

# A new connectivity index for container ports

John J Bartholdi III<sup>a</sup>, Pisit Jarumaneeroj<sup>b</sup> and Amar Ramudhin<sup>c</sup>

<sup>a</sup>Georgia Tech Panama Logistics Innovation and Research Center Panama, 765 Ferst Drive NW, Atlanta, GA 30332-0205, USA.

<sup>b</sup>Department of Industrial Engineering, Chulalongkorn University, Bangkok 10330, Thailand.

E-mail: [pisit.ja@chula.ac.th](mailto:pisit.ja@chula.ac.th)

<sup>c</sup>The Logistics Institute, University of Hull, Hull HU6 7RX, UK.

E-mail: [A.Ramudhin@hull.ac.uk](mailto:A.Ramudhin@hull.ac.uk)

**Abstract** We propose a new index, the Container Port Connectivity Index, to measure the trade connectivity of ports within the network of container shipping. This index is based on both economics and network topology, and a distinctive feature is that the strength of a port is based on its position within the global structure of the shipping network, and not just on local information such as the number of TEUs handled, or direct links to other ports. Furthermore, the index produces separate scores for inbound and outbound container movements and in so doing it supports more detailed analyses. We explore the usefulness of the index by analyzing the global network of scheduled mainline container-shipping services as it existed in September 2011.

*Maritime Economics & Logistics* (2016) **18**, 231–249. doi:10.1057/mel.2016.5; published online 23 June 2016

**Keywords:** container port; shipping network; connectivity

The online version of this article is available Open Access

---

## Measuring the Connectivity of Container Ports

What makes a container port important? From an operational point of view, it is *connectivity*: Are there convenient services to and from other important ports? Several measures borrowed from graph theory have been applied to measure connectivity, but these are typically based only weakly on economics. We suggest a new measure of importance with which to compare container ports, which we call the *Container Port Connectivity Index* (CPCI). This measure

is based on a richer model than used heretofore of the intensity of container movement among ports. Then we use this model to compute the importance of each port as if ranking Web pages. In this computation, the importance of a port is based not just on the importance of immediate neighbors but also on the importance of neighbors of those neighbors, and so on. We use these tools to analyze a model of the global network of mainline, scheduled, container-shipping services and argue that they offer a more nuanced and accurate reflection of the relative importance of ports.

## The Global Network of Container-Shipping

A number of recent papers have examined patterns of ship movement across the ocean by way of network models, but those models are different from ours, either in the type of ships followed (and so the ports) or in the details of the network model, or in the measures of importance computed – See, for example, Barigozzi *et al* (2011), Doshi *et al* (2012), Ducruet *et al* (2010), Ducruet and Notteboom (2012), Ducruet and Zaidi (2012), Hoffmann *et al* (2014), Wilmsmeier and Hoffmann (2008). Here we review the most notable works relevant to ours.

Kaluza *et al* (2010) and Kölzsch and Blasius (2011) were interested in the spread of invasive species in ship ballast, and so their networks represent the time-aggregated movements of all ships over a year. Both their models and ours represent each port as a vertex. Kaluza *et al* (2010) include a link (edge) directed from vertex (port) to vertex (port) if some container ship traveled directly from port to port at any time during 2007, as reported by [www.sea-web.com](http://www.sea-web.com). In our model the meaning of a link is different: There is a link directed from vertex (port)  $i$  to vertex (port)  $j$  if there was mainline, scheduled container service traveling directly from port  $i$  to port  $j$ , as reported by commercial data source Compar Data in September 2011. In other words our model is a snapshot of the network, as it would be engaged by a shipper, while theirs is more like a time-exposure, which includes ephemeral phenomena, such as seasonal feeder services. This makes sense for their concern, bio-invasion, but our focus is operational: What is the nature of the network on which a particular container might move?

Both Kaluza *et al* (2010) and Kölzsch and Blasius (2011) include many kinds of shipping besides container shipping, such as oil tankers, barges, or ferries and thus many other types of ports, including ferry terminals and ports specialized to handle chemicals, grain, ore, or other bulk products. In addition, they lump together some ports that are nearby, so that, for example, all four container ports of Panama (one on the Pacific and three on the Atlantic) are represented as a single port that spans both oceans.

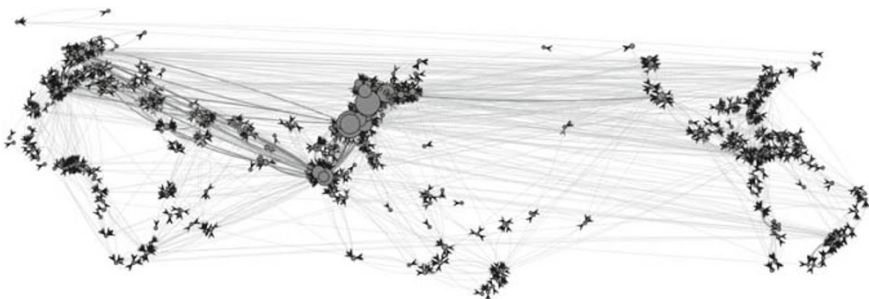


The network of Doshi *et al* (2012) is similar to ours in that it is based on scheduled container services, but they seemed interested primarily in data-mining and, in a heroic feat of parsing, scoured the Web to get schedules of ‘six randomly chosen shipping companies’, on which they based their network. We simply purchased a comprehensive list that included many additional details such as ship descriptions and distributions of actual transit times. Their network is also different from ours in that they connect one port to another if there is a service connecting one to the other (we require direct service).

