JOURNAL OF THE ROYAL SOCIETY Interface

FirstCite® e-publishing

J. R. Soc. Interface
doi:10.1098/rsif.2009.0495
Published online

The complex network of global cargo ship movements

Pablo Kaluza, Andrea Kölzsch, Michael T. Gastner and Bernd Blasius*

Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky Universität, Carl-von-Ossietzky-Straße 9-11, 26111 Oldenburg, Germany

Transportation networks play a crucial role in human mobility, the exchange of goods and the spread of invasive species. With 90 per cent of world trade carried by sea, the global network of merchant ships provides one of the most important modes of transportation. Here, we use information about the itineraries of 16 363 cargo ships during the year 2007 to construct a network of links between ports. We show that the network has several features that set it apart from other transportation networks. In particular, most ships can be classified into three categories: bulk dry carriers, container ships and oil tankers. These three categories do not only differ in the ships' physical characteristics, but also in their mobility patterns and networks. Container ships follow regularly repeating paths whereas bulk dry carriers and oil tankers move less predictably between ports. The network of all ship movements possesses a heavy-tailed distribution for the connectivity of ports and for the loads transported on the links with systematic differences between ship types. The data analysed in this paper improve current assumptions based on gravity models of ship movements, an important step towards understanding patterns of global trade and bioinvasion.

Keywords: complex network; cargo ships; bioinvasion; transportation

1. INTRODUCTION

The ability to travel, trade commodities and share information around the world with unprecedented efficiency is a defining feature of the modern globalized economy. Among the different means of transport, ocean shipping stands out as the most energy efficient mode of long-distance transport for large quantities of goods (Rodrigue *et al.* 2006). According to estimates, as much as 90 per cent of world trade is hauled by ships (International Maritime Organization 2006). In 2006, 7.4 billion tons of goods were loaded at the world's ports. The trade volume currently exceeds 30 trillion ton-miles and is growing at a rate faster than the global economy (United Nations Conference on Trade and Development 2007).

The worldwide maritime network also plays a crucial role in today's spread of invasive species. Two major pathways for marine bioinvasion are discharged water from ships' ballast tanks (Ruiz *et al.* 2000) and hull fouling (Drake & Lodge 2007). Even terrestrial species such as insects are sometimes inadvertently transported in shipping containers (Lounibos 2002). In several parts of the world, invasive species have caused dramatic levels of species extinction and landscape alteration, thus damaging ecosystems and creating hazards for human livelihoods, health and local economies (Mack *et al.* 2000). The financial loss owing to bioinvasion is estimated to be $120 billion per year in the USA alone (Pimentel *et al.* 2005).

Despite affecting everybody's daily lives, the shipping industry is far less in the public eye than other sectors of the global transport infrastructure. Accordingly, it has also received little attention in the recent literature on complex networks (Wei *et al.* 2007; Hu & Zhu 2009). This neglect is surprising considering the current interest in networks (Albert & Barabási 2002; Newman 2003*a*; Gross & Blasius 2008), especially airport (Barrat *et al.* 2004; Guimerà & Amaral 2004; Hufnagel *et al.* 2004; Guimerà *et al.* 2005), road (Buhl *et al.* 2006; Barthélemy & Flammini 2008) and train networks (Latora & Marchiori 2002; Sen *et al.* 2003). In the spirit of current network research, we take here a large-scale perspective on the global cargo ship network (GCSN) as a complex system defined as the network of ports that are connected by links if ship traffic passes between them.

Similar research in the past had to make strong assumptions about flows on hypothetical networks with connections between all pairs of ports in order to approximate ship movements (Drake & Lodge 2004; Tatem *et al.* 2006). By contrast, our analysis is based on comprehensive data of real ship journeys allowing us to construct the actual network. We show that it has a small-world topology where the combined cargo capacity of ships calling at a given port (measured in

*Author for correspondence (blasius@icbm.de).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2009.0495 or via http://rsif.royalsocietypublishing.org.

Received 11 November 2009
Accepted 21 December 2009

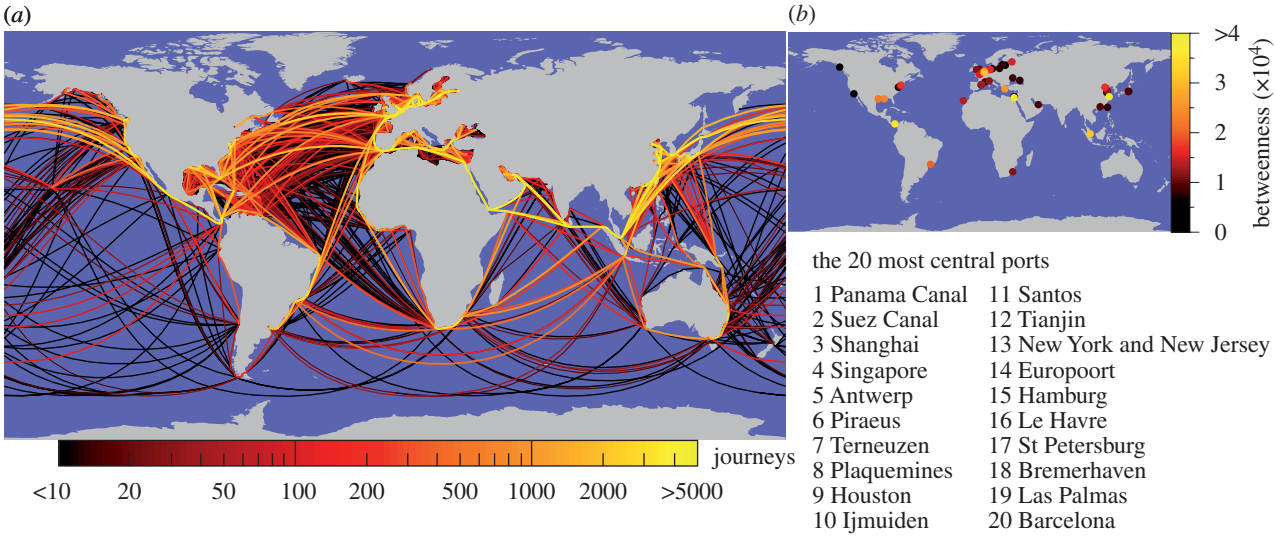

Figure 1. Routes, ports and betweenness centralities in the GCSN. (*a*) The trajectories of all cargo ships bigger than 10 000 GT during 2007. The colour scale indicates the number of journeys along each route. Ships are assumed to travel along the shortest (geodesic) paths on water. (*b*) A map of the 50 ports of highest betweenness centrality and a ranked list of the 20 most central ports.

gross tonnage (GT)) follows a heavy-tailed distribution. This capacity scales superlinearly with the number of directly connected ports. We identify the most central ports in the network and find several groups of highly interconnected ports showing the importance of regional geopolitical and trading blocks.

A high-level description of the complete network, however, does not yet fully capture the network's complexity. Unlike previously studied transportation networks, the GCSN has a multi-layered structure. There are, broadly speaking, three classes of cargo ships—container ships, bulk dry carriers and oil tankers—that span distinct subnetworks. Ships in different categories tend to call at different ports and travel in distinct patterns. We analyse the trajectories of individual ships in the GCSN and develop techniques to extract quantitative information about characteristic movement types. With these methods, we can quantify that container ships sail along more predictable, frequently repeating routes than oil tankers or bulk dry carriers. We compare the empirical data with theoretical traffic flows calculated by the gravity model. Simulation results, based on the full GCSN data or the gravity model, differ significantly in a population-dynamic model for the spread of invasive species between the world's ports. Predictions based on the real network are thus more informative for international policy decisions concerning the stability of worldwide trade and for reducing the risks of bioinvasion.

