WPS5722

Policy Research Working Paper 5722

The Air Connectivity Index

Measuring Integration in the Global Air Transport Network

Jean-François Arvis
Ben Shepherd

The World Bank
Poverty Reduction and Economic Management Network
International Trade Department
June 2011

Abstract

The authors construct a new measure of connectivity in the global air transport network, covering 211 countries and territories for the year 2007. It is grounded in network analysis methods, and is based on a gravity-like model that is familiar from the international trade and regional science literatures. It is a global measure of connectivity, in the sense that it captures the full range of interactions among all network nodes, even when there is no direct flight connection between them. The best connected countries are the United States, Canada, and Germany; the United States' score is more than two-thirds higher than the next placed country's, and connectivity overall follows a power law distribution that is fully consistent with the hub-and-spoke nature of the global air transport network. The measure of connectivity is closely correlated with important economic variables, such as the degree of liberalization of air transport markets, and the extent of participation in international production networks. It provides a strong basis for future research in areas such as air and maritime transport, as well as international trade.

This paper is a product of the International Trade Department, Poverty Reduction and Economic Management Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at jarvis1@worldbank.org, and ben@developing-trade.com.

The Air Connectivity Index:

Measuring Integration in the Global Air Transport Network

Jean-François Arvis[1] and Ben Shepherd[2,3]

Keywords: Air transport; Network industries; Connectivity; International trade; Gravity model; Spatial interaction models.

JEL codes: L93; F15.

[1] Senior Economist, World Bank. Contact: jarvis1@worldbank.org.

[2] Principal, Developing Trade Consultants Ltd. Contact: Ben@Developing-Trade.Com.

[3] The authors are grateful to the following for many helpful discussions, and for sharing data and product classifications: Mitsuyo Ando, Henry Bofinger, Olivier Cadot, the late Michel Iches, Fukunari Kimura, Charles Kunaka, Pierre Latrille, Aaditya Mattoo, Ayako Obashi, Marcelo Olarreaga, and Charles Schlumberger. We also received useful suggestions from participants at a World Bank seminar. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not represent the view of the World Bank, its Executive Directors, or the countries they represent.

1 Introduction

Air transport is increasingly important to the global economy. Be it for exports of perishable goods, such as Kenyan cut flowers, or the movement of highly specialized component parts within transnational production networks, the ability to move goods quickly, reliably, and at a reasonable price is a crucial ingredient in the trade performance of a wide range of countries. The economic importance of air transport was highlighted by the eruption of Iceland's Eyjafjoell volcano in April 2009, which severely disrupted air traffic—including freight—into and out of Europe. Press reports indicate that Kenyan farmers were forced to dump stocks of fresh food and flowers destined for European markets, at a cost of some $3.8m per day.[4]

Our aim in this paper is to analyze one aspect of the global air transport system that has received relatively little attention from economists and policy experts to date: its character as a network industry, and thus the importance of *connectivity* as an overall measure of the level of service—frequency of flights, reliability, and diversity of destinations—available to end users. Connectivity is increasingly emerging as a key policy concept, particularly in international forums such as ASEAN and the Asia-Pacific Economic Cooperation, yet it currently lacks a rigorous empirical framework that would allow policymakers to undertake cross-country comparisons, or track progress through time. Our approach aims to help fill that void, using air transport as an example.

This paper builds on and extends two recent efforts at measuring connectivity in economically important sectors. Whereas UNCTAD (2007) adopts an intuitive but ad hoc approach to developing a connectivity indicator for liner shipping based on a weighted average of capacity and utilization data, we embed our approach in a rigorous network analysis framework. In addition to being more consistent, our approach also has the advantage of taking account of the hub-and-spoke nature of the global air transport

[4] http://news.bbc.co.uk/2/hi/8629623.stm.

network in a way that the country aggregates used by UNCTAD (2007) do not. Pearce (2007) also takes an intuitive approach to analyzing air transport connectivity, and his measure better captures the hub-and-spoke nature of the network. However, our approach is more rigorous in the sense that it is embedded in a consistent model of the global air transport network. We also apply our method to a much larger dataset: over 200 countries and territories, as compared to the 47 considered by Pearce (2007).

To capture the various dimensions of performance in a single numerical indicator—the *Air Connectivity Index (ACI)*—we define connectivity as the importance of a country[5] as a node within the global air transport system. A country is considered to be better connected the stronger is the overall "pull" it exerts on the rest of the network. A country's connectivity score is higher if the cost of moving to other countries in the network is relatively low. It is considered to be less well connected if the dispersion of those costs is high. These two factors are both important in the context of a network industry. Combining them means that, for example, a spoke country with a strong connection to only one other country cannot have a high connectivity score: although the cost of moving from one country to the other is low, the dispersion of costs across all markets is very high. A regional hub with strong connections to a moderate number of destinations receives an intermediate connectivity score, since costs are relatively dispersed across the remainder of the network, although to a lesser extent than for a pure spoke country. Finally, a global hub with strong connections to many other countries in the network receives a relatively high connectivity score.

This approach to defining and measuring connectivity flows naturally from a generalized gravity model framework familiar from the regional science literature, since the pioneering work of A.G. Wilson (Wilson, 1970; Roy, 2004) which proposed a general bilinear formulation of bilateral flows. It also nests a

[5] Given the breadth of the dataset we use for the empirical analysis in this paper, the term "country" is used loosely to refer to both countries in the strict legal sense and other territories with independent air services.

variety of recent models from the international trade literature that follow a similar path to Wilson's, including the "gravity with gravitas" model of Anderson and Van Wincoop (2003), the Ricardian model of Eaton and Kortum (2002), and the heterogeneous firms model of Chaney (2008). We show that origin and destination fixed effects from a simple gravity regression can be rescaled to produce a measure of connectivity that corresponds to the total pull exercised by each country on the rest of the network. This measure captures the level of costs on bilateral routes, as well as the dispersion of costs across routes. It takes account of all possible nodal connections, even those for which no direct flights are in fact observed. As such, it provides a highly tractable yet informative indicator of air transport connectivity. It has the added advantage of being closely related to the concepts of multilateral resistance and market potential, which are familiar from the international trade and economic geography literatures (Anderson and Van Wincoop, 2003; Head and Mayer, 2011); indeed, our analysis makes it possible to link such measures to the deep structure of networks with bilateral flows, such as air transport or trade.

Although we focus on air transport in this paper, the methods we develop can easily be applied more broadly, in areas such as maritime transport, and even international production networks for manufactured goods. The concept of connectivity that we develop measures "before-the-border" effects, whether endogenous (e.g., transport policies), or exogenous (e.g., geography, and the hub-and-spoke structure of the network). In addition to explaining how countries effectively connect to international freight transport and logistics networks, it complements existing indicators of trade and transport facilitation outcomes ("beyond-the-border" metrics such as the *Logistics Performance Index* or the Doing Business *Trading Across Borders* data), which are primarily endogenous, and not tied to the geography of the network (Arvis et al., 2007 and 2010).

We implement our model empirically to produce an internationally comparable ACI for the year 2007 as an input into future research work, and a useful policy benchmarking tool. We show that, as expected, the ACI is strongly correlated with the degree of liberalization in air services markets, which suggests that policy can play an important role in shaping connectivity. In addition, we find that better connected countries tend to be more specialized in trade in machinery parts and components, which is consistent with their being more deeply integrated into international production networks that rely heavily on air transport.

The paper proceeds as follows. The next section reviews existing work on measuring connectivity in air transport and related contexts. We draw on the policy literature, as well as contributions from the applied mathematics and network analysis fields. Section 3 discusses some important features of the international air transport network that need to be kept in mind when analyzing connectivity. Section 4 presents our measure of connectivity, and derives it from a general gravity model framework. Section 5 presents empirical results, and discusses robustness issues. In Section 6, we show that the ACI is closely correlated with indicators that we would expect to be related to connectivity, particularly the degree of liberalization in air transport markets, and the percentage of exports traded within international production networks. Section 7 concludes with a discussion of policy implications, and possible directions for future research.

2 Previous Attempts at Measuring Connectivity

The previous literature does not disclose any comprehensive attempt to measure air transport connectivity at the national level using rigorous network analysis methods. However, a number of important contributions deal with related or complementary areas.

The most closely related work to ours is by Pearce (2007). He defines connectivity as summarizing the scope of access between an individual airport or country and the global air transport network. On that basis, his connectivity indicator for each international airport combines information on the number of destinations served, the frequency of service, the number of seats per flight, and the size of the destination airport. Using a dataset similar to the one we use here (see further below), Pearce (2007) calculates connectivity measures for airports in 47 countries. He finds suggestive evidence of a relationship between connectivity thus defined, and important economic outcome measures such as labor productivity, and competitiveness of the travel and tourism sector.

In addition to the connectivity index developed by Pearce (2007), the applied mathematics and physics literature discloses a number of examples of the application of network methods to the air transport context, although they do not focus specifically on connectivity. For example, Guimerà et al. (2008) analyze the network structure of global air transport at the city level, focusing on the description of its mathematical properties and the identification of separate communities—an abstract analogue of hub-and-spoke systems—within the global network. There is also a number of applications of similar techniques to national or regional air traffic: Bagler (2008) examines India, and Li and Cai (2003) analyze China.

Another strand of the existing literature on air transport focuses on the impact of bilateral liberalization (Gönenç and Nicoletti, 2000; Doove et al., 2001; Intervistas, 2007; and Geloso Grosso and Shepherd, 2011). The general approach of these papers is to measure the restrictiveness of policy settings in a group of countries, and then to relate these policies to observed outcomes, such as bilateral traffic levels or international trade in goods. Although they thus provide important insight into the impacts of policy on sector performance, they do not explicitly consider the role that network structure and connectivity might simultaneously play.

Outside the air transport context, UNCTAD is continuing to develop a *Liner Shipping Connectivity Index* (UNCTAD, 2007; Hoffmann and Wilmsmeier, 2008). They define connectivity in terms of access to regular and frequent transport services, then use factor analysis to bring together data on capacity and utilization in the liner shipping sector. In particular, they consider the number of carriers, the number of bilateral links between national ports, total vessel capacity and number, the number of direct services, and the size of the largest vessel on each route. Their methodology produces a connectivity index that is a weighted average of those data, with the weights determined statistically. Hoffman and Wilmsmeier (2008) show that connectivity defined in this way is an important determinant of observed freight rates in the Caribbean.

Our approach differs from previous contributions in three main ways. First, we extend the connectivity work of Pearce (2007) and UNCTAD (2007) by embedding it in a systematic model. This approach avoids the charge that the resulting measure of connectivity is ad hoc. Second, we are interested in using connectivity as a policy tool, rather than simply a means of describing network properties, as in the applied mathematics literature. Third, we focus on the country as the level of analysis—again, this is important from a policy point of view, since many of the economic outcome variables that we would expect to be correlated with air connectivity are measured at the national, not airport, level. Trade data are an important example.

3 Characteristics of the Global Air Transport Network

Before describing our measure of air connectivity and its derivation, it is useful to consider some salient characteristics of the network we are analyzing. To do this, we use data from the *Schedules Reference Service (SRS) Analyzer* platform (www.srsanalyser.com). SRS is the leading industry database of flight information, covering passenger and cargo links among over 200 countries and territories. We use data aggregated to the country level, but SRS can also provide much more detailed information on links between individual airports; Pearce (2007) uses SRS data at the airport level, for example.

Although SRS represents the best currently available data source on international flight connections, it suffers from one major limitation: it provides information on bilateral capacity, not true origin-destination traffic. For example, SRS reports that there are 74 flights per week from the United States to Australia. However, it does not allow us to identify the proportion of passengers on those flights who actually stay in Australia, and those who use Australia as a transit point only.[6] Although IATA compiles data on passenger numbers based on true origin and true destination, we prefer SRS as a data source because it is more comprehensive in terms of country coverage. It also includes cargo traffic, which is crucial for present purposes because of our interest in examining the links between air connectivity and trade. Indeed, the lack of true origin-destination flows is less problematic than it at first appears in light of the fact that for policy purposes, we are primarily interested in relating connectivity to cargo, not passenger, flows.

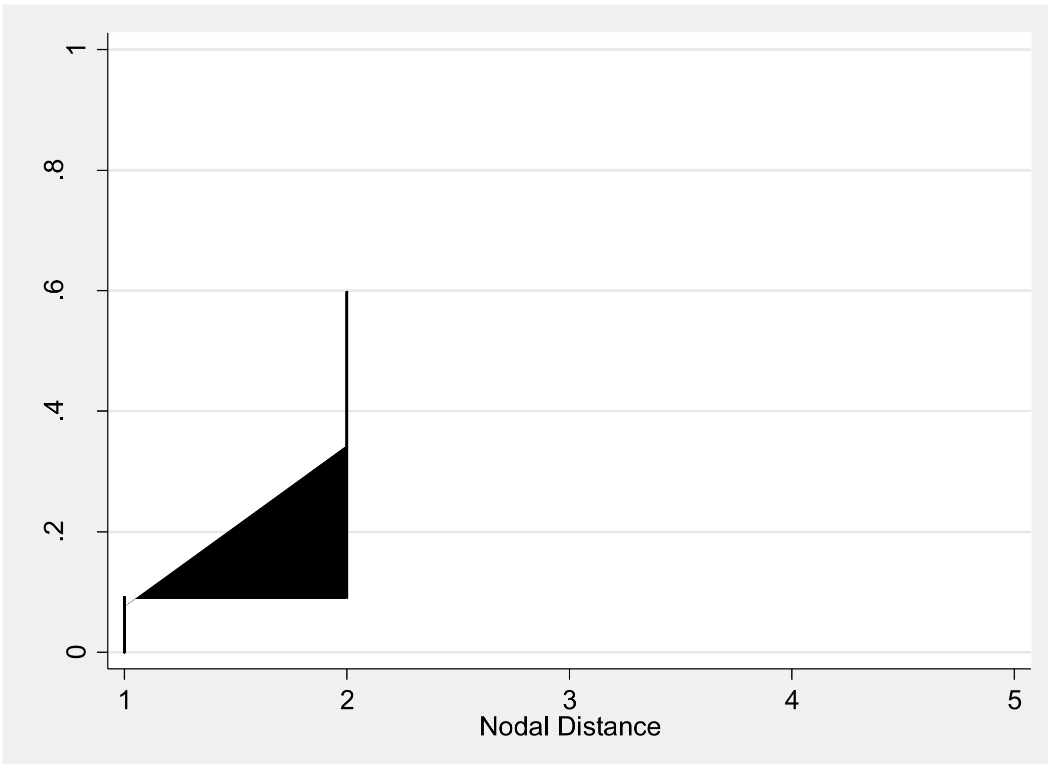
Previous work has shown that the international air transport network has many of the characteristics of a "small world network" in the terminology of the applied mathematics literature (e.g., Guimerà et al., 2008). Typically, small world networks have a relatively short average distance between nodes. Due to the presence of hubs which facilitate short connections between nodes, the number of direct connections between nodes tends to follow a power law distribution.

We find both of these properties in the SRS data. The average distance between nodes is 2.3, and the mode is 2. Over 80% of nodes can be reached from any other node in three steps or less (Figure 1), which is consistent with a general rule of thumb in the airline industry. (A three step journey corresponds to two intermediate stopovers between the origin and the destination.) Moreover, a

[6] There is an obvious analogy here with the trade literature, in which re-exports and re-imports pose frequent measurement problems in highly open economies.

