REGULAR ARTICLE

The international-trade network: gravity equations and topological properties

Giorgio Fagiolo

Abstract   This paper begins to explore the determinants of the topological properties of the international-trade network (ITN). We fit bilateral-trade flows using a standard gravity equation to build a "residual" ITN where trade-link weights are depurated from geographical distance, size, border effects, trade agreements, and so on. We then compare the topological properties of the original and residual ITNs. We find that the residual ITN displays, unlike the original one, marked signatures of a complex system, and is characterized by a very different topological architecture. Whereas the original ITN is geographically clustered and organized around a few large-sized hubs, the residual ITN displays many small-sized but trade-oriented countries that, independently of their geographical position, either play the role of local hubs or attract large and rich countries in relatively complex trade-interaction patterns.

Keywords   International trade network · Gravity equation · Weighted network analysis · Topological properties · Econophysics

JEL Classification   F10 · D85

1 Introduction

The last years have witnessed the emergence of a large body of contributions addressing international-trade issues from a complex-network perspective (Li et al. 2003; Serrano and Boguñá 2003; Garlaschelli and Loffredo 2004, 2005; Reichardt and White 2007; Serrano et al. 2007; Bhattacharya et al. 2008, 2007, Garlaschelli et al. 2007;

G. Fagiolo (✉)
Sant'Anna School of Advanced Studies, Laboratory of Economics and Management,
Piazza Martiri della Libertà 33, I-56127 Pisa, Italy
e-mail: giorgio.fagiolo@sssup.it

Tzekina et al. 2008; Fagiolo et al. 2008; Fagiolo et al. 2009b; Reyes et al. 2008).
The International trade network (ITN), aka World-trade web (WTW) or World-trade
network (WTN), is defined as the graph of import/export relationships between world
countries in a given year.

Understanding the topological properties of the ITN from a complex-network per-
spective (Albert and Barabási 2002; Dorogovtsev and Mendes 2003), and their evo-
lution over time, acquires a fundamental importance in explaining issues such as
economic globalization and internationalization, the spreading of international cri-
ses, and the transmission of economic shocks (Helliwell and Padmore 1985; Artis
et al. 2003; Forbes 2002). Indeed, the standard approach to the empirics of interna-
tional-trade employs indicators and statistics that fully characterize the profile of a
country in the system by referring only to its bilateral-trade direct linkages, i.e. direct
import-export relationships with its trade partners.[1] While direct bilateral trade link-
ages are known to be one of the most important channels of interaction between world
countries (Krugman 1995), recent studies show that they can only explain a small
fraction of the impact that an economic shock originating in a given country can have
on another one, which is not among its direct-trade partners (Abeysinghe and Forbes
2005). In order to fully account for such an impact, a more detailed knowledge of
the local and global topological properties of the network is required. This means, in
other words, acquiring a more in-depth understanding of the presence and importance
of trade paths connecting any pair of non-direct trade partners and, more generally,
of topological indicators that can proxy the likelihood that economic shocks might
be transmitted between the two countries (Kali and Reyes 2007) and possibly help
explaining macroeconomic dynamics (Kali et al. 2007).

The first stream of studies that have explored the properties of the ITN has employed
a *binary-network analysis*, where a (possibly directed) link between any two coun-
tries is either present or not according to whether the value of the associated trade
flow is larger than a given threshold (Serrano and Boguñá 2003; Garlaschelli and
Loffredo 2004, 2005). More recently, however, a growing number of contributions
(Li et al. 2003; Bhattacharya et al. 2008, 2007; Garlaschelli et al. 2007; Fagiolo et al.
2008; Fagiolo et al. 2009b) have adopted a *weighted-network approach* (Barrat et al.
2004; Barthélemy et al. 2005) to the study of the ITN. There, a link between any two
countries is weighted by the (deflated) value of trade (imports plus exports) that has
occurred between these countries in a given time interval. A set of weighted-network
topological measures (Barthélemy et al. 2005) is then computed to characterize the
architecture of the weighted ITN.

The motivation for a weighted-network analysis is that a binary approach cannot
fully extract the wealth of information about the trade intensity flowing through each
link and therefore might dramatically underestimate the role of heterogeneity in trade
linkages. Interestingly, Fagiolo et al. (2008); Fagiolo et al. (2009b) show that the sta-
tistical properties of the ITN viewed as a weighted network crucially differ from those
exhibited by its weighted counterpart, and that a weighted-network analysis is able to
provide a more complete and truthful picture of the ITN than a binary one.

---

[1] An example is "openness to trade" statistics, traditionally defined as the ratio of exports plus imports to
country's gross domestic product (GDP). See Reyes et al. (2008) for a discussion on this and related points.

Notwithstanding much is known about the topological properties of the ITN—in both its binary and weighted versions—and how they have evolved in the recent past, a set of fundamental questions remains to be answered: What are the determinants of such properties? Are there relevant node or link characteristics (other than the ones related to trade flows) that can explain the peculiar topological patterns actually observed in the ITN? Such questions might be in principle addressed from a theoretical perspective, i.e. looking for models of network formation and evolution that have as their equilibria graphs with properties similar to those actually observed in the ITN (cf. for example Albert and Barabási 2002; Jackson et al. 2003). In this paper we take an applied approach and attempt to explore the foregoing issues from a more empirical perspective.

To begin with, note that all network-related topological variables are univocally obtained from the link-weight distribution.[2] Hence, if (as typically happens) one weights the link between country i and country j in a given year by the sum of the deflated values of imports of i from j and exports of i to j in that year, then all topological properties of the weighted ITN defined in that way will entirely depend on the matrix of international bilateral-trade flows observed in the year under study, which plays the role of sufficient statistics. This is straightforward for the weighted-version of the ITN, but is also true for its binary representations, as the probability that a given link is present still depends, given the chosen threshold, on the distribution of observed bilateral-trade flows. As a result, one might safely conclude that much of what we know about the topological properties of the ITN can be empirically accounted for by the set of statistically-significant independent variables that explain international bilateral-trade flows.

In the context of the ITN literature, this issue has been initially addressed by Garlaschelli and Loffredo (2004), Garlaschelli et al. (2007), who have shown that the probability that any two countries are connected, as well as the value of the trade flow between them, is well explained by (the product of) their current GDP, which plays the role of a "hidden" node variable or fitness. More generally, we know from the huge empirical literature on gravity equations (Leamer and Levinsohn 1995; Overman et al. 2003) that international-trade flows can be almost entirely explained by a multiplicative model featuring as independent variables the GDPs of the two countries involved in the trade link, their geographical distance (as a proxy of all factors that might create trade resistance, e.g. transport costs), and a series of dummies accounting for other geographical, social, historical and political factors (e.g., existence of common borders, religion and languages, colonial ties, trade agreements, and so on). By fitting a gravity equation to the original international bilateral-trade data one may then account for explanatory variables of link weights.

Such an exercise might be interesting for two related reasons. First, from a gravity-equation approach, one may identify what are the main determinants of international-trade flows in the data used to build trade networks. Second, and more importantly, one might think to remove all the existing structure from the data to check whether

---

[2] For example, node strength, clustering coefficient, centrality, etc. are simple manipulations that require as inputs the link-weight distribution only. See "Appendix C" for formal definitions.

the residual weighted ITN exhibits topological features comparable to those of the original ITN.