Ducruet and Notteboom (2012) also studied network models of container shipping that are constructed, like that of Kaluza *et al* (2010), from the time-aggregated movements of all container ships over a year. In one of their models two ports are connected if a container ship has traveled directly from one to the other any time during the reference year. In their other model, two ports are connected if both appeared anywhere on the same scheduled container service during that year, so that each service is represented by a complete subgraph of undirected links, which ignores the direction of container movement.

## A Snapshot of Global Scheduled Container Services

Figure 1 shows the network described by our data set. Several large patterns are immediately evident, including the importance of the East Asian ports and the intensity of trade between East Asia and Europe, through the great transshipment ports of Southeast Asia and through the Suez Canal. Similarly, it is clear that services along the west coast of Africa or the east coast of South America are primarily local connections, from port to nearby port.



**Figure 1:** The network of scheduled container services among 457 ports of the world.

*Note:* Each arrow indicates scheduled container service from origin to destination port (but not the actual geography of the shipping route). Darker links are of greater trade intensity according to a computation based on the LSCI. Ports represented by larger disks scored proportionally higher according to the new measure of port connectivity described herein.

This network has 457 ports and 2479 links and is strongly connected (that is, for any two ports, each is reachable from the other by some directed path). The mean degree of ports is 10.85, and link-diameter of the network is 11 links (The diameter of a network is the length of the ‘longest shortest path’). The link-diameter is the diameter when distance is interpreted as the number of links to be traversed and represents an upper bound on the number of transshipments required. The longest shortest path in our network is that traversed by a container traveling from Maizuru (Japan) to Fortaleza (Brazil), passing en route through the container ports of Niigata (Japan), Tomakomai (Japan), Hachinohe (Japan), Busan (South Korea), Savannah (the United States), Kingston (Jamaica), Port of Spain (Trinidad and Tobago), Degrade des Cannes (French Guiana) and Belem (Brazil).

The mean degree is smaller in our network and the diameter larger than those of the time-aggregated networks, probably because they contain additional links such as seasonal and opportunistic changes to shipping routes.

Because our data set includes transit times, we can also report that the travel-time diameter of our network is 56 days, not counting time in port. To ship from Honiara (Solomon Islands) to Sortland (Norway) requires 56 days and traverses 9 links. Any container must pass en route through Shanghai (China), Busan (South Korea), Cristobal (Panama), Manzanillo (Panama), New York (the United States), Halifax (Canada), Argentia (Canada) and Reykjavik (Iceland).

### **A link-weight based on economics**

Our network differs most significantly from others in the choice of link weights. Ducruet and Zaidi (2012) ignore weights on links, so each link records merely the fact of direct service. But this treats as equally important services between major ports and services between minor ports. Also, in their network links are undirected, which ignores the direction of ship travel and so makes no distinction between inbound and outbound container movement. Doshi *et al* (2012) provide a slightly richer model by assigning link weights based on ‘the total number of times a trip is made between a set of two ports’ within the schedules they considered. However, setting aside the somewhat arbitrary nature of the schedules chosen, this weight ignores TEU capacity of the ships and so makes no distinction between a feeder vessel and a post-Panamax vessel. Kaluza *et al* (2010) defined the weight of a link to be the sum of gross tonnage of all shipping traversing that link. Kölzsch and Blasius (2011) are closer to us in choosing link weights proportional to cumulative cargo capacity along that link, which they use as a proxy for amount of ballast water (that might harbor invasive species). In contrast, we chose a weight designed specifically to reflect intensity of trade.

It is hard to get trade data at the level of containers and ports, but there is a well-established measure of trade available in the *Liner Shipping Connectivity*



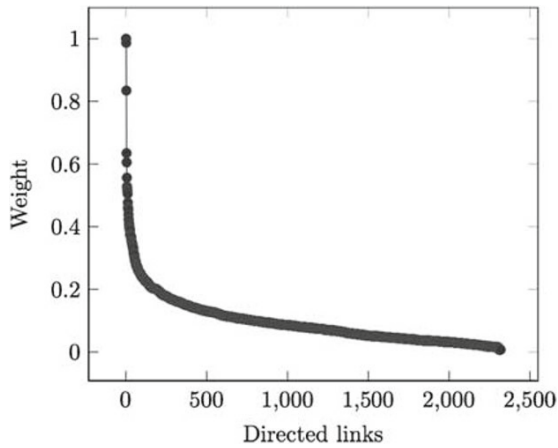
*Index* (LSCI). The LSCI was developed by the United Nations Conference on Trade and Development (UNCTAD) to compare the trade competitiveness of countries with respect to logistics and transport (See Hoffman (2005) UNCTAD's Review of Maritime Transport (2014) and links therein, especially stats.unctad.org/lsci. It is also worth noting that Hoffman *et al* (2014) have recently extended the idea of the LSCI to bilateral trade, though maintaining their focus on trade between nations).

