



UNIVERSITY
OF TRENTO - Italy

Department of Economics

Global Networks of Trade and Bits

Massimo Riccaboni, Alessandro Rossi, Stefano Schiavo

n.
8/2011



UNIVERSITY
OF TRENTO - Italy

Department of Economics

Global Networks of Trade and Bits

Massimo Riccaboni, Alessandro Rossi, Stefano Schiavo

n.
8/2011

Global Networks of Trade and Bits

Massimo Riccaboni* Alessandro Rossi† Stefano Schiavo‡

September 13, 2011

Abstract

Considerable efforts have been made in the recent years to produce detailed topologies of the Internet. While Internet topology data have brought to the attention of a wide and somehow diverse audience of scholars they have been so far overlooked by economic analyses. In this paper we suggest that such data could be effectively treated as a proxy to characterize the size of the “digital economy” activities at national country level: we therefore analyze the topological structure of the network of trade in digital services (trade in bits) and compare it with that of the more traditional flow of manufactured goods across countries. To perform a meaningful comparison across networks with different characteristics we define a common null model for the number of connections among each country-pair, based on the hypergeometric distribution. Original data are thus filtered using different thresholds so that we focus our attention on the strongest links only, i.e., on links that represent a significant departure from the stochastic benchmark. We find that trade in bits displays a more sparse and less hierarchical network structure, which is more similar to trade in high-skill manufactured goods than total trade. Last, distance plays a more prominent role in shaping the network of international trade in physical goods than trade in digital services.

JEL Codes: F14, L86, O33

Keywords: Internet, hypergeometric, international trade, network analysis, distance

The authors thank participants to the Trade NetWorkShop 3.0 held in Trento in December 2010 for insightful comments on an earlier draft of the paper. They blame each other for all remaining mistakes.

*LIME, IMT Institute of Advanced Studies, Lucca. e-mail: mriccaboni@gmail.com

†DISA, University of Trento. e-mail: alessandro.rossi@unitn.it

‡DECO, University of Trento and OFCE-DRIC. e-mail: stefano.schiavo@unitn.it

1 Introduction

The impact of the Internet on the global economy has been largely debated. Two main effects have been considered: (a) the death of distance conjecture (Cairncross 1997, Friedman 2005, Leamer 2007); (b) the global digital divide. To test these two claims concerning the impact of distance on information and information-intensive goods as well as access of less developed countries to the digital web and the consequences of the digital divide on trade and economic growth, scholars should focus on the coevolution of the trade and computer networks. Such analysis would also contribute to shed new light on the impact of foreign direct investments (FDIs) and offshoring on trade. Indeed, an increase of horizontal and vertical FDIs implies a surge in the coordination and information costs. Thus, we should expect that information flows should increase and partially substitute trade flows (Navaretti & Venables 2004).

The standard explanation on why distance matters for trade is that transport costs are increasing in distance. Access to information about foreign markets and potential trading partners is also facilitated by proximity. This implies that major improvements in information and communication technologies (ICTs) should reduce the importance of distance for trade. The death of distance conjecture has been scrutinized and largely questioned since there is evidence that the negative impact of distance on trade has not decreased after the Internet revolution (Freund & Weinhold 2004, Disdier & Head 2008, Berthelon & Freund 2008). Conversely, gravity equation estimates on long time horizons find an increase of the negative impact of distance on trade (Coe 2002, Brun et al. 2005). Moreover, it has been found that Internet relationships are also constrained by distance (Blum & Goldfarb 2006, Mok et al. 2007, Hortaçsu et al. 2009).

Computer scientists have analyzed the structural properties and the growth of the world network of computers since the late 1990s (Faloutsos et al. 1999, Huberman & Adamic 1999, Pastor-Satorras & Vespignani 2004). More recently, the same network analysis techniques have been applied to study the world trade network (Serrano & Boguná 2003, Garlaschelli & Loffredo 2004, 2005, Fagiolo et al. 2008). Despite the two streams of literature adopting the same methodological approach, the theoretical discussion about the impact of the Internet on world trade has not been tested so far. The main reason why the world networks of computers and trade have never been jointly analyzed so far has to do with the different levels of aggregation at which the two webs are analyzed and the lack of methodological tools to compare networks and to study networks of networks. The Internet topology has been typically investigated at the level of IP addresses. A complete coverage of the world network is still lacking and only recently aggregate data at the level of Autonomous Systems (AS) and country level data have been released. Conversely, international statistics on the world trade of goods have long been compiled, and in recent years disaggregated data at the level of single exporting firms have become

available. Moreover, only recently the literature on complex network analysis has start focusing on weighted networks (Newman 2004, Fagiolo et al. 2008, Bhattacharya et al. 2008) and the coevolution of multiple interdependent networks (Buldyrev et al. 2010).

AS topology data seems to represent a very promising data source to measure the economic involvement in the digital economy of the various national economic systems, in a field where traditional data are difficult to collect or not available at all. For instance, while some statistics on the pervasiveness of ICT and exposure to Internet traffic do exist, figures on the trade of information and telecommunication services (or even services *tout court*) are lacking (Head et al. 2009). In this respect, complementing the existing economic figures with Internet topology data inferred at the level of AS can be a way to characterize the relative involvement (and its evolution over time) of the various countries in the digital economy, and to map the international trade of digital goods.

It has to be noted that the services sector is the largest contributor to a country's economy and is correlated to a country's level of development (ranging from about 50 per cent of GDP, in the case of low-income countries, up to 70 per cent, in the case of high-income countries). In addition, over the past two decades the growth of trade in services has grown almost fivefold, while trade in goods "only" by a factor of 3.5 (Lennon 2009). The growing importance of services in national economies and in international trade is largely due to an increase in the production of intermediate services (i.e., outsourcing). Besides the economic importance of services activity in general, and services outsourcing in particular, this phenomenon has received a huge amount of attention in the media and in political debates (Hamilton & Quinlan 2007) and the sector has increasingly been included in the framework of multilateral negotiations and regional trade agreements (WTO 1998).

Furthermore, in the economic literature there is a lively debate on the role of ICT investments and capital in contributing to productivity, growth and competitiveness of national systems both in the case of developed and developing countries. Some authors have argued that the relative importance of the ICT-producing sector in different countries, and its growth over time, might be a factor contributing to the differences in growth performance that have been observed in several OECD countries in recent years (Pilat et al. 2002). As such, these are sound arguments advocating the need for policies explicitly targeting ICT for development. Such concerns have gathered momentum also in the agenda of many governments: in response to the recent economic crisis, many countries have implemented policies aimed at containing the cyclical impact of the recession and to revive the economy. Compared to the past, where interventions gave priority to physical infrastructures, recent policies distinguish themselves for greater attention to public intervention in the realms of digital infrastructures and how such stimulus can affect productivity, innovation, sustainability and quality of life (Andes & Castro 2009). At the present state, though, many studies on the relationship of ICT investment with

growth and productivity raise more questions than answers. For instance, while evidence at the sectoral level on the effects of ICT on productivity and growth for some major countries such as US and Australia are incontrovertible, for many other OECD countries observed improvements are less straightforward (van Ark et al. 2008). Disparities in accessing to the Internet are more and more viewed as an important source of poverty and inequality and leveling the playing field in broadband network access is viewed as a key factor for reducing inequalities (Mehra et al. 2004). A possible explanation of such different returns on ICT investment might be related to the existence of different national capabilities in leveraging those investments in terms of competitiveness and involvement in international trade. While there is a considerable amount of data on ICT investment and on the international trade of physical goods, almost no data or research focus on issues such as the flow of digital bits, the international trade in ICT service industries, or how Internet penetration and broadband connectivity affect the competitiveness of economic systems or the international provision of services (but see Freund & Weinhold 2002, 2004, for notable exceptions). Starting from such limitations in the literature, in this paper we aim at using the data on Internet interconnectivity as a proxy measure for the International Trade in Service (ITS), and we will evaluate such data to measure the involvement of the various countries in the international trade of digital goods and services.