This paper explores these two lines of inquiry by performing the following simple exercise. In line with the recent literature (Li et al. 2003; Bhattacharya et al. 2008, 2007; Garlaschelli et al. 2007; Fagiolo et al. 2008; Fagiolo et al. 2009b), we start from a weighted-network representation of the ITN in a given year where a link between any two countries is weighted by the (deflated) value of their total trade (import plus exports). This is the standard technique employed in the literature to build a weighted undirected representation of the international trade network. Our aim is to "filter" this representation using existing empirical knowledge about the relation between bilateral trade flows (or total trade) and (a) country-specific variables (e.g., size, etc.); (b) other relational bilateral variables that are known to characterize trade interactions between any two countries (geographical distance, and other geographical, social, historical and political factors affecting trade). To do so, we fit to our original link weights a standard gravity-equation specification where the dependent variable is total trade between the two nodes, and the independent variables are country-specific and relational variables as in (a) and (b) above. This allows us to identify the relative impact of size, geographical distance, and other geographical, social, historical and political factors, on the weighted-network representation of the ITN under study. Next, we define the filtered or "residual" version of the ITN as the network where each link is now weighted by the associated residual of the fitted gravity regression equation. In other words, each link in the residual network is weighted by a number that reflects what is left in the after having explained the original trade flow by the natural candidates suggested in the gravity literature. This allows us to remove much of the structure that is conceivably present in the original data and to obtain a filtered representation of the international trade flow where topological structure are no longer dependent on obvious size effects, geographical and political factors, etc.

This simple exercise leaves us with two complementary representations of the ITN, the original one and the residual one. It is then natural to ask whether—and to what extent—the topological properties of the residual ITN (as defined above) mimic those of the original version of the ITN (as studied, e.g., in Fagiolo et al. 2009b). Notice that a similar approach has been already used in Fagiolo et al. (2009a). They find that, by and large, the ITN architecture remains unaltered if ones removes the GDP dependence only from link weights (e.g., if one employs as link weights the ratio between bilateral-trade flows divided by the GDP of *either* the importer *or* the exporter country). The gravity-equation fit allows us to generalize this approach and obtain a weighted ITN that depends on underlying factors, either unobserved or not accounted for in the regression, related e.g. to country technological similarities, degree of specialization, etc.

Our results show that the residual ITN is characterized by power-law shaped distributions of link weights and node statistics (e.g., strength, clustering, random-walk betweenness centrality). Hence, the underlying architecture of the weighted ITN seem to display signatures of complexity. This must be contrasted with the original ITN, where log-normal distributions were ubiquitous. We also find that the correlation structure among node statistics, and between node statistics and country per-capita GDP, changes substantially when comparing the original and residual weighted networks. Whereas the original ITN is geographically clustered and organized around a few

large-sized hubs, the residual ITN features many small-sized but trade-oriented countries that, independently of their geographical position, either play the role of local hubs or attract large and rich countries in relatively complex trade-interaction patterns.

The rest of the paper is organized as follows. Section 2 briefly presents the data sets employed in the paper. Section 3 contains the results of gravity-equation fits. A comparison of the topological properties of the original and the "residual" version of the ITN is carried out in Sect. 4. Finally, Sect. 5 concludes discussing the implications of our results for modeling and suggesting extensions of the present work.

2 Data and definitions

We employ international-trade data provided in Gleditsch (2002) to build a time-sequence of weighted directed ITNs. Our balanced panel refers to $T = 20$ years (1981–2000) and $N = 159$ countries (see "Appendix A" for the list of acronyms and countries in the panel). For each country and year, data report directed trade flows (e.g., exports) in current US dollars, which we properly deflate. Weight matrices are built following the flow of goods. This means that rows represent exporting countries, whereas columns stand for importing countries. Following Li et al. (2003), Bhattacharya et al. (2008, 2007), Garlaschelli et al. (2007), we begin by initially defining the weight $\tilde{w}_{ij}^t$ of a link from i to j in year t as total exports from i to j. However, a simple statistical analysis of weighted matrices suggests that $\tilde{W}^t$ are sufficiently symmetric to justify an undirected analysis for all t (Fagiolo et al. 2009b). We therefore symmetrize the network by defining the entries of the new weight matrix W^t as the arithmetic average of import and export flows, i.e. $w_{ij}^t = \frac{1}{2}[\tilde{w}_{ij}^t + \tilde{w}_{ji}^t]$.[3] Finally, in order to have $w_{ij}^t \in [0, 1]$ (and to remove all trend-related factors), for all (i, j) and t, we re-normalize all entries in W^t by their maximum value $w_*^t = \max_{i,j=1}^N \{w_{ij}^t\}$. This means that the symmetrized weight w_{ij}^t is proportional to total trade (imports plus exports) flowing through the link ij in a given year.

Note that Fagiolo et al. (2009b) have shown that all topological properties of the ITN have remained fairly stable in the period 1981–2000. Therefore, in what follows, we shall focus on year 2000 only for the sake of exposition (all results robustly hold in other years). Our relevant ITN data is then the weight $N \times N$ matrix $W = W^{2000}$, which is symmetric by construction and features only zeroes in its main diagonal. More formally, we have:

Definition 1 The original weighted-undirected network representation of the ITN in year 2000 is the graph where the nodes are the N countries in the sample and links are fully characterized by the $N \times N$ symmetric weight matrix $W = W^{2000}$.

For each country in the panel, we also gather a long list of variables traditionally employed in gravity-equation exercises, see "Appendix B" for labels, sources and explanations. These variables can be grouped in three classes. First, there are variables related to trade resistance factors, as the geographical distance between countries and

---

[3] Following Baldwin and Taglioni (2006), we have also replicated our exercises by using a geometric average of bilateral flows without noticing relevant differences in the results.

the degree of country remoteness (Bhavnani et al. 2002). Second, country-size effects are controlled by country GDP, population and geographical area. Third, a number of country variables or link-related dummies control for geographical (common border, landlocking, continent), economic (trade agreements, exchange rates, consumer price indices) and social/ political/ historical effects (common languages and religion, trade agreements, common currency). Together, these factors have been shown to success-fully explain, in a way or in the other, international-trade flows in gravity-equation econometric exercises. Therefore, in the next section, we shall employ a standard gravity-equation setup to fit the entries of the original ITN in year 2000 defined above.

3 Fitting gravity equations to ITN data

The gravity equation has become the workhorse setup to study the determinants of bilateral international-trade flows (Overman et al. 2003; Fratianni 2009). Its basic sym-metric original formulation, inspired by Newton's gravity equation, states that total trade between any two countries in the world is directly proportional to the product of country masses (e.g., their GDP) and inversely proportional to their geographic distance.[4]

A strong gravity-like dependence of bilateral-trade flows on geography and size is evident in our data. For example, Table 1 reports trade-flow shares within and between macro geographical areas. It is easy to see that areas that trade more, mostly trade within the same area, or with geographically-close areas. Conversely, those that trade less, tend to trade with less distant areas and/or with areas where there are countries historically and culturally tied.

Furthermore, a gravity-law structure clearly emerges if one studies how bilateral trade flows w_{ij} correlate with the product of i's and j's GDPs and their geographical distance. Figure 1 plots in a log-log scale w_{ij} against $M_{ij} = \text{GDP}_i \cdot \text{GDP}_j$, DIST_{ij} and M_{ij}/DIST_{ij}. Non-parametric estimates of w_{ij} conditional means, superimposed on the clouds of points together with %95 confidence bounds, clearly indicate that both the correlation structure and the functional form linking w_{ij} to size and geo-graphical effects are in line with the prediction of the basic gravity law.[5] Notice how if one only considers country size (GDPs) or geographical distance between countries as explaining variables, conditional means appear to be only slightly correlated with w_{ij}, and one typically picks up a lot of oscillations in non-parametric estimates (with relatively large error bounds especially where the number of observations is relatively small). Conversely, using together masses and distances with a gravity-like functional

---

[4] This empirically-inspired law has been found to be consistent with a number of theoretical foundations, e.g. trade specialization models, monopolistic-competition frameworks with intra-industry trade, Heckscher-Ohlin models, etc.; see, e.g., Anderson (1979), Bergstrand (1985), Deardorff (1998), Anderson and van Wincoop (2003).