UNCTAD computes the LSCI for a country as an aggregation of five statistics: number of liner services calling, number of liner companies providing those services, number of ships in those services, combined container capacity of those ships (in TEUs), and capacity of the largest ship calling. Despite the somewhat arbitrary method of aggregating the component statistics, the LSCI is based on hard numbers and is felt to accurately reflect levels of trade. Indeed, the LSCI has been observed to be strongly correlated with other measures, such as the Logistics Performance Index (LPI); a comprehensive survey of perceptions that is reported annually by the World Bank (Arvis *et al*, 2007; Ojala and Hoffmann, 2010).

The LSCI implicitly treats each country of concern as if it were a single location, and the entire rest of the world a single trading partner. In effect, the world container network is reduced to two vertices. The five statistics on which the LSCI is based to describe the container capacity connecting the country to the rest of the world and so we may interpret the LSCI as a measure of the strength of the link between two vertices.

We follow the idea of the LSCI to compute, for each pair of ports  $i$  and  $j$ , a weight reflecting the intensity of container capacity moving from  $i$  directly to  $j$ . The computation is exactly that of the LSCI, except for ports rather than for countries, and for directed transit (that is, from port  $i$  to port  $j$  rather than totals between ports). With regard to each statistic in turn, the ports are ranked and the value of the statistic is normalized so that the maximum value equals 1.0. Then for each port in turn, the five scores are summed and the result normalized, again to 1.0. As a result, every link in the network is assigned a weight in  $(0, 1)$ .

Figure 2 shows the resultant distribution of weights for all direct links in our network, and Table 1 lists the 20 links of greatest weight. The most distinctive pattern is that all but one of these links are intra-Asian. Notably, Shanghai figures in six of these links, three times as an origin and three times as a destination. Hong Kong appears seven times and always as a destination, reflecting its role as a marshaling point for exports. This dominance of East Asian container flows is consistent with the statistics reported by Global Insight, quoted in Hayuth (2012), which also observes 'Particularly striking is the fact that in 2010, the volume of trade in the intra-Asia market is four times higher than the volume of trade in the transatlantic'.



**Figure 2:** Distribution of LSCI weights of directed links between ports.

**Table 1:** The 20 links of greatest weight, as determined by an LSCI-like computation

<i>From</i>	<i>To</i>	<i>Weight</i>
Shanghai	Ningbo	1.000
Ningbo	Shanghai	0.987
Hong Kong	Yantian, Shenzhen	0.834
Port Klang	Singapore	0.635
Busan	Shanghai	0.605
Singapore	Hong Kong	0.556
Yantian, Shenzhen	Hong Kong	0.528
Shanghai	Busan	0.523
Hong Kong	Shekou, Shenzhen	0.515
Singapore	Port Klang	0.519
Qingdao	Shanghai	0.505
Ningbo	Hong Kong	0.477
Shanghai	Hong Kong	0.462
Kaohsiung	Hong Kong	0.459
Rotterdam	Hamburg	0.454
Yantian, Shenzhen	Tanjung Pelepas	0.438
Chiwan, Shenzhen	Hong Kong	0.423
Shekou, Shenzhen	Hong Kong	0.421
Qingdao	Ningbo	0.410
Qingdao	Busan	0.408

## Clustering and Communities

A *community* within a network is a collection of vertices with dense and strong connections among themselves but sparser and weaker connections to



other vertices. Barigozzi *et al* (2011) identified communities among countries trading several important commodities. Here we take a more granular look and identify natural trading communities among container ports, as revealed by LSCI-weighted links.

To recognize communities, we rely on an objective function termed *modularity*. The idea is that the modularity  $Q$  of a group of communities  $\{c_i\}$  is large when there is more total weight contributed by edges within the communities than might be expected by chance (Newman, 2004). More formally,

$$Q = \frac{1}{m} \sum_{ij} \left( A_{ij} - \frac{(\sum_k A_{ik})(\sum_k A_{kj})}{m} \right) \delta_{c(i)} \delta_{c(j)},$$

where  $A_{ij}$  has value  $w_{ij}$  if there is a link of weight  $w_{ij}$  from vertex (port)  $i$  to vertex (port)  $j$ ,  $m$  is the sum over all  $A_{ij}$ ,  $\delta$  is the Kronecker delta symbol, and  $c_{(i)}$  is the index of the community to which vertex (port)  $i$  is assigned.

To identify communities in a network one must search over all partitions  $\{c_i\}$  of the vertices to find one that maximizes modularity  $Q$ . We used the heuristic search method of Newman and Girvan (2004), which is known to work well, under which our network resolved into eight communities based on links weighted by LSCI. The results, shown in Figure 3, clearly recognized important global patterns, including trans-Pacific trade (Figure 3a), as well as trans-Atlantic (Figure 3b) and intra-American trade (Figure 3e). Moreover, our communities, like most container services, generally span at most two continents. Furthermore, obvious idiosyncrasies can be explained.

*Figure 3a:* The Asia-Pac and trans-Pacific community is the most strongly-defined in the sense that it includes the ports that contribute most to the total modularity, such as the giants Shanghai, Ningbo, and Hong Kong, and these Asian ports are the anchors of this community. It may seem surprising that this Pacific-spanning community also includes the Caribbean port of Colon, Panama (all the other Panamanian ports are, as would be expected, in the Caribbean community of Figure 3e). But this makes sense because many services from Asia to the US East Coast find it convenient to transship at Colon for subsequent distribution throughout the Caribbean.