The contribution of the paper is threefold. First of all, to the best of our knowledge this is the first time the network of international Internet connectivity is analyzed at the country level and used to proxy the amount of trade in digital services. Second, we provide a detailed comparison between trade in digital and physical goods, and analyze the different role played by distance in shaping the two networks. The third contribution is methodological and has to do with the comparison of networks featuring different characteristics. To perform a meaningful comparison we propose a proper stochastic benchmark based on the hypergeometric distribution.

The paper is organized as follows: in the next Section we introduce the various datasets used in this study and, due to their differences (e.g., type of data, methods used for the data collection), we describe the methodology used to increase the comparability between them. Next, in Section three, we review the major findings arising from the network analysis of the Bits and the various Trade datasets. Finally, in Section four we discuss our results and provide some further conclusive remarks.

2 Data and Methodology

2.1 Internet Data

In this paper we make use of data regarding the high-level architecture of the Internet collected within the research activities of the DIMES (Distributed Internet Measurement and Simulations) project.¹

The basic unit of observation is represented by an Autonomous System (AS), which Provider (ISP) or other large subnetworks in which the Internet is divided (these are part of the network belonging to a single administrative control, such as in the case of a university, a business enterprise, or a business division). ASs can be viewed as a high-level node, so that data exchanges among ASs are edges of a high-level Internet map. Information related to interconnectivity between ASs has remained largely unknown until the past few years since ISPs and other business operators are generally not willing to publicly disclose their connectivity strategies. Nevertheless, gathering a truthful map of the Internet is viewed by network researchers as a very important task for routing efficiency, scalability, security and many other reasons. As a result, in the past 10 years there have been many efforts towards providing reliable measures of ASs interconnectivity.

Among such attempts, active monitoring of packets traveling across the network via traceroute and ping algorithms has allowed to reproduce significant parts of the network through random sampling. and the DIMES collaboration has implemented such techniques using a distributed multi-point infrastructure (composed of lightweight measurement software hosted by volunteers on computers all over the globe) which allowed to increase both the number of edges and nodes discovered and to improved the representativeness of the data collected with the respect to the whole Internet. Since the beginning of the project on September 1st, 2004, the project has collected 10,833,966,958 independent measurements (records on data exchange between two ASs), occurring between 29,404 AS distributed over 224 countries and along 204,204 edges, using a total of 24,421 software agents distributed in over 121 countries of the world.

As far as the preparation of the data is concerned, the original dataset describes an edge in terms of: (i) city and country of the source and destination nodes and (ii) the number of distinct couples of ASs that were observed exchanging traffic between the two cities during a given month. In order to derive the country-country adjacency matrix for year 2007, we dropped intra-country exchanges edges were consolidated at the country and year level. The result of this aggregation process is a 224 X 224 matrix in which the

¹DIMES is a distributed research collaboration aimed at studying how the Internet looks like and evolve overtime. The project, coordinated by a research team located at Tel Aviv University, employed in the past years thousands of volunteers in over a hundred countries to gather data on the Internet topology and has generated one of the most extensive and reliable dataset freely available for research purposes. See Shavitt & Shir (2005) and <http://www.netdimes.org> for additional information on the project.

i, j element shows the number of distinct ASs that were observed communicating during the year between country i and country j .

There are several remarks regarding the quality of such dataset that are worth making. As previously mentioned, the data are collected via sampling techniques and do not provide the complete set of interconnections among ASs. These caveats notwithstanding, the DIMES data are based on a large and distributed set of observation points, which with respect to the techniques based on smaller sets of agents or other passive monitoring techniques should improve the data at least in three different way: enlarging the absolute number of discovered nodes and edges, increasing the share of discovered peer-to-peer links (thus diminishing the bias towards customer-provider links), minimizing the so called “small observer population” bias, according to which sampling from few nodes might result in power-law distributions of links even in the case of ordinary random graphs.

Regarding the previously defined “intensity of exchange” measure, such figures build upon the aggregation of different observations on the interconnections which take place between different couples of ASs in the same couple of countries, in a given period of time. In this sense, while these figures simply signal the existence of distinct connections, we treat them as a rough proxy of the intensity of data exchange (and by extension, as an even rougher proxy for digital trade). It has to be stressed that we are not able to observe the amount of data exchanged (which would be a proper measure of magnitude of traffic flow), nor can we distinguish between the type of data which is exchanged in terms of protocol used. Moreover, these figures can be further biased by different practices employed by different ASs regarding the assignment of IP numbers to computers (just to name one example: an AS might decide to make extensive use of virtual IP addressing, which result in the same machine responding to a plurality of IP numbers, which would result in increasing the intensity figures of our dataset). Finally, aggregating monthly values into yearly ones (by simply adding up the figures) might partially counterbalance our inability to distinguish between more and less intense traffic exchange, since stable and recurring links are weighted more than unstable ones in the resulting measure of exchange intensity.

Existing international service trade dataset suffer from specific limitations: for instance, The World Bank’s World Development Indicators, while providing the time and country coverage, does not offer bilateral trade data. Of the bilateral datasets, Eurostat cover only 31 countries, and the OECD just covers four large categories of service, which does not allow to focus on the sectoral specificities of the digital economy (Head et al. 2009).

2.2 Trade Data

We use trade data taken from the BACI dataset maintained by CEPII (Gaulier & Zignago 2010), which collects information on bilateral trade flows among over 200 countries, at the 6-digit level of the Harmonized System (HS) classification. In the analysis we focus on the year 2007, though choosing a different year does not alter the results. Total bilateral trade is obtained by aggregating 6-digit HS flows for each country-pair. Data are expressed in thousands of US dollars, and display a lower cutoff at 1,000 dollars. Furthermore, to enhance comparability with data on the Internet topology we look at the number of HS-6 products traded by each country pair, and we perform a decomposition of goods according to the skill intensity associated with their production. Starting from Peneder (2007) we define two broad classes of goods, namely high-skill and low-skill ones, based on the ISIC sector to which each HS-6 product can be associated.²

When merging the two datasets we retain only the countries that are present in both samples, so as to enhance comparability. We end up with a common set of 189 nodes populating both the trade and the Internet networks.³

2.3 Methodology

The sampling nature of the DIMES data makes a direct comparison between the Bits and Trade networks unreliable. Indeed, strong heterogeneity in the number of relationships in the two networks makes it very difficult to discriminate links that are just reflecting the data collection method from those that provide useful information on the properties of the system. To address this issue we use a stochastic benchmark for normalization purpose, similar to the one recently proposed by Micciche et al. (2011). This methodology has been regularly used in genetics to identify statistically significant relationships (Tavazoie et al. 1999), and a similar approach has been applied to identify statistically significant cliques and clusters in networks (Alba 1973, Wuchty et al. 2006).

Normalization is the typical strategy used to compare networks: it may be applied to either rows or columns, or it may be applied to the entire adjacency matrix (for example, rescaling all trade flows as percentages of the total amount of trade flow in the whole matrix). Normalization may also be applied to both rows and columns, iteratively. For example, if we wanted an “average” number to put in each cell of the trade flow matrix, so that both the rows and the columns summed to 1, we could apply the iterative row and column approach. This is sometimes used when we want to give “equal weight” to each node, and to take into account both outflow (row) and inflow (column) structure.

²Peneder (2007) provides an international classification of sectors that ranges between 1 (very high) to 7 (very low) and refers to the educational intensity of the sector. In our work, high-skill products are associated to sectors ranging from 1 to 3, i.e., 20 out of the 56 sectors that are classified by Peneder (2007).

³The complete list of countries analyzed is available in Appendix A.

There are a number of alternative approaches to rescale the data. Rescaling can be helpful in highlighting structural features of the data, but, obviously, different normalizing approaches highlight very different features.

