteristics as well as connectivity of each port. Thus, it provides a sufficient general measure for spread in transportation networks. In future work, the relations of R_0 centrality to other network characteristics, e.g. spectral properties, should be worked out. It would be especially important to find out whether the R_0 centrality can be easily interpreted or possibly analytically assessed in particular network topologies, such as random or scale-free graphs.

From the here determined different types of centralities of the GCSN, the very importance and prospective effectiveness of installing ballast water and antifouling management facilities in the named most central ports are obvious. It may be surprising that Rotterdam, the largest port in Europe, is not listed in Table 3. By developing the GCSN Rotterdam's different port terminals were not merged and therefore count as own, single ports. Otherwise, it would be among the most central ports.

The discussion about how to most effectively manage marine bioinvasion [23,24] has been clarified somewhat. Which of the several techniques and treatments to clean ballast water and the large selection of antifouling-paints

is most effective to slow down bioinvasion? Drake and Lodge [32] have directly compared if the exclusion of ports, in-port treatment, or the reduction of transition rates per ship, onboard treatment, more effectively decreased numbers of infected ports in their approximate ship network. Their study indicated that a decrease in the transmission rates of organisms by ships has a stronger effect than the deletion of single ports from the network. The analysis was rather case study like, only comparing three specific spread scenarios. We find that this is not sufficient for drawing general conclusions, however fits in the general framework that we propose.

We applied distinct, general and quantitative methods for the examination of effects of node deletion and the modification of the transmission probability on network connectivity and bioinvasive proliferation properties. Our results of network robustness to node deletion are not port specific, but quantitative and allow for general conclusions. We found that the global cargo ship network cannot be controlled by the application of treatment techniques in a random set of ports. To this treatment it is highly robust. However, similar to many other scale-free networks [58] the GCSN is very vulnerable to the deletion of the most strongly connected nodes. In line with this result one could claim that the "deletion" of at least 40% of the largest ports could substantially decrease spread of bioinvasive organisms. However, this only means that most ports cannot be reached from all other ones (strong connectedness), but local and global spread is still possible. This was revealed by our study of threshold percolation transmission rates (see below). Ballast water treatment strategies might furthermore be improved by using refined strategies to restructure the network, such as graph-partitioning [67].

We showed that there is a steep increase of spread intensity from transmission rates that are insignificant to those that facilitate global spread (percolation). The percolation threshold is quite low, but less than an order of magnitude below the value estimated by Drake and Lodge [32]. Therefore, in their study a decrease of r had a relatively pronounced effect on spread. This is one of the differences that might ultimately be related to the simplification of network topology imposed by gravity models, i.e. fundamental differences in a network structure which is approximated by gravity models (as done by Drake and Lodge [32]) compared to the true network. Similar differences were found in [34] where the community structure of a transportation network could not be reproduced by using the gravity law ansatz; this might point to a general non-applicability of gravity law in transportation networks (see also [38]).

Here, we systematically quantified how much decrease of the transmission probability per ship is necessary for the inhibition of global spread. At least 80% of all invasive organisms in ballast tanks have to be extinguished to push r^{DL} below r^* . Whether onboard ballast treatment is that efficient remains to be seen. Therefore, it may be useful to couple in-port treatment and onboard treatment of bioinvasive stowaways. Our simulations of this situation show that a surprisingly high number of ports have

to be treated for the percolation threshold to considerably decrease. Selective treatment of strongest connected ports has only a slight advantage, above 40% deletion being not better than some random cases. Thus, it is not sufficient to only apply ballast water and hull fouling treatment in the 40% biggest ports, but a combination of in port and onboard treatment is most promising. Another option may be to change the network's topology. Such has recently been proposed for mitigation of attack vulnerability of e.g. traffic or electric power networks [68]. Of course, topological changes of the GCSN may not be practicable, especially because optimality of trade structures has to be retained.

As a note of caution one has to add that here no information on habitat characteristics in the ports and geographical port distances are considered. On the one hand, between very close ports many species may be exchanged, but are often also native in the recipient port. On the other hand, if conditions in very distant ports are extremely different, establishment of alien species may be difficult. Thus, local homogenisation of species assemblies and global bioinvasion are treated similarly here. In this study, only the first stage of invasion, the introduction, but not establishment and proliferation were evaluated. As noted above, at this first stage it is most likely that one is able to hinder bioinvasive spread and exert effective control measures.