*Figure 3b:* Rotterdam and Hamburg are the core ports of the trans-Atlantic community.

*Figure 3c:* The Mideast community is based on trade through the Suez Canal. It includes East Africa above the ports of Tanzania and the Comoros and Seychelles Islands.

*Figure 3d:* The East African ports below Tanzania, including the large ports of South Africa, are better connected to the West African trading community than to others. Tanjung Pelepas is the easternmost member, reflecting its role as point of



**Figure 3:** Trading communities identified by maximizing total LSCI weight of links within groups. (a) Asia-Pac and trans-Pacific; (b) Trans-Atlantic; (c) South Asia and Mideast; (d) West and south Africa; (e) Gulf of Mexico, Caribbean, Pacific South America; (f) Southeastern Latin America; (g) New Zealand; (h) Mediterranean Europe.

distribution of manufactured goods from East Asia to Africa. The few European members are connected primarily through the ports of Tanger or Algeciras.

*Figure 3e:* The Caribbean community includes two outliers inviting comment. Wilmington, Delaware, in the US, has strong ties to Central America because of its specialization in the handling of tropical fruits and fruit juices. On the west coast, San Diego is more strongly connected to Latin America than to East Asia because the Asian services prefer to call at Los Angeles or Long Beach for their larger regional market and superior hinterland storage and transportation infrastructure.





*Figure 3f:* Port-of-Spain (Trinidad) is the northernmost member. Services travel from it into this community.

*Figure 3g:* This community is an artifact of the isolation of New Zealand. It consists of the regional ports of Lyttelton, Napier, Port Chalmers, and Wellington, which have very few direct international connections. They are better connected amongst themselves than to the rest of the world. The international connections to New Zealand call mainly at Auckland and Tauranga, which are members of the Asia-Pac and trans-Pacific trading community.

*Figure 3h:* This is another community determined by geography. These ports are locally connected but all significant connections to the outside world are mainly through a few ports near the Straits of Gibraltar, at the cusp of the Atlantic Ocean and the Mediterranean Sea.

The ports that contribute most to the modularity score of a community are, in a sense, the anchors of those communities. Those of highest modularity are overwhelmingly Asian and especially Chinese, with the top 10 being Shanghai, Ningbo, Hong Kong, Busan, Rotterdam, Yantian, Hamburg, Singapore, Port Klang (Malaysia) and Qingdao. The ports that contribute most within the Trans-Atlantic community are Rotterdam, Hamburg and Savannah; within the South Asia/Mideast community: Port Klang, Jeddah and Dubai; within the West/South Africa: Tanjung Pelepas, Cape Town and Durban; and within the Gulf of Mexico, Caribbean and Pacific South American community: Callao (Peru), Manzanillo (Panama) and Balboa (Panama).

It is worth noting that Singapore is only the sixth largest contributor to modularity in the powerful Asia-Pac and trans-Pacific community. This is because it does not have dense local connections as do the big China ports. Instead, it serves more as a transshipment hub, with services to and from other ports that may not be directly connected themselves. This is reflected in that the *clustering coefficient* of Singapore, which measures how connected to each other are its immediate neighbors (Watts and Strogatz, 1998), is the very lowest among all container ports, followed by other important transshipment hubs such as Port Klang, Algeiras, Kingston and Cartagena. These ports send containers to and receive them from many other ports, but their immediate neighbors do not ship much directly to each other.

Others have searched for natural communities within shipping networks, but it is hard to compare results because of differences in the network models as described above and in what is meant by 'community'. For example, the network of Ducruet and Zaidi (2012) ignored directions of freight flow and they identified communities by successively pruning the network of any node of degree  $k$  or less. The eventual communities are independent of link weights, and therefore did not reflect intensities of trade. Furthermore, they are highly dependent on choice of  $k$ , for which no guidance was given.



Kaluza *et al* (2010) also computed communities based on minimizing modularity, but their network resolved into 12 communities rather than the 8 we found, and with some peculiarities that are hard to explain by trade data (for example, the largest community spanned multiple continents, while some ports in southern California seem to appear in a community based in West Africa.) The authors offer no interpretation of such results, but presumably it is because of the different network (time-aggregated) and different definition of link weight. Their computation was performed in a version of the network in which all weights were identically 1, and so ignored trade intensity. In comparison, the LSCI-based link weight incorporates much more economic information and so provides a richer discriminator of community membership (this is, consistent with the observation of Fan *et al* (2007) that link weights improve the quality of community detection).

Kölzsch and Blasius (2011) also identified communities by modularity minimization, but found that the movement of cargo ships resolved into two main communities, the trans-Pacific and the trans-Atlantic. We speculate that this lack of resolution reflects the fact that their network, by including all types of cargo shipping, had many more links than ours and so bound ports more tightly. In contrast, we found eight distinct communities, and they seem explainable both in aggregate and in detail.