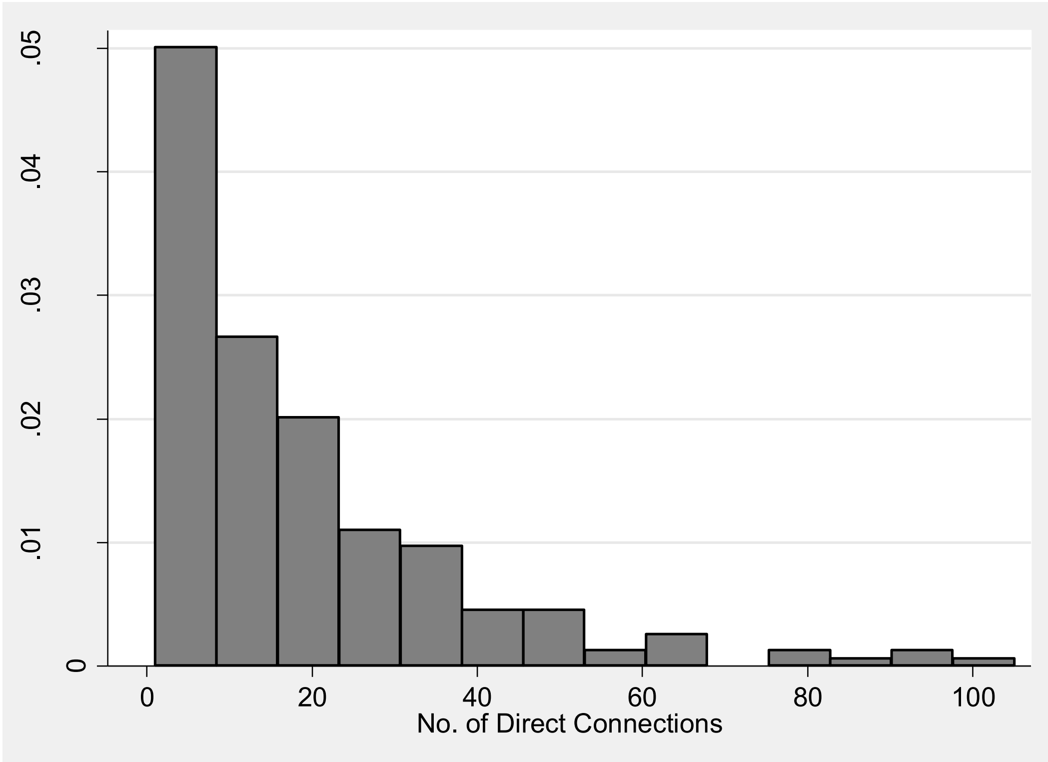
histogram of the number of direct connections of each node in the network is clearly suggestive of a power law (Figure 2), an observation which is confirmed by a simple regression.[7]

Figure 1: Cumulative distribution of the number of steps required for each origin-destination link, 2005.



[7] Following Gabaix and Ibragimov (2011), we regress log(rank-0.5) on a constant and log(no. of direct connections). The coefficient is -0.82, and it is 1% significant based on robust standard errors. However, the hypothesis that the coefficient is equal to unity—i.e., that the data follow a Zipf law—is rejected at the 1% level. Gabaix and Ibragimov (2011) show that the regression we use has superior properties to the more common one in which log(rank) is the dependent variable.

Figure 2: Histogram of the number of direct connections from each node, 2005.



These findings suggest that it is extremely important to account for hub and spoke interactions when measuring connectivity in the air transport network. Of course, this accords closely with thinking in the industry.[8] In the next section, we develop a measure of connectivity that does just that.

Graphical software developed for network analysis is also a popular tool for grasping the structure of a network, including hub and spoke structure and the centrality of nodes. However, any two-dimensional graphical illustration of the air transport network must be arbitrary, at least in part. Figure 3—prepared using the Pajek[9] freeware program—uses the air transport adjacency matrix to plot the most central nodes in the middle of the figure, based on closeness centrality (see Appendix 3 for a discussion of that concept). Nodes with less than 20 connections have been dropped. Unsurprisingly, this approach

[8] It is outside the scope of this paper to discuss the factors that have given rise to the emergence of a hub-and-spoke system in the air transport market. See, for example, Dempsey (1990) for a discussion of the role of policy and deregulation.

[9] http://pajek.imfm.si/doku.php

provides a rather Western-centric view of the world, in which the most central countries are the USA and European hubs, with other regions revolving around them.

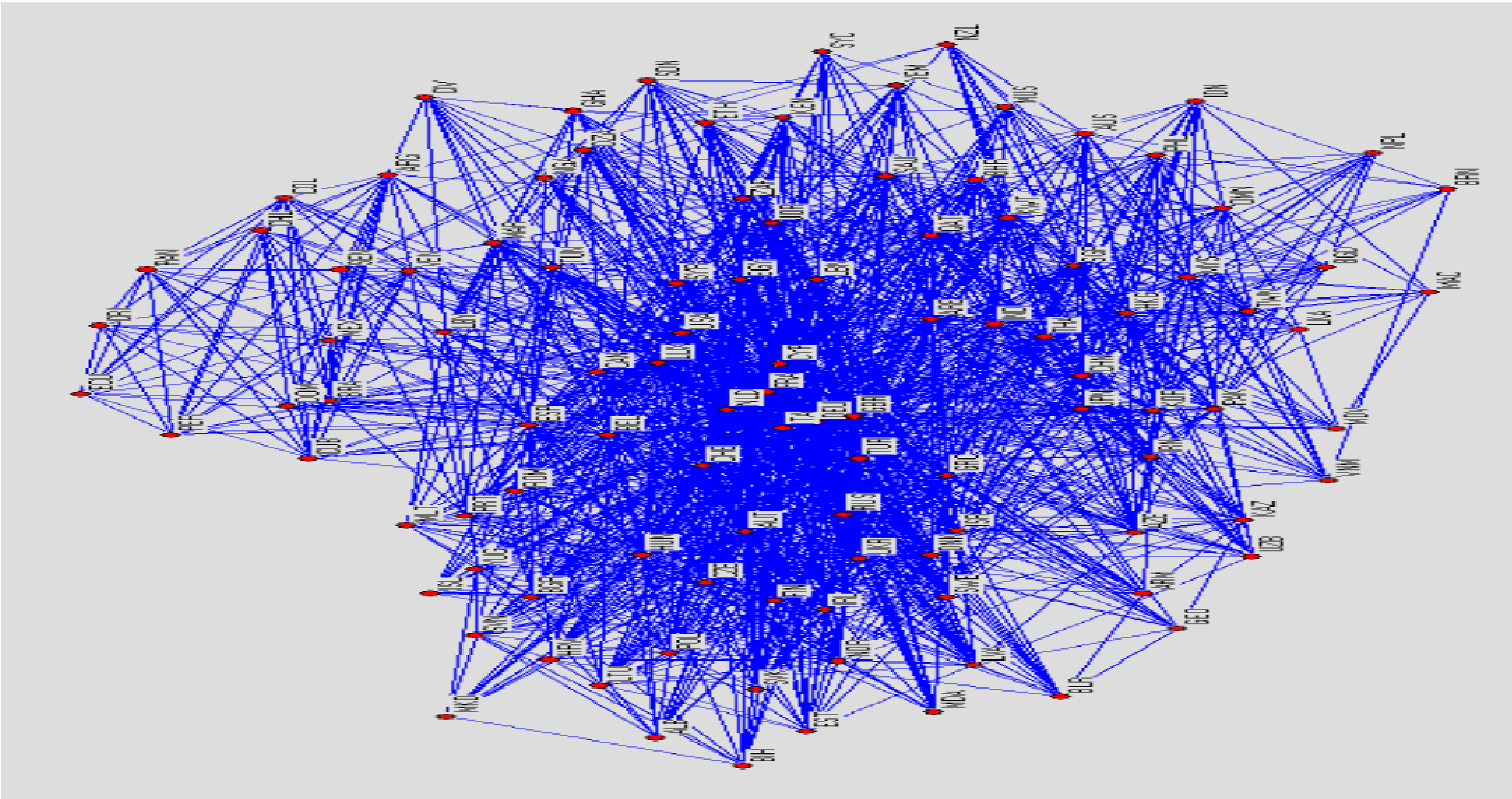
4 The *Air Connectivity Index* (ACI): Theoretical Basis and Definition

Connectivity is an intuitive concept, but one without a generally agreed definition. We envisage it as an attribute of individual nodes in a network. Any measure of connectivity should provide a consistent and robust definition of how well the node is connected to the rest of the network. In the case of air transport, we are interested in assessing how well connected each country (node) is to the rest of the world. Intuitively a good definition of connectivity should have the following properties:

1. It should be *realistic*, in the sense of being supported by an underlying model of transportation flow within the network, where shipments move depending on the size of each node, and the attributes of individual links and nodes. Depending on its degree of realism, the model should take account of the costs associated with moving between nodes on a given link, as well as the finiteness of trips. Ideally, the model should be linked with mainstream models in transportation economics, such as the gravity-type model.[10]
2. It should be *intensive*, in the sense of being independent of the size of a node. Ideally, two nodes with the same connections to the rest of the world should have the same connectivity even if they have different size. For this reason, measures such as the total number of passengers or amount of traffic do not themselves qualify as indicators of connectivity. This is not to say, however, that size does not matter at all. Indeed, it is widely expected, and frequently observed in practice, that larger nodes are also better connected, but this should not come from the definition of connectivity itself.

[10] In fact, one of the earliest uses of a gravity-type model includes air transport: see Zipf (1946).

Figure 3: Graphical representation of the global air transport network, based on unweighted closeness centrality (nodes with less than 20 connections excluded).



3. It should be *dimensionless* and *normalized*. Connectivity should be a number without dimension, preferably with bounds (a finite or semi-finite predefined interval such as 0-1). The motivation for non-dimensionality comes from the fact that a good definition should be irrespective of the objects described, and should be based on normalized flows (e.g., making flows through the transport network sum to one).This requirement is a stronger one than being intensive. For instance, the number of passengers per unit of GDP is intensive but not dimensionless, while the number of links from a node is neither. The share of total flows is both dimensionless and intensive.
4. An ideal definition of connectivity should be *global*, in the sense that it directly or recursively incorporates information on the full network, and not only the properties of a given node and its immediate neighborhood (i.e., those other nodes to which it is directly connected). Thus the connectivity of node i should depend on the connectivity of the full set of nodes to which it is connected.

There are essentially four groups of connectivity measures applied so far to transport and economic problems: intuitive metrics; concentration indicators; clustering techniques; and centrality indices such as closeness centrality or PageRank. (We review their implementation in detail in Appendix 3.) The last two sets of indicators are rooted in recent developments in network theory. Indeed, network theory has been a very active field of knowledge over the last two decades, at the juncture of statistical physics and the social sciences. Although a wide variety of tools and models is now available, we did not find an "off-the-shelf" framework that could be directly applied to describing the problems posed by spatial economic interaction networks, such as air transportation or shipping. We therefore develop our own model in this section. The value added of the solution proposed here is that it combines some of the previous ideas in a spatial interaction framework, and makes connectivity a natural metric in gravity modeling.

4.1 A Minimalist Gravity Model

In the problem of interest, nodes are not just abstract individuals or entities with a zero-one linkage between them. They are interacting spatial objects (countries, airports, ports) with differentiated levels of activity and measurable costs of interaction between them, such as the time it takes to fly from one point to another in the case of air transport. The essence of the gravity interpretation of spatial interaction models, such as those in the trade literature, is a bi-proportional structure. On the one hand, the flow between the origin and the destination is proportional to the size or potential of the origin and destination. It is also inversely proportional to an "impedance" between them, which is independent of their potentials and incorporates information about bilateral interactions between the nodes, such as the distance, time, and cost involved in moving from one node to another.

The generic bi-proportional gravity model takes the following form:

$$(1)\ X_{ij} = A_i B_j K_{ij}$$

where A_i is the repulsive potential of node i, and B_j is the attractive potential of node j; the flow X_{ij} is "pushed" from i and "pulled" to j. The bilateral impedance K_{ij} is exogenous and decreasing in the cost, distance, or disutility of movement or interaction between origin and destination.[11] Depending on the model, the impedance may be exponential or a power of trade costs. For instance, the entropy based model of spatial interaction (Wilson, 1967, 1970; Roy, 2004) refers to an exponential law, i.e. $K_{ij} = \exp(-\beta c_{ij})$ where $c_{ij} \geq 0$ is the bilateral cost of transportation. In contrast, neo-classical models of trade based on CES preferences tend to yield a power dependence on trade costs (e.g., Anderson, 1979;

[11] This is a very general form of gravity, but it can easily be adapted to meet specific theoretical assumptions from the trade literature. Setting the potential terms equal to exporter and importer GDP adjusted by appropriately defined multilateral resistance gives the "gravity with gravitas" model of Anderson and Van Wincoop (2003). The Ricardian gravity model of Eaton and Kortum (2002) also takes this basic form. If bilateral impedance is assumed to include fixed and variable cost elements, then the heterogeneous firms model of Chaney (2008) also fits this general pattern. In each case, of course, the economic interpretation of the trade cost exponent is different. See Head and Mayer (2011) for a similar, though less general, approach.

Anderson and Van Wincoop, 2003). In this formulation, $K_{ij} \cong \left(1 + c_{ij}\right)^{-\sigma+1}$ where σ is the elasticity of substitution, and $c_{ij} \geq 0$ is the bilateral ad valorem cost of trade. In practice, we can normalize cost and impedance so that the lowest possible bilateral cost is null and corresponds to an impedance of one, and thus we can refer to $-\log K_{ij} \geq 0$ as "cost".

Wilson (1967) shows that there are two broad categories of spatial interaction models depending on whether the potentials A and B are endogenous or exogenous. The exogenous potentials models are by far the most used in international economics, where the potentials are econometrically estimated functions of macro-variables capturing the size (GDP or population) and perhaps degree of development or performance of the country (GDP per capita). With endogenous potentials models, by contrast, the potentials are estimated from the requirement that row and column totals in the gravity model estimates must equal the total outflows or inflows of the nodes. Thus:

$$2(a)\ X_{i.} = \sum_j A_i B_j K_{ij}$$

$$2(b)\ X_{.j} = \sum_i A_i B_j K_{ij}$$

This non-linear problem yields the potentials up to the trivial scale transformation $A_i \to \lambda A_i$ and $B_j \longrightarrow \frac{1}{\lambda} B_j$. That total flows apparently have one degree of freedom less than the potentials simply results from the fact that the totals in rows and columns are not independent since $\sum_i X_{i.} = \sum_j X_{.j} = X_{..}$.

The endogenous gravity model has several interesting properties. First, once bilateral costs are known or econometrically estimated, only two variables per node are needed: total outflows and inflows. Furthermore, as a consequence of the non-linear nature of the model, the potential of a node does not depend upon its own variables, but on every other interaction in the network. A third property is that by construction, the model conserves row and column totals. This problem is one that arose early on in the gravity literature (Linnemann, 1966; Tinbergen, 1962): it was observed that regressing flows against distance and size parameters does not guarantee conservation of the relevant sums. Simple convexity considerations lead to the conclusion that the sum of regressed flows by node (country) exceeds the actual total flow. This discrepancy is even more serious when a country has larger and closer partners, which tend to divert trade from more distant partners. The same phenomenon happens with transport flows and is referred to as "intervening opportunities" in the transport literature (Roy, 2004). Arvis and Shepherd (Forthcoming) show that the discrepancy is quantitatively significant in the international trade context.

In the trade literature, this question is addressed using the concept of multilateral resistance (MR), which corrects for origin and destination interactions with the rest of the world. Anderson and Van Wincoop (2003) use standard trade theory to motivate a rigorous approach to MR in place of the ad hoc "remoteness" corrections used in some traditional gravity modeling. The same idea is built into the Wilson (1967) framework, which provides a simpler, purely mechanical explanation—i.e., independent of the nature of the underlying economics—in which MR is the pull or push exercised by the rest of the world. Let D_i be the pull exercised by destinations in the rest of the world on origin node i, and let O_j be the push exercised by origins in the rest of the world on destination node j. We can then define:

$$3(a)\ X_{i.} = A_i D_i, where\ D_i = \sum_j B_j K_{ij}$$

$$3(b)\ X_{.j} = B_j O_j, where\ O_j = \sum_i A_i K_{ij}$$

from which it follows (Wilson, 1967) that:

$$(4)\ X_{ij} = X_{i.} D_i^{-1} X_{.j} O_j^{-1} K_{ij}$$

Total outflows are, thus, as expected roughly proportional to total outflows/inflows of the origin/destination multiplied by an impedance factor, as in the most naïve gravity formulation. This conclusion is maintained notwithstanding the correction for the pull and push from the rest of the world, or adjustment for multilateral resistance. In other words, $X_{ij} \approx X_{i.}X_{.j}K_{ij}$.