2. DATA

An analysis of global ship movements requires detailed knowledge of ships' arrival and departure times at their ports of call. Such data have become available in recent years. Starting in 2001, ships and ports have begun installing the automatic identification system (AIS) equipment. AIS transmitters on board of the ships automatically report the arrival and departure times to the port authorities. This technology is primarily used to avoid collisions and increase port security, but arrival and departure records are also made available by Lloyd's Register Fairplay for commercial purposes as part of its Sea-web database (www.sea-web.com). AIS devices have not been installed in all ships and ports yet, and therefore there are some gaps in the data. Still, all major ports and the largest ships are included, thus the database represents the majority of cargo transported on ships.

Our study is based on Sea-web's arrival and departure records in the calendar year 2007 as well as Sea-web's comprehensive data on the ships' physical characteristics. We restrict our study to cargo ships bigger than 10 000 GT that make up 93 per cent of the world's total capacity for cargo ship transport. From these, we select all 16 363 ships for which AIS data are available, taken as representative of the global traffic and long distance trade between the 951 ports equipped with AIS receivers (for details see electronic supplementary material). For each ship, we obtain a trajectory from the database, i.e. a list of ports visited by the ship sorted by date. In 2007, there were 490 517 non-stop journeys linking 36 351 distinct pairs of arrival and departure ports. The complete set of trajectories, each path representing the shortest route at sea and coloured by the number of journeys passing through it, is shown in figure 1*a*.

Each trajectory can be interpreted as a small directed network where the nodes are ports linked together if the ship travelled directly between the ports. Larger networks can be defined by merging trajectories of different ships. In this article, we aggregate trajectories in four different ways: the combined network of all available trajectories and the subnetworks of container ships (3100 ships), bulk dry carriers (5498) and oil tankers (2628). These three subnetworks combinedly cover 74 per cent of the GCSN's total GT. In all four networks, we assign a weight w_{ij} to the link from port i to j equal to the sum of the

available space on all ships that have travelled on the link during 2007 measured in GT. If a ship made a journey from i to j more than once, its capacity contributes multiple times to w_{ij}.

3. THE GLOBAL CARGO SHIP NETWORK

The directed network of the entire cargo fleet is noticeably asymmetric, with 59 per cent of all linked pairs of ports being connected only in one direction. Still, the vast majority of ports (935 out of 951) belongs to one single strongly connected component, i.e. for any two ports in this component, there are routes in both directions, though possibly visiting different intermediate ports. The routes are intriguingly short: only few steps in the network are needed to get from one port to another. The shortest path length l between two ports is the minimum number of non-stop connections one must take to travel between origin and destination. In the GCSN, the average over all pairs of ports is extremely small, $\langle l \rangle = 2.5$. Even the maximum shortest path between any two ports (e.g. from Skagway, Alaska, to the small Italian island of Lampedusa) is only of length $l_{\max} = 8$. In fact, the majority of all possible origin–destination pairs (52%) can already be connected by two steps or less.

Comparing these findings to those reported for the worldwide airport network (WAN) shows interesting differences and similarities. The high asymmetry of the GCSN has not been found in the WAN, indicating that ship traffic is structurally very different from aviation. Rather than being formed by the accumulation of back-and-forth trips, ship traffic seems to be governed by an optimal arrangement of unidirectional, often circular routes. This optimality also shows in the GCSN's small shortest path lengths. In comparison, in the WAN, the average and maximum shortest path lengths are $\langle l \rangle = 4.4$ and $l_{\max} = 15$, respectively (Guimerà *et al.* 2005), i.e. about twice as long as in the GCSN. Similar to the WAN, the GCSN is highly clustered: if a port X is linked to ports Y and Z, there is a high probability that there is also a connection from Y to Z. We calculated a clustering coefficient C (Watts & Strogatz 1998) for directed networks and found $C = 0.49$, whereas random networks with the same number of nodes and links only yield $C = 0.04$ on average. Degree-dependent clustering coefficients C_k reveal that clustering decreases with node degree (see electronic supplementary material). Therefore, the GCSN—like the WAN—can be regarded as a small-world network possessing short path lengths despite substantial clustering (Watts & Strogatz 1998). However, the average degree of the GCSN, i.e. the average number of links arriving at and departing from a given port (in-degree plus out-degree), $\langle k \rangle = 76.5$, is notably higher than in the WAN, where $\langle k \rangle = 19.4$ (Barrat *et al.* 2004). In the light of the network size (the WAN consists of 3880 nodes), this difference becomes even more pronounced, indicating that the GCSN is much more densely connected. This redundancy of links gives the network high structural robustness to the loss of routes for keeping up trade.

The degree distribution $P(k)$ shows that most ports have few connections, but there are some ports linked to hundreds of other ports (figure 2*a*). Similar right-skewed degree distributions have been observed in many real-world networks (Barabási & Albert 1999). While the GCSN's degree distribution is not exactly scale-free, the distribution of link weights, $P(w)$, follows approximately a power law $P(w) \propto w^{-\mu}$ with $\mu = 1.71 \pm 0.14$ (95% CI for linear regression, figure 2*b*, see also electronic supplementary material). By averaging the sums of the link weights arriving at and departing from port i, we obtain the node strength s_i (Barrat *et al.* 2004). The strength distribution can also be approximated by a power law $P(s) \propto s^{-\eta}$ with $\eta = 1.02 \pm 0.17$, meaning that a small number of ports handle huge amounts of cargo (figure 2*c*). The determination of power law relationships by line fitting has been strongly criticized (e.g. Newman 2005; Clauset *et al.* 2009), therefore, we analysed the distributions with model selection by Akaike weights (Burnham & Anderson 1998). Our results confirm that a power law is a better fit than an exponential or a lognormal distribution for $P(w)$ and $P(s)$, but not $P(k)$ (see electronic supplementary material). These findings agree well with the concept of hubs–spokes networks (Notteboom 2004) that were proposed for cargo traffic, for example, in Asia (Robinson 1998). There are a few large, highly connected ports through which all smaller ports transact their trade. This scale-free property makes the ship trade network prone to the spreading and persistence of bioinvasive organisms (e.g. Pastor-Satorras & Vespignani 2001). The average nearest-neighbour degree, a measure of network assortativity, additionally underlines the hubs–spokes property of cargo ship traffic (see electronic supplementary material).

Strengths and degrees of the ports are related according to the scaling relation $\langle s(k) \rangle \propto k^{1.46 \pm 0.1}$ (95% CI for standardized major axis regression; Warton *et al.* 2006). Hence, the strength of a port grows generally faster than its degree (figure 2*d*). In other words, highly connected ports not only have many links, but their links also have a higher than average weight. This observation agrees with the fact that busy ports are better equipped to handle large ships with large amounts of cargo. A similar result, $\langle s(k) \rangle \propto k^{1.5 \pm 0.1}$, was found for airports (Barrat *et al.* 2004), which may hint at a general pattern in transportation networks. In the light of bioinvasion, these results underline empirical findings that big ports are more heavily invaded because of increased propagule pressure by ballast water of more and larger ships (Williamson 1996; Mack *et al.* 2000; see Cohen & Carlton 1998).