## **A New Index of Strength for Container Ports**

We have defined the weight of a link to be the value of its LSCI; now we use these weights to compute a new index of port connectivity. We compute the CPCI by the 'HITS' algorithm (Hyperlink-Induced Topic Search), which is an eigenvector-based method originally developed to rank Web pages (Kleinberg, 1999) (see the Appendix for details). The HITS algorithm computes two scores for each vertex of a network of directed edges. In the context of container shipping, we refer to these as inbound and outbound scores. Roughly speaking, a port with a high inbound score has greater power to aggregate goods; and a port with a high outbound score has greater power to distribute them.

There are other eigenvector-based measures of centrality that we might have chosen, such as PageRank (Page *et al*, 1998). We prefer the HITS algorithm because it distinguishes between inbound and outbound connectivity, which can reveal something about the role of a port. A port will be assigned a high inbound score if container capacity flows to it from ports with high outbound score, or if it is not too far downstream from such a port. Similarly, a port will be assigned a high outbound score if container capacity flows from it to ports with high inbound score, or if it is not too far upstream from such a port.

**Table 2:** The 20 highest-scoring ports by CPCI (inbound)

<i>Rank</i>	<i>Port</i>	<i>Inbound</i>	<i>Outbound</i>	<i>Country</i>	<i>By volume</i>
1	Hong Kong	0.4080	0.4035	China	3
2	Shanghai	0.3726	0.3475	China	1
3	Ningbo	0.3040	0.3419	China	6
4	Yantian	0.2861	0.2646	China	—
5	Busan	0.2515	0.2415	South Korea	5
6	Singapore	0.2456	0.3420	Singapore	2
7	Kaohsiung	0.2054	0.2009	Taiwan	12
8	Chiwan	0.1973	0.1905	China	—
9	Xiamen	0.1933	0.1841	China	19
10	Shekou	0.1859	0.1614	China	—
11	Port Klang	0.1748	0.1935	Malaysia	13
12	Qingdao	0.1725	0.1593	China	8
13	Nansha	0.1369	0.1109	China	—
14	Tanjung Pelepas	0.1289	0.1134	Malaysia	16
15	Gwangyang	0.1289	0.1182	South Korea	—
16	Los Angeles CA	0.0927	0.0286	USA	17
17	Long Beach CA	0.0917	0.0187	USA	18
18	Xingang/Tianjin	0.0890	0.0802	China	11
19	Da Chan Bay	0.0829	0.0809	China	—
20	Laem Chabang	0.0818	0.0633	Thailand	22

*Note:* Volume rankings are based on the number of TEUs transported through the port in 2010 as reported by the World Shipping Council.

St Vincente, a container port in Cape Verde, is an example. It is unusual in that region to have a relatively high inbound score, which arises because it receives service directly from Algeciras, a regional hub with a relatively high outbound score. That service continues on to Praia, which has a lower inbound score because it is further removed from Algeciras. Similarly, the next few down-service ports have still lower inbound scores.

The CPCI thus combines economics with network topology. Economics is reflected in the weight of the links, which are scored by an adaptation of the LSCI. Network topology is reflected in the recursive ranking of the HITS algorithm. A port scores well under the CPCI if it has strong trade connections; but it also inherits some of the importance of its neighbors, and – with diminishing effect – their neighbors, and so on.

## Ranking Ports by the CPCI

As measured by the CPCI the best-connected ports are not necessarily those with the most links. For example, Cartagena receives services from 20 different ports, which is more than the 18 received by Yantian, but no one will suggest that Cartagena is a more significant container port. The CPCI corresponds to common

**Table 3:** The 20 highest-scoring ports by CPCI (outbound)

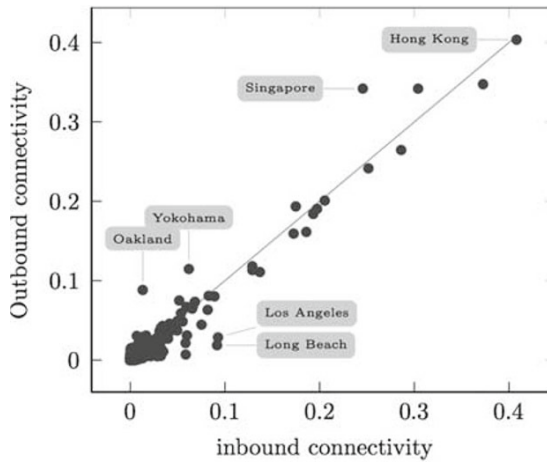
<i>Rank</i>	<i>Port</i>	<i>Inbound</i>	<i>Outbound</i>	<i>Country</i>	<i>By volume</i>
1	Hong Kong	0.4080	0.4035	China	3
2	Shanghai	0.3726	0.3475	China	1
3	Singapore	0.2456	0.3420	Singapore	2
4	Ningbo	0.3040	0.3419	China	6
5	Yantian	0.2861	0.2646	China	—
6	Busan	0.2515	0.2415	South Korea	5
7	Kaohsiung	0.2054	0.2009	Taiwan	12
8	Port Klang	0.1748	0.1935	Malaysia	13
9	Chiwan	0.1973	0.1905	China	—
10	Xiamen	0.1933	0.1841	China	19
11	Shekou	0.1859	0.1614	China	—
12	Qingdao	0.1725	0.1593	China	8
13	Gwangyang	0.1289	0.1182	South Korea	—
14	Yokohama	0.0619	0.1147	Japan	36
15	Tanjung Pelepas	0.1289	0.1134	Malaysia	16
16	Nansha	0.1369	0.1109	China	—
17	Oakland	0.0132	0.0883	USA	—
18	Da Chan Bay	0.0829	0.0809	China	—
19	Xingang/Tianjin	0.0890	0.0802	China	11
20	Cai Mep	0.0517	0.0750	Viet Nam	28

sense by ranking Yantian much higher with regard to inbound (0.286 versus 0.0036).