Concluding, we provide and apply a number of measures for the characterisation of directed and weighted networks in the light of invasive spread. We specified them for examining the efficiency, structure and robustness of transportation vectors, for the evaluation and prediction of bioinvasion and, possibly, the spread of infectious diseases. In the case of marine bioinvasion by cargo ships we are able to point out which ports are most vulnerable to and which are most strongly promoting invasive spread. Furthermore, we quantified the robustness of network connectivity and spread. These results provide first indications for bioinvasion risk management. However, in the light of the obtained insights one should now be able to develop general population dynamics models including species' establishment and proliferation in the ports (but see [62]). Additionally, one can examine the effects of ballast water treatment in more detail and integrate specific environmental conditions into modelling. The global cargo ship network provides a sensible framework for such an examination of bioinvasion risk for management and decision making.

We want to thank J. Drake and M. Gastner for providing data and H. Rosenthal and H. Seebens for helpful discussion. This work was supported by the German VW-Stiftung and the BMBF.

References

1. C.S. Elton, *The ecology of invasions by animals and plants* (Chapman and Hall, London, 1958)
2. G.J. Vermeij, *Biol. Conserv.* **78**, 3 (1996)
3. R.N. Mack, D. Simberloff, W.M. Lonsdale, H. Evans, M. Clout, F.A. Bazzaz, *Ecol. Appl.* **10**, 689 (2000)
4. C.S. Kolar, D.M. Lodge, *TREE* **16**, 199 (2001)
5. D.M. Lodge, *TREE* **8**, 133 (1993)
6. O.E. Sala, F.S. Chapin III, J.J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L.F. Huenneke, R.B. Jackson, A. Kinzig, R. Leemans, D.M. Lodge, H.A. Mooney, M. Oesterheld, N.L. Poff, M.T. Sykes, B.H. Walker, M. Walker, D.H. Wall, *Science* **287**, 1770 (2000)
7. D. Pimentel, R. Zuniga, D. Morrison, *Ecol. Econ.* **52**, 273 (2005)
8. G.M. Ruiz, T.K. Rwlings, F.C. Dobbs, L.A. Drake, T. Mullady, A. Huq, R.R. Colwell, *Nature* **408**, 49 (2000)
9. A.J. Tatem, S.I. Hay, D.J. Rogers, *Proc. Natl. Acad. Sci.* **103**, 6242 (2006)
10. M.H. Williamson, *Biological invasions* (Chapman and Hall, London, 1996)
11. L.M. Puth, D.M. Post, *Ecol. Lett.* **8**, 715 (2005)
12. J.T. Carlton, *Biol. Conserv.* **78**, 97 (1996)
13. G.M. Ruiz, J.T. Carlton, E.D. Grosholz, A.H. Hines, *Am. Zool.* **37**, 621 (1997)
14. A. Meinesz, J. deVaugelas, B. Hesse, X. Mari, *J. Appl. Phycol.* **5**, 141 (1993)
15. GESAMP, *GESAMP Rep. Stud.* **58**, 1 (1997)
16. J.T. Carlton, *Oceanography and Marine Biology* **23**, 313 (1985)
17. J.T. Carlton, J.B. Geller, *Science* **261**, 78 (1993)
18. S. Gollasch, J. Lenz, M. Dammer, H.-G. Andres, *J. Plankton Res.* **22**, 923 (2000)
19. M.H. Williamson, A. Fitter, *Biol. Conserv.* **78**, 163 (1996)
20. J.M. Drake, D.M. Lodge, *Aquat. Invas.* **2**, 121 (2007)
21. International Maritime Organisation, *International shipping and world trade. Facts and figures* (IMO library service, 2006), Available at <http://www.imo.org>
22. United Nations, *Review of maritime transport* (United Nations, New York and Genua, 2007), Available at http://www.unctad.org/en/docs/rmt2007_en.pdf
23. S. Gollasch, M. David, M. Voigt, E. Dragsund, C. Hewitt, Y. Fukuyo, *Harmful Algae* **4**, 585 (2007)
24. G.A. Hopkins, B.M. Forrest, *ICES J. Mar. Sci.* **65**, 811 (2008)
25. M.E.J. Newman, *SIAM Rev.* **45**, 167 (2003)
26. M.E.J. Newman, A.-L. Barabasi, D.J. Watts, *The structure and dynamics of networks* (Princeton University Press, Princeton, 2006)
27. D.L. Urban, E.S. Minor, E.A. Treml, R.S. Schick, *Ecol. Lett.* **12**, 260 (2009)
28. L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci.* **101**, 15124 (2004)
29. L. da F. Costa, F.A. Rodrigues, G. Travieso, P.R. Villas Boas, *Adv. Phys.* **56**, 167 (2007)
30. D.J. Watts, S.H. Strogatz, *Nature* **393**, 440 (1998)
31. M. Barthélemy, *Phys. Rep.* **499**, 1 (2011)
32. J.M. Drake, D.M. Lodge, *Proc. R. Soc. Lond. B* **271**, 575 (2004)
33. K.E. Haynes, A.S. Fotheringham, *Gravity and spatial interaction models* (Sage, Beverly Hills, 1984)
34. C. Thiemann, F. Theis, D. Grady, R. Brune, D. Brockmann, *PLoS ONE* **5**, e15442 (2010)