Given two countries A and B, let N_A be the number of goods exported by country A and N_B the number of goods imported by country B. The total number of traded goods is N_k and the observed number of goods exported from A to B is N_{AB} . Under the null hypothesis of random co-occurrence, that is to say customer in country B are indifferent to the nationality of the exporter, the probability of observing X goods traded is given by the hypergeometric distribution (Feller 1950)

$$H(X|N_k, N_A, N_B) = \frac{\binom{N_A}{X} \binom{N_k - N_A}{N_B - X}}{\binom{N_k}{N_B}}. \quad (1)$$

We can associate a p -value to the observed N_{AB} as $p(N_{AB}) = 1 - \sum_{X=0}^{N_{AB}-1} H(X|N_k, N_A, N_B)$. It is worth noting that the described null hypothesis directly takes into account the heterogeneity of countries with respect to the number of goods traded. For each pair of countries, we separately evaluate the p -value for each trade relationship and then use different cutoffs to select only those links that represent a significant departure from the hypergeometric ($p < .05, p < 10^{-12}$) the resulting matrices are then dichotomized. The hypergeometric multi-urn benchmark is equivalent to a degree-preserving network rewiring numerical procedure (Maslov & Sneppen 2002).⁴

3 Results

In this section we first describe the results on the topological properties of the Internet and Trade networks (Section 3.1), then move to analyze the role of distance in shaping bilateral links (Section 3.2)

3.1 Topological properties

Density Due to the fact that data on Internet connectivity are collected via sampling and do not represent the universe of AS links, we expect the Bits network to be less dense with respect to the trade one. In fact this is one of the issues that lead us to use a stochastic benchmark before running the comparative analysis. Table 1 summarizes the relevant information on the density of the various networks, plus the share of bilateral (i.e., reciprocal) connections (in brackets).

Density is indeed much higher for Trade than for Bits in the original data (64 Vs 13 percent). Albeit reduced, this difference remains considerably large even after applying

⁴Alternative normalization procedure has been applied to trade networks (Serrano et al. 2007). In particular, departures from the Gravity benchmark have been investigated (Krempel & Plumper 2003).

the two cutoffs based on the hypergeometric distribution (3.5 Vs 14 percent, and 2.1 Vs 8.6 percent respectively).

When considering the distinction between high- and low-skill products we do observe that the density of the first network is lower, in line with our hypothesis that products associated with higher skill requirements are more similar to trade in digital services. It is interesting to note that when focusing on the most demanding threshold (i.e., when concentrating only on those links taking values whose probability is below 10^{-12}) the difference between high- and low-skill products becomes particularly relevant, as density for the former is half that of the latter, which remains very similar to the density of the whole trade network. Still, the Bits network remains always the less dense.

Table 1: Network density and share of bilateral links

	Bits	Trade	Trade-Hi	Trade-Lo
real	13.382	64.055	49.983	61.511
	(69.480)	(86.760)	(79.890)	(85.210)
cut 0.05	3.462	13.999	11.393	13.610
	(43.250)	(61.080)	(58.790)	(60.170)
cut 10^{-12}	2.142	8.646	4.889	8.190
	(36.790)	(60.220)	(51.240)	(59.450)

Data expressed in percentage points

Figures in parentheses: share of bilateral links.

Bilateral links represent the majority of connections, and this is particularly true for trade in goods, where less than 15 percent of links are not reciprocated. This share doubles for Internet connections, and once again the difference remains sizable after the hypergeometric filtering.

This finding, the Bits network being less reciprocal than the Trade one, is not surprising if one consider that there are global content providers located in few countries (e.g., the USA) which distribute contents worldwide, generating large asymmetries in the balance of the flow of bits with respect to countries more marginal in the production and distribution of bits. Thus, in case of such disparities, sampling measurements might be unable to report any activity in one direction of flow between two countries.⁵

Network density figures suggest that many more country pairs are involved in world trade than in Internet connectivity. While it is probably true that more countries participate to trade in goods, not least because this is a much older activity, a lower density for the AS network also suggests that the Internet is more polarized, with strong connections occurring among a smaller set of actors. To further investigate this conjecture we turn now to other network statistics.

Node Degree Average total degree is much higher for the Trade network than for Bits, as one would expect from the higher density of the former. Again, the hypergeometric

⁵Also, in the case of the Bits network there is not anything similar to the trade balance, so disparities in fluxes in principle do not create any imbalance for a country.

filtering reduces this difference but does not alter substantially the picture, and we keep observing that while the network of low-skilled products is very close to the whole trade network, high-skilled goods display a behavior that is more similar to the Internet. This similarity is reinforced by the hypergeometric filtering procedure.

On average each country has roughly 136 trade partners, but it connects to AS located in just over 30 foreign countries, as we see from Table 2. When we focus on the strongest links only, we find an average number of 22.72 trade partners vis-à-vis only 6.57 connections with foreign AS. Remarkably, as mentioned above, countries involved in trade of high-skill products on average only exchange them with 13.67 other countries.

Table 2: Summary Information on Degree Connectivity

	Average Total Degree			
	Bits	Trade	Trade-Hi	Trade-Lo
real	32.836	136.370	112.868	132.741
cut 0.05	10.201	36.561	30.243	35.778
cut 10^{-12}	6.571	22.720	13.672	21.640

	Average In/Out Degree			
	Bits	Trade	Trade-Hi	Trade-Lo
real	25.159	120.423	93.968	115.640
cut 0.05	6.508	26.318	21.418	25.587
cut 10^{-12}	4.027	16.254	9.191	15.397

	Max/Min Total Degree			
	Bits	Trade	Trade-Hi	Trade-Lo
real	171 / 1	188 / 28	188 / 14	188 / 26
cut 0.05	111 / 0	125 / 8	116 / 5	127 / 7
cut 10^{-12}	80 / 0	106 / 2	54 / 1	103 / 2

Differentiating between in- and out- degree (imports and exports) does not alter substantially the picture. On average the number of in- and out- partners coincide because for the world as a whole every exporting relationship needs to have an importing counterpart.

The bottom panel of Table 2 displays the maximum and minimum number of partners in each of the networks. Here the difference is less severe as both the Trade and the Bits networks have “extreme” cases of very well-connected and very-poorly connected players. Still, in the Bits network none of the countries is fully connected (while this happens in the case of goods), whereas the sharper differences concern the minimum degree, with the least connected country in the world trade web still exchanging goods with some 28 partners. The effect of the thresholds becomes very evident here: by pruning all the trivial links the hypergeometric filtering makes the minimum total degree very similar across the different networks.

Degree Distribution To better grasp the different behavior of node degree across the various networks, Figure 1 displays the distribution of node total degree for the four networks. The Figure shows both the estimated kernel density (left-hand side), and the counter-cumulative distribution function (CCDF) in double-log scale (right-hand side).

The Bits network is characterized by a much more skewed distribution at all levels of observation, while trade in low-skilled goods is almost indistinguishable from the whole trade network, and the network of high-tech goods lying somewhat in between. From the comparison of the top row (original data) with the distributions obtained after applying the filtering procedure based on the hypergeometric benchmark, it appears the importance of the latter to adequately compare networks with different features. In fact, the shape of the degree distribution for the different trade networks changes substantially once we apply the cutoffs, and the difference with the Bits network is reduced.

None of the CCDF approaches the straight line that represents a power-law distribution, though the right tail of the CCDF for Bits is characterized by portions that resembles a power-law.

Assortativity So far we have only looked at single-node characteristics, from which we cannot infer much information on the structure of the links. Now we turn to second-order features of the graphs, by looking at the assortative structure of the networks.

Table 3 reports the simplest characterization of assortativity, i.e., Pearson’s degree correlation. We can see that all the networks display a disassortative structure whereby nodes with low degree tend to link with high-degree partners.⁶ We can also appreciate the importance of the filtering procedure as in the original data the Bits network has the larger (in absolute value) correlation, while the ranking is reversed once we focus on the strongest links only.

Such a result implies that in the Bits network many of the core-periphery links (i.e. links between high- and low-degree nodes) tend to be weak and do not withstand the hypergeometric filtering.

Table 3: Pearson’s degree correlation

	Bits	Trade	Trade-Hi	Trade-Lo
real	-0.407	-0.311	-0.355	-0.312
cut 0.05	-0.177	-0.270	-0.285	-0.273
cut 10^{-12}	-0.165	-0.289	-0.270	-0.288

Average K-neighbors degree A less synthetic way to look at assortativity is K-neighbors degree (Knn), i.e., the average degree of partners of nodes with a given degree. Figure 2 plots Knn for the different networks. While the three trade networks are very

⁶This is consistent with previous results. See for instance Mahadevan et al. (2006) on the Internet topology, and Fagiolo et al. (2008, 2009) on trade.