4.2 A Gravity-Based Definition of Connectivity

Given their fundamental nature and meaning, it is tempting to use the pull or push from the rest of the world as a natural implementation of the connectivity of an origin or destination. However, in view of the scale invariance property of potentials, some normalization is needed for a proper definition. It is natural to compare the push and pull to what they would be if all the bilateral impedances were one, which yields a first candidate for connectivity:

$$(5)\ C_i = \frac{D_i}{\sum_j B_j - B_i} = \frac{\sum_{j \neq i} K_{ij} B_j}{\sum_{j \neq i} B_j} \leq 1$$

The above expression applies to the connectivity of outflows, but permuting A, B, D, and O gives the corresponding value for inflows. It is reminiscent of the market potential measure developed recently by Head and Mayer (2011) in the economic geography literature, but is derived from a more general framework in which no adding up constraints are imposed on the flows and the potentials.

This expression can be interpreted in two ways. First, connectivity is equal to the average impedance K, weighted by the potential of each partner. An alternative interpretation is that the numerator summarizes the pull or push of all partners, and the denominator represents the maximum possible pull or push.

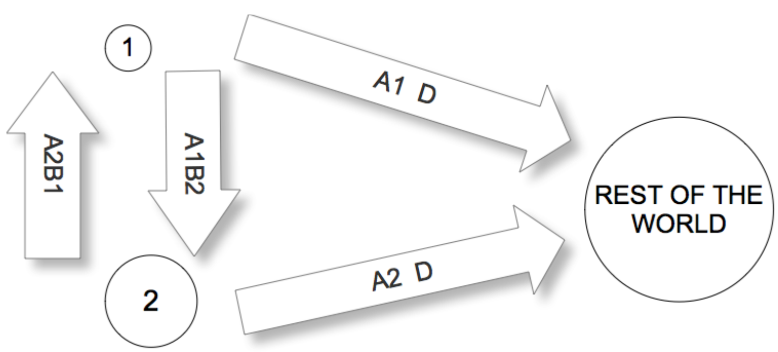
4.2.1 Self-Interaction and the Definition of Connectivity

The denominator in equation (5) excludes a country's own potential from the sum $\sum_j B_j$. This choice leads to some inconsistency, as is apparent in rather common cases, such as comparing equation (5) for close neighbors. Take an idealized Canada-USA situation with two countries, 1 and 2, with different sizes (2 being the bigger one), close interaction (impedance one) between them, and the same pull/push with the rest of the world (Figure 4). Let $A_{1,}A_2$ and $B_{1,}B_2$ be the potentials for the two nodes, and D be the pull from the rest of the world. Then equation (5) yields $\frac{D+B_2}{\sum_{j\neq 1,2} B_j+B_2}$ for node 1 , and $\frac{D+B_1}{\sum_{j\neq 1,2} B_j+B_2}$ for node 2. Because $B_1 < B_2$, the smaller node has the higher connectivity value, which is expected since 1 has a bigger pull from its neighbor than 2 does from 1. This is not totally intuitive, however. The inclusion of the potential of the nodes in the formula to take account of this problem and equalize the connectivity of the two nodes, as in the expression $\frac{D+B_1+B_2}{\sum_{j\neq 1,2} B_j+B_1+B_2}$, is more consistent with the expected properties of connectivity. Hence a consistent definition of connectivity should also include a country's own contribution to push and pull.

$$(6)\ C_i = \frac{D_i+B_i}{\sum_j B_j} = \frac{{}^{X_{i.}}/_{A_i} + B_i}{\sum_j B_j}$$

where $\sum_{j\neq i} K_{ij} = D_i = {}^{X_{i.}}/_{A_i}$

Figure 4: Connectivity of two countries with similar flow patterns.



This improved definition amounts to including in the flow matrix X a diagonal term, which corresponds to the effective flow between each country and itself with an impedance of one, i.e. $X_{ii} = A_i B_i$ or $K_{ii} = 1$. This factor represents the intensity of the flow with the closest possible neighbor. This effective flow may not be the real one, for instance domestic air transportation, but rather an equivalent level of domestic traffic that would exhibit the same pattern as international traffic. Such an approach of re-creating a diagonal of economic flows for consistency purposes is not new, and was indeed introduced half a century ago by Deutsch and Savage (1960).

4.2.2 Final Definition and Properties

For a symmetric definition of connectivity, we can take the geometric average of the connectivities of i as origin and destination. (The rationale for the geometric average becomes apparent below, when we interpret the log of connectivity in terms of entropy and trade costs.) This approach gives:

$$(7)\ \bar{C}_i = \sqrt{\frac{X_{i.}/A_i + B_i}{\sum_j B_j}} \times \sqrt{\frac{X_{.i}/B_i + A_i}{\sum_j A_j}}$$

The above construction produces a consistent definition of connectivity with the desirable properties outlined earlier. It is rooted not only in the topology of the network, but also in a fundamental understanding of spatial interactions among the nodes. It can also be seen to be closely related to the idea of multilateral resistance that underlies the recent gravity literature in international trade.

Under this definition, connectivity is a non-dimensional number between zero and one. The non-linear construction means that the concept is indeed global: a country's connectivity depends not only on its neighbors, but also on all of the interactions among the other countries in the network (just as multilateral resistance depends on trade costs across all potential trading partners). The size or potential of the node does not enter directly into our measure of connectivity, which represents the pull and push of the rest of the world.

Finally, our proposed connectivity concept is consistent with the intuitive property of propagation of connectivity: i.e., a country's connectivity increases if that of its closest neighbors increases. Because the construction is non-linear, the mechanism of propagation is less evident than with linear diffusion models like eigenvector centrality or PageRank, for instance (see Appendix 3). But the connectivity of node i indeed increases with the potential of close neighbors for which the bilateral impedance K_{ij} is above the connectivity C_i, because from (6):

$$(8)\ \frac{\partial C_i}{\partial B_j} = \frac{K_{ij} - C_i}{\sum_j B_j}$$

Another useful property of our measure emerges from a simple decomposition. The flow-weighted average "cost", measured by log-impedance, for origin node i is:

$$(9)\ \bar{c}_i = \sum_j \frac{X_{ij}}{X_{i.}}\left(-\log K_{ij}\right) = -\sum_j \frac{X_{ij}}{X_{i.}} \log \frac{X_{ij}}{A_i B_j} = -\sum_j \frac{X_{ij}}{X_{i.}} \log \frac{X_{ij}}{X_{i.}} \frac{D_i + B_i}{B_j}$$

$$= -\log \frac{D_i + B_i}{\sum_j B_j} - \sum_j \frac{X_{ij}}{X_{i.}} \log \frac{\frac{X_{ij}}{X_{i.}}}{\frac{B_j}{\sum_j B_j}}$$

where the effective diagonal term $X_{ii} = A_i B_i$ is included. Combining this expression with the definition of connectivity in (6) allows us to write:

$$(10)\ \log C_i = \underbrace{-\sum_j \frac{X_{ij}}{X_{i.}}\left(-\log K_{ij}\right)}_{Average\ Cost} - \underbrace{\sum_j \frac{X_{ij}}{X_{i.}} \log \frac{\frac{X_{ij}}{X_{i.}}}{\frac{B_j}{\sum_j B_j}}}_{Relative\ Entropy}$$

where the relative entropy of the flows from node i is the Kullback-Leibler distance with reference to the destination potentials B.[12] Average costs and entropy are both strictly positive, which means that connectivity is the product of a cost factor—which measures remoteness or location in the network—

[12] In the trade literature, Straathof (2008) notes a connection between multilateral resistance in the Anderson and Van Wincoop (2003) model and Shannon's measure of entropy, which is a generalization of the Kullback-Leibler measure used here. However, Straathof (2008) interprets it as a measure of product diversity.

and the exponential of the negative of entropy, which measures the diversity of a node's connections. Connectivity can thus be summarized as the product of remoteness and diversification.[13] In the context of air transport, the first factor has a lot to do with geography, while the second is likely also influenced by policies that tend to establish more connections with the rest of the world through a more "open sky".

Table 1 provides a comparison of the *Air Connectivity Index* (ACI; equation 7) with the other measures of connectivity discussed in Appendix 3, in terms of the criteria set out at the beginning of this section. As can be seen, the ACI is the only measure that responds to all five desirable criteria for a measure of connectivity.

Table 1: Comparison of the ACI with alternative measures of connectivity.

	Realistic	Intensive	Normalized	Local	Global
UNCTAD (2007)	Y	N	N	N	N
Pearce (2007)	Y	N	N	Y	N
Kullback-Leibler distance	N	Y	Y	Y	Y
Clustering Coefficient	N	Y	Y	Y	N
PageRank or Eigenvector Centrality	partial	Y	N	Y	Y
Closeness centrality	partial	Y	Y	Y	Y
ACI	Y	Y	Y	Y	Y

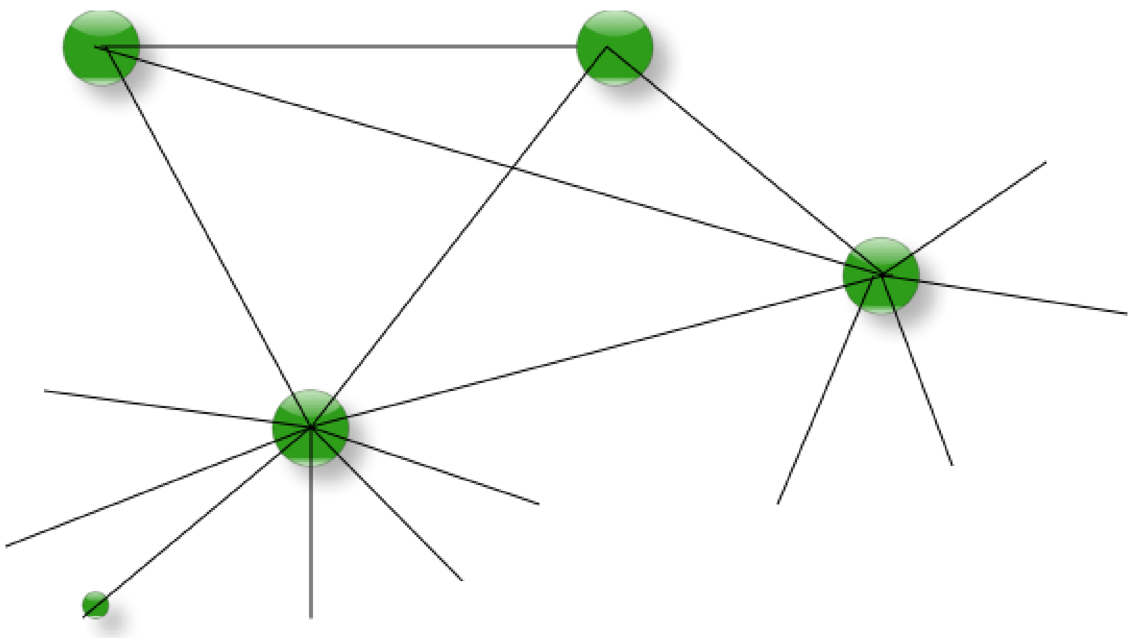
4.3 A Simple Example

To see how our measure of connectivity works in practice, we take a simple example that is of some relevance to the air transport setting. We assume a network consisting of two layers: an inner layer of N fully connected hubs, and an outer layer of M spokes attached to the hubs (Figure 5). All links have the

[13] We note in passing that the decomposition proposed here retains data on observations where no actual flow is observed, due to the presence of summation operators prior to taking logarithms.

same impedance $K < 1$, and the flow matrix X is symmetric. Let A_h and A_s be respectively the potential of hubs and spokes. Then it follows that the pull from the rest of the world is $NA_h + MA_s$ for hubs and $A_h + A_s$ for spokes.

Figure 5: A simple two layer hub and spoke network.



From (6), we obtain normalized connectivity scores as follows:

$$(11a)\ C_h = K\frac{N + M\frac{A_s}{A_h}}{N + MN\frac{A_s}{A_h}}$$

$$(11b)\ C_s = K\frac{1 + \frac{A_s}{A_h}}{N + MN\frac{A_s}{A_h}}$$

As $\frac{A_s}{A_h}$ decreases from ∞ to 0, C_h increases from $\frac{K}{N}$ to K, and C_s increases linearly from $\frac{K}{MN}$ to $\frac{K}{N}$. Thus, the connectivity of the spokes has two important properties. As expected, it is less than that of the hubs.

Moreover, the connectivity of a spoke improves with that of the hub, in this case linearly (after elimination of $\frac{A_s}{A_h}$ in (11)):

$$(12)\ C_s = \frac{1}{N-1}\left[C_h - \frac{K}{N} + (\frac{K - C_h}{M})\right]$$

4.4 Empirical Implementation

The model presented above, and its application to air transport, is parametric. The dependent variable measures the bilateral interaction between nodes, i.e. a measure such as the number of flights or total seat capacity. Since the potentials are endogenous, the independent variables are the bilateral impedances. The bilateral impedance variable is not a "cost" in the strict sense, but an adequate measure of how far apart nodes are, using air transport. It can be proxied by data derived from geographical distance, as described below.

To implement the model empirically, we need estimates of the potentials (the A and B terms in equation 19). They can be consistently obtained by using origin and destination fixed effects, as in much of the trade literature (e.g., Anderson and Van Wincoop, 2003). This approach determines the A and B terms up to a multiplicative constant. Santos Silva and Tenreyro (2006) argue that heteroskedasticity concerns, as well as the need to account for zeros in the flow matrix, make the Poisson pseudo-maximum likelihood estimator a good candidate for a workhorse estimator of the fixed effects and other parameters in gravity models. In a companion paper (Arvis and Shepherd, Forthcoming), we show that Poisson has an additional desirable property: it is the only unconstrained (pseudo-)maximum likelihood estimator to preserve equality between actual and estimated total trade flows. For these reasons, we use Poisson to obtain estimates of the gravity model potentials.[14]

[14] Helpman et al. (2008) show that sample selection can also be an issue in gravity models with zero trade flows. However, the literature does not yet disclose an estimator that combines the desirable properties of those

5 Estimation of the ACI for 2007

The previous developments can be applied quite generally to many spatial interaction networks. However, as stated earlier, this paper focuses on an application to the global air transportation network. We estimate an *Air Connectivity Index* (ACI) for 2007, using data for the month of June. The flow variables are sourced directly from data in the *SRS Analyzer* system. The SRS database provides the number of direct flights per week among over 200 origin and destination countries, including passenger and cargo flights.[15] In the remainder of this section, we present results from the empirical implementation of the model using the approach outlined in the previous section, and provide a comparison with alternative metrics.

5.1 Cost Function Specification

As discussed in Section 4, we derive the potential terms (A and B) as fixed effects using a Poisson estimator. In the trade literature, it is standard to include the simple logarithm of distance as a measure of trade costs (impedance). In the context of air transport, it is also natural to posit that the bilateral impedance is a function of distance, or equivalently time of flight. However, we take a more refined approach to the parameterization of this dependence for two reasons.

First, it is important to allow for a richer set of shapes than the simple logarithm to account for the network structure of air transport. The commercial implications of short- versus long-haul flights are very different. Airplanes can only safely fly a limited distance, thereby making it technically impossible

developed by Santos Silva and Tenreyro (2006) and Helpman et al. (2008). We prefer Poisson because it is consistent under much weaker assumptions than the Heckman-like estimator used by Helpman et al. (2008).

[15] Although from an international trade perspective we are primarily interested in movements of goods, not people, we include passenger traffic in addition to pure cargo flights because a significant proportion, perhaps 50%, of air freight travels as cargo on passenger flights (Geloso Grosso and Shepherd, 2011). However, it is beyond the scope of this paper to examine possible differences in connectivity for passenger or mixed versus pure cargo flights.

for some country pairs to have a direct flight between them. In practice, trade costs are therefore unlikely to be a simple logarithmic function of distance without any breaks or other nonlinearities.