A further indication of the importance of a node is its betweenness centrality (Freeman 1979; Newman 2004). The betweenness of a port is the number of topologically shortest directed paths in the network that pass through this port. In figure 1*b*, we plot and list the most central ports. Generally speaking, centrality and degree are strongly correlated (Pearson's correlation coefficient: 0.81), but in individual cases other factors can also play a role. The Panama and Suez canals, for instance, are shortcuts to avoid long passages around

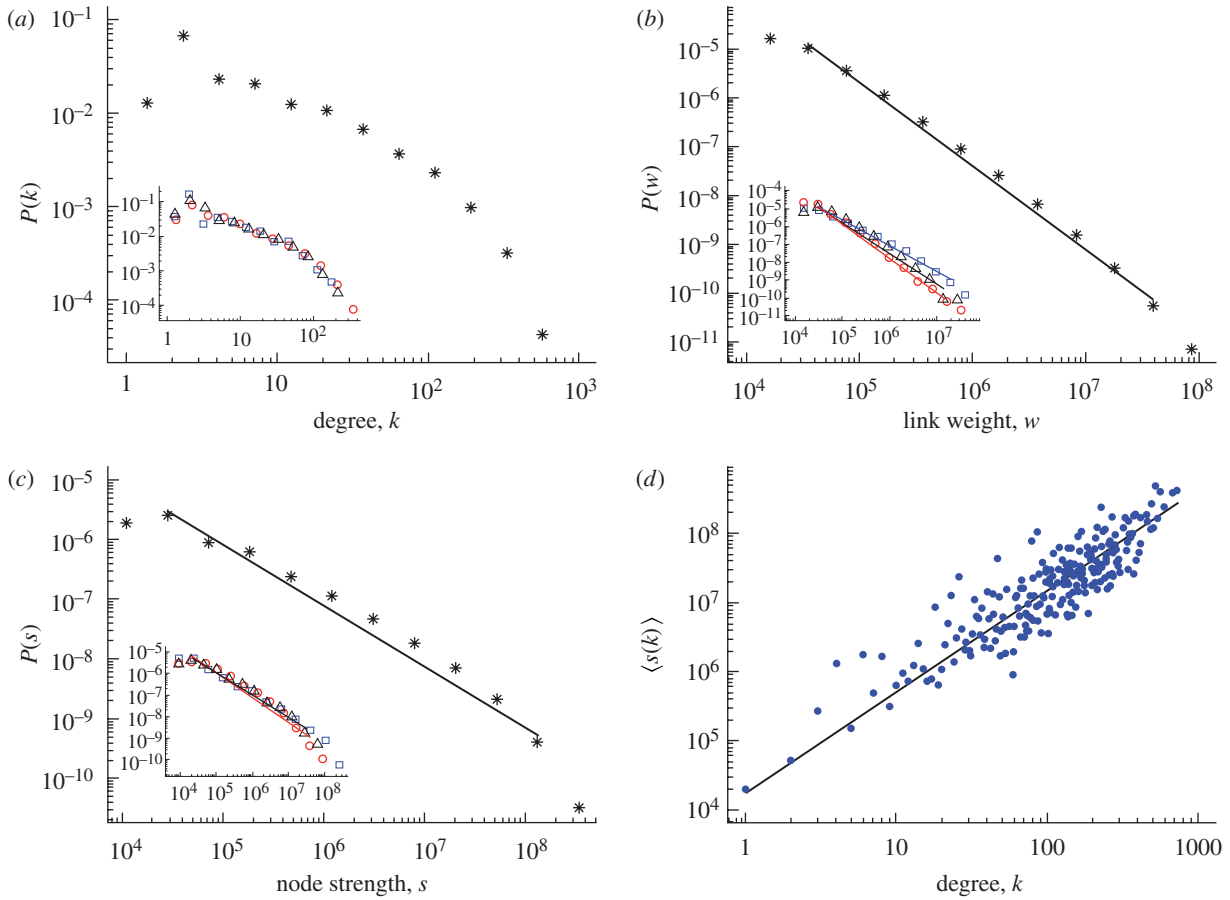

Figure 2. Degrees and weights in the GCSN. (*a*) The degree distributions $P(k)$ are right-skewed, but not power laws, neither for the GCSN nor its subnetworks. The degree k is defined here as the sum of in- and out-degree, thus $k = 1$ is rather rare (asterisk, all ship types; square, container ships; circle, bulk dry carriers; triangle, oil tankers). (*b*) The link weight distributions $P(w)$ reveal clear power law relationships for the GCSN and the three subnetworks, with exponents μ characteristic for the movement patterns of the different ship types (asterisk, $\mu = 1.71 \pm 0.14$; square, $\mu = 1.42 \pm 0.15$; circle, $\mu = 1.93 \pm 0.11$; triangle, $\mu = 1.73 \pm 0.25$). (*c*) The node strength distributions $P(s)$ are also heavy tailed, showing power law relationships. The stated exponents are calculated by linear regression with 95% confidence intervals (similar results are obtained with maximum likelihood estimates, see electronic supplementary material) (asterisk, $\eta = 1.02 \pm 0.17$; square, $\eta = 1.05 \pm 0.13$; circle, $\eta = 1.13 \pm 0.21$; triangle, $\eta = 1.01 \pm 0.16$). (*d*) The average strength of a node $\langle s(k) \rangle$ scales superlinearly with its degree, $\langle s(k) \rangle \propto k^{1.46 \pm 0.1}$, indicating that highly connected ports have, on average, links of higher weight.

South America and Africa. Other ports have a high centrality because they are visited by a large number of ships (e.g. Shanghai), whereas others gain their status primarily by being connected to many different ports (e.g. Antwerp).

4. THE NETWORK LAYERS OF DIFFERENT SHIP TYPES

To compare the movements of cargo ships of different types, separate networks were generated for each of the three main ship types: container ships, bulk dry carriers and oil tankers. Applying the network parameters introduced in the previous section to these three subnetworks reveals some broad-scale differences (table 1). The network of container ships is densely clustered, $C = 0.52$, has a rather low mean degree, $\langle k \rangle = 32.44$, and a large mean number of journeys (i.e. number of times any ship passes) per link, $\langle J \rangle = 24.26$. The bulk dry carrier network, on the other hand, is less clustered, has a higher mean degree and fewer journeys per link ($C = 0.43$, $\langle k \rangle = 44.61$, $\langle J \rangle = 4.65$). For the oil tankers, we find intermediate values ($C = 0.44$, $\langle k \rangle = 33.32$, $\langle J \rangle = 5.07$). Note that the mean degrees $\langle k \rangle$ of the subnetworks are substantially smaller than that of the full GCSN, indicating that different ship types use essentially the same ports but different connections.

A similar tendency appears in the scaling of the link weight distributions (figure 2*b*). $P(w)$ can be approximated as power laws for each network, but with different exponents μ. The container ships have the smallest exponent ($\mu = 1.42$) and bulk dry carriers the largest ($\mu = 1.93$) with oil tankers in between ($\mu = 1.73$). In contrast, the exponents for the distribution of node strength $P(s)$ are nearly identical in all three subnetworks, $\eta = 1.05$, $\eta = 1.13$ and $\eta = 1.01$, respectively.