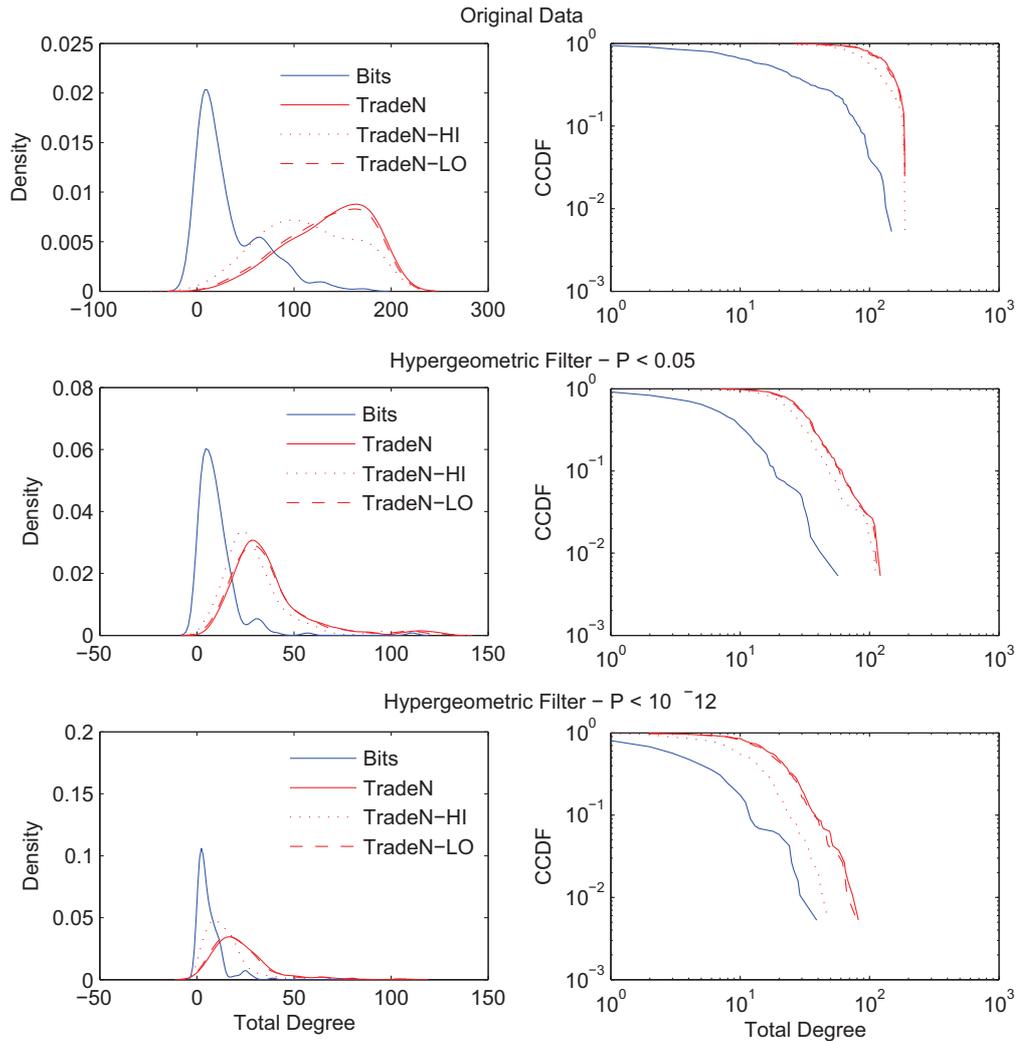


Figure 1: Degree distribution. Left panels: kernel density; right panels: counter cumulative distribution (log-log scale). Top row: original data; middle row: hypergeometric filtering with $P < 0.05$; bottom row: hypergeometric filtering with $P < 10^{-12}$

similar, the top-left panel shows that, when looking at the strongest link only (using the most stringent threshold), the Bits network has a flatter Knn profile, signalling that the average degree of each country’s partners is less dependent on its degree than in the case of trade.

Degree – Average Nearest Neighbor Degree Correlation This feature also emerges from yet another measure of assortativity, namely the correlation between node degree and the average degree of its neighbors (ANND). In Table 4 we distinguish between in- and out- degree and report correlation between the out-degree of a node and the average in-degree of its partners (top panel), as well as the one between the in-degree of a country and the average out-degree of its partners (bottom panel). These measures look at

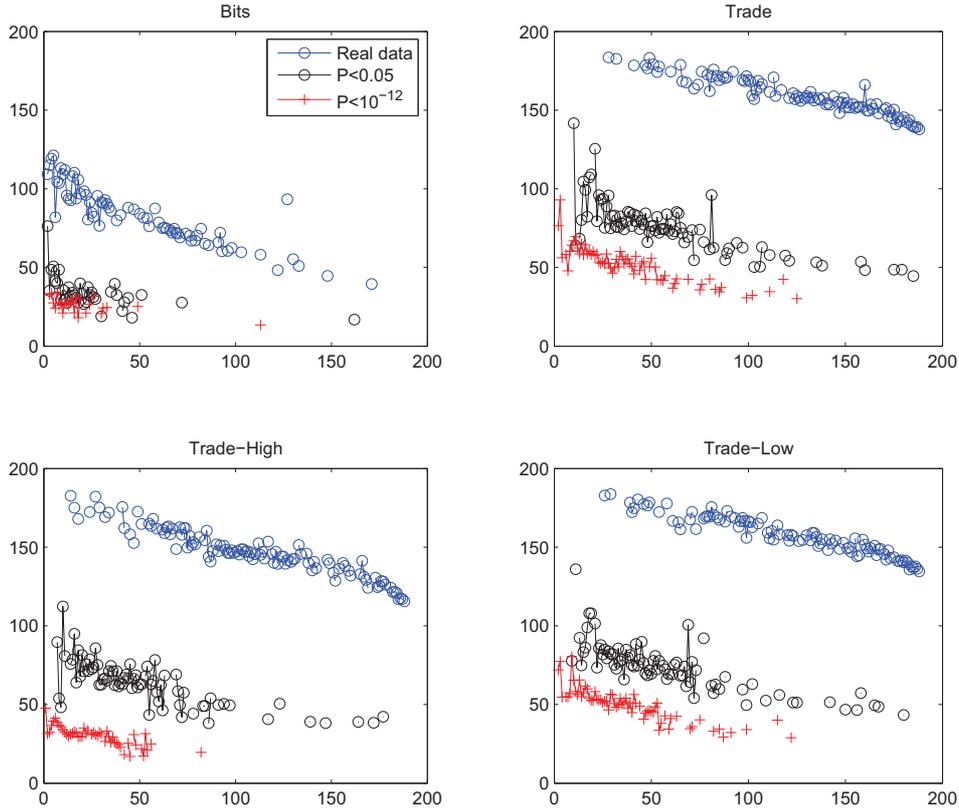


Figure 2: K-neighbors degree (Knn): the average degree of nodes with a given degree. The four panels report Knn for different networks: from left to right and from top to bottom Bits, Trade, Trade in high-skills goods and Trade in low-skills goods.

Table 4: Correlation Node Degree – ANND

	Out Degree – $ANND_{in}^{out}$			
	Bits	Trade	Trade-Hi	Trade-Lo
real	-0.271	-0.973	-0.961	-0.968
cut 0.05	-0.141	-0.647	-0.651	-0.624
cut 10^{-12}	0.018	-0.325	-0.067	-0.276
	In Degree – $ANND_{out}^{in}$			
	Bits	Trade	Trade-Hi	Trade-Lo
real	-0.629	-0.982	-0.972	-0.978
cut 0.05	-0.235	-0.684	-0.620	-0.705
cut 10^{-12}	-0.075	-0.704	-0.522	-0.713

All correlations are significant at 1%.

whether heavy exporters tend to ship goods to countries that import from many sources, and vice-versa whether countries that import from many suppliers do so prevalently from countries that serve many customers.