[5] Conditional means have been estimated using the local-linear non-parametric method proposed in Li and Racine (2004), employing cross-validated bandwidth selection using the procedure of Hurvich et al. (1998). Estimation has been performed using the package np (Hayfield and Racine 2008) under R (http://www.r-project.org/).

Table 1 Total trade within and between macro geographical areas

	N America (%)	C America (%)	S America (%)	Cont Europe (%)	East Europe (%)	Middle East (%)	C Asia (%)	China & E Asia (%)	N Africa (%)	C Africa (%)	S Africa (%)	Pacific (%)	% Total trade
N America	31.87	12.35	3.66	18.52	0.71	2.68	0.98	26.57	0.44	0.68	0.49	1.05	20.49
C America	74.19	3.96	3.37	8.50	0.44	0.59	0.15	8.06	0.15	0.15	0.15	0.29	3.41
S America	29.88	4.58	23.90	22.11	1.00	1.99	0.80	12.35	1.20	1.00	0.60	0.60	2.51
Cont Europe	9.75	0.75	1.43	67.53	3.75	3.17	0.76	9.44	1.45	0.72	0.64	0.60	38.91
East Europe	6.43	0.67	1.11	64.75	10.64	3.99	0.89	9.76	0.67	0.67	0.22	0.22	2.26
Middle East	13.87	0.50	1.26	31.15	2.27	9.08	4.67	32.53	1.51	0.63	1.26	1.26	3.97
C Asia	16.53	0.41	1.65	24.38	1.65	15.29	4.55	29.34	1.24	2.48	0.83	1.65	1.21
China & E Asia	23.08	1.17	1.31	15.58	0.93	5.47	1.50	46.82	0.32	0.57	0.59	2.65	23.59
N Africa	10.06	0.56	3.35	63.13	1.68	6.70	1.68	8.38	2.79	1.12	0.00	0.56	0.90
C Africa	17.95	0.64	3.21	35.90	1.92	3.21	3.85	17.31	1.28	12.18	1.92	0.64	0.78
S Africa	15.04	0.75	2.26	37.59	0.75	7.52	1.50	21.05	0.00	2.26	9.77	1.50	0.67
Pacific	16.17	0.75	1.13	17.67	0.38	3.76	1.50	46.99	0.38	0.38	0.75	10.15	1.33

Columns 1–12: Percentage of total trade of countries in the area that is exchanged with countries of the same or other areas. Column 13: Share of total trade of the area to world total trade. Year: 2000
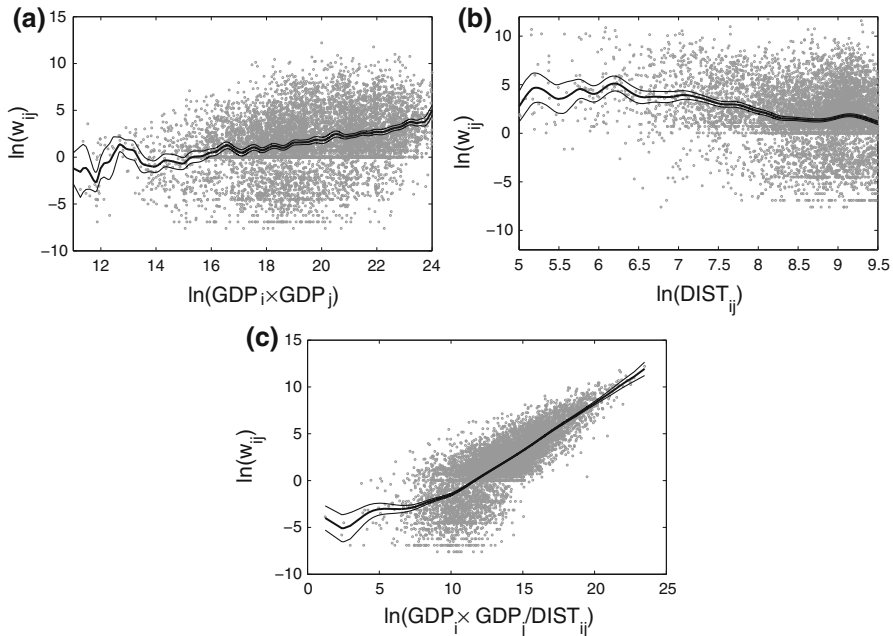
Fig. 1 Bilateral-trade flows vs. (**a**) product of country GDPs; (**b**) geographical distance; (**c**) product of country GDPs divided by geographical distance. *Solid curves*: Non-parametric estimate of bilateral-trade flow conditional mean with confidence bounds at 95%

form allows one to get a much sharper and smoother positive relation with bilateral trade flows.

From an empirical perspective, the basic gravity equation has been expanded in the literature to improve the fit by taking into account country- or trade-specific characteristics that may influence bilateral trade flows in addition to masses and distance (see previous section). Overall, this strategy has proven itself to be extremely successful in order to explore the determinants of bilateral trade flows and we shall replicate it in what follows.[6]

For our estimation purposes, it is convenient to start with the most general multiplicative formulation of the gravity equation (Santos Silva and Tenreyro 2006), which reads:

$$w_{ij} = \alpha_0 [\text{GDP}_i^{\alpha_1} \text{GDP}_j^{\alpha_2} \text{DIST}_{ij}^{\alpha_3}] \left[\prod_{k=1}^{K} X_{ik}^{\beta_{1k}} X_{jk}^{\beta_{2k}} \right]$$

$$\cdot \exp \left\{ \sum_{h=1}^{H} \theta_h D_{ijh} + \sum_{l=1}^{L} (\delta_{1l} Z_{il} + \delta_{2l} Z_{jl}) + \gamma_i C_i + \gamma_j C_j \right\} \eta_{ij}. \quad (1)$$

[6] On this huge empirical literature, see for example Overman et al. (2003), Frankel (1997), Rose and Spiegel (2002), Rose (2000), Glick and Rose (2001), Eichengreen and Irwin (1996), Leamer and Levinsohn (1995), Feenstra et al. (2001), among others.

Here $w_{ij}, i, j = 1, \ldots, N$ (with $i > j$) is the $N(N-1)/2$-dimensional vector of symmetric bilateral trade flows (i.e. the upper-diagonal entries of the weight matrix W defining the original ITN, see Definition 1); $(\alpha, \beta, \theta, \delta, \gamma)$ are unknown (vectors of) parameters to be estimated; X_i are K-dimensional vectors of country-specific continuous or categorical variables (other than GDP); Z_i are L-dimensional vectors of country-specific dummy variables; D_{ijh} are H-dimensional vectors of link-specific dummy variables (other than distance); C_i are country-specific fixed effects (see below); and η_{ij} is an error term, statistically independent on the regressors, s.t. its mean conditional on all regressors is equal to 1. In our exercises below, we shall start by the most general case where $X = $ (AREA, POP, CPI, EXC, RM), $Z = $ (LL, CONT), whereas $D = $ (CTG, COMC, COML, COL, TA, COMR), see Appendix B.