Similarly, the best-connected ports are not necessarily the busiest. Table 2 shows the CPCI scores of the 20 ports that scored highest with respect to our measure of inbound connectivity (where, for comparison, we have included ranking by TEUs handled in 2010). Similarly, Table 3 gives the highest ranking ports by outbound score. The ports of East Asia dominate with respect to either measure, inbound or outbound. Even though Shanghai handled more TEUs, Hong Kong ranks higher by CPCI, presumably because it is better connected within the global container-shipping network (The ranking by volume combines several of the Shenzhen ports, including Yantian, Chiwan, Shekou and Da Chan Bay, into one, ranked fourth by volume).

On the other hand, our ranking appears to neglect the high-volume European ports such as Rotterdam, Antwerp and Hamburg, as well as the busy Mideast port of Dubai, but this is because they are more isolated from other big ports. In contrast, the big East Asian ports are well-connected with the rest of the world – and with each other, which further increases their scores.

Figure 4 plots scores of all 457 ports. Several stand out for their significant differences between inbound and outbound scores, and these differences illustrate how the CPCI can make structural distinctions about the position of a port in the network.



**Figure 4:** Inbound and outbound CPCI scores of the world's major container ports.

Los Angeles and Long Beach have inbound scores that are relatively high in comparison with outbound scores. This reflects the fact that these are the two main ports of entry for product manufactured in East Asia. To reduce in-transit inventory, powerful retailers in North America insist that their freight be the last loaded out of Asia and the first unloaded in North America, and so there are many direct links from big Asian ports into Los Angeles and Long Beach. Services that have traversed the Pacific Ocean to call at Los Angeles or Long Beach then typically call at Oakland, a big exporter of agricultural products, before returning to the large ports of Asia. Consequently, Oakland has a high outbound score in comparison with its inbound score. This is a general pattern that may be observed along many service loops: Ports that are immediately downstream from important ports tend to have higher inbound scores, while ports toward the end of the loop tend to have higher outbound scores.

## North American Ports

Table 4 shows that, among the ports of North America, the west coast ones, led by Los Angeles and Long Beach, dominate by the measure of inbound connectivity, reflecting the many services that come directly from the great manufacturing centers of East Asia. Moreover, many of the west coast ports score much higher with respect to inbound connectivity than to outbound.

New York is the only port on the east coast to score highly with respect to inbound scores, presumably because of the great population density of

**Table 4:** The top 10 North American ports by CPCI (inbound)

<i>Rank</i>	<i>Port</i>	<i>Inbound</i>	<i>Outbound</i>
1	Los Angeles CA	0.0927	0.0286
2	Long Beach CA	0.0917	0.0187
3	New York NY/NJ	0.0586	0.0070
4	Lazaro Cardenas, MEX	0.0349	0.0108
5	Manzanillo, MEX	0.0325	0.0406
6	Tacoma WA	0.0305	0.0084
7	Prince Rupert BC	0.0285	0.0048
8	Vancouver BC	0.0260	0.0249
9	Seattle WA	0.0231	0.0207
10	Oakland CA	0.0132	0.0883

**Table 5:** The top 10 North American ports ranked by CPCI (outbound)

<i>Rank</i>	<i>Port</i>	<i>Inbound</i>	<i>Outbound</i>
1	Oakland CA	0.0132	0.0883
2	Manzanillo, MEX	0.0325	0.0406
3	Savannah GA	0.0122	0.0289
4	Los Angeles CA	0.0927	0.0286
5	Vancouver BC	0.0260	0.0249
6	Seattle WA	0.0231	0.0207
7	Long Beach CA	0.0917	0.0187
8	Boston MA	0.0007	0.0150
9	Wilmington NC	0.0008	0.0133
10	Miami FL	0.0030	0.0125

**Table 6:** Ranks among those countries represented by a single port

<i>Port</i>	<i>by LSCI</i>	<i>by CPCI (inbound)</i>	<i>by CPCI (outbound)</i>
<i>Singapore</i>	1	1	1
<i>Rotterdam</i>	2	2	2
<i>Colombo</i>	3	3	3
<i>Malta</i>	4	6	11
<i>Beirut</i>	5	4	10
<i>Piraeus</i>	6	7	6
<i>Buenos Aires</i>	7	18	26
<i>Karachi</i>	8	8	5
<i>Gothenburg</i>	9	35	20
<i>Gdansk</i>	10	15	57

the region. But Table 5 shows that east coast ports such as Savannah are more competitive with respect to outbound scores.

It will be interesting to see how these rankings change after the widening of the Panama Canal is complete.



## Comparison of the CPCI with the LSCI

The LSCI is defined for countries, while the CPCI is defined for ports. Nevertheless, we can directly compare the rankings of those countries with a single dominant port. We identified 64 container ports that were, within our data set, unique within their country, and then compared rankings by the 2011 LSCI and by each of the inbound and outbound versions of the CPCI. The results appear in Table 6 and are generally consonant in that those ranked among the top 10 by LSCI are among the top 20 by CPCI, either inbound or outbound.