We find that the Trade networks feature a much stronger negative correlation between node degree and ANND, but there is also a relevant difference between the top and the

bottom panels of Table 4. In the former case, when looking at the relation between the number of export markets served and the average number of suppliers these have, the negative correlation for high-skill goods is very close to zero, while such figure for the Bits network turns slightly positive. Hence, for both digital services and trade in high-skill goods, we do not find a strong link between the number of export partners and the number of import relationships these maintain.

The same is not completely true when we reverse the direction of the edges and look at the correlation between in-degree and $ANND_{out}^{in}$ (bottom panel of Table 4). In this case, Bits maintain a weak (negative) correlation, whereas Trade-Hi displays a rather strong one. Furthermore, the correlations are always stronger in the bottom panel with respect to the top one.

Clustering Assortativity provides us with some information about the structure of the network, by telling us whether nodes with high degree tend to connect to partners with a similar number of connections. To understand how these partners interact among themselves we can look at clustering, which tells us how likely it is for someone’s partners to be also linked.

On average, the clustering coefficient is rather high in all networks built from the original data, but its value decreases markedly once we control for the hypergeometric benchmark, meaning that many of the triangles we observed in the network feature weak ties that do not stand the hypergeometric test (see Table 5). This appears to be particularly true for the Bits network. Interestingly, here there is no much difference across good types, with high- and low-skill goods differing very little, whereas the Bits network is much less clustered.

Table 5: Average clustering coefficient

	Bits	Trade	Trade-Hi	Trade-Lo
real	0.726	0.835	0.798	0.827
cut 0.05	0.288	0.448	0.440	0.439
cut 10^{-12}	0.234	0.465	0.444	0.460

The correlation between node degree and clustering is again negative and significant, much stronger for the three Trade networks than for the Bits case. When looking at the results pertaining to the stricter threshold, the correlation for Bits is close to zero. Once again, this confirms that trade in digital services, as opposed to trade in physical goods, is characterized by a flatter, less hierarchical structure.

RICH CLUB A simple way to characterize the hierarchical structure of a network is provided by the rich-club coefficient (Zhou & Mondragon 2004, Colizza et al. 2006).

Figure 3 displays the results for the rich-club coefficient computed following Colizza

Table 6: Correlation Node Degree – Clustering

	Bits	Trade	Trade-Hi	Trade-Lo
real	-0.349	-0.930	-0.957	-0.932
cut 0.05	-0.180	-0.701	-0.639	-0.663
cut 10^{-12}	-0.024	-0.562	-0.391	-0.531

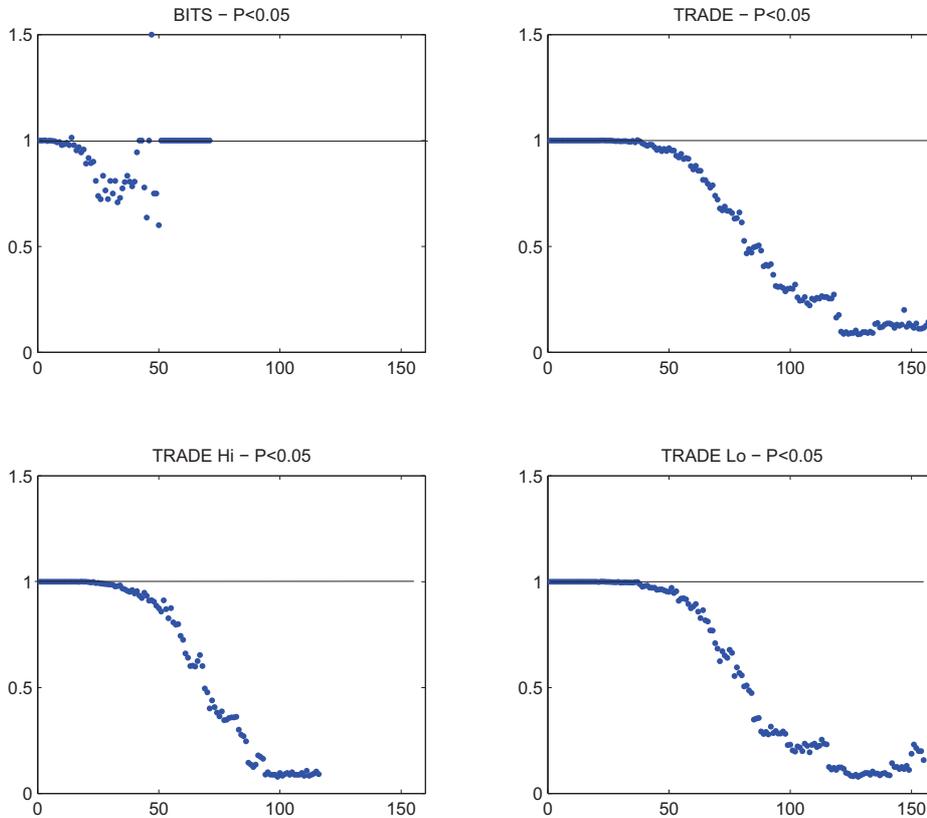


Figure 3: Rich-club coefficient on the networks obtained by applying the 5% cutoff.

et al. (2006) for the various networks obtained using the 5 percent hypergeometric cutoff.⁷ The original rich-club coefficient is compared to that obtained for a random network with a similar degree sequence, and a meaningful rich-club ordering is observed whenever the normalized coefficient is greater than 1. In all cases we find a normalized coefficient that is smaller than what we would expect for a random network, suggesting that high-degree nodes (hubs) provide connectivity to regional/local hubs, which in turn are not tightly interconnected. Remarkably, this behavior is much more pronounced in trade than in Bits, suggesting that stronger links (compared with our null model based on the hypergeometric distribution) occurs mainly between the core and the periphery in trade, whereas a larger number of them occurs among hubs in the case of the Internet.

We conjecture that this behavior may depend on the different role played by distance

⁷Results for the original data and the other cutoff value are not shown but are available upon request.

in hindering link formation in the two types of networks (physical versus digital trade) and we now turn to explore this issue.

3.2 The Role of Distance

We investigate the role of distance in shaping international trade relationship in different ways. The simplest is to compare average distance across connected pairs in the different domains, as summarized in Table 7.⁸ We can see that in the original data trade partners are, on average, further away from each other than country-pairs connected by the exchange of Bits. However, once we control for the null model, and focus on the strongest links only, the hierarchy is reversed, and this is more apparent when moving from the 0.05 threshold to the more stringent cutoff of 10^{-12} . In the last row of Table 7 we indeed see that average distance is smaller in the three trade networks than in the case of Bits. Hence, stronger links appear to occur at shorter distances, as one would expect, but this phenomenon is especially marked in the case of trade in physical goods. Note also that in this case the similarity between Bits and trade in high-skill goods that we have found previously, does not hold, as Trade-Hi experiences a very sharp drop in average distance between connected pairs when moving from the complete dataset to the most stringent threshold.

Table 7: Average distance between connected country-pairs

	Bits	Only bits	Trade	Only trade	Trade-Hi	Trade-Lo
real	6904	8113	7430	7570	7145	7366
cut 0.05	4594	6900	4460	4662	4132	4472
cut 10^{-12}	4128	6288	3671	3794	2948	3632

Figures in kilometres. Average distance between all countries: 8311.

Figure 4 provide a graphic snapshot of the different role of distance in the Trade and Bits networks (using the most stringent hypergeometric threshold, blue edges are links which resist filtering only in the Bits network but are canceled in the Trade network, vice versa for red edges while green edges are persisting links in both networks) and shows how the strongest links have a geographical (regional) component for trade in physical goods.

To test that average distances actually differ we revert to a Wilcoxon signed-rank test (Wilcoxon 1945), as we cannot assume the population to be normally distributed and use the standard t -test. We find that the null hypothesis of equality of means is rejected for the original data, whereas cannot be rejected when looking at the strongest links only, consistently with the observation that average distance for trade links goes down markedly when we trim the trivial links.