Furthermore, at the core of the generalized gravity model is the fact that the impedance should be normalized to one for the closest neighboring nodes, with zero cost between them. A simple logarithmic dependence diverges for zero distance, and therefore needs to be adapted. One solution to this problem is to introduce a regularized cost function which is zero for zero distance, but which has the same overall shape as the classic logarithmic dependence. The simplest modification is the shifted logarithmic function:

$$(13)\ \tilde{X}_{ij} = A_i B_j \exp\left(-\beta f(d_{ij})\right)$$

where the dependence of distance is given by a shifted log:

$$(14)\ f(d) = a * [Log(a + d) - Log(a)]$$

In this expression, the constant a represents the "natural scale" of the network. The intuitive interpretation of this scale is that there is a fixed minimum cost in the interaction between nodes, or in the context of air transport the time to take off and land when moving from terminal to terminal. From a theoretical standpoint, this expression is also consistent with the CES foundation of gravity (Anderson, 1979), where $Log(K_{ij}) = -(\sigma - 1) * Log\left(1 + c_{ij}\right)$. Identification of costs with distance means that the corresponding elasticity of substitution would be:

$$(15)\ \ \sigma = 1 + \beta a$$

5.2 Econometric Results

The primary question for implementing the shifted log estimator in distance is the choice of scale parameter a. A natural approach is to determine it endogenously in the regression by selecting a and β

in (14) and (15) so as to maximize the log-likelihood of the Poisson regression function. To do this, we adopt a grid search approach and run a series of fixed effects Poisson regressions (Figure 6). We run the grid search at 100km intervals over the range 100-19,000km. The value of the log-likelihood is maximized at approximately $a = 3{,}900$. Interestingly, this value corresponds to an operational threshold in the industry, between medium- and long-haul flights (5 hours or 4,000 kilometers).The beta coefficient for the cost function is 1% significant, and the model provides a strong overall fit with the data (Table 3). As measured by R2, model fit improves from 0.74 to 0.85 when we use the shifted log function as opposed to the simple logarithm of distance, which suggests that our approach is capturing an important feature of the data.

Figure 6: Grid search results using the shifted log distance function.

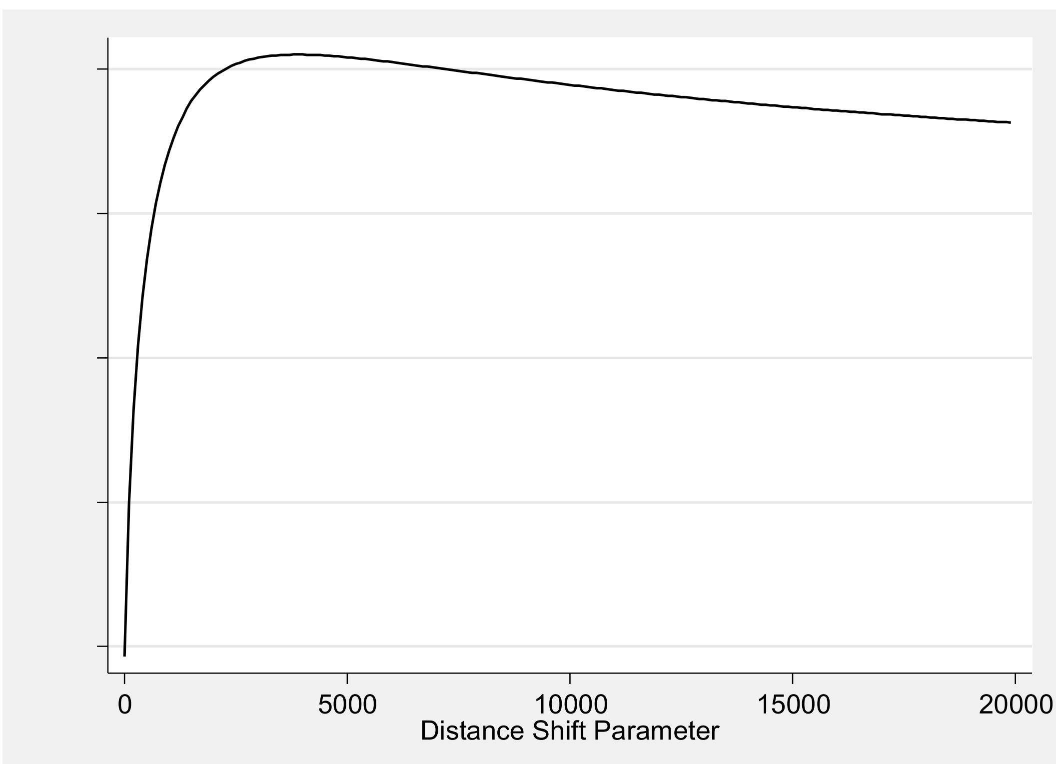


Table 2: Gravity model regression results using the shifted log distance function.

Dependent Variable:	**No. of Flights per Week**
$3900 * [Log(3900 + distance) - Log(3900)]$	-0.001***
	[0.000]
Observations	44,313
R2	0.854

*Prob. values based on robust standard errors corrected for clustering by country-pair are in brackets below the coefficient estimates. Statistical significance is indicated by * (10%), ** (5%), and *** (1%). The model includes fixed effects by origin and destination country (estimates suppressed for brevity), and is estimated by Poisson. R2 is calculated as the squared correlation coefficient between the actual and fitted values.*

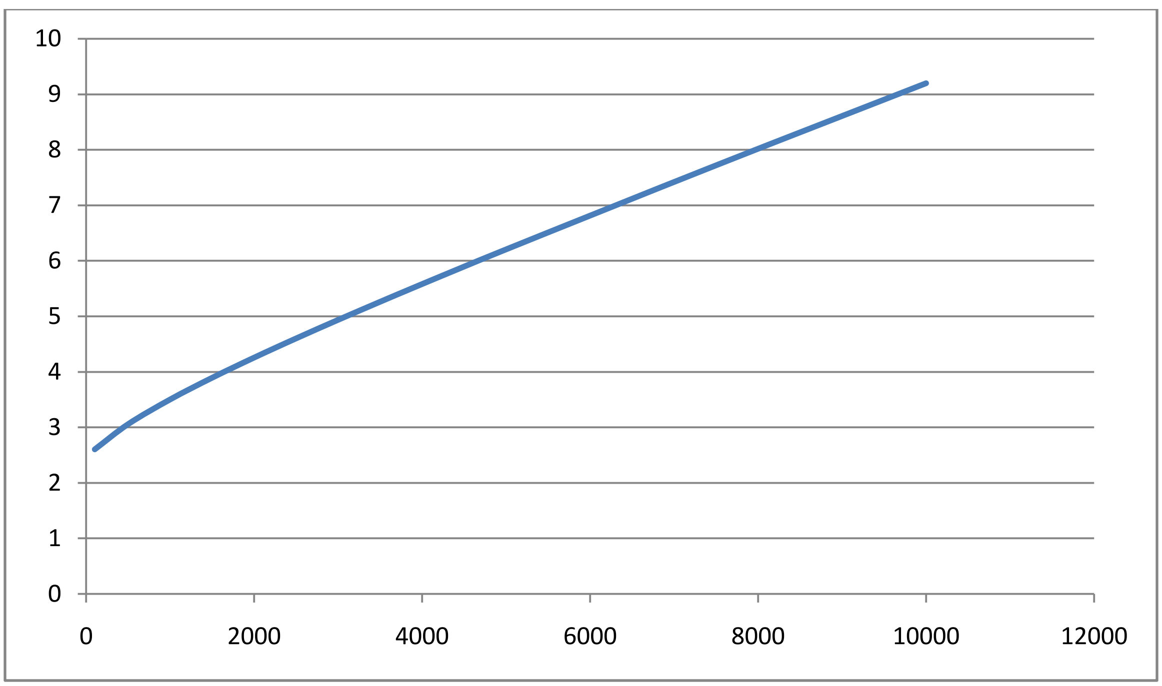
The elasticity of substitution implied by the model (see equation 15) is about 5.5, which confirms a high impact of distance on transport flows. Although sectoral characteristics are obviously very different, it is useful to compare this order of magnitude with estimates of the elasticity of substitution from goods markets. Anderson and Van Wincoop (2003), for example, use a benchmark of $\sigma = 5$ for their counterfactual gravity model simulations. The simple average of the sectoral elasticities of substitution estimated by Broda and Weinstein (2006) ranges from four to 17, depending on the time period and aggregation scheme used. Our estimate might therefore be considered to be towards the low end of estimates for goods, which would be consistent with the intuition that different transport routes are less substitutable than products from different origins.

Another way of looking at the distance result is by converting the regression coefficient to a simple elasticity, familiar from standard gravity models in the trade literature. Evaluated at the mean, a one percent increase in distance is associated with an approximately 3% decrease in the number of flights per week. This elasticity is much stronger than is typically found in gravity models of international trade, particularly those estimated by Poisson (Santos Silva and Tenreyro, 2006). Nonetheless, this finding sits well with the nature of the transport industry, in which distance is obviously a primary commercial concern due to the importance of fuel in total operational costs.

5.3 Robustness of the Cost Specification and Dependence on the Distance Shift Parameter

The cost function used above depends on the shift parameter a. For a given distance d the cost increases from zero to d as the shift goes from zero to infinity. Hence, the dependence of distance on potential is comparatively less with a lower shift parameter, changing from a power dependence with zero shift to a negative exponential dependence with infinite shift. This phenomenon can be seen in the value of the implied elasticity of substitution, which increases with the shift parameter (Figure 7). The lower value is simply one plus the negative exponent of the log regression of aviation flows, known since Zipf (1946) to be slightly higher than one.

Figure 7: Implied elasticity of substitution as a function of the distance shift parameter.



The choice of distance shift parameter has a direct impact on potential and connectivity. With a higher shift parameter, a country that is relatively closer to its partners—like many in Europe, for example—would have comparatively higher bilateral impedances, and hence higher connectivity (an average of impedances) than a country in a region where distances are longer, as in East Asia for instance. Although results are robust around the value of the shift parameter that maximizes the log-likelihood, we nonetheless observe that a lower value of the shift parameter improves the score of regional hubs such as the UAE, or those in East Asia, as compared to European countries. An extreme case of the phenomenon is to use nodal distance instead of geographical distance. This amounts to using the adjacency matrix as the impedance matrix, i.e. $K_{ij} = L_{ij}$.[16] Intuitively, this measure uses the same information as closeness centrality and is also based on a definition that uses moves on the network (through the iteration in this case). For comparative purposes, this nodal distance connectivity is given in Appendix 2 along with the connectivity computed with a lower value of the shift than the one that maximizes the log-likelihood, namely 100km instead of 3,900km.

5.4 Robustness and Comparisons with Other Connectivity/Centrality Metrics

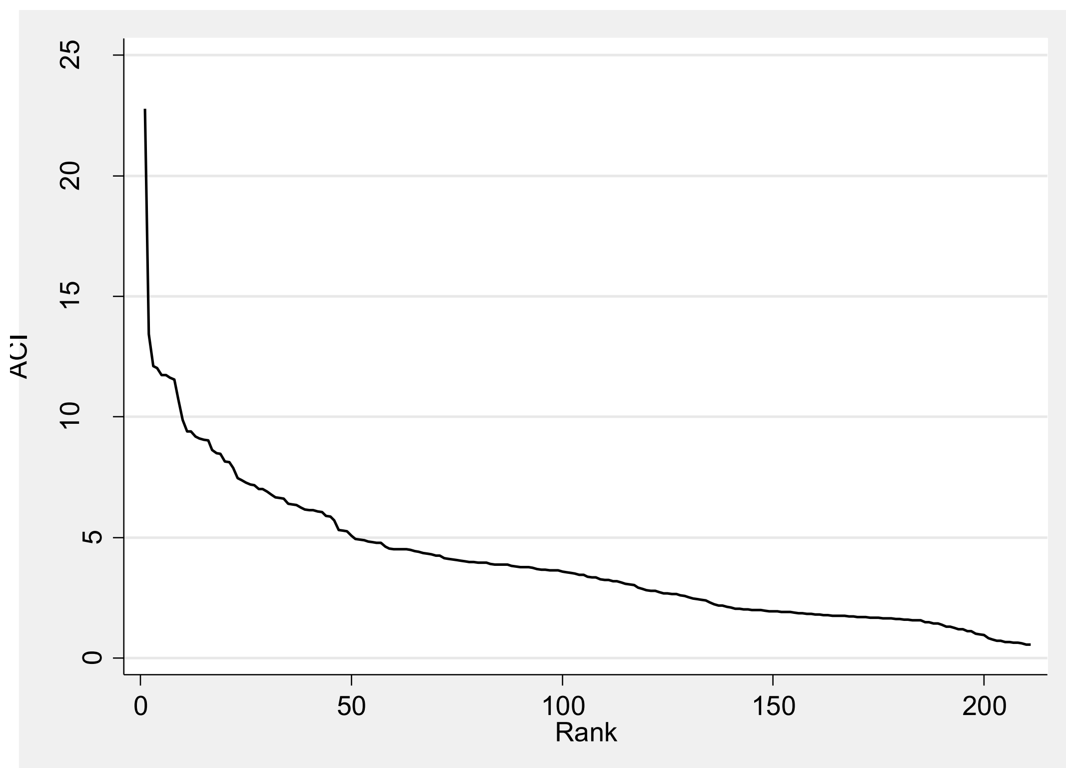
Full details of the ACI calculated using the regression results from Table 3 substituted into equation (7) are presented in Appendix 1. Results are intuitively appealing, and generally accord well with thinking in the industry. The most connected country is the USA (22%), followed by Canada (13%), and Germany (12%). A cluster of European countries makes up the top ten, with scores ranging from about 10% to 12%. This positioning is consistent with their role as regional hubs, and their close connections with Germany and the UK as major international gateways. Partly as a result of the optimal distance shift parameter suggested by the data, Asian countries—including regional hubs such as China (6%), Hong Kong SAR, China (5%), Japan (5%), Korea (5%), and Thailand (4%)—fall into the middle range of connectivity scores. The same is true for the Middle Eastern hubs of the UAE (5%), Bahrain (4%), and Qatar (4%). The bottom end of the rank table is made up of isolated countries in Oceania, such as the

[16] In practice no regression is needed, as the potentials A and B can be determined iteratively from any starting value (e.g. total outflows): $A_i^{(n+1)} = \frac{X_{i.}}{\sum_j K_{ij} B_j^{(n)}}$, and conversely for B. This iteration converges quadratically.

Cook Islands, French Polynesia, Niue, Kiribati, and the Marshall Islands, as well as African countries including Zimbabwe, Mauritius, Madagascar, and Angola.

Figure 8 shows that connectivity drops off sharply from the most connected country (USA, 22%) to the second ranked country (Canada, 13%). The mean ACI score is about 4%, but the median is 3.4%, which suggests that the distribution is significantly left-skewed. Both characteristics are suggestive of a power law distribution, as is the case for the number of direct air connections of each country (Figure 2). A simple regression (Gabaix and Ibragimov, 2011) confirms this finding.[17] Intuitively, it is not surprising given that the air transport network is widely known to be composed of a relatively small number of well-connected hubs, and a large number of less well-connected spokes.

Figure 8: ACI score vs. rank.



[17] The ACI (in logarithms) has a coefficient of -0.33, which is significant at the 1% level based on robust standard errors. The hypothesis of a Zipf law is, however, strongly rejected at the 1% level.

One interesting feature of the ACI is that the most connected country, the US, still receives only a fairly low score (22%). The reason is that although the US is very well connected globally, it still only has direct air links to 101 out of a possible 210 countries. Some of the links it does have—in particular with countries that are themselves poorly connected—are very weak. For instance, there are only a few flights per week between the USA and some destinations in Africa, Oceania, and even Latin America, compared with thousands of flights per week to neighboring Canada and Mexico. So although the US as an origin provides a wide range of possible destinations, many other countries can only be reached either indirectly, or using relatively irregular services. Most commonly, passengers and goods will pass through another regional hub on their way to a poorly connected destination. Because of this hub and spoke structure, even the most connected country in our sample has a relatively low overall connectivity score.