Our strategy is extremely simple: using the information from observed variables in X, Z and D (see Sect. 2), we shall fit the dependent variable w_{ij} employing properly modified maximum-likelihood techniques to get estimates of $(\alpha, \beta, \theta, \delta, \gamma)$ and, consequently, predicted trade-flow values $\hat{w}_{ij}$. This will allow us not only to assess the impact of country-specific and relational variables X, Z and D on the original ITN structure, but also to obtain a measure of how much of the original ITN can be explained using traditional gravity-like regressors. Accordingly, the estimates for the residuals, i.e. $\hat{\eta}_{ij}$, will represent a measure of what is left out in w_{ij} after most of natural candidates to explain them have been accounted for. It is then natural to define, as we shall see more formally below, the residual ITN as the weighted undirected network where the link between any two countries is weighted by the correspondent residual estimate $\hat{\eta}_{ij}$.

The estimation of Eq. (1) is not easy and it potentially entails many difficulties. Among them, we recall the treatment of zero-valued flows (Santos Silva and Tenreyro 2006; Linders and de Groot 2006; Burger et al. 2009), non-linearity and heteroscedasticity (Santos Silva and Tenreyro 2006), endogeneity and omitted-term biases (Baldwin and Taglioni 2006). In order to simultaneously deal with all these problems, we estimate Eq. (1) using a zero-inflated Poisson pseudo-maximum likelihood (ZIPPML) model (Burger et al. 2009) with country-fixed effects C_i (Baldwin and Taglioni 2006) to overcome endogeneity due to omitted terms (e.g., related to non-observable variables as landed prices of origin goods in destination country, which may in principle be correlated with trade-cost terms, as proxied by distance).[7] To double-check our results, we also estimate Eq. (1) with alternative econometric approaches discussed in the relevant literature (Linders and de Groot 2006), including standard OLS on the log-linearized form (omitting zero-valued flows or substituting zero-valued flows by a small constant), log-normal pseudo-maximum likelihood estimators as in Santos Silva and Tenreyro (2006), and zero-inflated negative-binomial pseudo-maximum likelihood (ZINBMML) techniques (Burger et al. 2009). Overall, we obtain the best fit using the ZIPPML estimator, although our results are not dramatically different under alternative estimation strategies, both in terms of model selection and correlation of residuals, a tendency already documented in Linders and de Groot (2006).

---

[7] Note that all variables have been original deflated, but this does not create any difficulty here as we undertake a cross-section study.

Table 2 Estimated gravity equation for year 2000 Poisson regression

Regressor (coefficient)	Estimate (Rob. SE)	Regressor (coefficient)	Estimate (Rob. SE)
GDP$_i$ ($\hat{\alpha_1}$)	1.471*** (0.000)	LL$_i$ ($\hat{\delta_1}$)	−0.336*** (0.000)
GDP$_j$ ($\hat{\alpha_2}$)	1.338*** (0.000)	LL$_j$ ($\hat{\delta_2}$)	−0.019*** (0.000)
DIST ($\hat{\alpha_3}$)	−0.727*** (0.000)	CTG ($\hat{\theta_1}$)	0.553*** (0.000)
AREA$_i$ ($\hat{\beta_{11}}$)	−0.144*** (0.000)	COML ($\hat{\theta_2}$)	0.242*** (0.000)
AREA$_j$ ($\hat{\beta_{21}}$)	−0.187*** (0.000)	COL ($\hat{\theta_3}$)	0.007*** (0.001)
POP$_i$ ($\hat{\beta_{12}}$)	−0.504*** (0.000)	TA ($\hat{\theta_4}$)	0.024*** (0.000)
POP$_j$ ($\hat{\beta_{22}}$)	−0.413*** (0.000)	–	–
Constant ($\hat{\alpha_0}$)	NO	Wald χ^2	17600000
Country Dummies ($\hat{\gamma_i}, \hat{\gamma_j}$)	YES***	Prob $> \chi^2$	0.00***
Adj. R^2	0.93	Vuong Z	5.73
Log Likelihood	−631000	Prob $>$Z	0.00***

Dependent variable $= w_{ij}$. No. observations $= 12{,}561$. Estimates obtained using a ZIPPML procedure. Robust standard errors in parentheses. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 2 reports the ZIPPML fit to year-2000 data, together with usual goodness-of-fit statistics for our preferred (selected) model.[8] The final list of regressors in Table 2 has been selected using a general-to-specific procedure. In other words, we have initially fitted Eq. (1) using all available independent variables, i.e. setting $X = $(AREA, POP, CPI, EXC, RM), $Z = $(LL, CONT), and $D = $(CTG, COMC, COML, COL, TA, COMR). Next, we have successively removed from the list those regressors in X, Z and D that turned out to be statistically not significant. Only GDP, DIST, AREA, POP, LL, CTG, COML, COL, TA resisted this successive selection procedure.[9] Therefore, in the final model specification (see Table 2), $X^* = $(AREA, POP), $Z^* = $(LL), and $D^* = $(CTG, COML, COL, TA). Accordingly, the regressors CONT, RM, CPI, COMC, EXC and COMR were not considered in the estimation of the residual network, i.e. are set to zero in Eq. (1), see below. For example, the common religion (COMR) dummy and nominal exchange rates (EXC) were dropped because they did not appear to be statistically different from zero (at 5%). Conversely, consumer-price index (CPI), both common-currency (COMC) and continent (CONT) dummy, and the remoteness indicator (RM) were discarded because, notwithstanding a weakly statistically-significant coefficient, the absolute values of the estimates were very close to zero.

---

[8] The ZIPPML is a two-stage procedure (Wooldridge 2001). The first stage features a logit regression estimating the probability that there is no bilateral trade at all. In the second stage, a Poisson regression is run to fit trade flows of the non-zero group of link (as estimated in the first stage). We use the same set of regressors to model both the first and second stage. Here we show only the estimation results for the Poisson final stage for the sake of exposition (the complete set of regression results is available on request from the Author). All exercises have been performed using Stata 9 (StataCorp 2007).

[9] This result is also consistent with a sequence of standard likelihood-ratio tests aimed at removing, in each step, only the regressor that causes a not significant improvement in the likelihood function.

The final gravity equation thus reads:

$$w_{ij} = \alpha_0 [\text{GDP}_i^{\alpha_1} \text{GDP}_j^{\alpha_2} \text{DIST}_{ij}^{\alpha_3}] \cdot [\text{AREA}_i^{\beta_{11}} \text{AREA}_j^{\beta_{21}} \text{POP}i^{\beta_{12}} \text{POP}j^{\beta_{22}}]$$
$$\cdot \exp \left\{ \theta_1 \text{CTG}_{ij} + \theta_2 \text{COML}_{ij} + \theta_3 \text{COL}_{ij} + \theta_4 \text{TA}_{ij} \right.$$
$$\left. + \delta_1 \text{LL}_i + \delta_2 \text{LL}_j + \gamma_i C_i + \gamma_j C_j \right\} \eta_{ij}. \qquad (2)$$

A quick look at diagnostic statistics in Table 2 indicates that the final preferred model in Eq. (2) seems well-specified, according to both Wald (Wooldridge 2001) and Vuong (1989) test statistics, and achieves an extremely large adjusted R^2 (0.93)—a typical performance of trade gravity equations. All signs and magnitudes of estimated coefficients are in line with those found in the relevant literature. Trade flows are positively affected by country GDP, existence of trade agreements, common borders, common language and colony relationships. They are negatively affected by distance, area and population (net of GDP effects), as well as the probability of being landlocked. Country remoteness, continental position, religion, exchange rates and CPI-effects do not appear instead to significantly affect trade flows in our data as often documented in the gravity-equation literature.