35. R. Guimera, S. Mossa, A. Turttschi, L.A.N. Amaral, Proc. Natl. Acad. Sci. **102**, 7794 (2005)
36. X. Xu, J. Hu, P. Liu, Chaos **17**, 023129 (2007)
37. Y. Hu, D. Zhu, Physica A **388**, 2061 (2009)
38. P. Kaluza, A. Kölzsch, M.T. Gastner, B. Blasius, J.R. Soc. Interface **7**, 1093 (2010)
39. M. Zachcial, C. Heideloff, *ISL shipping statistics year-book 2001* (Institute of Shipping Economics and Logistics, Bremen, 2001)
40. J.I.L. Miguéns, J.F.F. Mendes, Lect. Notes Comput. Sci. **5151**, 145 (2008)
41. V. Latora, M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001)
42. R. Pastor-Satorras, A. Vazquez, A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001)
43. A. deMontis, M. Barthelemy, A. Chessa, A. Vespignani, Environ. Plann. B **34**, 905 (2007)
44. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
45. M.E.J. Newman, Phys. Rev. E **67**, 026126 (2003a)
46. S. Redner, Nature **453**, 47 (2008)
47. M.E.J. Newman, Proc. Natl. Acad. Sci. **103**, 8577 (2006a)
48. E.A. Leicht, M.E.J. Newman, Phys. Rev. Lett. **100**, 118703 (2008)
49. R. Guimera, L.A.N. Amaral, Nature **433**, 895 (2005a)
50. R. Guimera, L.A.N. Amaral, J. Stat. Mech. Theor. Exp. P02001, 1 (2005b)
51. M.E.J. Newman, *The mathematics of networks*, in *The New Palgrave Encyclopedia of Economics* (Palgrave Macmillan, Basingstoke, 2007), pp. 1–12
52. P. Bonacich, J. Math. Soc. **2**, 113 (1972)
53. G.S. Canright, K. Engo-Monsen, Complexus **3**, 131 (2006)
54. S. Wassermann, K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994)
55. L. Freeman, Sociometry **40**, 35 (1977)
56. D. Stauffer, A. Aharony, *Introduction to percolation theory* (CRC-press, 1994)
57. D.S. Callaway, M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. Lett. **85**, 5468 (2000)
58. R. Albert, H. Jeong, A.-L. Barabasi, Nature **406**, 378 (2000)
59. R.M. Anderson, R.M. May, *Infectious diseases of humans – dynamics and control* (Oxford University Press, New York, 1991)
60. V. Latora, M. Marchiori, Physica A **314**, 109 (2002)
61. A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. **101**, 3747 (2004)
62. R.P. Keller, J.M. Drake, M.B. Drew, D.M. Lodge, Divers. Distrib. **17**, 93 (2011)
63. J.L. Molnar, R.L. Gamboa, C. Rvenga, M.D. Spalding, Front. Ecol. Environ. **6**, 485 (2008)
64. C. Simkanin, I. Davidson, M. Falkner, M. Sytsma, G. Ruiz, Mar. Pollut. Bull. **58**, 366 (2009)
65. O. Floerl, G.J. Inglis, K. Dey, A. Smith, J. Appl. Ecol. **46**, 37 (2009)
66. D. Heng, Bibliosia **1**, 12 (2007)
67. Y. Chen, G. Paul, S. Havlin, F. Liljeros, H.E. Stanley, Phys. Rev. Lett. **101**, 058701 (2008)
68. C.M. Schneider, A.A. Moreira, J.S. Andrade Jr., S. Havlin, H.J. Herrmann, Proc. Natl. Acad. Sci. **108**, 3838 (2011)