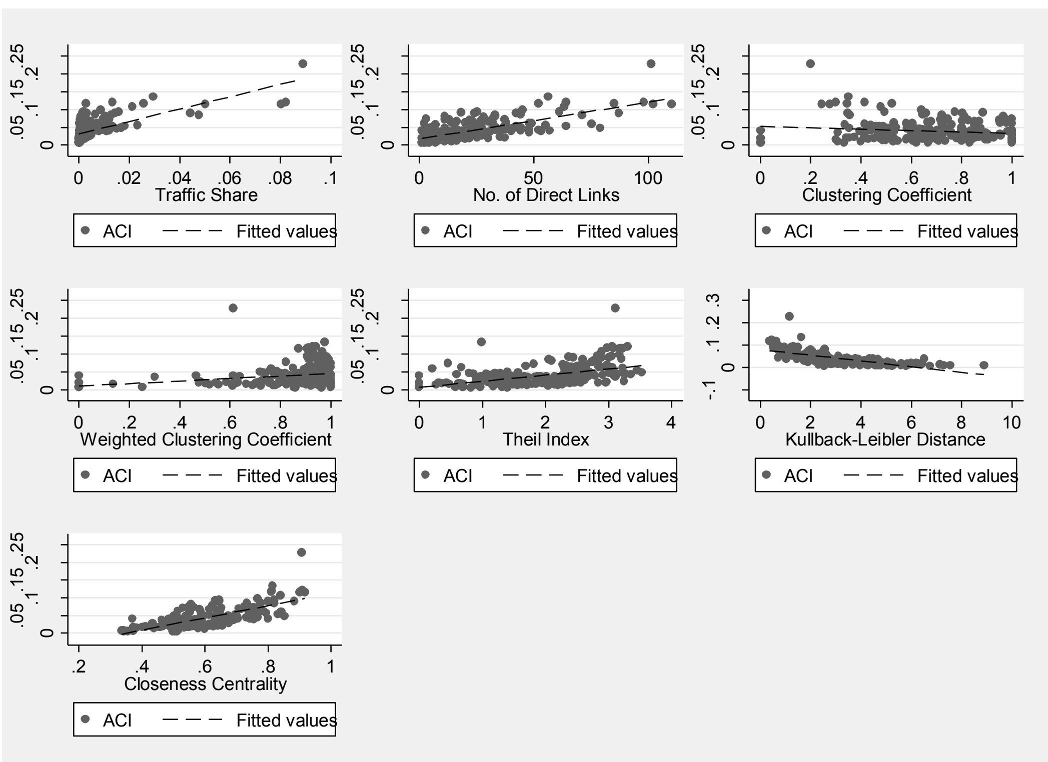
In Table 4, we use Spearman's rank correlation coefficient to compare the ACI with the other possible measures of connectivity discussed in Appendix 3. Figure 9 presents the same information visually. With the exception of the two clustering coefficients, the ACI correlates strongly with all of these measures. Interestingly, the strong association with traffic share emerges even though our definition of connectivity is independent of market size: it is a consequence of the position of large countries in the network, rather than a conclusion that flows directly from the definition of connectivity. In addition, the fact that the ACI is strongly associated with the number of direct air links a country has captures the intuitive importance of direct versus indirect links in assessing connectivity: the implication is that indirect links are effectively down weighted, which would be consistent with a model in which transfers are penalized due to the additional costs involved.

The reason for the weak correlation with the weighted and unweighted clustering measures is undoubtedly that they capture local connectivity only, whereas the ACI is a global measure, as discussed above. In light of the decomposition presented above in which the ACI is shown to be negatively related to the Kullback-Leibler entropy measure, the negative rank correlation in that case is entirely expected. In sum, these results confirm that the ACI captures an important part of the intuitive concept of connectivity in the air transport context.

Table 3: Rank correlations between the ACI and alternative connectivity indicators.

	Spearman's Rho
Traffic Share	0.7590
No. of Direct Links	0.6704
Clustering Coefficient	-0.0831
Weighted Clustering Coefficient	0.2476
Theil Index	0.5537
Kullback-Leibler Distance	-0.8400
Closeness Centrality	0.7295

Figure 9: Correlations between the ACI and alternative connectivity indicators.

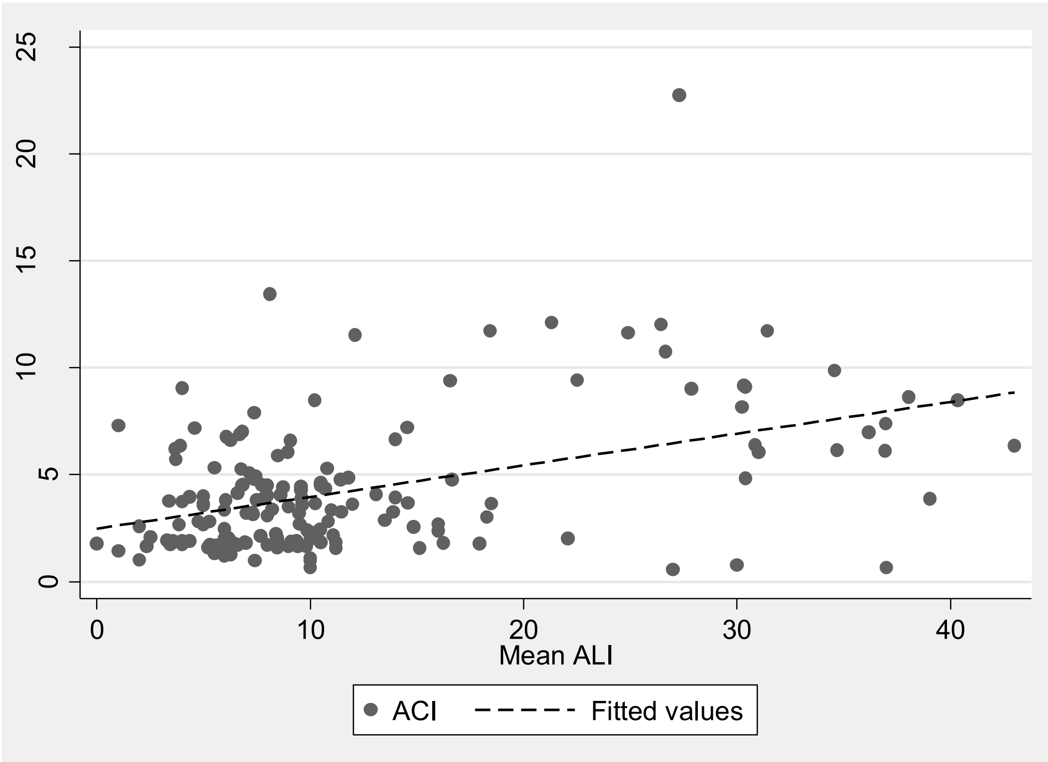


One important aspect of the association between the ACI and simpler, but intuitive, measures of connectivity such as traffic share and the number of direct links, is that the link is much weaker in low income countries than for the sample as a whole: Spearman's rho is only around 0.4 in both cases. This finding suggests that although the ACI captures similar overall tendencies to those summarized in simpler metrics, its main value added is in dealing differently with smaller, less connected countries. We believe this makes the ACI particularly useful from a development policy point of view.

6 The ACI, Policy, and Trade Outcomes

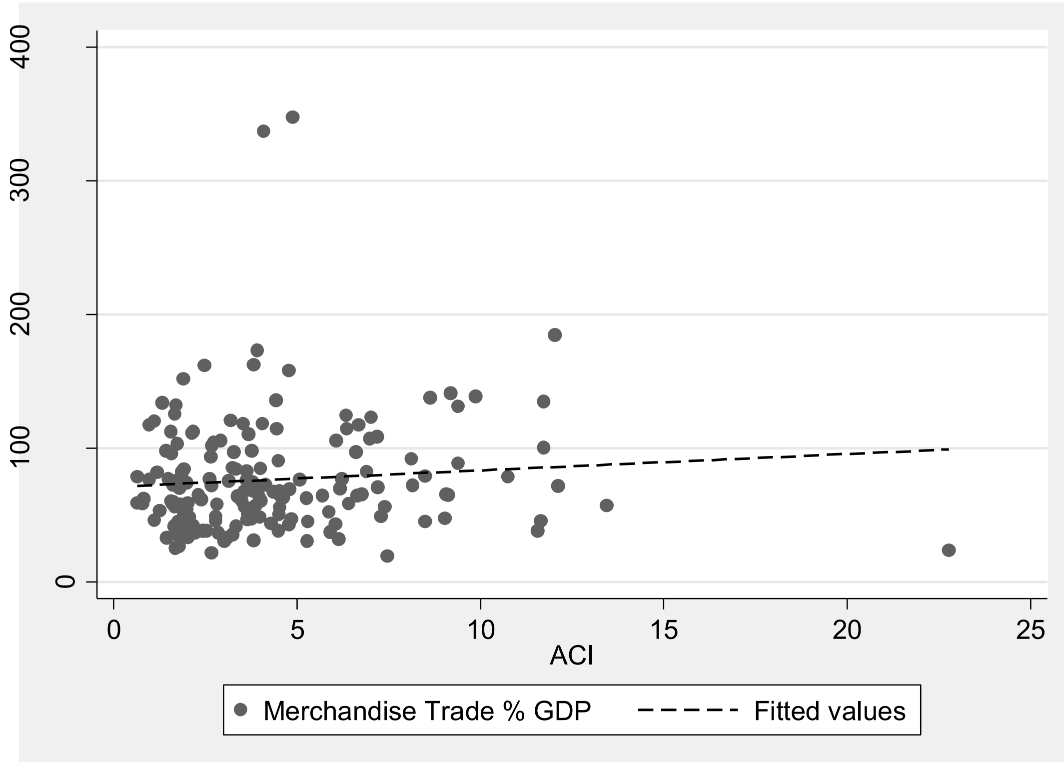
It is important to check the external validity of the ACI by comparing it with input and output indicators that we would expect to be correlated with it. We start by considering the input side. We expect that liberalization of air transport markets tends to promote better connectivity. By reducing the cost of moving from one country to another, Bilateral Air Services Agreements (BASAs) should lead to more and stronger connections between countries. More liberal countries should therefore have a higher ACI score. Figure 10 shows that this is clearly the case (correlation coefficient = 0.45, 1% significant). We use the World Trade Organization's *Air Liberalization Index* (ALI), averaged across all partner countries, to measure how liberal a country's air transport policy environment is. The clear implication of this result, should it be borne out by more detailed analysis, is that countries seeking to improve their air transport connectivity would do well to move towards more liberal policy regimes, in addition to potentially considering measures such as lower landing or airport fees.

Figure 10: ACI vs. ALI.



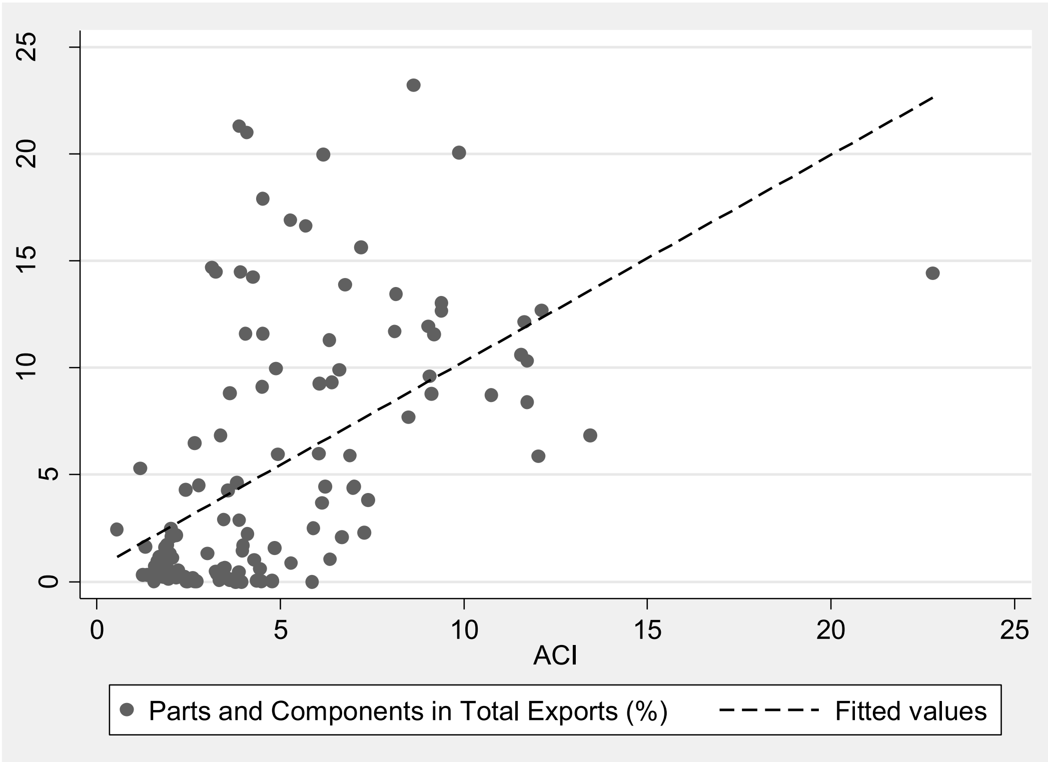
We also expect the ACI to be an important determinant of economic outcomes, particularly in the area of trade performance. As an increasingly important mode of transport, better air connectivity should be associated with a greater degree of trade integration. Figure 11 investigates this hypothesis. Interestingly, we find a positive but statistically insignificant association between the ACI and trade openness, as measured by merchandise trade as a percentage of GDP. One possible explanation for this finding is that it reflects the fact that air transport is only an important mode for some sectors. Goods with a relatively low value to weight ratio cannot be profitably moved by air, and tend to rely to a greater degree on maritime shipping. Since high value to weight products make up only a relatively small percentage of total trade, this would explain why the overall correlation is weak (correlation coefficient = 0.09, not statistically significant).

Figure 11: ACI vs. merchandise trade as a percentage of GDP.



We can test the explanation advanced in the previous paragraph by focusing on goods traded within international production networks, which tend to have a high value to weight ratio and rely heavily on air transport due to the need to move goods quickly and maintain low inventories. Production networks need to move intermediate parts and components across borders multiple times in the course of producing final goods. The process can only work profitably if transport is relatively quick and cost-effective. Because of these factors, we expect to see a particularly strong correlation between a country's degree of integration into international production networks, and its air connectivity.

Figure 12: ACI vs. trade in parts and components.



This relationship is exactly what we see in Figure 12 (correlation coefficient = 0.54, 1% significant). The upward sloping line indicates that parts and components—a common proxy for goods circulating within network structures (Ando and Kimura, 2005)—represent a higher share of total exports in countries that are better connected with the international air transport network. The contrast with the result for overall merchandise trade is striking, and indicates that air transport is particularly important for promoting trade in high value to weight sectors, such as those traded within international production networks. This finding potentially has important implications for developing countries seeking to deepen their level of involvement in this type of trade.

7 Conclusion

This paper has developed a simple but consistent measure of the degree to which countries are connected to the international air transport network: the *Air Connectivity Index* (ACI). Our measure captures important features of that network, such as its hub and spoke structure, and the dual importance of the number and strength of flight connections. As expected, the ACI correlates strongly with important economic measures on both the input and output sides, including the degree of policy liberalization in air services markets, and specialization in parts and components trade as a proxy for trade openness in high value to weight sectors.

There are a number of ways in which future work can expand on our findings. As a starting point, we hope to update the index on a regular basis, to give policymakers and analysts consistent information on connectivity over time. This approach will enable them to track performance, and examine the impacts of policies designed to improve the air transport environment, and boost connectivity.

Second, there is scope for detailed econometric work looking at the extent to which air connectivity determines trade outcomes, and the pattern of specialization across countries. We would expect countries with strong connectivity to specialize in industries such as perishable goods and networked components, which are intensive in their use of high speed transport services. The results presented in this paper can only be regarded as suggestive, and should be investigated more thoroughly in the future.

Third, we expect air transport connectivity to influence exporters' choice of transport mode. This is an issue that can be investigated using data on imports by mode for two major markets, namely the US and the EU. To the extent that a country's ACI score is influenced by its policy environment—and particularly the degree of liberalization of air services markets—there may be scope to reduce distortions in the modal choices of exporters and transport operators by improving connectivity.

Fourth, there is scope to apply and adapt the proposed implementation of connectivity to other modes of transport or international networks. Maritime transport is a particular priority, because it still represents the dominant mode of transport for international trade transactions. This position is particularly true for developing countries. UNCTAD's *Liner Shipping Connectivity Index* has already made critical contributions in this area. However, as explained in Section 2, there are essential differences in the way information is incorporated in the two methodologies. The two approaches should prove to be strongly complementary in future policy work.