These numbers give a first indication that different ship types move in distinctive patterns. Container ships typically follow set schedules visiting several ports in a fixed sequence along their way, thus providing regular services. Bulk dry carriers, by contrast,

Table 1. Characterization of different subnetworks. Number of ships, total GT (10^6 GT) and number of ports n in each subnetwork; together with network characteristics: mean degree $\langle k \rangle$, clustering coefficient C, mean shortest path length $\langle l \rangle$, mean journeys per link $\langle J \rangle$, power law exponents μ and η; and trajectory properties: average number of distinct ports $\langle N \rangle$, links $\langle L \rangle$, port calls $\langle S \rangle$ per ship and regularity index $\langle p \rangle$. Some notable values are highlighted in bold.

ship class	ships	MGT	n	$\langle k \rangle$	C	$\langle l \rangle$	$\langle J \rangle$	μ	η	$\langle N \rangle$	$\langle L \rangle$	$\langle S \rangle$	$\langle p \rangle$
whole fleet	16 363	664.7	951	**76.4**	0.49	2.5	13.57	1.71	1.02	10.4	15.6	31.8	0.63
container ships	3100	116.8	**378**	32.4	0.52	2.76	**24.25**	**1.42**	1.05	11.2	**21.2**	**48.9**	**1.84**
bulk dry carriers	5498	196.8	616	44.6	0.43	2.57	4.65	1.93	1.13	8.9	10.4	12.2	0.03
oil tankers	2628	178.4	505	33.3	0.44	2.74	5.07	1.73	1.01	9.2	12.9	17.7	0.19

appear less predictable as they frequently change their routes on short notice depending on the current supply and demand of the goods they carry. The larger variety of origins and destinations in the bulk dry carrier network ($n = 616$ ports, compared with $n = 378$ for container ships) explains the higher average degree and the smaller number of journeys for a given link. Oil tankers also follow short-term market trends, but, because they can only load oil and oil products, the number of possible destinations ($n = 505$) is more limited than for bulk dry carriers.

These differences are also underlined by the betweenness centralities of the three network layers (see electronic supplementary material). While some ports rank highly in all categories (e.g. Suez Canal, Shanghai), others are specialized on certain ship types. For example, the German port of Wilhelmshaven ranks tenth in terms of its worldwide betweenness for oil tankers, but is only 241st for bulk dry carriers and 324th for container ships.

We can gain further insight into the roles of the ports by examining their community structure. Communities are groups of ports with many links within the groups but few links between different groups. We calculated these communities for the three subnetworks with a modularity optimization method for directed networks (Leicht & Newman 2008) and found that they differ significantly from modularities of corresponding Erdös–Renyi graphs (figure 3; Guimerà *et al.* 2004). The network of container trade shows 12 communities (figure 3*a*). The largest ones are located (i) on the Arabian, Asian and South African coasts, (ii) on the North American east coast and in the Caribbean, (iii) in the Mediterranean, the Black Sea and on the European west coast, (iv) in Northern Europe and (v) in the Far East and on the American west coast. The transport of bulk dry goods reveals seven groups (figure 3*b*). Some can be interpreted as geographical entities (e.g. North American east coast, trans-Pacific trade), while others are dispersed on multiple continents. Especially interesting is the community structure of the oil transportation network that shows six groups (figure 3*c*): (i) the European, north and west African market, (ii) a large community comprising Asia, South Africa and Australia, (iii) three groups for the Atlantic market with trade between Venezuela, the Gulf of Mexico, the American east coast and Northern Europe and (iv) the American Pacific coast. It should be noted that the network includes the transport of crude oil as well as commerce with already refined oil products so that oil-producing regions do not appear as separate communities. This may be due to the limit in the detectability of smaller communities by modularity optimization (Fortunato & Barthelemy 2007), but does not affect the relevance of the revealed ship traffic communities. Because of the, by definition, higher transport intensity within communities, bioinvasive spread is expected to be heavier between ports of the same community. However, in figure 3, it becomes clear that there are no strict geographical barriers between communities. Thus, spread between communities is very likely to occur even on small spatial scales by shipping or ocean currents between close-by ports that belong to different communities.

Despite the differences between the three main cargo fleets, there is one unifying feature: their motif distribution (Milo *et al.* 2002). Like most previous studies, we focus here on the occurrence of three-node motifs and present their normalized Z score, a measure for their abundance in a network (figure 4). Strikingly, the three fleets have practically the same motif distribution. In fact, the Z scores closely resemble those found in the World Wide Web and different social networks that were conjectured to form a superfamily of networks (Milo *et al.* 2004). This superfamily displays many transitive triplet interactions (i.e. if $X \rightarrow Y$ and $Y \rightarrow Z$, then $X \rightarrow Z$); for example, the overrepresented motif 13 in figure 4, has six such interactions. Intransitive motifs, like motif 6, are comparably infrequent. The abundance of transitive interactions in the ship networks indicates that cargo can be transported both directly between ports as well as via several intermediate ports. Thus, the high clustering and redundancy of links (robustness to link failures) appear not only in the GCSN but also in the three subnetworks. The similarity of the motif distributions to other humanly optimized networks underlines that cargo trade, like social networks and the World Wide Web, depends crucially on human interactions and information exchange. While advantageous for the robustness of trade, the clustering of links as triplets also has an unwanted side effect: in general, the more clustered a network, the more vulnerable it becomes to the global spread of alien species, even for low invasion probabilities (Newman 2003*b*).

5. NETWORK TRAJECTORIES

Going beyond the network perspective, the database also provides information about the movement