Define $\hat{\eta}_{ij}$ as the estimated residual from Eq. (2), i.e. obtained by substituting unknown parameters with estimated ones from Table 2. Note that $\hat{\eta}_{ij}$ can be interpreted as the weight of the ij trade link once all structural effects related to country size, geographical, social, historical and political factors have been removed from the original weight w_{ij}. It is then straightforward to define the "residual" weighted ITN by simply re-weight links using the $\hat{\eta}_{ij}$s.[10] More formally:

Definition 2 The residual weighted-undirected network representation of the ITN in year 2000 is the graph where the nodes are the N countries in the sample and links are fully characterized by the $N \times N$ symmetric weight matrix $E = \{e_{ij}\}$, where $e_{ij} = \hat{\eta}_{ij}$ for $i > j$, $e_{ji} = e_{ij}$ for $i < j$, $e_{ii} = 0$ for all i, $\hat{\eta}_{ij} = w_{ij}/\hat{w}_{ij}$, where

$$\hat{w}_{ij} = \hat{\alpha_0} [\text{GDP}_i^{\hat{\alpha}_1} \text{GDP}_j^{\hat{\alpha}_2} \text{DIST}_{ij}^{\hat{\alpha}_3}] \cdot [\text{AREA}_i^{\hat{\beta}_{11}} \text{AREA}_j^{\hat{\beta}_{21}} \text{POP}i^{\hat{\beta}_{12}} \text{POP}j^{\hat{\beta}_{22}}]$$
$$\cdot \exp \left\{ \hat{\theta}_1 \text{CTG}_{ij} + \hat{\theta}_2 \text{COML}_{ij} + \hat{\theta}_3 \text{COL}_{ij} + \hat{\theta}_4 \text{TA}_{ij} \right.$$
$$\left. + \hat{\delta}_1 \text{LL}_i + \hat{\delta}_2 \text{LL}_j + \hat{\gamma}_i C_i + \hat{\gamma}_j C_j \right\}. \qquad (3)$$

and estimated coefficients are as in Table 2.

In what follows, we will study the topological properties of the residual ITN and compare them with those of original ITN (Definition 1).

4 Topological properties of gravity-equation residual network

One of the most puzzling stylized facts emerging from the study of the topological properties of W is that the distribution of link weights w_{ij} is well approximated

---

[10] See Krempel and Pluemper (2003) for a germane approach aimed at visualizing the properties of the ITN.

Fig. 2 Residual link-weight distribution. Size-rank plot with power-law fit, estimated equation and R^2 of the fit. Note: Estimated coefficients refer to original values and not to the rank transformation
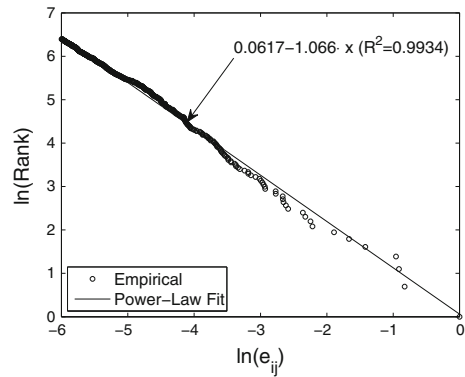
by a log-normal density (Li et al. 2003; Fagiolo et al. 2009b). Indeed, this militates against the view that the ITN is a complex network, as log-normality can be simply the limit outcome of uncorrelated link-weight multiplicative growth processes (e.g., Gibrat laws; see Sutton 1997). In other words, original link weights are markedly heterogeneous but display exponentially-decaying upper tails, without the typical fat-tailed behavior that is known to be the signature of complexity (Mitzenmacher 2004; Rosser 2008). The present analysis shows that, once all gravity-equation dependence is removed from the original data, the residual ITN is characterized by power-law distributed link weights, see the size-rank plot (Stanley et al. 1995) in Fig. 2. In fact, the correlation between original and residual weights is not statistically-significantly different from zero (-0.009, with a p-value 0.9171). We shall go back to the implications of this finding on modeling below. For the moment, let us stress the fact that power-law behavior of residual link weights hints to an inherent complex behavior of trade flows, possibly due to deep similarities between countries, which are somewhat hidden by the standard determinants of trade accounted for in the empirical literature.

We now explore how the other topological properties of the residual ITN compare to original ones. As it is customary in this literature, we shall focus on node strength (NS), defined as the sum of all link weights of a node; node average nearest-neighbor strength (ANNS), i.e. the average strength of the trade partners of that node; weighted node-clustering (WCC), measured as the relative weighted intensity of trade triangles with that node as one of the vertices; and random-walk betweenness centrality (RWBC), accounting for global centrality of a node in the weighted network.[11]

As happens for link weights, power-law shapes characterize in the residual ITN all node statistics (NS, ANNS, and WCC) that were originally well-proxied by log-normal densities (Fagiolo et al. 2009b), see Fig. 3 for NS. Interestingly, the only topological property that was power-law shaped in the original ITN (RWBC) keeps the same shape also in the residual network. RWBC is actually a peculiar statistics, because, unlike

---

[11] See "Appendix C" for formal definitions. In the text, we will not make use of subscripts indicating the country to which the statistics applies in order to simplify the notation. Notice also that the ZIPPML estimator seems to consistently predict the density of the residual ITN. Indeed, the fraction of all possible links with positive weight is 0.63 under W and 0.62 in the residual graph E.

Fig. 3 Residual node-strength (NS) distribution. Size-rank plot with power-law fit, estimated equation and R^2 of the fit. *Note*: Estimated coefficients refer to original values and not to the rank transformation
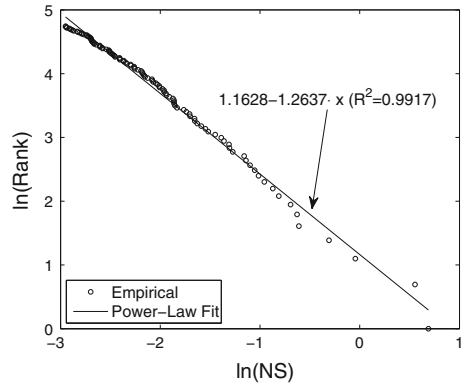

Table 3 Correlation structure between topological properties, and country per-capita GDP (pcGDP)

		W				E				pcGDP
		NS	ANNS	WCC	RWBC	NS	ANNS	WCC	RWBC	
W	NS	-	-0.3453	0.9484	0.5741	**-0.0881**	0.2902	0.0331	-0.0909	0.5170
	ANNS	-	-	-0.4774	-0.3759	**0.0051**	-0.7753	-0.2967	-0.0887	-0.4590
	WCC	-	-	-	0.5437	**-0.1133**	0.3985	0.0698	-0.1011	0.5968
	RWBC	-	-	-	-	**-0.0797**	0.3178	0.0554	-0.0673	0.4975
E	NS	-	-	-	-	-	0.1155	0.8363	0.5202	-0.1678
	ANNS	-	-	-	-	-	-	0.3834	0.1574	0.3312
	WCC	-	-	-	-	-	-	-	0.5193	-0.0961
	RWBC	-	-	-	-	-	-	-	-	-0.1908

In boldface not-statistically significant correlation coefficients
W original weighted ITN, *E* residual weighted ITN

the others, each of its node values somewhat reflect the whole structure of the network. The fact that RWBC is still power-law distributed suggests that complexity is really an intrinsic feature of the ITN.