Finally, the theoretical foundation of our approach, which bridges spatial modeling and trade theory, makes it suitable for a broad range of applications to spatial economic interactions between countries or other nodes. Applications to international trade immediately come to mind, in particular in measuring connectivity in relation to international production networks, for instance. In this respect, there are areas for further research and improvement, for example when it comes to the robustness of the specification for the impedances and the cost function; the dependence on distance used here may not be suitable for other problems, and will need to be addressed on a case by case basis.

In particular, the implementation of the model departs from the simplicity of the solutions promoted in the network literature (Appendix 3), such as clustering and centrality indicators. In the absence of economic interactions between nodes, generic network analysis does not need to refer to bilateral costs on the network and indicators do not depend on exogenous independent variables, in this case bilateral distance. Developing a fully endogenous model would mean that not only the potentials, as here, but also the impedances would be determined from the data, in a way which is consistent with the generalized gravity framework. It is an apparently open and seemingly formidable problem.

References

Anderson, J.E. 1979. "A Theoretical Foundation for the Gravity Equation." *American Economic Review*, 69(1): 106-116.

Anderson, J.E., and E. Van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170-192.

Ando, M., and F. Kimura. 2005. "The Formation of International Production and Distribution Networks in East Asia." In T. Ito and A.K. Rose eds. *International Trade in East Asia*. Chicago: University of Chicago Press.

Arvis, J.-F., M.A. Mustra, J. Panzer, L. Ojala, and T. Naula. 2007. *Connecting to Compete: Trade Logistics in the Global Economy*. Washington, D.C.: World Bank.

Arvis, J.-F., M.A. Mustra, L. Ojala, B. Shepherd, and D. Saslavsky. 2010. *Connecting to Compete: Trade Logistics in the Global Economy*. Washington, D.C.: World Bank.

Arvis, J.-F., and B. Shepherd. Forthcoming. "The Poisson Pseudo-Maximum Likelihood Estimator: A Solution to the 'Adding Up' Problem in Gravity Models." Working Paper.

Bagler, G. 2008. "Analysis of the Airport Network of India as a Complex Weighted Network." Working Paper, http://arxiv.org/abs/cond-mat/0409773v2.

Barrat, A., M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. 2004. "The Architecture of Complex Weighted Networks." *Proceedings of the National Academy of Sciences*, 101(11): 3747-3752.

Brin, S., and L. Page. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine." *Computer Networks and ISDN Systems*, 33: 107-117.

Broda, C., and D. Weinstein. 2006. "Globalization and the Gains from Variety." *Quarterly Journal of Economics*, 121(2): 541-585.

Bryan, K., and T. Liese. 2006. "The $25,000,000,000 Eigenvector: The Linear Algebra Behind Google." *SIAM Review*, 48(3): 569-581.

Chaney, T. 2008. "Distorted Gravity: The Intensive and Extensive Margins of International Trade." *American Economic Review*, 98(4): 1707-1721.

Dempsey, P. 1990. *Flying Blind: The Failure of Airline Deregulation*. Washington, D.C.: Economic Policy Institute.

Doove, S., O. Gabbitas, D. Nguyen-Hong, and J. Owen. 2001. *Price Effects of Regulation: International Air Passenger Transport, Telecommunications, and Electricity Supply*. Canberra: Productivity Commission.

Eaton, J., and S. Kortum. 2002. "Technology, Geography, and Trade." *Econometrica*, 70(5): 1741-1779.

Gabaix, X., and R. Ibragimov. 2011. "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents." *Journal of Business Economics and Statistics*, 29(1): 24-39.

Geloso Gross, M., and B. Shepherd. 2011. "Air Cargo Transport in APEC: Regulation and Effects on Merchandise Trade." *Journal of Asian Economics*, 22(3): 203-212.

Gönenç, R., and G. Nicoletti. 2000. "Regulation, Market Structure, and Performance in Air Passenger Transportation." Economics Working Paper No. 254, OECD.

Gourieroux, C., A. Monfort, and A. Trognon. 1984. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica*, 52(3): 701-720.

Guimerà, R., S. Mossa, A. Turtschi, and L.A.N. Amaral. 2008. "The World-Wide Air Transport Network: Anomalous Centrality, Community Structure, and Cities' Global Roles." Working paper, http://arxiv.org/abs/cond-mat/0312535.

Head, K., and T. Mayer. 2011. "Gravity, Market Potential, and Economic Development." *Journal of Economic Geography*, 11: 281-294.

Helpman, E., M. Melitz, and Y. Rubinstein. 2008. "Estimating Trade Flows: Trading Partners and Trading Volumes." *Quarterly Journal of Economics*, 123(2): 441-487.

Hoffman, J., and G. Wilmsmeier. 2008. "Liner Shipping Connectivity and Port Infrastructure as Determinants of Freight Rates in the Caribbean." *Maritime Economics and Logistics*, 10(1): 130-151.

Intervistas. 2007. "The Economic Impact of Air Service Liberalization." Report accessible at http://www.intervistas.com/4/reports/2006-06-07_EconomicImpactOfAirServiceLiberalization_FinalR

Li, W., and X. Cai. 2003. "Statistical Analysis of Airport Network of China." Working Paper, http://arxiv.org/abs/cond-mat/0309236v1.

Linnemann, H. 1966. *An Econometric Study of International Trade Flows*. Amsterdam: North Holland.

Newman, M.E.J. 2008. "The Mathematics of Networks." In L. Blume and S. Durlauf ed. *The New Palgrave Encyclopedia of Economics*, 2nd edn. Basingstoke: Palgrave Macmillan.

Pearce, B. 2007. "Investing in Air Transport Connectivity to Boost National Productivity and Economic Growth." In J. Blanke and T. Chiesa ed. *The Travel and Tourism Competitiveness Report 2007: Furthering the Process of Economic Development*. Davos: World Economic Forum.

Roy, J.R. 2004. *Spatial Interaction Modeling: A Regional Science Context*. Berlin: Springer-Verlag.

Santos Silva, J.M.C., and S. Tenreyro. 2006. "The Log of Gravity." *Review of Economics and Statistics*, 88(4): 641-658.

Savage, I.R., and K.W. Deutsch. 1960. "A Statistical Model of the Gross Analysis of Transaction Flows." *Econometrica*, 28(3): 551-572.

Straathof, B. 2008. "Gravity with Gravitas: A Comment." Discussion Paper 111, CPB Netherlands Bureau for Economic Policy Analysis.

Tinbergen, J. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: Twentieth Century Fund.

UNCTAD. 2007. *Transport Newsletter No. 36.* Geneva: UNCTAD.

Wilson, A.G. 1967. "A Statistical Theory of Spatial Distribution Models." *Transport Research*, 1: 253-269.

Wilson, A.G. 1970. *Entropy in Urban and Spatial Modeling*. London: Pion.

Zipf, G.K. 1946. "The P1 P2/D Hypothesis: On the Intercity Movement of Persons." *American Sociological Review*, 11(6): 677-686.

Appendix 1: The *Air Connectivity Index* (ACI) 2007

Country	ACI	Rank	Country	ACI	Rank
Afghanistan	3.08%	115	Chile	1.79%	162
Albania	7.28%	25	China	5.70%	46
Algeria	6.64%	33	Christmas Island	1.99%	145
Angola	1.42%	189	Cocos (Keeling) Islands	1.48%	187
Anguilla	4.11%	73	Colombia	3.02%	117
Antigua and Barbuda	3.95%	82	Comoros	1.43%	188
Argentina	2.41%	133	Congo	1.57%	184
Armenia	3.99%	79	Cook Islands	0.54%	211
Aruba	3.88%	84	Costa Rica	3.24%	110
Australia	5.90%	44	Cote d'Ivoire	1.84%	158
Austria	9.40%	11	Croatia	9.06%	15
Azerbaijan	3.63%	98	Cuba	4.92%	52
Bahamas	5.87%	45	Cyprus	4.84%	54
Bahrain	4.42%	66	Czech Republic	9.87%	10
Bangladesh	2.79%	121	Dem. Republic of the Congo	1.64%	178
Barbados	3.37%	106	Denmark	9.11%	14
Belarus	6.67%	32	Djibouti	2.32%	135
Belgium	12.03%	4	Dominica	3.77%	91
Belize	3.78%	90	Dominican Republic	4.50%	62
Benin	1.93%	150	East Timor	1.30%	192
Bermuda	7.47%	23	Ecuador	2.39%	134
Bhutan	2.74%	123	Egypt	4.29%	69
Bolivia	1.56%	185	El Salvador	3.44%	105
Bosnia and Herzegovina	8.13%	21	Equatorial Guinea	1.73%	169
Botswana	1.84%	159	Eritrea	2.52%	130
Brazil	2.67%	125	Estonia	6.35%	37
British Virgin Islands	4.24%	71	Ethiopia	2.22%	136
Brunei Darussalam	2.57%	129	Falkland Islands	0.72%	203
Bulgaria	7.02%	28	Fiji	0.97%	199
Burkina Faso	2.02%	143	Finland	6.16%	39
Burundi	1.71%	170	France	11.64%	7
Cambodia	3.67%	95	French Polynesia	0.54%	210
Cameroon	1.73%	168	French Guiana	1.92%	151
Canada	13.44%	2	Gabon	1.61%	179
Cape Verde	1.77%	163	Gambia	1.91%	152
Cayman Islands	4.33%	68	Georgia	3.96%	80
Central African Republic	1.68%	173	Germany	12.11%	3
Chad	1.98%	147	Ghana	1.86%	156

Country	ACI	Rank	Country	ACI	Rank
Gibraltar	6.14%	40	Malaysia	3.91%	83
Greece	6.13%	41	Maldives	1.66%	175
Grenada and South Grenadines	3.46%	104	Mali	1.96%	148
Guadeloupe	3.87%	87	Malta	6.07%	42
Guatemala	3.50%	103	Marshall Islands	0.64%	207
Guinea	1.67%	174	Martinique	3.69%	94
Guinea Bissau	1.81%	161	Mauritania	2.13%	139
Guyana	2.46%	131	Mauritius	1.18%	195
Haiti	4.50%	63	Mexico	4.52%	60
Honduras	3.54%	102	Micronesia	0.82%	201
Hong Kong SAR, China	4.88%	53	Moldova	6.37%	36
Hungary	8.63%	17	Mongolia	2.68%	124
Iceland	3.87%	86	Montserrat	3.95%	81
India	3.82%	88	Morocco	5.27%	49
Indonesia	2.79%	122	Mozambique	1.65%	176
Iran	3.74%	93	Myanmar	3.19%	112
Iraq	3.99%	78	Namibia	1.74%	167
Ireland	8.48%	19	Nauru	0.71%	204
Israel	4.52%	61	Nepal	2.86%	119
Italy	9.03%	16	Netherlands	11.73%	6
Jamaica	4.35%	67	Netherlands Antilles	3.87%	85
Japan	5.28%	48	New Caledonia	1.37%	190
Jordan	4.44%	65	New Zealand	2.03%	142
Kazakhstan	2.61%	128	Nicaragua	3.35%	107
Kenya	2.05%	141	Niger	2.09%	140
Kiribati	0.63%	208	Nigeria	1.94%	149
Kuwait	4.14%	72	Niue	0.62%	209
Kyrgyzstan	2.65%	127	Norfolk Island	1.61%	180
Laos	3.64%	97	Norway	7.39%	24
Latvia	6.90%	30	Oman	3.76%	92
Lebanon	4.63%	58	Pakistan	3.23%	111
Lesotho	1.90%	154	Palau	1.49%	186
Liberia	1.58%	182	Panama	3.33%	108
Libya	5.07%	50	Papua New Guinea	1.10%	197
Lithuania	6.99%	29	Paraguay	1.86%	157
Luxembourg	11.74%	5	Peru	1.81%	160
Macau	4.77%	56	Philippines	3.13%	114
Macedonia	7.18%	27	Poland	8.16%	20
Madagascar	1.25%	193	Portugal	6.41%	35
Malawi	1.60%	181	Qatar	4.50%	64

Country	ACI	Rank	Country	ACI	Rank
Reunion	1.19%	194	Tuvalu	0.65%	206
Romania	6.77%	31	Uganda	1.89%	155
Russian Federation	5.30%	47	Ukraine	6.23%	38
Rwanda	1.76%	164	United Arab Emirates	4.77%	57
Saint Kitts and Nevis	4.03%	76	United Kingdom	11.55%	8
Saint Lucia	3.62%	99	United States	22.78%	1
Saint Vincent and Grenadines	3.56%	101	Uruguay	2.17%	137
Sao Tome and Principe	1.57%	183	Uzbekistan	2.82%	120
Saudi Arabia	4.00%	77	Vanuatu	1.11%	196
Senegal	1.98%	146	Venezuela	3.65%	96
Serbia	7.88%	22	Vietnam	3.80%	89
Seychelles	1.31%	191	Wallis and Futuna Islands	0.67%	205
Sierra Leone	1.65%	177	Western Samoa	3.06%	116
Singapore	4.09%	74	Yemen	2.65%	126
Slovakia	9.18%	13	Zambia	1.71%	171
Slovenia	9.39%	12	Zimbabwe	1.01%	198
Solomon Islands	0.96%	200			
Somalia	1.75%	165			
South Africa	3.57%	100			
South Korea	4.79%	55			
Spain	8.49%	18			
Sri Lanka	2.02%	144			
Sudan	2.44%	132			
Suriname	2.17%	138			
Swaziland	1.70%	172			
Sweden	7.20%	26			
Switzerland	10.76%	9			
Syria	4.54%	59			
Taiwan, China	4.25%	70			
Tajikistan	2.92%	118			
Tanzania	1.75%	166			
Thailand	4.06%	75			
Togo	1.90%	153			
Tonga	0.78%	202			
Trinidad and Tobago	3.26%	109			
Tunisia	6.61%	34			
Turkey	6.05%	43			
Turkmenistan	3.19%	113			
Turks and Caicos Islands	4.93%	51			

Appendix 2: Alternative Connectivity/Centrality Measures

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Afghanistan	0.27%	5.09%	0.03%	9	0.83	0.96	1.63	3.10	0.49
Albania	0.61%	17.10%	0.11%	13	0.97	0.99	1.82	1.92	0.56
Algeria	0.58%	23.32%	0.20%	23	0.68	0.98	1.68	1.85	0.61
Angola	0.17%	9.93%	0.02%	13	0.49	0.52	2.15	3.68	0.53
Anguilla	0.36%	5.59%	0.04%	4	1.00	1.00	1.12	3.45	0.50
Antigua and Barbuda	0.46%	19.48%	0.15%	19	0.43	0.46	2.53	3.78	0.58
Argentina	1.25%	20.35%	0.43%	21	0.49	0.84	2.10	3.42	0.58
Armenia	0.35%	18.59%	0.08%	17	0.84	0.98	1.55	2.79	0.59
Aruba	0.44%	8.35%	0.12%	7	0.81	1.00	1.32	3.14	0.52
Australia	3.55%	22.18%	0.67%	32	0.34	0.76	2.53	2.61	0.59
Austria	1.34%	49.04%	1.59%	61	0.49	0.92	3.05	0.76	0.82
Azerbaijan	0.37%	25.41%	0.11%	30	0.66	0.91	2.68	2.06	0.64
Bahamas	0.82%	12.39%	0.55%	7	0.76	0.99	0.20	2.33	0.55
Bahrain	0.65%	27.85%	0.33%	34	0.62	0.91	2.75	2.21	0.65
Bangladesh	0.41%	16.95%	0.15%	21	0.72	0.93	2.50	2.35	0.57
Barbados	0.44%	15.19%	0.17%	15	0.51	0.81	2.24	3.56	0.58
Belarus	0.53%	19.83%	0.08%	16	0.87	0.97	2.32	2.07	0.58
Belgium	1.74%	48.51%	1.34%	64	0.41	0.94	3.06	0.51	0.81
Belize	0.31%	6.44%	0.04%	5	1.00	1.00	0.98	2.75	0.50
Benin	0.26%	4.97%	0.03%	14	0.58	0.61	2.32	4.93	0.50
Bermuda	0.46%	10.19%	0.07%	3	1.00	1.00	0.45	2.05	0.55
Bhutan	0.21%	1.39%	0.01%	3	1.00	1.00	0.90	4.71	0.43
Bolivia	0.23%	9.62%	0.05%	7	0.90	0.91	1.86	3.37	0.52
Bosnia and Herzegovina	0.67%	15.26%	0.06%	11	0.96	0.96	2.10	2.55	0.55