⁸Distance is defined as the geodesic distance between two countries' main city. We use the data provided by CEPII and available at <http://www.cepii.fr/anglaisgraph/bdd/distances.htm> .

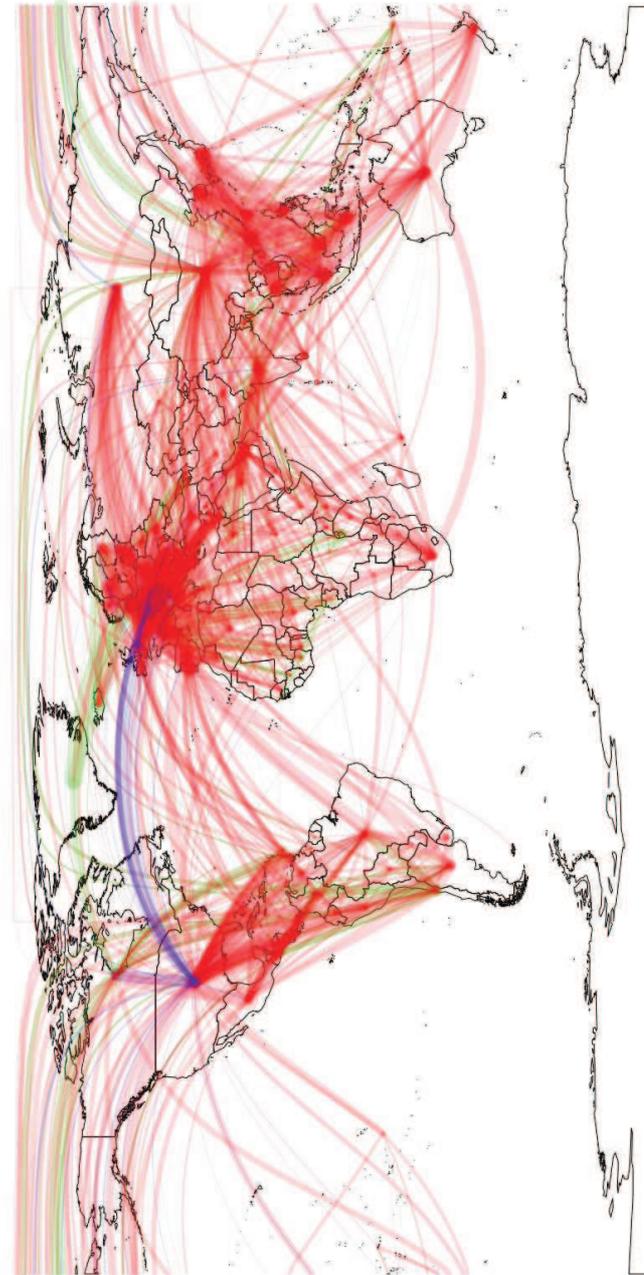


Figure 4: Global computer and trade networks, statistically significant linkages, cut 10^{-12} .

In Figure 5 we show that as distance increases, so does the share of statistically significant links in the Bits networks, while the relative importance of Trade-only connections goes down.

As a further step, instead of looking at average distance only, we take a broader view

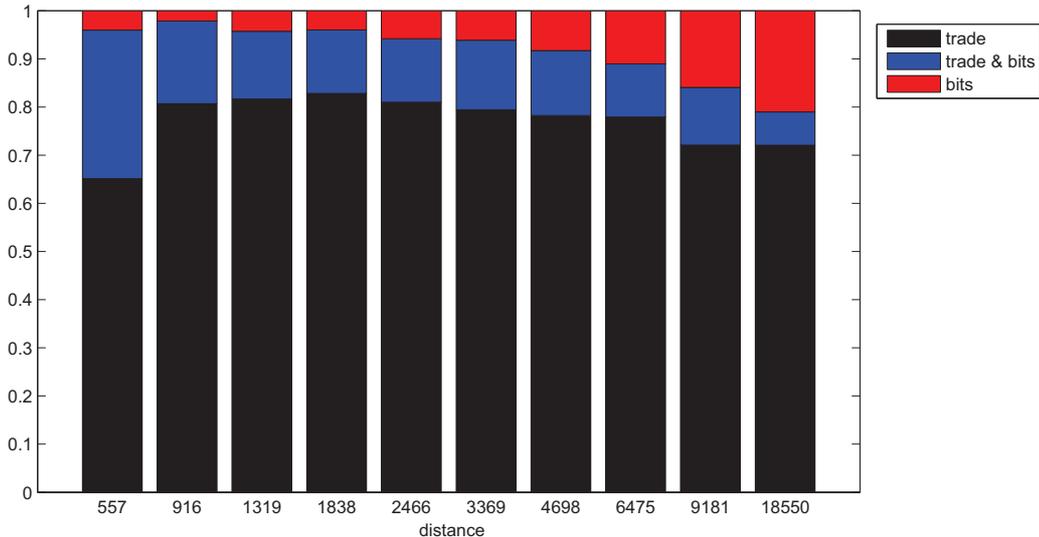


Figure 5: The distribution of statistically significant links by type and distance, cut 10^{-12} .

and compare the distribution of distances across the various networks by means of a Kolmogorov-Smirnov test. Once again we find evidence of an inversion in the distance ranking between Bits and Trade when moving from the original data to those obtained after the hypergeometric filtering procedure. In fact, in the former case we cannot reject the null hypothesis of the distribution of distances for Bits being smaller than the one for Trade, whereas the opposite holds true for the 10^{-12} threshold. Figure 6 summarizes these findings, by displaying a comparison of the distribution of distances in the Bits and Trade networks.

Finally, another way to test for the correlation between the network structure and distance across countries is via a quadratic assignment procedure (Krackhardt 1987, Hubert & Arabie 1989, QAP, see for instance). This entails computing the correlation between the matrix of hypergeometric cumulative probabilities P and the matrix of physical distances across countries, and comparing it with the values obtained once the rows and columns of P are randomly shuffled.

Results show that such correlation is positive and significant for all the four networks analyzed. This means that as distance grows, so does the cumulative hypergeometric probability associated with the link intensity between the two countries. Out of jargon, a positive correlation implies that more distant countries are connected by links whose intensity is more and more similar to what our null model would predict. Another way to look at the same result is to say that links that are significantly stronger than the hypergeometric benchmark are associated with lower distances. This correlation is 0.153 for the Bits network, while is 0.351 in the case of Trade.⁹ Hence, the role played by distance in hindering strong connections is more evident in the case of trade in physical

⁹Results do not differ substantially for trade in high-skill (0.360) and low-skill goods (0.349).

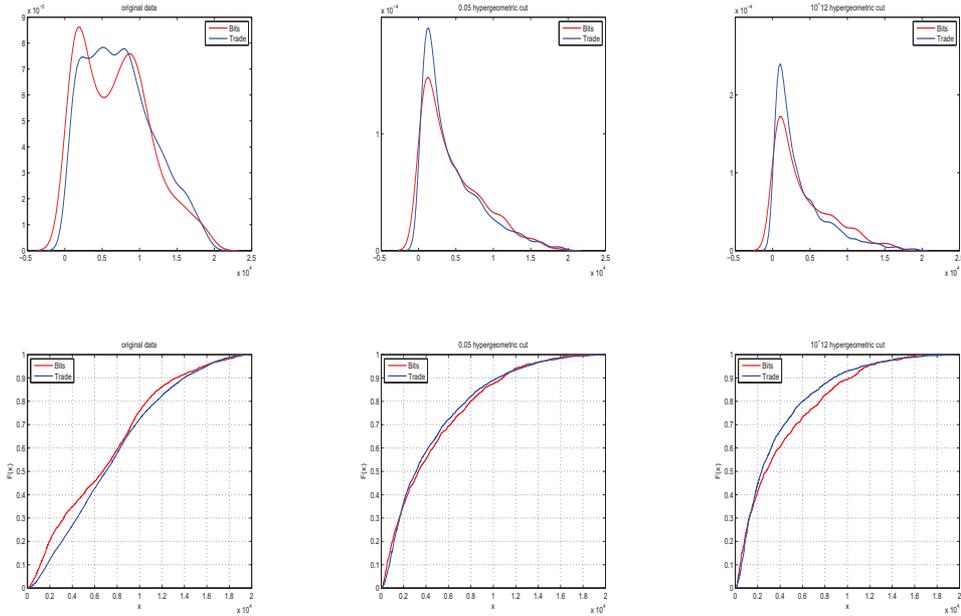


Figure 6: Empirical cumulative distribution for distance between connected pairs. Top row: probability distribution functions; bottom row: cumulative distribution functions. Left panel: original data; center panel: hypergeometric filtering with $P < 0.05$; right panel: hypergeometric filtering with $P < 10^{-12}$.

goods than for trade in digital services.