Another robust set of stylized facts regarding the weighted ITN concerns the correlation structure among node-distributions of topological statistics, and between node topological statistics and country per-capita GDP (as a proxy for country income). Table 3 reports such a correlation structure for both the original ITN (W) and the residual one (E). Coefficients not statistically-significantly different from zero are marked in boldface. The original ITN, where all correlation coefficients were statistically different from zero, hinted to a trade structure where countries that trade more intensively are also high-income ones, they are more clustered and central, but tend to trade with relatively-less connected partners. This configures a relatively disassortative pattern for the ITN. Once size, geographically-related and other determinants of trade have been removed from the data, however, the topological properties of the residual ITN are almost uncorrelated with their original counterpart. The only exception is ANNS, which displays a strong and negative correlation. This suggests that countries that in W typically traded with intensively connected partners (i.e., small and poor countries) exhibit in E small ANNS values, i.e. tend to trade with poorly-connected partners, a pattern that can be intuitively explained by recalling that trade flows in E do not reflect any size, geographic, or colonial preferential relationship. Note also that
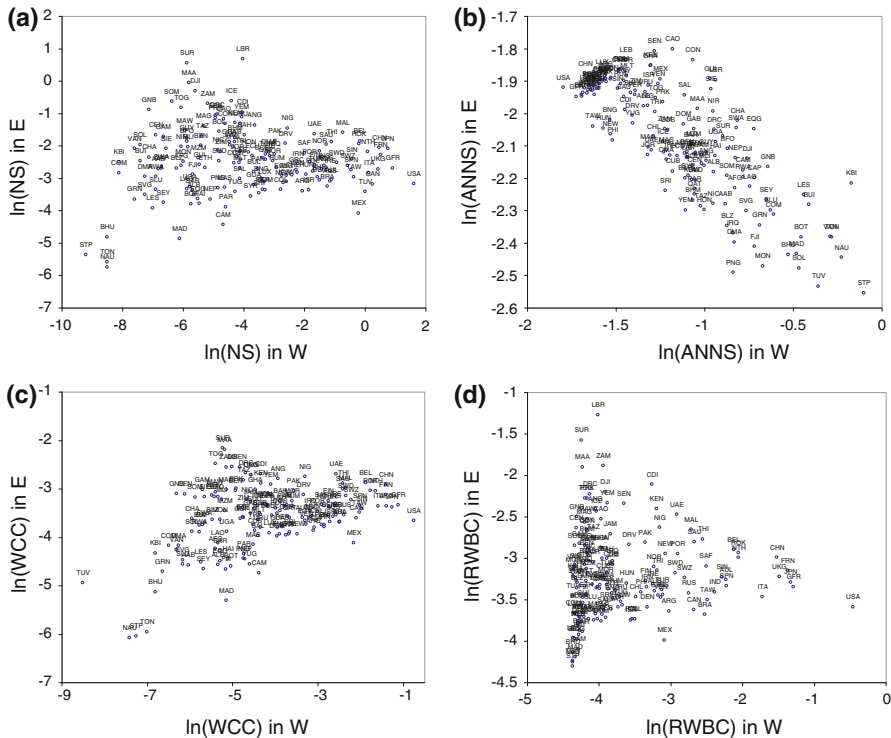
Fig. 4 Topological properties of the ITN in original vs. residual network. (**a**) NS; (**b**) ANNS; (**c**) WCC; (**d**) RWBC. Year: 2000

in E countries that trade more intensively (i.e., high NS) still are more central and clustered, but do not display any assortative or disassortative pattern anymore, as the NS-ANNS correlation in E is not significantly-different from zero. Furthermore, the removal of size effects (GDP and population) naturally destroys any positive correlation between income and trade intensity, centrality and clustering: now high-income countries tend to trade relatively less intensively and occupy less central positions (and trade with relatively more connected partners).

This result is confirmed by looking at the Spearman's rank-correlation coefficient between country-statistic rankings. If one correlates, e.g., the rank of countries under W and E according to node statistics, it emerges that rankings made with respect to NS, WCC and RWBC are only weakly (positively) correlated, whereas the correlation for ANNS rankings is equal to -0.7676. This suggests that if a country was scoring high in terms of intensity of trade (NS), clustering (WCC) and centrality (RWBC) in the original ITN, it is not likely to keep its top position in the residual network, and with high probability will appear at the bottom of the list in the ANNS ranking, see Fig. 4. For example, whereas the United Staetes (USA), Germany (GFR) and Japan (JPN) used to occupy top positions in the rankings of trade intensity, clustering and centrality, in the residual network such positions are now filled by relatively small but very dynamic countries like Iceland (ICE) and Belgium (BEL), as well
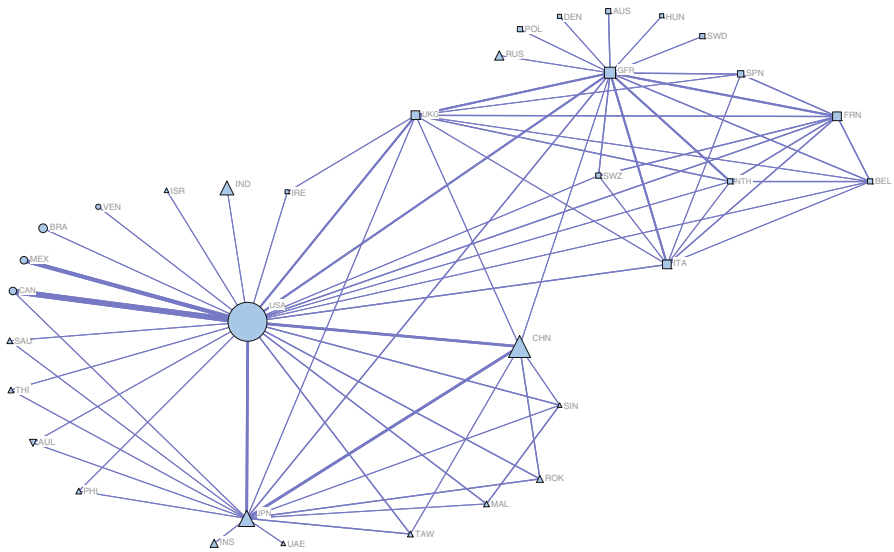
Fig. 5 A partial visualization of the original weighted ITN (W). Thickness of links is proportional to their weight. Only the largest 1% of links are shown. Node sizes are proportional to country's GDP. *Node shapes* represent the continent which the country belongs to (*Circles*: America; *Empty Squares*: Europe; *Upright Triangles*: Asia; *Crossed Squares*: Africa; *Reversed Triangles*: Pacific)

as middle-east oil-related countries (United-Arab Emirates, UAE) and many African ones. Furthermore, the ANNS ranking, which was topped in W by micro-economies (e.g., Pacific islanders) now features in the first positions relatively large and active countries like Cameroon (CAO), Senegal (SEN), Kenya (KEN), Rumania (RUM), Colombia (COL), Morocco (MOR), South Africa (SAF), Mexico (MEX), as well as China (CHN) and Indonesia (INS).

All in all, the foregoing analysis shows that only few of the stylized facts found for the weighted ITN apply to its residual version, and that a great extent of ITN topological features can be explained by the control variables employed in the gravity-equation regression. Our results also indicate that, removing such structure from the data and interpreting residuals as proxies for the underlying trade similarities between countries, interesting patterns emerge about the inherent complex structure of the ITN and how countries are interconnected and play their roles in the network.