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Botswana	0.23%	2.44%	0.04%	3	0.67	0.97	0.44	5.05	0.41
Brazil	1.66%	33.24%	0.48%	27	0.47	0.73	2.53	2.21	0.68
British Virgin Islands	0.41%	6.05%	0.13%	4	1.00	1.00	0.58	2.42	0.50
Brunei Darussalam	0.27%	8.94%	0.06%	10	0.96	0.99	2.05	2.70	0.49
Bulgaria	0.66%	28.59%	0.25%	27	0.83	0.99	2.61	1.08	0.65
Burkina Faso	0.20%	4.43%	0.02%	10	0.80	0.84	2.10	4.80	0.49
Burundi	0.17%	1.57%	0.02%	5	0.90	0.97	1.19	6.41	0.46
Cambodia	0.45%	8.38%	0.17%	9	0.86	0.96	1.87	3.13	0.48
Cameroon	0.23%	6.31%	0.05%	15	0.53	0.62	2.39	4.28	0.50
Canada	1.84%	53.10%	2.93%	56	0.35	0.98	0.98	1.65	0.82
Cape Verde	0.18%	15.56%	0.02%	10	0.64	0.89	1.69	3.26	0.58
Cayman Islands	0.33%	8.37%	0.06%	6	0.80	0.97	0.85	2.43	0.51
Central African Republic	0.14%	3.25%	0.00%	5	0.60	0.72	1.56	4.52	0.49
Chad	0.17%	3.32%	0.01%	8	0.50	0.75	1.62	3.76	0.49
Chile	0.79%	19.19%	0.23%	20	0.51	0.90	2.15	3.04	0.57
China	4.08%	49.96%	2.33%	55	0.44	0.89	2.56	1.72	0.79
Christmas Island	0.17%	1.92%	0.00%	3	0.67	0.71	1.08	6.00	0.44
Cocos (Keeling) Islands	0.12%	1.05%	0.00%	2	1.00	1.00	0.69	7.14	0.37
Colombia	0.73%	18.59%	0.35%	17	0.70	0.96	2.10	2.46	0.57
Comoros	0.15%	1.51%	0.01%	5	0.50	0.55	1.49	5.90	0.39
Congo	0.24%	5.55%	0.03%	11	0.58	0.70	2.01	5.13	0.51
Cook Islands	0.10%	5.01%	0.01%	4	0.83	0.96	0.82	5.23	0.50
Costa Rica	0.50%	15.27%	0.22%	15	0.67	0.89	1.94	2.43	0.54
Coted'Ivoire	0.85%	28.49%	0.29%	24	0.80	0.96	2.45	1.23	0.64
Croatia	0.38%	33.28%	0.10%	27	0.46	0.72	2.69	1.73	0.68
Cuba	0.59%	27.66%	0.29%	32	0.75	0.97	2.53	1.82	0.64
Cyprus	1.09%	41.84%	0.86%	45	0.64	0.93	3.18	0.72	0.77

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Czech Republic	0.26%	6.12%	0.03%	11	0.71	0.76	2.08	4.54	0.50
Democratic Republic of the Congo	1.28%	44.47%	1.43%	42	0.66	0.97	2.90	0.97	0.80
Denmark	0.23%	4.91%	0.03%	9	0.67	0.76	1.71	5.10	0.51
Djibouti	0.41%	5.94%	0.04%	6	0.80	0.87	1.66	3.93	0.50
Dominica	0.59%	26.35%	0.30%	22	0.52	0.93	1.28	1.88	0.63
Dominican Republic	0.14%	2.11%	0.02%	2	1.00	1.00	0.26	5.03	0.37
East Timor	0.42%	12.75%	0.15%	13	0.83	0.98	1.92	2.74	0.53
Ecuador	0.78%	43.45%	0.48%	52	0.49	0.79	3.35	1.49	0.77
Egypt	0.42%	10.08%	0.15%	9	0.72	0.94	1.73	2.89	0.51
El Salvador	0.17%	11.51%	0.01%	8	0.68	0.72	1.95	2.82	0.54
Equatorial Guinea	0.20%	4.47%	0.01%	7	0.76	0.88	1.69	3.97	0.50
Eritrea	0.68%	27.37%	0.16%	20	0.95	0.98	2.64	1.85	0.64
Estonia	0.41%	21.01%	0.14%	35	0.37	0.48	3.25	2.82	0.60
Ethiopia	0.07%	0.53%	0.00%	2	1.00	1.00	0.69	5.07	0.37
Falkland Islands	0.29%	9.39%	0.05%	14	0.31	0.72	2.04	3.97	0.52
Fiji	0.91%	43.70%	0.72%	38	0.72	0.95	2.91	1.08	0.78
Finland	2.46%	65.34%	5.03%	110	0.25	0.87	3.23	0.44	0.92
France	0.18%	3.58%	0.02%	4	0.67	0.62	1.18	4.56	0.50
French Guiana	0.16%	7.63%	0.02%	7	0.62	0.86	1.40	2.52	0.51
French Polynesia	0.18%	3.99%	0.02%	8	0.54	0.79	1.66	5.60	0.49
Gabon	0.22%	9.56%	0.01%	7	0.52	0.79	1.40	5.18	0.54
Gambia	0.33%	17.74%	0.05%	15	0.80	0.92	2.35	2.36	0.58
Georgia	3.62%	61.12%	8.22%	98	0.30	0.93	3.30	0.46	0.91
Germany	0.31%	20.86%	0.09%	22	0.45	0.63	2.48	3.83	0.63
Ghana	0.55%	7.16%	0.02%	2	1.00	1.00	0.66	2.06	0.51
Gibraltar	1.09%	43.24%	0.95%	48	0.60	0.95	2.85	0.88	0.78
Greece	0.36%	6.67%	0.06%	9	0.69	0.90	1.65	4.67	0.50

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Grenada and South Grenadines	0.49%	11.39%	0.18%	13	0.47	0.86	1.45	4.34	0.54
Guadeloupe	0.43%	10.02%	0.16%	10	0.76	0.93	1.84	2.73	0.53
Guatemala	0.18%	4.11%	0.02%	7	0.62	0.84	1.72	4.75	0.49
Guinea	0.17%	1.14%	0.01%	2	1.00	1.00	0.30	6.20	0.43
Guinea Bissau	0.22%	6.68%	0.03%	5	0.80	1.00	1.03	5.07	0.51
Guyana	0.37%	9.22%	0.07%	11	0.67	0.91	1.20	2.67	0.51
Haiti	0.36%	8.24%	0.09%	9	0.72	0.96	1.55	3.17	0.50
Honduras	2.11%	43.01%	1.66%	44	0.54	0.90	2.61	1.64	0.74
Hong Kong SAR, China	0.95%	42.72%	0.66%	44	0.64	0.91	3.20	0.82	0.78
Hungary	0.41%	30.59%	0.15%	16	0.98	0.98	2.37	1.03	0.69
Iceland	2.08%	42.96%	1.08%	53	0.47	0.80	3.09	1.49	0.75
India	1.11%	17.57%	0.49%	16	0.76	0.92	1.88	2.86	0.56
Indonesia	0.53%	31.36%	0.25%	35	0.61	0.94	2.28	2.16	0.66
Iran	0.32%	3.56%	0.05%	7	0.86	0.94	1.20	4.45	0.49
Iraq	1.21%	37.96%	1.28%	33	0.72	0.99	2.05	0.97	0.73
Ireland	0.62%	40.46%	0.32%	41	0.62	0.92	3.26	0.73	0.76
Israel	3.26%	56.83%	4.41%	87	0.35	0.93	3.00	0.52	0.88
Italy	0.28%	7.37%	0.07%	22	0.42	0.54	2.57	4.46	0.51
Jamaica	0.48%	19.91%	0.20%	16	0.48	0.93	1.06	2.03	0.59
Japan	3.96%	42.85%	1.79%	42	0.49	0.94	2.49	1.39	0.75
Jordan	0.58%	37.42%	0.27%	38	0.64	0.74	3.14	2.10	0.73
Kazakhstan	0.43%	23.60%	0.16%	25	0.64	0.93	2.49	1.90	0.61
Kenya	0.64%	27.23%	0.24%	37	0.35	0.59	2.97	3.06	0.61
Kiribati	0.09%	4.67%	0.00%	4	0.33	0.25	1.31	6.75	0.50
Kuwait	0.60%	30.79%	0.31%	33	0.70	0.90	2.74	2.28	0.69
Kyrgyzstan	0.27%	5.13%	0.03%	9	0.83	0.96	1.46	3.32	0.50
Laos	0.33%	3.76%	0.04%	4	1.00	1.00	1.20	3.82	0.46

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Latvia	0.66%	34.46%	0.26%	28	0.80	0.92	2.98	1.35	0.71
Lebanon	0.54%	28.38%	0.19%	40	0.58	0.87	3.00	1.76	0.65
Lesotho	0.17%	1.53%	0.02%	1	0.00	0.00	0.00	5.25	0.41
Liberia	0.16%	1.25%	0.01%	5	0.70	0.66	1.51	6.02	0.40
Libya	0.43%	24.46%	0.12%	33	0.48	0.78	3.06	1.95	0.61
Lithuania	0.61%	29.20%	0.18%	22	0.95	0.98	2.83	1.32	0.65
Luxembourg	1.38%	41.84%	0.26%	52	0.52	0.91	2.97	0.65	0.81
Macau	1.17%	11.32%	0.41%	11	0.87	1.00	1.56	2.69	0.54
Macedonia	0.58%	11.77%	0.05%	11	0.96	0.98	2.27	2.51	0.53
Madagascar	0.20%	7.01%	0.04%	8	0.61	0.78	1.72	4.72	0.51
Malawi	0.18%	7.01%	0.02%	8	0.89	0.93	1.80	4.28	0.51
Malaysia	1.02%	36.69%	0.78%	40	0.52	0.83	2.79	2.01	0.67
Maldives	0.25%	16.96%	0.05%	14	0.85	0.90	2.03	3.11	0.58
Mali	0.23%	5.08%	0.04%	13	0.62	0.72	2.15	4.50	0.50
Malta	0.52%	28.76%	0.14%	25	0.85	0.98	2.30	1.07	0.65
Marshall Islands	0.10%	4.22%	0.01%	3	0.67	0.85	0.96	4.65	0.50
Martinique	0.42%	9.30%	0.09%	12	0.53	0.64	1.59	4.33	0.54
Mauritania	0.20%	6.82%	0.02%	8	0.89	0.99	1.69	3.75	0.51
Mauritius	0.28%	21.42%	0.06%	16	0.71	0.69	2.20	3.37	0.60
Mexico	3.29%	29.99%	1.51%	24	0.53	0.98	0.79	1.99	0.65
Micronesia	0.09%	4.19%	0.00%	2	1.00	1.00	0.68	4.80	0.50
Moldova	0.48%	15.80%	0.05%	14	0.98	1.00	2.36	2.11	0.58
Mongolia	0.23%	4.88%	0.02%	5	1.00	1.00	1.36	2.77	0.49
Montserrat	0.41%	2.36%	0.02%	1	0.00	0.00	0.00	6.51	0.37
Morocco	0.74%	33.59%	0.48%	37	0.47	0.91	2.42	1.27	0.70
Mozambique	0.31%	3.56%	0.08%	7	0.43	0.95	0.62	4.83	0.45
Myanmar	0.31%	6.05%	0.06%	7	0.95	1.00	1.36	3.12	0.48

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Namibia	0.26%	10.79%	0.06%	5	0.60	0.96	0.64	4.61	0.53
Nauru	0.08%	0.51%	0.00%	2	0.00	0.00	0.68	8.88	0.34
Nepal	0.29%	8.62%	0.06%	13	0.78	0.93	2.09	2.98	0.50
Netherlands	1.89%	59.65%	2.56%	85	0.34	0.92	3.07	0.38	0.90
Netherlands Antilles	0.64%	14.64%	0.36%	20	0.39	0.61	2.26	3.37	0.55
New Caledonia	0.14%	3.54%	0.02%	7	0.62	0.83	1.62	4.22	0.44
New Zealand	1.46%	15.91%	0.30%	21	0.34	0.81	1.50	3.72	0.55
Nicaragua	0.33%	8.46%	0.07%	8	0.89	0.93	1.70	3.20	0.51
Niger	0.18%	4.14%	0.01%	8	0.93	0.99	1.94	4.84	0.49
Nigeria	0.36%	26.48%	0.11%	30	0.48	0.70	2.84	2.36	0.66
Niue	0.07%	0.62%	0.00%	1	0.00	0.00	0.00	5.81	0.35
Norfolk Island	0.11%	1.68%	0.01%	2	1.00	1.00	0.47	4.68	0.37
Norway	1.04%	36.49%	0.94%	32	0.75	0.99	2.59	1.18	0.73
Oman	0.54%	12.13%	0.25%	18	0.76	0.99	1.92	2.82	0.54
Pakistan	0.57%	26.38%	0.27%	29	0.66	0.95	2.21	2.34	0.64
Palau	0.15%	6.57%	0.01%	4	1.00	1.00	1.29	2.79	0.51
Panama	0.61%	14.55%	0.30%	21	0.50	0.82	2.44	2.73	0.54
Papua New Guinea	0.15%	8.00%	0.02%	6	0.73	0.95	0.86	4.27	0.52
Paraguay	0.23%	3.55%	0.05%	5	1.00	1.00	1.25	4.43	0.50
Peru	0.64%	16.67%	0.20%	15	0.71	0.87	2.45	2.63	0.56
Philippines	0.75%	22.47%	0.36%	23	0.60	0.92	2.59	1.74	0.59
Poland	1.08%	38.58%	1.03%	33	0.81	0.98	2.77	0.69	0.74
Portugal	1.08%	39.67%	1.02%	38	0.53	0.96	2.42	0.85	0.74
Qatar	0.73%	38.42%	0.43%	48	0.51	0.80	3.25	1.72	0.70
Reunion	0.25%	5.13%	0.05%	11	0.55	0.77	1.55	5.06	0.51
Romania	0.75%	34.90%	0.47%	30	0.85	0.99	2.57	1.09	0.72
Russian Federation	1.66%	50.56%	1.35%	64	0.46	0.76	3.45	1.12	0.83

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Rwanda	0.19%	2.12%	0.02%	6	0.80	0.96	1.14	5.86	0.49
Saint Kitts and Nevis	0.41%	6.27%	0.08%	6	0.73	0.99	1.29	3.77	0.50
Saint Lucia	0.42%	12.43%	0.09%	11	0.64	0.86	2.01	3.54	0.55
Saint Vincent and Grenadines	0.38%	6.53%	0.06%	8	0.89	0.99	1.69	4.65	0.50
Sao Tome and Principe	0.14%	0.83%	0.00%	5	0.30	0.14	1.47	6.28	0.43
Saudi Arabia	0.74%	35.62%	0.45%	44	0.52	0.79	3.13	2.11	0.70
Senegal	0.39%	26.38%	0.12%	25	0.41	0.48	2.95	2.96	0.64
Serbia	0.68%	25.90%	0.23%	28	0.72	0.82	2.73	2.00	0.62
Seychelles	0.15%	16.72%	0.02%	11	0.84	0.90	2.26	2.32	0.58
Sierra Leone	0.17%	2.16%	0.01%	7	0.52	0.50	1.82	5.63	0.46
Singapore	1.81%	43.00%	1.21%	48	0.46	0.83	2.91	1.76	0.74
Slovakia	1.02%	26.15%	0.14%	18	0.88	0.98	2.37	1.51	0.63
Slovenia	0.85%	28.50%	0.16%	25	0.77	0.91	2.83	1.33	0.65
Solomon Islands	0.11%	1.90%	0.01%	5	0.40	0.61	1.42	5.59	0.37
Somalia	0.17%	2.25%	0.02%	5	0.90	0.83	1.29	5.62	0.47
South Africa	2.15%	36.72%	0.53%	48	0.31	0.30	3.11	3.32	0.69
South Korea	1.73%	43.64%	1.32%	41	0.51	0.94	2.30	1.73	0.76
Spain	4.11%	52.19%	4.77%	71	0.36	0.95	2.59	0.66	0.82
Sri Lanka	0.51%	19.62%	0.17%	18	0.86	0.96	2.03	2.87	0.60
Sudan	0.31%	10.13%	0.09%	19	0.68	0.85	2.44	3.11	0.53
Suriname	0.17%	1.96%	0.01%	4	0.67	0.88	0.95	4.29	0.48
Swaziland	0.20%	1.93%	0.03%	2	1.00	1.00	0.36	5.10	0.41
Sweden	1.19%	40.25%	1.20%	37	0.72	0.98	2.84	1.02	0.75
Switzerland	1.56%	51.07%	2.13%	63	0.48	0.96	2.90	0.50	0.84
Syria	0.52%	31.01%	0.16%	37	0.63	0.82	3.11	1.96	0.66
Taiwan, China	1.37%	29.92%	0.91%	27	0.60	0.87	2.33	1.99	0.65
Tajikistan	0.28%	5.02%	0.04%	9	0.86	0.95	1.20	3.63	0.49