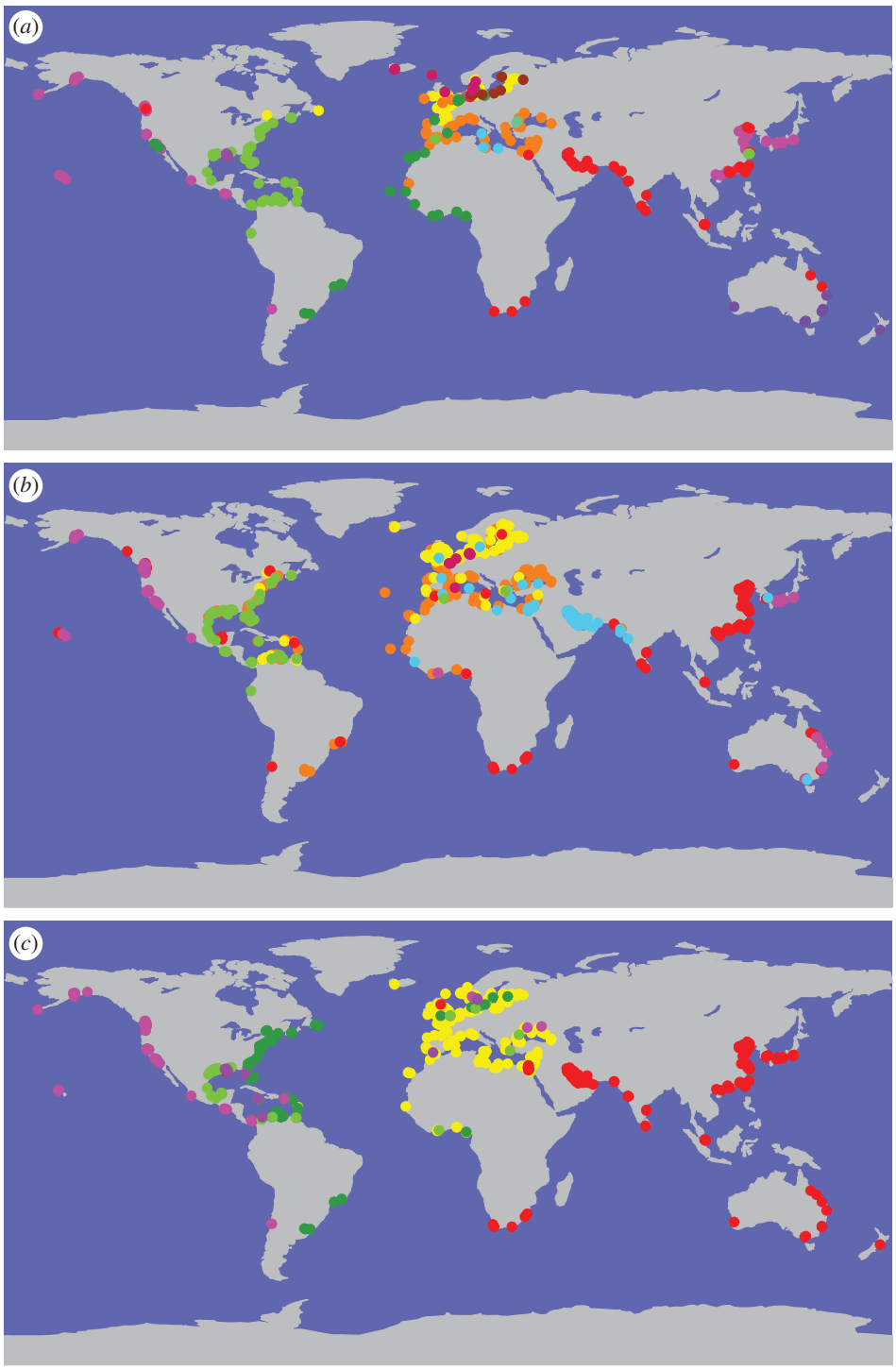

Figure 3. Communities of ports in three cargo ship subnetworks. The communities are groups of ports that maximize the number of links within the groups, as opposed to between the groups, in terms of the modularity Q (Leicht & Newman 2008). In each map, the colours represent the c distinct trading communities for the goods transported by (*a*) container ships ($c = 12$, $Q = 0.605$), (*b*) bulk dry carriers ($c = 7$, $Q = 0.592$) and (*c*) oil tankers ($c = 6$, $Q = 0.716$). All modularities Q of the examined networks differ significantly from modularities in Erdös–Renyi graphs of the same size and number of links (Guimerà *et al.* 2004). For the networks corresponding to (*a*), (*b*) and (*c*), values are $Q_{\mathrm{ER}} = 0.219$, $Q_{\mathrm{ER}} = 0.182$ and $Q_{\mathrm{ER}} = 0.220$, respectively.

characteristics per individual ship (table 1). The average number of distinct ports per ship $\langle N \rangle$ does not differ much between different ship classes, but container ships call much more frequently at ports than bulk dry carriers and oil tankers. This difference is explained by the characteristics and operational mode of these ships. Normally, container ships are fast (between 20 and 25 knots) and spend less time (1.9 days on average in our data) in the ports for cargo operations. By contrast, bulk dry carriers and oil tankers move more slowly (between 13 and 17 knots) and stay longer in the ports (on average 5.6 days for bulk dry carriers and 4.6 days for oil tankers).

The speed at sea and of cargo handling, however, is not the only operational difference. The topology of

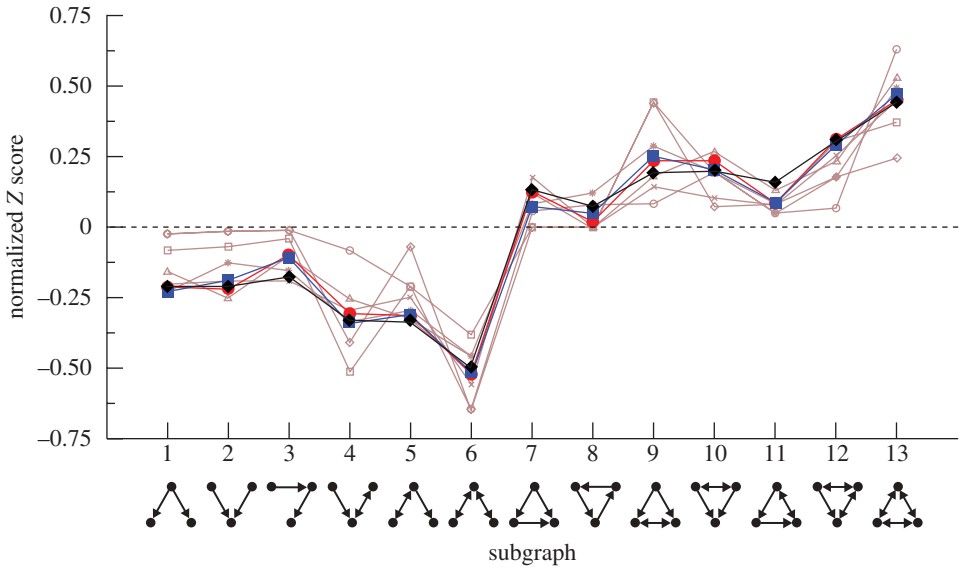

Figure 4. Motif distributions of the three main cargo fleets. A positive (negative) normalized Z score indicates that a motif is more (less) frequent in the real network than in random networks with the same degree sequence. For comparison, we overlay the Z scores of the World Wide Web and social networks. The agreement suggests that the ship networks fall in the same superfamily of networks (Milo *et al.* 2004). The motif distributions of the fleets are maintained even when 25, 50 and 75% of the weakest connections are removed. Line with open circles, www-1; line with open squares, www-2; line with open diamonds, www-3; line with triangles, social-1; line with crosses, social-2; line with asterisks, social-3; line with red circles, bulk dry carriers; line with blue squares, container ships; line with black diamonds, oil tankers.

the trajectories also differs substantially. Characteristic sample trajectories for each ship type are presented in figure 5*a–c*. The container ship (figure 5*a*) travels on some of the links several times during the study period, whereas the bulk dry carrier (figure 5*b*) passes almost every link exactly once. The oil tanker (figure 5*c*) commutes a few times between some ports, but by and large also serves most links only once.

We can express these trends in terms of a 'regularity index' p that quantifies how much the frequency with which each link is used deviates from a random network. Consider the trajectory of a ship calling S times at N distinct ports and travelling on L distinct links. We compare the mean number of journeys per link $f_{\mathrm{real}} = S/L$ to the average link usage f_{ran} in an ensemble of randomized trajectories with the same number of nodes N and port calls S. To quantify the difference between real and random trajectories, we calculate the Z score $p = (f_{\mathrm{real}} - f_{\mathrm{ran}})/\sigma$ (where σ is the standard deviation of f in the random ensemble). If $p = 0$, the real trajectory is indistinguishable from a random walk, whereas larger values of p indicate that the movement is more regular. Figure 5*d–f* present the distributions of the regularity index p for the different fleets. For container ships, p is distributed broadly around $p \approx 2$, thus supporting our earlier observation that most container ships provide regular services between ports along their way. Trajectories of bulk dry carriers and oil tankers, on the other hand, appear essentially random with the vast majority of ships near $p = 0$.

6. APPROXIMATING TRAFFIC FLOWS USING THE GRAVITY MODEL

In this article, we view global ship movements as a network based on detailed arrival and departure records. Until recently, surveys of seaborne trade had to rely on far fewer data: only the total number of arrivals at some major ports was publicly accessible, but not the ships' actual paths (Zachcial & Heideloff 2001). Missing information about the frequency of journeys, thus, had to be replaced by plausible assumptions, the gravity model being the most popular choice. It posits that trips are, in general, more likely between nearby ports than between ports far apart. If d_{ij} is the distance between ports i and j, the decline in mutual interaction is expressed in terms of a distance deterrence function $f(d_{ij})$. The number of journeys from i to j then takes the form $F_{ij} = a_i b_j O_i I_j f(d_{ij})$, where O_i is the total number of departures from port i and I_j the number of arrivals at j (Haynes & Fotheringham 1984). The coefficients a_i and b_j are needed to ensure $\sum_j F_{ij} = O_i$ and $\sum_i F_{ij} = I_j$.

How well can the gravity model approximate real ship traffic? We choose a truncated power law for the deterrence function, $f(d_{ij}) = d_{ij}^{-\beta}\exp(-d_{ij}/\kappa)$. The strongest correlation between model and data is obtained for $\beta = 0.59$ and $\kappa = 4900$ km (see electronic supplementary material). At first sight, the agreement between data and model appears indeed impressive. The predicted distribution of travelled distances (figure 6*a*) fits the data far better than a simpler non-spatial model that preserves the total number of journeys, but assumes completely random origins and destinations.