Additional insights on these relationships can be gathered from a simple weighted-graph visualization of the original vs. residual ITN. Figures 5 and 6 depict a partial view (largest 1% links in terms of weights) of the original and residual ITN in year 2000 as an undirected graph where the thickness of a link is proportional to its weight, node size is proportional to country's GDP, and node shapes represent the continent which the country belongs to. It is easy to see that, as discussed in the previous sections, large countries tend to trade more intensively with each other, with contiguous/close countries or with partners that are relevant in terms of history and trade agreements. The original ITN is also characterized by a core-periphery structure where large influential countries (USA, Japan, China, Germany, France, Italy, etc.) play the role of
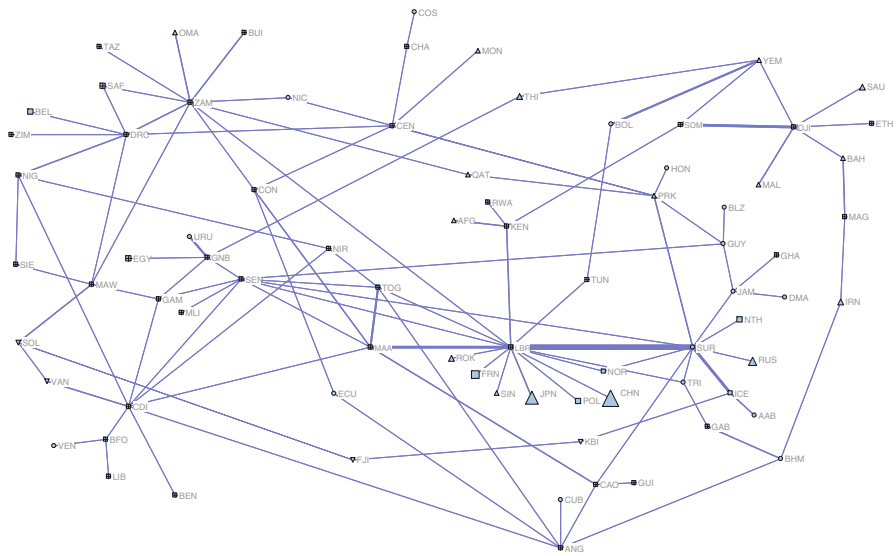
Fig. 6 A partial visualization of the residual weighted ITN (E). Thickness of links is proportional to their weight. Only the largest 1% of links are shown. Node sizes are proportional to country's GDP. *Node shapes* represent the continent which the country belongs to (*Circles*: America; *Empty Squares*: Europe; *Upright Triangles*: Asia; *Crossed Squares*: Africa; *Reversed Triangles*: Pacific)

hubs, attached to many other countries by relatively weaker trade links. Geographical clustering is also evident, as same-continent countries tend to be tightly interconnected.

Simple inspection of the residual weighted ITN graph (links with largest 1% of weight in E) hints instead to a much more interconnected pattern, characterized by much more spread links, without large-sized hubs, nor excessive geographical clustering. All large-GDP countries disappear from the graph (e.g., USA, Germany, Italy, France). Small and medium-sized African, South-American and Asian countries are most represented, possibly because of their endowments in tradable natural resources and/or over-specialization of their export profile.

Some countries, like Liberia (LBR), Surinam (SUR), Mauritania (MAA) and Djibuti (DJI), play an unexpectedly prominent role in the network. These are countries with relatively low per-capita GDPs, which in the original network were holding many trade relationships (between 52 and 88) and scored around the median in all node-statistic rankings (except centrality). The majority of their trade relationships were weak but occurred outside trade agreements. Unlike Pacific islanders, such flows did not necessarily concern large-sized partners, were not related to colony relationships and not necessarily geographically close (they are not landlocked). Therefore, it is not a surprise that their residual link weights end up being relatively large for many existing relationships, thus promoting such countries to more important positions in the network.

Finally, to get a more precise feeling of the role of large-sized countries in the residual network, we plot the complete minimal spanning trees (MSTs) associated to
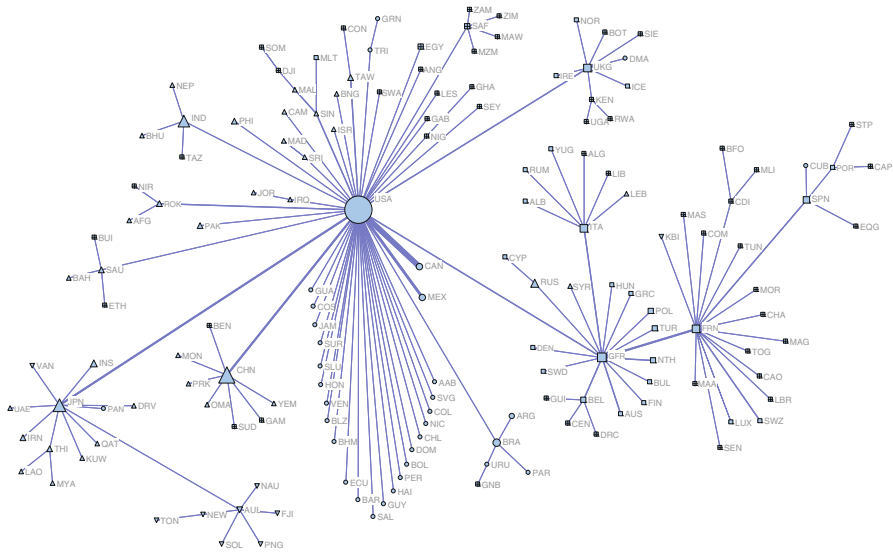
Fig. 7 Complete minimal spanning tree (MST) of the original weighted ITN (W). Thickness of a link is proportional to one minus the weight associated to that link in the MST returned by the Kruskal's algorithm (Cormen et al. 2001, Chap. 23.2). Node sizes are proportional to country's GDP. *Node shapes* represent the continent which the country belongs to (*Circles*: America; *Empty Squares*: Europe; *Upright Triangles*: Asia; *Crossed Squares*: Africa; *Reversed Triangles*: Pacific)

W and E.[12] This graphical analysis may complement the former because, by simplifying the structure of the network, allows one to simultaneously visualize all existing nodes and their relevant interactions, a thing that was simply impossible to do with W and E, given the extremely large density of the two graphs. The plot of W-MST (Fig. 7) magnifies the hub-role of the largest countries, especially the US, and better accounts for the position of other peripheral countries in the ITN. Conversely, the plot of E-MST (Fig. 8) hints to a relatively more central role of very small countries like Liberia and Surinam (see above); downplays the importance of hubs and geography; and relegates large countries at the periphery. For instance, US and China now link to small countries in the ITN, whereas India keeps having a marginal position.