Country	ACI (a=100)	ACI (Nodal Dist.)	Traffic Share	No. of Links	Clustering Coefficient	Weighted Clustering Coefficient	Theil Index	Kullback-Leibler Distance	Closeness Centrality
Tanzania	0.35%	12.53%	0.10%	15	0.52	0.89	1.76	4.23	0.55
Thailand	1.96%	50.93%	1.13%	57	0.44	0.78	3.25	1.46	0.80
Togo	0.23%	4.04%	0.02%	9	0.78	0.72	1.88	4.97	0.50
Tonga	0.11%	2.03%	0.01%	4	1.00	1.00	1.17	5.27	0.37
Trinidad and Tobago	0.42%	16.56%	0.15%	16	0.51	0.81	2.21	3.65	0.59
Tunisia	0.60%	29.42%	0.22%	32	0.67	0.95	2.50	1.36	0.65
Turkey	1.17%	53.29%	1.04%	75	0.41	0.85	3.22	0.89	0.84
Turkmenistan	0.23%	1.81%	0.01%	4	1.00	1.00	1.29	4.15	0.47
Turks and Caicos Islands	0.36%	6.80%	0.05%	5	0.80	0.91	1.40	3.11	0.50
Tuvalu	0.07%	0.38%	0.00%	1	0.00	0.00	0.00	7.56	0.34
Uganda	0.24%	9.05%	0.04%	12	0.65	0.87	1.97	4.27	0.53
Ukraine	0.74%	41.91%	0.47%	49	0.58	0.91	3.05	1.52	0.77
United Arab Emirates	2.32%	55.64%	1.66%	79	0.35	0.73	3.54	1.57	0.85
United Kingdom	3.91%	63.02%	8.01%	102	0.28	0.91	3.18	0.45	0.90
United States	19.42%	69.60%	8.88%	101	0.20	0.61	3.10	1.16	0.91
Uruguay	0.38%	5.78%	0.11%	6	0.87	1.00	0.91	4.51	0.47
Uzbekistan	0.35%	25.29%	0.11%	25	0.69	0.88	2.33	2.37	0.63
Vanuatu	0.13%	2.27%	0.01%	5	0.80	0.93	1.39	5.09	0.37
Venezuela	0.69%	29.09%	0.39%	31	0.42	0.84	2.15	1.99	0.65
Vietnam	0.67%	19.36%	0.36%	17	0.71	0.91	2.44	2.34	0.57
Wallis and Futuna Islands	0.07%	0.39%	0.00%	2	1.00	1.00	0.56	7.27	0.34
Western Samoa	0.25%	6.14%	0.03%	5	0.90	0.94	0.91	2.59	0.50
Yemen	0.29%	19.83%	0.07%	22	0.64	0.82	2.62	2.89	0.58
Zambia	0.36%	6.41%	0.10%	9	0.81	0.95	1.30	4.49	0.50
Zimbabwe	0.29%	9.09%	0.07%	11	0.53	0.92	1.30	4.56	0.51

Appendix 3: Alternative Measures of Connectivity

The air transport network can be envisaged as a collection of nodes, each of which represents a country.[18] The structure of the network is given by the so-called adjacency or link matrix L, where a typical element $L_{ij} = 1$ if there is a direct air link going from country i to country j; otherwise, it is equal to zero. In practice the adjacency matrix is very close to symmetric for the air transportation network, so we can assume $L_{ij} = L_{ji}$.

The activity on each link of the network is measured in terms of the flow on the network, which is summarized in the matrix X. A typical element X_{ij} is the total flow from country i to country j, and X_{ji} is the flow in the other direction from j to i (transport networks are bi-directional). There are two primary choices for X in the case of air transport, namely frequency (number of flights per unit time) or capacity (number of seats per unit time).

Intuitive Metrics

A first intuitive approach to connectivity is simply counting the number of connections by node, often referred to as degree centrality.[19] For instance, using the elements of the adjacency matrix:

$$(A1)\ L_i = \sum_j L_{ij}$$

An alternative approach is to use the total outflow or inflow from each node:[20]

$$(A2)\ X_{i.} = \sum_j X_{ij}\ \ or\ X_{.i} = \sum_j X_{ji}$$

[18] In reality, of course, many countries have multiple international airports. In principle, our analysis could be replicated at the level of individual airports, so as to obtain connectivity scores at the airport level. From a policy point of view, however, it is primarily country scores that are of the most interest, so we abstract from the airport dimension in this paper.

[19] In the context of approximately symmetric link and flow matrices, the distinction commonly made between in- and out-degree centrality is of limited relevance, and so we present only one measure.

[20] We use the dot convention to designate summation over an index.

Such simple output indicators obviously contain significant information, but by definition remain local rather than global, in the sense that they do not incorporate information from the full structure of the network. Counts of total outflow or total destinations do not incorporate information on the network beyond the immediate neighborhood of each node. That is, they do not capture indirect connections, which are of particular importance in air transport where connecting flights are common. Nor do these types of measures reflect the interactions among nodes. A further shortcoming of the most common form of degree centrality—which is based on the adjacency matrix—is that it does not combine information on the number and intensity (flow strength) of connections. Both are important, however, for measuring connectivity from a policy perspective.

In order to extract information on connectivity from simple output indicators, UNCTAD (2007) and Pearce (2007)—see further in the main text—combine them with other node-specific outcomes, like GDP or GDP per capita. On the one hand, these compound indicators are arbitrary definitions, and thus not entirely satisfactory from a formal standpoint: there is a size bias—larger nodes are considered to be better connected—and these indicators are not global metrics rooted in network modeling. On the other hand, they are legitimate indicators based on expert knowledge and a professional understanding of the factors that influence the flows measured. The rankings produced are consistent with analytical priors.

Concentration Metrics

A more sophisticated approach than simple counting, and one which makes use of more information from the matrix, is to use concentration indices such as the Herfindhal or Theil indices of the flows to and from a node in the network. The Theil index is the entropy of the relative weight of the outflows or inflows from node i to the neighboring node js, and is given by the following expression:

$$(A3)\ Theil\ Index = -\sum_j x_{ij}\ln x_{ij} = -\sum_j \frac{X_{ij}}{X_{i.}}\ln\frac{X_{ij}}{X_{i.}}$$

These kinds of concentration indicators have no dimension. However, they do not incorporate information about the structure of the network. They are essentially local measures, because they only use information from a single node. Relative entropy measures are better in this respect, since they compare the composition of flows to or from a node to an average composition. The most commonly used is the Kullback-Leibler distance, which is a modification of the Theil index. In this case, the reference composition should be the relative weight of each node (excluding the contribution of node i since there is no flow from a node to itself), or $\frac{X_{.j}}{X_{..}-X_{.i}}$. Then:

$$(A4)\ KL\ Distance = \sum_{j\neq i}\frac{X_{ij}}{X_{i.}}\ln\frac{\frac{X_{ij}}{X_{i.}}}{\frac{X_{.j}}{X_{..}-X_{.i}}}$$

A variation of this definition will be significant in the approach to connectivity developed in this paper—see Section 4 of the main text.

Clustering

Clustering is an important concept in network theory. It is essentially a topological concept. The clustering coefficient of node i is an intuitive measure of how well connected the nodes in the neighborhood of i are. This number, comprised between 0 and 1, is given by:

$$(A5)\ Clustering_i = \frac{No.\ of\ triangles\ attached\ to\ i}{No.\ of\ possible\ triangles\ attached\ to\ i}$$

Or using the notation above for the adjacency matrix:

$$(A6)\ Clustering_i = \frac{\sum_{j\neq k} L_{ij}L_{jk}L_{ki}}{L_{i.}L_{.i} - \sum_j L_{ij}L_{ji}};\ 0 \leq Clustering_i \leq 1$$

A hub or a bridge between clusters will have a small clustering coefficient. A node connected to several hubs will have a clustering coefficient of close to one. This definition applies equally to directional and non-directional networks; that is, a triangle is counted once for each direction it can be run, ijk and ikj. A variant of this definition applies to weighted networks (Barrat et al., 2004). It gives higher weight to triangles with higher flows to the node:

$$(A7)\ Clustering_i^w = \frac{\sum_{jk\,j\neq k} x_{ij} L_{jk} x_{ki}}{1 - \sum_j x_{ij} x_{ji}};\ 0 \leq Clustering_i^w \leq 1$$

This definition of clustering has already been applied to airports (Barrat et al., 2004). By construction, it not only includes information about interactions with neighboring nodes, but also about interactions between neighboring nodes. However, the analysis of interactions does not go beyond the immediate partners within a cluster. It is not clear what high or low clustering means for a node's global connectivity: it is rather a measure of local connectivity, and the degree of belonging to a local cluster. A hub will tend to exhibit local clustering, but so does an isolated node bridging between clusters. Visual inspection of the results of Barrat et al. (2004) and our computations in the main text do not support clustering as a candidate for connectivity in a broader policy context.

Centrality and PageRank

Centrality measures the importance of a node in relation to all the other nodes in the network. The more important a destination is for its neighbors, the more central it is. There are several implementations of the concept. All of them are based on the concept of random movement through the network from one node to the next, an image directly relevant to transportation on a hub and spoke network. The most central nodes are the ones most often encountered when moving randomly through the network. The "pointing" is defined by the transition probabilities matrix from a modified adjacency matrix A:

$$(A8a)\ P(j \to i|j) = p_{ji} = \frac{A_{ji}}{A_{j.}}\ (outward\ probability)$$

$$(A8b)\ P(j \to i|i) = q_{ij} = \frac{A_{ji}}{A_{.i}}\ (inward\ probability)$$

Depending on the nature of a given problem, the modified adjacency matrix can be the adjacency matrix with elements equal to unity if a link is active, or the flow matrix X.

Eigenvector centrality, one of the most elegant tools in network theory, is a linear implementation of this concept. It has been made famous by the Google algorithm called PageRank (Brin and Page, 1988; Bryan and Liese, 2006). PageRank gives the highest ranking to the most probable pages when clicking on relevant hyperlinks from web pages. Formally, ranking pages is analogous to the current problem (replace page by country, hyperlink by air route, and web-surfer by passenger).

Eigenvector centrality Y_i is the average occupation of individual nodes when moving randomly through the network:

$$(A9)\ Y_i^{n+1} = \sum_{j \neq i} p_{ji} Y_j^n$$

From an initial occupation, Y converges to the main eigenvector of the transposed transition matrix. According to the Frobenius theorem, for a matrix with positive coefficients, the leading eigenvalue is positive, and the corresponding leading eigenvector coordinates have the same sign. Furthermore, since the probability of node occupation is conserved when moving randomly on the network:

$$(A10)\ \sum_i Y_i^{n+1} = \sum_{ij} p_{ji} Y_j^n = \sum_j \sum_i p_{ji} Y_j^n = \sum_j Y_j^n$$

The eigenvalue is therefore one, and the procedure converges to a mean occupation Y_i. A generalized version of the algorithm introduces a damping factor d, which corresponds to a probability of stopping at an intermediate node. In the transport context, a rationale for including such a damping factor could be to impose a cost penalty for transit or transshipment at a hub. This approach gives:

$$(A11)\ Y_i^{n+1} = (1-d)\sum_{j\neq i} p_{ji} Y_j^n + dY_i^0$$

For practical purposes, d would be the inverse of the "diameter" of the network, i.e. the number of links in the shortest path between the furthest pair of nodes. Transport networks typically have a diameter of around four, which gives $d = \frac{1}{4}$. The PageRank algorithm has been implemented for web pages and journal citations using the link matrix as the adjacency matrix, i.e. $p_{ij} = \frac{L_{ij}}{L_{i.}}$.

The networks considered by measures such as eigenvector centrality and PageRank are unidirectional, i.e. $L_{ij} = 0\ if\ L_{ji} = 1$. In such cases, the algorithm is very successful at identifying the nodes which are most referenced eventually in cascade by the others. In the case when the activity between nodes is measured by bi-directional flows, as in transportation or trade, $A_{ij} = X_{ij}$. When flows are symmetric, as in the case of transportation capacity, or just balanced (total inflows equal total outflows), PageRank only produces the trivial result that centrality is simply proportional to total out(in)flow or "market share". To see this, let connectivity C_i be the ratio of eigenvector centrality to total outflow, i.e. $Y_i = X_{.i}C_i$. Then:

$$(A12)\ C_i^{n+1} = \sum_{j\neq i} \frac{X_{ji}}{X_{.i}}\frac{X_{.j}}{X_{j.}} C_j^n = \sum_{j\neq i} q_{ij}\frac{X_{.j}}{X_{j.}} C_j^n$$

When flows are balanced, as is the case for a symmetric flow matrix, $X_{ij} = X_{ji}$ and $X_{.j} = X_{j.}$. As a result, $C_i^{n+1} = \sum_{j\neq i} q_{ij} C_j^n$, for which a solution is $C_i = 1$ and $Y_i = X_{i.} = X_{.i}$. In the damped version of the model, connectivity is essentially a rescaling of the step zero value, i.e. $C_i \approx (1-d) + dC_i^0$.

Although eigenvector centrality is very appealing as a measure of connectivity, there is no easy way to adapt the model to produce non-trivial results for an approximately symmetric flow matrix, as in the case of air transport. (In the data—see main text— the correlation between inbound and outbound traffic is 0.998, which reflects an approximate bilateral balance in traffic.) Intuitively, linear procedures will not break the symmetry between inflows and outflows, and will produce the same centrality measures for inflows and outflows proportional to total flows.

Nodal Distance and Closeness Centrality

Another application of centrality, closeness centrality, defines distance or the separation between two nodes in the network $d(i,j)$ as the number of steps between them. One definition (Newman, 2008) sets distance in terms of the minimum number of steps required to reach j from i (nodal distance). An alternative definition (Newman, 2008) treats the distance between two nodes as the average number of steps required to reach j for the first time from i when moving randomly through the network.

Since bilateral separation is at least equal to unity, the inverse of average distance—referred to as accessibility or closeness centrality (Newman, 2008)—is a number between zero and one. A higher value indicates that a node is closer to the rest of the network, and is thus more connected. The nodes are weighted according to their outflows (excluding i), thus:

$$(A13)\ C_i = \frac{1}{\sum_{j \neq i} d(i,j) \frac{X_{.j}}{X_{..} - X_{.i}}}$$

This definition is relatively simple to implement, at least in the case of the geodetic distance given by the minimal separation.[21] It is more complex (Newman, 2008) when random separation is applied, given the obvious constraint of non-return. However, the concept of separation distance is topological; it

[21] This definition requires less detailed information than the adjacency matrix: just the link matrix, and total outflows.

imperfectly emulates real distance, time, or some other bilateral cost that is an important determinant of activity in real-life economic networks such as air transport. Furthermore, this concept becomes trivial (nodal distance of one) when the network is very connected and the adjacency matrix has mostly ones. This issue does not arise in the context of air transport, for which the adjacency matrix is sparse, but is more of a concern for possible applications in other areas, such as international trade.