4 Conclusions

The topological properties analysis shows that the Bits network is sparser, less dense, with fewer bilateral links with respect to Trade networks, with differences which are higher in comparison with low-skill than high-skill products trade. This suggests that the Internet is a more polarized network, with important fluxes occurring among a smaller set of countries. All networks share a substantial disassortative structure, but after applying the hypergeometric filtering, most of the core-periphery links fade out in the Bits network and the underlying structure becomes significantly less disassortative compared to Trade. Clustering and rich club analysis both confirm this evidence: after filtering the data we find that many triangles observed in Bits disappear, while rich club analysis suggests that stronger links occurs mainly between the core and the periphery in Trade, whereas a larger number of them occurs among hubs in the case of Bits. Finally, distance plays a much stronger role in shaping the structure of trade in physical goods than in the case of Bits. Indeed, when moving from the complete network to considering the strongest links only, we observe in the Bits network that the average distance between partners increases and becomes larger than the Trade network case.

The analysis leaves a number of interesting avenues for further research, related to

open questions. A first one concerns the filtering methodology based on the hypergeometric distribution and has to do with a way to better discriminate among “large” flows. The analysis has shown that fluxes tend to cluster around zero, so that beyond a certain threshold it becomes increasingly difficult to distinguish among them. We conjecture that by tinkering with the three parameters of the distribution it should be possible to increase the power of the hypergeometric test. A second open question relates to the use of “small” fluxes as well (i.e., links whose values is significantly smaller than what a hypergeometric benchmark would predict) to extract further information on the topological properties of the networks as well as on their economic and social implications.

References

- Alba, R. (1973), ‘A graph-theoretic definition of a sociometric clique’, *Journal of Mathematical Sociology* **3**(1), 113–126.
- Andes, S. & Castro, D. (2009), Driving a digital recovery: It investments in the g-20 stimulus plan. The Information Technology and Innovation Foundation, Washington D.C.
- Berthelon, M. & Freund, C. (2008), ‘On the conservation of distance in international trade’, *Journal of International Economics* **75**(2), 310–320.
- Bhattacharya, K., Mukherjee, G., Saramäki, J., Kaski, K. & Manna, S. (2008), ‘The international trade network: weighted network analysis and modelling’, *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P02002.
- Blum, B. S. & Goldfarb, A. (2006), ‘Does the internet defy the law of gravity?’, *Journal of International Economics* **70**(2), 384–405.
- Brun, J., Carrère, C., Guillaumont, P. & De Melo, J. (2005), ‘Has distance died? evidence from a panel gravity model’, *The World Bank Economic Review* **19**(1), 99.
- Buldyrev, S., Parshani, R., Paul, G., Stanley, H. & Havlin, S. (2010), ‘Catastrophic cascade of failures in interdependent networks’, *Nature* **464**(7291), 1025–1028.
- Cairncross, F. (1997), *The Death of Distance: How the Communications Revolution Will Change Our Lives*, Harvard Business School Press, Cambridge, MA.
- Coe, D. (2002), *The missing globalization puzzle*, number 2002-2171, International Monetary Fund.
- Colizza, V., Flammini, A., Serrano, M. & Vespignani, A. (2006), ‘Detecting rich-club ordering in complex networks’, *Nature Physics* **2**, 110–115.

- Disdier, A. & Head, K. (2008), ‘The puzzling persistence of the distance effect on bilateral trade’, *The Review of Economics and Statistics* **90**(1), 37–48.
- Fagiolo, G., Reyes, J. & Schiavo, S. (2009), ‘World-trade web: Topological properties, dynamics, and evolution’, *Physical Review E* **79**(3), 036115.
- Fagiolo, G., Schiavo, S. & Reyes, J. (2008), ‘On the topological properties of the world trade web: A weighted network analysis’, *Physica A: Statistical Mechanics and its Applications* **387**(15), 3868–3873.
- Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999), On power-law relationships of the internet topology, in ‘Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication’, ACM, pp. 251–262.
- Feller, W. (1950), ‘An introduction to probability theory and its applications.’.
- Freund, C. L. & Weinhold, D. (2004), ‘The effect of the internet on international trade’, *Journal of International Economics* **62**(1), 171–189.
- Freund, C. & Weinhold, D. (2002), ‘The internet and international trade in services’, *American Economic Review* **92**(2), 236–240.
- Friedman, T. (2005), ‘The world is flat: A brief history of the twenty-first century.’.
- Garlaschelli, D. & Loffredo, M. (2004), ‘Fitness-dependent topological properties of the world trade web’, *Physical Review Letters* **93**(18), 188701.
- Garlaschelli, D. & Loffredo, M. (2005), ‘Structure and evolution of the world trade network’, *Physica A: Statistical Mechanics and its Applications* **355**(1), 138–144.
- Gaulier, G. & Zignago, S. (2010), Baci: International trade database at the product-level. the 1994-2007 version, Working Papers 2010-23, CEPII research center.
- Hamilton, D. & Quinlan, J. (2007), *Awakening the Transatlantic Services Economy*, Center for Transatlantic Relations, Washington DC.
- Head, K., Mayer, T. & Ries, J. (2009), ‘How remote is the offshoring threat?’, *European Economic Review* **53**(4), 429–444.
- Hortaçsu, A., Martinez-Jerez, A. & Douglas, J. (2009), ‘The geography trade on ebay and mercadolibre’, *American Economic Journal: Microeconomics* **1**(1), 53–74.
- Huberman, B. & Adamic, A. (1999), ‘Growth dynamics of the world-wide web’, *Nature* **401**(6749), 131.

- Hubert, L. & Arabie, P. (1989), ‘Combinatorial data analysis: Confirmatory comparisons between sets of matrices’, *Applied Stochastic Models and Data Analysis* **5**, 273–325.
- Krackhardt, D. (1987), ‘QAP partialling as a test of spuriousness’, *Social Networks* **9**, 171–186.
- Krempel, L. & Plumper, T. (2003), ‘Exploring the dynamics of international trade by combining the comparative advantages of multivariate statistics and network visualizations’, *Journal of social structure* **4**(1), 1–22.
- Leamer, E. (2007), ‘A flat world, a level playing field, a small world after all, or none of the above? A review of Thomas L. Friedman’s *The World is Flat*’, *Journal of Economic Literature* **45**(1), 83–126.
- Lennon, C. (2009), Trade in services and trade in goods: Differences and complementarities, Working Paper 53, Vienna Institute for International Economic Studies.
- Mahadevan, P., Krioukov, D., Fomenkov, M., Dimitropoulos, X., Claffy, K. & Vahdat, A. (2006), ‘The internet as-level topology: three data sources and one definitive metric’, *SIGCOMM Comput. Commun. Rev.* **36**, 17–26.
- Maslov, S. & Sneppen, K. (2002), ‘Specificity and stability in topology of protein networks’, *Science* **296**(5569), 910.
- Mehra, B., Merkel, C. & Bishop, A. (2004), ‘The internet for empowerment of minority and marginalized users’, *New Media and Society* **6**, 781–802.
- Micciche, S., Tumminello, M., Lillo, F. & Mantegna, R. (2011), ‘Statistically validated networks in bipartite complex systems’, *PLoS ONE* .
- Mok, D., Wellman, B. et al. (2007), ‘Did distance matter before the internet? interpersonal contact and support in the 1970s’, *Social networks* **29**(3), 430–461.
- Navaretti, G. B. & Venables, A. J. (2004), *Multinational Firms in the World Economy*, Princeton University Press, Princeton, N.J.
- Newman, M. (2004), ‘Analysis of weighted networks’, *Physical Review E* **70**(5), 056131.
- Pastor-Satorras, R. & Vespignani, A. (2004), *Evolution and structure of the Internet: A statistical physics approach*, Cambridge Univ Press.
- Peneder, M. (2007), ‘A sectoral taxonomy of educational intensity’, *Empirica* **34**(3), 189–212.
- Pilat, D., Lee, F. & van Ark, B. (2002), ‘Production and use of ict: A sectoral perspective on productivity growth in the oecd area’, *OECD Economic Studies* **35**(2), 47–78.