A closer look at the gravity model, however, reveals its limitations. In figure 6*b*, we count how often links with an observed number of journeys N_{ij} are predicted to be passed F_{ij} times. Ideally, all data points would align along the diagonal $F_{ij} = N_{ij}$, but we find that the data are substantially scattered. Although the parameters β and κ were chosen to minimize the scatter,

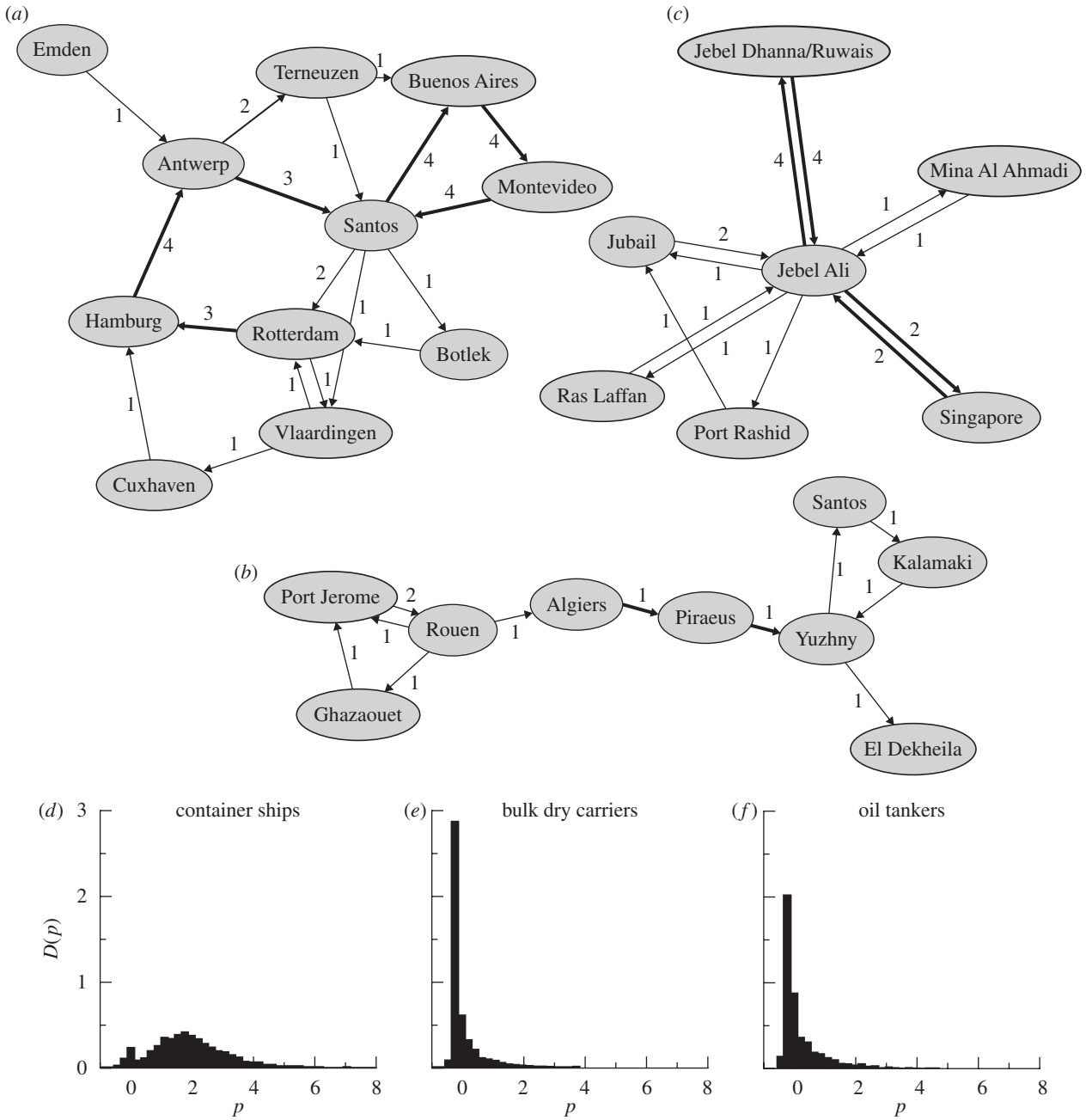

Figure 5. Sample trajectories of (*a*) a container ship with a regularity index $p = 2.09$, (*b*) a bulk dry carrier, $p = 0.098$ and (*c*) an oil tanker, $p = 1.027$. In the three trajectories, the numbers and the line thickness indicate the frequency of journeys on each link. (*d*–*f*) Distribution of p for the three main fleets.

the correlation between data and model is only moderate (Kendall's $\tau = 0.433$). In some cases, the prediction is off by several thousand journeys per year.

Recent studies have used the gravity model to pinpoint the ports and routes central to the spread of invasive species (Drake & Lodge 2004; Tatem *et al.* 2006). The model's shortcomings pose the question as to how reliable such predictions are. For this purpose, we investigated a dynamic model of ship-mediated bioinvasion where the weights of the links are either the observed traffic flows or the flows of the gravity model.

We follow previous epidemiological studies (Rvachev & Longini 1985; Flahault *et al.* 1988; Hufnagel *et al.* 2004; Colizza *et al.* 2006) in viewing the spread on the network as a metapopulation process where the population dynamics on the nodes are coupled by transport on the links. In our model, ships can transport a surviving population of an invasive species with only a small probability $p_{\mathrm{trans}} = 1\%$ on each journey between two successively visited ports. The transported population is only a tiny fraction s of the population at the port of origin. Immediately after arriving at a new port, the species experiences strong demographic fluctuations that lead, in most cases, to the death of the imported population. If, however, the new immigrants beat the odds of this 'ecological roulette' (Carlton & Geller 1993) and establish, the population P grows rapidly following the stochastic logistic equation $\mathrm{d}P/\mathrm{d}t = rP(1-P) + \sqrt{P}\xi(t)$ with growth rate $r = 1$

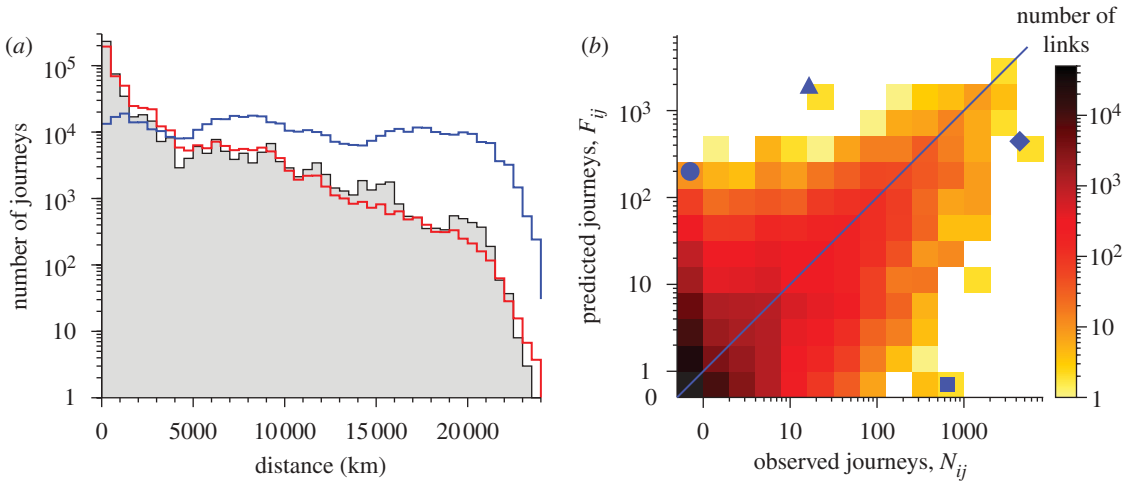

Figure 6. (*a*) Histogram of port-to-port distances travelled in the GCSN (navigable distances around continents as indicated in figure 1). We overlay the predictions of two different models. The gravity model (red), based on information about distances between ports and total port calls, gives a much better fit than a simpler model (blue) which only fixes the total number of journeys (black line, observed; red line, gravity model; blue line, random traffic). (*b*) Count of port pairs with N_{ij} observed and F_{ij} predicted journeys. The flows F_{ij} were calculated with the gravity model (rounded to the nearest integer). Some of the worst outliers are highlighted in blue. Circle, Antwerp to Calais ($N_{ij}=0$ versus $F_{ij}=200$); triangle, Hook of Holland to Europoort (16 versus 1895); diamond, Calais to Dover (4392 versus 443); square, Harwich to Hook of Holland (644 versus 0).

per year and Gaussian white noise ξ. For details of the model, see the electronic supplementary material.