The differences in ranking between Gothenburg and Gdansk again illustrate how our suggested index captures the structure of the network. Gdansk ranks relatively high in inbound strength because it receives shipments directly from Hamburg but ships only to the lesser port of Aarhus, which accounts for its relatively lower ranking in outbound strength. On the other hand, Gothenburg receives freight only from Aarhus, but it ships to the more significant port of Bremerhaven, from which it derives a higher outbound score.

## Comparison with Other Measures of Centrality

There are many measures of centrality in a network and each reflects something different about the network. One measure of the centrality of a vertex within a network is *degree centrality*, which in our context tells from how many other ports a port receives direct shipments (in-degree) or to how many others it sends direct shipments (out-degree). This measure neglects economic issues, as the volume of trade along each link and records merely the fact of trade.

Some measures of centrality incorporate distance. For example, Doshi *et al* (2012) use geodesic distance to compute *closeness* and *betweenness* for ports in their network (*Closeness* is the reciprocal of the sum of shortest distances from a port to all other ports; and *betweenness* is the number of times a port appears on the shortest path between two other ports). However, such measures ignore actual trade. By distance-based measures a port like Cartagena is centrally located in the global network, yet it is not at all central to global container movement.

Kölzsch and Blasius (2011) computed distance-based centrality, using an imaginative notion of distance meant to address the likelihood of transmission of invasive organisms. They defined distance to be the reciprocal of the cumulative shipping capacity along a link of their network. Presumably this captures the intuition that a link with more cargo capacity is more likely to transmit organisms, but the model is not made explicit.

Our model focuses on the flow of containers and assumes that shipping routes are based on maximizing profitability for the shipping company.



Therefore, connections reflect not just distance but, more importantly, how many containers are shipped, from where, and to where. Moreover, freight on board the vessel represents in-transit inventory and any delay is a cost to the owner of the inventory. One expects the structure of the network to reflect this, but this is missed, or at least distorted, by measures of centrality based purely on topology, such as degree or betweenness, or on distance, such as closeness centrality.

Like us, Doshi *et al* (2012) and Kölzsch and Blasius (2011) invoke a measure of centrality that is more than local. They each compute eigenvector centrality, which is one of several possible measures in which the strength of a port is higher if it is connected to ports that are themselves strong. Neither explains their choice. Doshi *et al* (2012) reach no particular conclusions other than to note that the high scores of Asian ports correlate with the economic growth of Asia. Kölzsch and Blasius (2011) observed that eigenvector centrality named the most central ports in the world to be in the Gulf of Mexico. This disagrees notably with their results for other measures of centrality, such as closeness and betweenness, which named more familiar ports in East Asia as most central.

## Conclusions

The CPCI is based on both economics and network theory and so, we believe, provides a better measure of trade-connectivity than alternative measures. Results based on a network model, such as communities or link rankings or port rankings, should be generally consonant with results from models that are not network based, such as TEUs handled per year or the LSCI for countries with a single dominant port. Furthermore, notable disagreements should be explainable. We believe our model passes these tests. Where our measure differs significantly from others, we generally find that it is because our measure reflects additional information neglected by other measures.

The CPCI is based on link weights that are computed just like the LSCI; and because the LSCI has been vetted by economists as capturing intensities of trade, our index inherits that descriptive power but exercises it at a more granular level. It summarizes something about how each port is connected to others within the larger network. Importantly, this expresses more than local connectivity to immediate neighbors but also to neighbors-of-neighbors, and so on. This is important because containers do not move only from one port to the next neighbor down-service, but, more generally, they move along paths. Furthermore, this allows inbound and outbound strengths to be studied separately, and this gives a more detailed look at the economic roles played by each port.





Any index of logistics performance is an attempt to summarize a complex environment. The LSCI may be criticized for the somewhat arbitrary way that data is agglomerated; and the LPI for its reliance on perception rather than measurement. The CPCI has weaknesses as well. In particular, because it uses an LSCI-like computation, it inherits any criticism of that. The LSCI represents shipping capacity but ideally one would like to assign weights to the links in some way that reflects the actual number of TEUs transported, rather than TEU capacity. Unfortunately, data at this level of detail is not generally available.

We expect the CPCI to be useful in some of the same ways as the LSCI. This may include explaining how the container-shipping network changes over time, or using the edge weights and port scores as explanatory variables for economic phenomena. We believe these finer-grained statistics will be easier to understand and to explain because they directly reflect immediate decisions of primary actors such as shipping companies.

It should be remarked that none of the network models discussed herein captures anything about transshipment. Even though there may be direct links from port A to port B and from port B to port C, to transport a container from A to C may require transshipment. In this case ports A and C are further apart in both time and cost than they might appear in these models. Unfortunately there is insufficient information available to piece together a useful view; but if that information were available, it could be incorporated into a model that explicitly represents the structure of scheduled liner services, along the lines of Ducruet and Notteboom (2012), albeit with directed links.