- Serrano, M. & Boguná, M. (2003), ‘Topology of the world trade web’, *Physical Review E* **68**(1), 015101.
- Serrano, M., Boguñá, M. & Vespignani, A. (2007), ‘Patterns of dominant flows in the world trade web’, *Journal of Economic Interaction and Coordination* **2**(2), 111–124.
- Shavitt, Y. & Shir, E. (2005), ‘DIMES - Let the internet measure itself’, *ACM SIGCOMM Computer Communication Review* **35**(5), 71–74.
- Tavazoie, S., Hughes, J., Campbell, M., Cho, R. & Church, G. (1999), ‘Systematic determination of genetic network architecture’, *Nature genetics* **22**, 281–285.
- van Ark, B., O’Mahony, M. & Timmer, M. (2008), ‘The productivity gap between Europe and the U.S.: Trends and causes’, *Journal of Economic Perspectives* **22**(1), 25–44.
- Wilcoxon, F. (1945), ‘Individual comparisons by ranking methods’, *Biometrics Bulletin* **1**(6), 80–83.
- WTO (1998), Electronic commerce and the role of WTO, Special Studies 2, World Trade Organization.
- Wuchty, S., Barabási, A. & Ferdig, M. (2006), ‘Stable evolutionary signal in a yeast protein interaction network’, *BMC evolutionary biology* **6**(1), 8.
- Zhou, S. & Mondragon, R. J. (2004), ‘The rich-club phenomenon in the internet topology’, *IEEE Communications Letters* **8**(3), 180–182.

Elenco dei papers del Dipartimento di Economia
pubblicati negli ultimi due anni

2009.1 *Income Shocks, Coping Strategies, and Consumption Smoothing. An Application to Indonesian Data*, Gabriella Berloffa e Francesca Modena

2009.2 *Clusters of firms in space and time*, Giuseppe Arbia, Giuseppe Espa, Diego Giuliani e Andrea Mazzitelli

2009.3 *A note on maximum likelihood estimation of a Pareto mixture*, Marco Bee, Roberto Benedetti e Giuseppe Espa

2009.4 *Job performance and job satisfaction: an integrated survey*, Maurizio Pugno e Sara Depedri

2009.5 *The evolution of the Sino-American co-dependency: modeling a regime switch in a growth setting*, Luigi Bonatti e Andrea Fracasso

2009.6 *The Two Triangles: What did Wicksell and Keynes know about macroeconomics that modern economists do not (consider)?* Ronny Mazzocchi, Roberto Tamborini e Hans-Michael Trautwein

2009.7 *Mobility Systems and Economic Growth: a Theoretical Analysis of the Long-Term Effects of Alternative Transportation Policies*, Luigi Bonatti e Emanuele Campiglio

2009.8 *Money and finance: The heterodox views of R. Clower, A. Leijonhufvud and H. Minsky*, Elisabetta de Antoni

2009.9 *Development and economic growth: the effectiveness of traditional policies*, Gabriella Berloffa, Giuseppe Folloni e Ilaria Schnyder

2009.10 *Management of hail risk: insurance or anti-hail nets?*, Luciano Pilati, Vasco Boatto

2010.1 *Monetary policy through the "credit-cost channel". Italy and Germany pre- and post-EMU*, Giuliana Passamani, Roberto Tamborini

2010.2 *Has food price volatility risen?* Christopher L. Gilbert e C. Wyn Morgan

2010.3 *Simulating copula-based distributions and estimating tail probabilities by means of Adaptive Importance Sampling*, Marco Bee

2010.4 *The China-US co-dependency and the elusive costs of growth rebalancing*, Luigi Bonatti e Andrea Fracasso

2010.5 *The Structure and Growth of International Trade*, Massimo Riccaboni e Stefano Schiavo

2010.6 *The Future of the Sino-American co-Dependency*, Luigi Bonatti e Andrea Fracasso

2010.7 *Anomalies in Economics and Finance*, Christopher L. Gilbert

2010.8 *Trust is bound to emerge (In the repeated Trust Game)*, Luciano Andreozzi

2010.9 *Dynamic VaR models and the Peaks over Threshold method for market risk measurement: an empirical investigation during a financial crisis*, Marco Bee e Fabrizio Miorelli

2010.10 *Creativity in work settings and inclusive governance: Survey based findings from Italy*, Silvia Sacchetti, Roger Sugden e Ermanno Tortia

2010.11 *Food Prices in Six Developing Countries and the Grains Price Spike*, Christopher L. Gilbert

2010.12 *Weighting Ripley's K-function to account for the firm dimension in the analysis of spatial concentration*, Giuseppe Espa, Diego Giuliani, Giuseppe Arbia

2010.13 *I Would if I Could: Precarious Employment and Childbearing Intentions in Italy*, Francesca Modena e Fabio Sabatini

2010.14 *Measuring industrial agglomeration with inhomogeneous K-function: the case of ICT firms in Milan (Italy)*, Giuseppe Espa, Giuseppe Arbia e Diego Giuliani

2010.15 *Dualism and growth in transition economies: a two-sector model with efficient and subsidized enterprises*, Luigi Bonatti e Kiryl Haiduk

2010.16 *Intermediaries in International Trade: Direct Versus Indirect Modes of Export*, Andrew B. Bernard, Marco Grazzi e Chiara Tomasi

2010.17 *Production and Financial Linkages in inter-firm Networks: Structural Variety, Risk-sharing and Resilience*, Giulio Cainelli, Sandro Montresor e Giuseppe Vittucci Marzetti

2010.18 *Organizational Capital and Firm Performance. Empirical Evidence for European Firms*, Claudia Tronconi e Giuseppe Vittucci Marzetti

2010.19 *The Fund-flow Approach. A Critical Survey*, Giuseppe Vittucci Marzetti

2010.20 *Is Open Source about Innovation? How Interaction with the Open Source Community Impact on the Innovative Performances of Entrepreneurial Ventures*, Evila Piva, Francesco Rentocchini e Cristina Rossi Lamastra.

2011.1 *Leaving home and housing prices. The Experience of Italian youth emancipation*, Francesca Modena e Concetta Rondinelli.

2011.2 *Pareto versus lognormal: a maximum entropy test*, Marco Bee, Massimo Riccaboni e Stefano Schiavo.

2011.3 *Does a virtuous circle between social capital and CSR exist? A "network of games" model and some empirical evidence*, Giacomo Degli Antoni e Lorenzo Sacconi

2011.4 *The new rules of the Stability and Growth Pact. Threats from heterogeneity and interdependence*, Roberto Tamborini

2011.5 *Chinese reserves accumulation and US monetary policy: Will China go on buying US financial assets?* Luigi Bonatti e Andrea Fracasso

2011.6 *Taking Keller seriously: trade and distance in international R&D spillovers*, Andrea Fracasso e Giuseppe Vittucci

2011.7 *Exchange Rate Exposure under Liquidity Constraints*, Sarah Guillou e Stefano Schiavo

2011.8 *Global Networks of Trade and Bits*, Massimo Riccaboni, Alessandro Rossi e Stefano Schiavo

The Discussion Paper series provides a means for circulating preliminary research results by staff of or visitors to the Department. Its purpose is to stimulate discussion prior to the publication of papers.

Requests for copies of Discussion Papers and address changes should be sent to:

Dott. Luciano Andreozzi

e-mail: luciano.andreozzi@unitn.it

Department of Economics

University of Trento

via Inama, 5 - 38122 Trento (IT)