Starting from a single port at carrying capacity $P=1$, we model contacts between ports as Poisson processes with rates N_{ij} (empirical data) or F_{ij} (gravity model). As shown in figure 7*a*, the gravity model systematically overestimates the spreading rate, and the difference can become particularly pronounced for ports that are well connected, but not among the central hubs in the network (figure 7*b*). Comparing typical sequences of invaded ports, we find that the invasions driven by the real traffic flows tend to be initially confined to smaller regional ports, whereas in the gravity model the invasions quickly reach the hubs. The total out- and in-flows at the ship journeys' origin and departure ports, respectively, are indeed more strongly positively correlated in reality than in the model ($\tau=0.157$ versus 0.047). The gravity model thus erases too many details of a hierarchical structure present in the real network. That the gravity model eliminates most correlations is also plausible from simple analytic arguments (see electronic supplementary material for details). The absence of strong correlations makes the gravity model a suitable null hypothesis if the correlations in the real network are unknown, but several recent studies have shown that correlations play an important role in spreading processes on networks (e.g. Boguña & Pastor-Satorras 2002; Newman 2002). Hence, if the correlations are known, they should not be ignored.

While we observed that the spreading rates for the AIS data were consistently slower than for the gravity model even when different parameters or population models were considered, the time scale of the invasion is much less predictable. The assumption that only a small fraction of invaders succeed outside their native habitat appears realistic (Mack *et al.* 2000). Furthermore, the parameters in our model were adjusted so

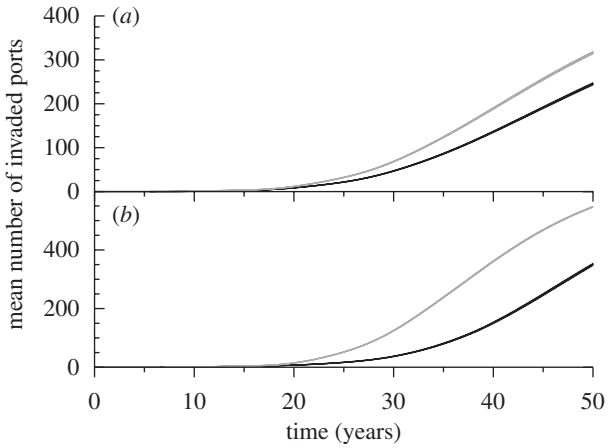

Figure 7. Results from a stochastic population model for the spread of an invasive species between ports. (*a*) The invasion starts from one single, randomly chosen port. (*b*) The initial port is fixed as Bergen (Norway), an example of a well-connected port (degree $k=49$) which is not among the central hubs. The rate of journeys from port i to j per year is assumed to be N_{ij} (real flows from the GCSN) or F_{ij} (gravity model). Each journey has a small probability of transporting a tiny fraction of the population from origin to destination. Parameters were adjusted ($r=1$ per year, $p_{\mathrm{trans}}=0.01$ and $s=4\times10^{-5}$) to yield a per-ship-call probability of initiating invasion of approximately 4.4×10^{-4} (Drake & Lodge 2004; see electronic supplementary material for details). Plotted are the cumulative numbers of invaded ports (population number larger than half the carrying capacity) averaged over (*a*) 14 000 and (*b*) 1000 simulation runs (standard error equal to line thickness). Black line, real traffic flows; grey line, gravity model.

that the per-ship-call probability of initiating invasion is approximately 4.4×10^{-4}, a rule-of-thumb value stated by Drake & Lodge (2004). Still, too little is empirically known to pin down individual parameters with sufficient accuracy to give more than a qualitative impression. It is especially difficult to predict how a

potential invader reacts to the environmental conditions at a specific location. Growth rates certainly differ greatly between ports depending on factors such as temperature or salinity, with respect to the habitat requirements of the invading organisms. Our results should, therefore, be regarded as one of many different conceivable scenarios. A more detailed study of bioinvasion risks based on the GCSN is currently underway (H. Seebens & B. Blasius 2009, unpublished data).

7. CONCLUSION

We have presented a study of ship movements based on AIS records. Viewing the ports as nodes in a network linked by ship journeys, we found that global cargo shipping, like many other complex networks investigated in recent years, possesses the small-world property as well as broad degree and weight distributions. Other features, like the importance of canals and the grouping of ports into regional clusters, are more specific to the shipping industry. An important characteristic of the network are differences in the movement patterns of different ship types. Bulk dry carriers and oil tankers tend to move in a less regular manner between ports than container ships. This is an important result regarding the spread of invasive species because bulk dry carriers and oil tankers often sail empty and therefore exchange large quantities of ballast water. The gravity model, which is the traditional approach to forecasting marine biological invasions, captures some broad trends of global cargo trade, but for many applications its results are too crude. Future strategies to curb invasions will have to take more details into account. The network structure presented in this article can be viewed as a first step in this direction.

We thank B. Volk, K. H. Holocher, A. Wilkinson, J. M. Drake and H. Rosenthal for stimulating discussions and helpful suggestions. We also thank Lloyd's Register Fairplay for providing their shipping database. This work was supported by German VW-Stiftung and BMBF.

REFERENCES

Albert, R. & Barabási, A.-L. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)

Barabási, A.-L. & Albert, R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)

Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. 2004 The architecture of complex weighted networks. *Proc. Natl Acad. Sci. USA* **101**, 3747–3752. (doi:10.1073/pnas.0400087101)

Barthélemy, M. & Flammini, A. 2008 Modeling urban street patterns. *Phys. Rev. Lett.* **100**, 138 702. (doi:10.1103/PhysRevLett.100.138702)

Boguña, M. & Pastor-Satorras, R. 2002 Epidemic spreading in correlated complex networks. *Phys. Rev. E* **66**, 047 104. (doi:10.1103/PhysRevE.66.047104)

Buhl, J., Gautrais, J., Reeves, N., Solé, R. V., Valverde, S., Kuntz, P. & Theraulaz, G. 2006 Topological patterns in street networks of self-organized urban settlements. *Eur. Phys. J. B* **49**, 513–522. (doi:10.1140/epjb/e2006-00085-1)

Burnham, K. P. & Anderson, D. R. 1998 *Model selection and multimodel inference. A practical information-theoretic approach*. New York, NY: Springer.