## References

- Arvis, J.-F., Mustra, M.A., Panzer, J., Ojala, L. and Naula, T. (2007) Connecting to Compete: Trade Logistics in the Global Economy. Technical Report, The World Bank.
- Barigozzi, M., Gagiolo, B. and Mangioni, G. (2011) Identifying the community structure of the international-trade multi-network. *Journal of Physics A* 390(11): 2051–2066.
- Doshi, D., Malhotra, B., Bressan, S. and Lam, J. (2012) Mining maritime schedules for analyzing global shipping networks. *International Journal of Business Intelligence and Data Mining* 7(3): 186–202.
- Ducruet, C., Lee, S.W. and Ng, A.K.Y. (2010) Centrality and vulnerability in liner shipping networks: Revisiting the Northeast Asian port hierarchy. *Martin Policy & Management* 37(1): 17–36.
- Ducruet, C. and Notteboom, T. (2012) The worldwide maritime network of container shipping: Spatial structure and regional dynamics. *Global Networks* 12(3): 395–423.
- Ducruet, C. and Zaidi, F. (2012) Maritime constellations: A complex network approach to shipping and ports. *Maritime Policy & Management* 39(2): 151–168.
- Fan, Y., Li, M., Zhang, P., Wu, J. and Di, Z. (2007) The effect of weight on community structure of networks. *Journal of Physics A* 378(2): 583–590.
- Hayuth, Y. (2012) Shifting World Trade Patterns: The Intra-Asia Case. Technical Report, Wydra Shipping and Aviation Institute, University of Haifa, Israel.
- Hoffman, J. (2005) *Liner shipping connectivity*. Transport Newsletter of the United Nations Conference on Trade and Development, Issue 27: 4–12.

- Hoffman, J., van Hoogenhuizen, J.W. and Wilmsmeier, G. (2014) *Developing an Index for Bilateral Liner Shipping Connectivity*. Technical Report, presented at the 2014 Conference of the International Association of Maritime Economists, Norfolk, VA.
- Kaluza, P., Kölzsch, A., Gastner, M.T. and Blasius, B. (2010) The complex network of global cargo ship movement. *Journal of Royal Society Interface* 7(48): 1093–1103.
- Kleinberg, J.M. (1999) Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery* 46(5): 604–632.
- Kölzsch, A. and Blasius, B. (2011) Indications of marine bioinvasion from network theory. *The European Physical Journal B* 84(4): 601–612.
- Newman, M.E.J. (2004) Analysis of weighted networks. *Physical Review E* 70(5): 1–9.
- Newman, M.E.J. and Girvan, M. (2004) Finding and evaluating community structures in networks. *Physical Review E* 69(2): 1–16.
- Ojala, L. and Hoffmann, J. (2010) *A comparison of the LPI and the LSCI*. Transport Newsletter of the United Nations Conference on Trade and Development, Issue 46: 7–8.
- Page, L., Brin, S., R. Motwani, R. and Winograd, T. (1998) *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report, Stanford University, CA.
- Perra, N. and Fortunato, S. (2008) Spectral centrality measures in complex networks. *Physical Review E* 78(3): 1–11.
- United Nations Conference on Trade and Development. (2014) Liner shipping connectivity. *Review of Maritime Transport* pp. 110–114.
- Watts, D.J. and Strogatz, S. (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684): 440–442.
- Wilmsmeier, G. and Hoffmann, J. (2008) Liner shipping connectivity and port infrastructure as determinants of freight rates in the Caribbean. *Maritime Economics & Logistics* 10(1–2): 131–151.
- World Shipping Council. Top 50 world container ports, <http://www.worldshipping.org/about-the-industry/global-trade/top-50-world-container-ports>, accessed 4 December 2014.

## Appendix

### The HITS algorithm

The HITS algorithm was originally developed to rank the Web pages for a search engine. It computes two scores for each Web page, a hub score and an authority score, where a good authority page is a page with many incoming links, while a good hub page is a page with many outgoing links. The idea of the HITS algorithm is that any page that is cited by important hub pages should be considered an authority. Similarly, any page that cites important authority pages should be considered a hub.

In the context of container shipping, we interpret an authority as a port that receives shipments from many ports and so it is good at aggregating shipments, and so worthy of a high score for inbound traffic. Similarly, a hub is for us a port that sends shipments to many other ports, and so it is good at distribution, which results in a high score for outbound traffic.



More generally, the HITS algorithm can be exercised on any directed network. Let  $E$  be the set of directed edges of a network and  $\lambda$  a constant. Then the authority and hub scores of vertex are the solutions  $x_i$  and  $y_i$  to

$$\lambda x_i = \sum_{j:e_{ji} \in E} y_j$$

$$\lambda y_i = \sum_{j:e_{ij} \in E} x_j$$

If  $A$  is the adjacency matrix of the network, the equations for vertices  $i = 1, \dots, n$  can be written in matrix form as

$$\lambda x = A^T y \Rightarrow \lambda^2 x = A^T A x$$

$$\lambda y = A^T x \Rightarrow \lambda^2 y = A A^T y$$

Each of the above systems of equations is equivalent to the problem of finding eigenpairs satisfying constraints defined by the system of equations itself, and the importance scores are the principal eigenvectors corresponding to each of the system of equations. Such measures of centrality are known as spectral centrality measures (Perra and Fortunato, 2008).



This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/>

© The Authors(s) 2016