Carlton, J. T. & Geller, J. B. 1993 Ecological roulette: the global transport of nonindigenous marine organisms. *Science* **261**, 78–82. (doi:10.1126/science.261.5117.78)

Clauset, A., Shalizi, C. R. & Newman, M. E. J. 2009 Power-law distributions in empirical data. *SIAM Rev.* **51**, 661–703. (doi:10.1137/070710111)

Cohen, A. N. & Carlton, J. T. 1998 Accelerating invasion rate in a highly invaded estuary. *Science* **279**, 555–558. (doi:10.1126/science.279.5350.555)

Colizza, V., Barrat, A., Barthélemy, M. & Vespignani, A. 2006 The role of the airline transportation network in the prediction and predictability of global epidemics. *Proc. Natl Acad. Sci. USA* **103**, 2015–2020. (doi:10.1073/pnas.0510525103)

Drake, J. M. & Lodge, D. M. 2004 Global hot spots of biological invasions: evaluating options for ballast-water management. *Proc. R. Soc. Lond. B* **271**, 575–580. (doi:10.1098/rspb.2003.2629)

Drake, J. M. & Lodge, D. M. 2007 Hull fouling is a risk factor for intercontinental species exchange in aquatic ecosystems. *Aquat. Invasions* **2**, 121–131. (doi:10.3391/ai.2007.2.2.7)

Flahault, A., Letrait, S., Blin, P., Hazout, S., Ménarés, J. & Valleron, A. J. 1988 Modelling the 1985 influenza epidemic in France. *Stat. Med.* **7**, 1147–1155. (doi:10.1002/sim.4780071107)

Fortunato, S. & Barthelemy, M. 2007 Resolution limit in community detection. *Proc. Natl Acad. Sci. USA* **104**, 36–41. (doi:10.1073/pnas.0605965104)

Freeman, L. C. 1979 Centrality in social networks I: conceptual clarification. *Soc. Netw.* **1**, 215–239. (doi:10.1016/0378-8733(78)90021-7)

Gross, T. & Blasius, B. 2008 Adaptive coevolutionary networks: a review. *J. R. Soc. Interface* **5**, 259–271. (doi:10.1098/rsif.2007.1229)

Guimerà, R. & Amaral, L. A. N. 2004 Modeling the worldwide airport network. *Eur. Phys. J. B* **38**, 381–385. (doi:10.1140/epjb/e2004-00131-0)

Guimerà, R., Sales-Pardo, M. & Amaral, L. A. N. 2004 Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E* **70**, 025 101(R). (doi:10.1103/PhysRevE.70.025101)

Guimerà, R., Mossa, S., Turtschi, A. & Amaral, L. A. N. 2005 The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc. Natl Acad. Sci. USA* **102**, 7794–7799. (doi:10.1073/pnas.0407994102)

Haynes, K. E. & Fotheringham, A. S. 1984 *Gravity and spatial interaction models*. Beverly Hills, CA: Sage.

Hu, Y. & Zhu, D. 2009 Empirical analysis of the worldwide maritime transportation network. *Phys. A* **388**, 2061–2071. (doi:10.1016/j.physa.2008.12.016)

Hufnagel, L., Brockmann, D. & Geisel, T. 2004 Forecast and control of epidemics in a globalized world. *Proc. Natl Acad. Sci. USA* **101**, 15 124–15 129. (doi:10.1073/pnas.0308344101)

International Maritime Organization. 2006 International shipping and world trade. Facts and figures. See http://www.imo.org/.

Latora, V. & Marchiori, M. 2002 Is the Boston subway a small-world network? *Phys. A* **314**, 109–113. (doi:10.1016/S0378-4371(02)01089-0)

Leicht, E. A. & Newman, M. E. J. 2008 Community structure in directed networks. *Phys. Rev. Lett.* **100**, 118 703. (doi:10.1103/PhysRevLett.100.118703)

Lounibos, L. P. 2002 Invasions by insect vectors of human disease. *Annu. Rev. Entomol.* **47**, 233–266. (doi:10.1146/annurev.ento.47.091201.145206)

Mack, R. N., Simberloff, D., Lonsdale, W. M., Evans, H., Clout, M. & Bazzaz, F. A. 2000 Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.* **10**, 689–710. (doi:10.1890/1051-0761(2000)010[0689:BICEGC]2.0.CO;2)

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. 2002 Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827. (doi:10.1126/science.298.5594.824)

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. & Alon, U. 2004 Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542. (doi:10.1126/science.1089167)

Newman, M. E. J. 2002 Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701. (doi:10.1103/PhysRevLett.89.208701)

Newman, M. E. J. 2003*a* The structure and function of complex networks. *SIAM Rev.* **45**, 167–256. (doi:10.1137/S003614450342480)

Newman, M. E. J. 2003*b* Properties of highly clustered networks. *Phys. Rev. E* **68**, 026 121. (doi:10.1103/PhysRevE.68.026121)

Newman, M. E. J. 2004 Who is the best connected scientist? A study of scientific coauthorship networks. In *Complex networks* (eds E. Ben-Naim, H. Frauenfelder & Z. Toroczkai), pp. 337–370. Berlin, Germany: Springer.

Newman, M. E. J. 2005 Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351. (doi:10.1080/00107510500052444)

Notteboom, T. E. 2004 Container shipping and ports: an overview. *Rev. Netw. Econ.* **3**, 86–106. (doi:10.2202/1446-9022.1045)

Pastor-Satorras, R. & Vespignani, A. 2001 Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. (doi:10.1103/PhysRevLett.86.3200)

Pimentel, D., Zuniga, R. & Morrison, D. 2005 Update on the environmental costs associated with alien-invasive species in the United States. *Ecol. Econ.* **52**, 273–288. (doi:10.1016/j.ecolecon.2004.10.002)

Robinson, R. 1998 Asian hub/feeder nets: the dynamics of restructuring. *Marit. Policy Manag.* **25**, 21–40. (doi:10.1080/03088839800000043)

Rodrigue, J.-P., Comtois, C. & Slack, B. 2006 *The geography of transport systems*. London, UK: Routledge.

Ruiz, G. M., Rawlings, T. K., Dobbs, F. C., Drake, L. A., Mullady, T., Huq, A. & Colwell, R. R. 2000 Global spread of microorganisms by ships. *Nature* **408**, 49–50. (doi:10.1038/35040695)

Rvachev, L. A. & Longini Jr, I. M. 1985 A mathematical model for the global spread of influenza. *Math. Biosci.* **75**, 3–22. (doi:10.1016/0025-5564(85)90064-1)

Sen, P., Dasgupta, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G. & Manna, S. S. 2003 Small-world properties of the Indian railway network. *Phys. Rev. E* **67**, 036 106. (doi:10.1103/PhysRevE.67.036106)

Tatem, A. J., Hay, S. I. & Rogers, D. J. 2006 Global traffic and disease vector dispersal. *Proc. Natl Acad. Sci. USA* **103**, 6242–6247. (doi:10.1073/pnas.0508391103)

United Nations Conference on Trade and Development. 2007 Review of maritime transport. See http://www.unctad.org/en/docs/rmt2007_en.pdf.

Warton, D. I., Wright, I. J., Falster, D. S. & Westoby, M. 2006 Bivariate line-fitting methods for allometry. *Biol. Rev.* **81**, 259–291. (doi:10.1017/S1464793106007007)

Watts, D. J. & Strogatz, S. H. 1998 Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442. (doi:10.1038/30918)

Wei, T., Deng, G. & Wu, P. 2007 Analysis of network effect in port and shipping system characterized by scale-free network. In *Proc. Int. Conf. on Intelligent Systems and Knowledge Engineering*. Amsterdam, The Netherlands: Atlantic Press.

Williamson, M. 1996 *Biological invasions*. London, UK: Chapman and Hall.

Zachcial, M. & Heideloff, C. 2001 ISL shipping statistics yearbook 2001. Technical report, Institute of Shipping Economics and Logistics, Bremen, Germany.