---

[12] See Gower and Ross (1969). To build the MST of a weighted graph with link-weights in the unit interval we have employed the following procedure. First, assume that information from link weights can be treated the same way as correlation coefficients in Mantegna (1999), i.e. that higher (symmetric) link weights signal a higher trade similarity between countries. Second, transform weight (i.e., similarity) matrices entries w_{ij} and e_{ij} into, respectively, $w_{ij}^R = \sqrt{2(1 - w_{ij})}$ and $e_{ij}^R = \sqrt{2(1 - e_{ij})}$, and let $w_{ii}^R = e_{ii}^R = 0$. As explained in Mantegna (1999), this is an appropriate metric distance. Third, we have computed the MST associated to W^R and E^R, and re-scaled the link weights returned by the Kruskal's algorithm (Cormen et al. 2001, Chap. 23.2) by their maximum values, in line with what we have done for W and E. Finally, we have weighted any resulting link by $1 - w_{ij}^R$ and $1 - e_{ij}^R$, to get the right proportionality between weights and similarity.
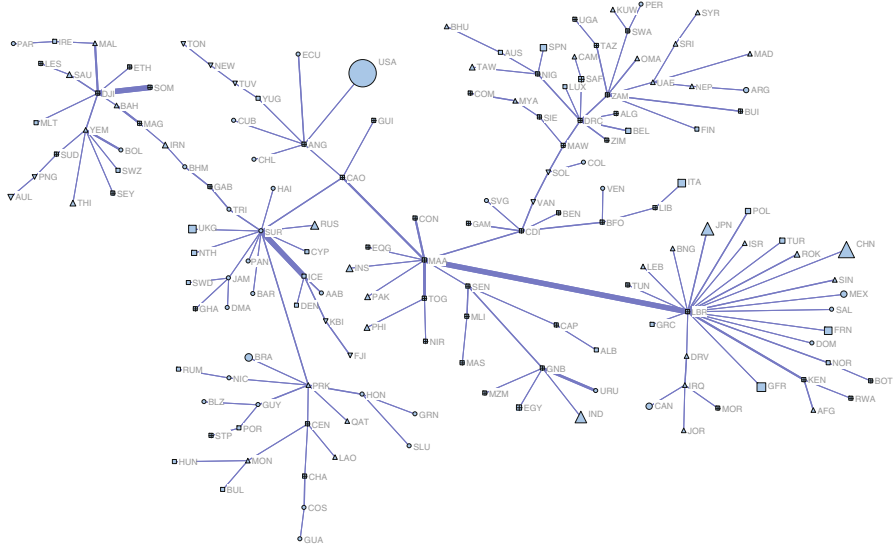
Fig. 8 Complete minimal spanning tree (MST) of the residual weighted ITN (E). Thickness of a link is proportional to one minus the weight associated to that link in the MST returned by the Kruskal's algorithm (Cormen et al. 2001, Chap. 23.2). Node sizes are proportional to country's GDP. *Node shapes* represent the continent which the country belongs to (*Circles*: America; *Empty Squares*: Europe; *Upright Triangles*: Asia; *Crossed Squares*: Africa; *Reversed Triangles*: Pacific)

5 Conclusions

In this paper we have begun to investigate the determinants of the statistical features of the weighted international-trade network. We have compared the topological properties of the ITN, as originally explored in a series of papers (Li et al. 2003; Bhattacharya et al. 2008, 2007; Garlaschelli et al. 2007; Fagiolo et al. 2008; Fagiolo et al. 2009b), to those of the residual weighted symmetric ITN. The latter has been defined as the estimated residuals of a gravity-equation regression where original bilateral-trade flows depend on country GDP, population and area; geographical distance between countries; and a series of link- or country-specific variables accounting for other factors related to geography (common border, landlocking, continent), the economy (trade agreements, exchange rates, consumer price indices), and social/ political/ historical effects (common language, religion, and currency). The residual ITN, therefore, can be considered as a "filtered" version of the original one, where all factors known to affect bilateral trade flows or total trade between any two countries have been removed. In other words, the structure and topology of the residual ITN can be considered as reflecting the variance that remains unexplained after having accounted for all the standard determinants of bilateral trade flows.

Such a filtering strategy also allows one to address the issue of how to define, in the first place, the most relevant weighting scheme for the international-trade network. Indeed, starting from a binary representation of the ITN, where only the presence/absence of links is accounted for, it is often useful to complement the analysis

with a weighted network approach, which fully exploits the existing among-link heterogeneity in terms of intensity of trade carried by any single link (Fagiolo et al. 2008). In such a case, one typically faces the problem of how selecting the correct (or most economically meaningful) weighting scheme, i.e. choosing the correct variable that will define the weight of any bilateral link. The standard solution, i.e. using exports or total trade, can be criticized because it is well-known that at the very least this variable strongly depends on the origin and target market sizes, but also on the geographical distances between the two countries, and all other factors discussed in the gravity-equation literature (see Sect. 3). By factoring out this dependence from original trade flows, one can then appreciate the topological properties of the ITN that cannot be explained by those "usual suspects".

Our findings indicate that the residual ITN is characterized by power-law shaped distributions of link weights and node statistics (e.g., strength, clustering, random-walk betweenness centrality). Hence, the underlying architecture of the weighted ITN seems to exhibit signatures of complexity, unlike its original counterpart. We have also found that the correlation structure among node statistics, and between node statistics and country per-capita GDP, changes substantially when comparing the original and residual weighted networks. Whereas the original ITN displays a structure with a few large-sized country hubs and a relatively strong connectivity among nearby countries (either geographically, socially, politically or economically), the residual ITN is organized around many relatively small-sized but trade-oriented countries that, independently of their geographical position, either play the role of local hubs or attract large and rich countries in complex trade-interaction patterns.

Notice also that, in principle, one may consider to take a more specific-to-general approach to gravity-equation modeling, starting with a minimal specification, and adding an increasing number of factors to the gravity equation to be estimated. For example, one may begin with country size only, and gradually introduce geography, trade agreements, etc. At each step a residual ITN can be defined and its properties can be accordingly compared to those of the original ITN. Preliminary results in this direction show that such a minimal set must contain at the very least country size effects (country GDP). In that case, a great deal of the topological properties of the residual ITN found above are preserved (e.g. correlation structure, power-law distributions, etc.). Adding geographical distance to this minimal set allows for a much better match with the residual ITN generated out of equation (1). Conversely, as Fig. 1 suggests, geographical distance alone is not able to generate a residual ITN that looks like those obtained using equation (1) or GDP only.

Our results have some implications for modeling purposes. Indeed, the strong gravity-like dependence that, in line with existing literature, can be detected in ITN data, as well as the impact of removing such structure on ITN topological properties, suggests that models of network formation and link-weight evolution aiming at replicating ITN topological properties should consider country size, geography, etc. among their building blocks. This may occur, for instance, by devising dynamic models of network formation and weight evolution where node size and position, contiguity and agreements, matter in the decision of the nodes to form/delete a link, or to change its weight. At the same time, models aiming at explaining and replicating ITN stylized facts should allow empirical calibration of their node- and link-related characteristics

(see Garlaschelli and Loffredo 2004, 2005; Garlaschelli et al. 2007; Bhattacharya et al. 2008 for preliminary exercises in this direction). This may be relevant for predictive and policy exercises. Furthermore, our findings call for models that are simultaneously able to replicate the set of stylized facts of both the original and the residual ITN (no matter the determinants accounted for in the gravity equation). For example, it would be important to have models of the ITN that recover log-normality of the (equilibrium) link-weight distribution when size and geography are not accounted for, and power-law shapes after having controlled for them.

To conclude, it may be worthwhile to notice that the simple and preliminary exercises described in this article may be extended in at least three directions. First, one could replicate the same experiment without symmetrizing the network and allowing for a weighted-directed representation of international trade flows. Second, one could play with expanded versions of the gravity equation that consider, e.g., country-specific natural endowments, industrial profile and trade specialization. This might shed some light on the peculiar structure of the residual network and explain trade patterns that otherwise could remain obscure. Finally, one might want to check the robustness of the foregoing results across different years. This might involve either comparing estimates from several cross-sectional gravity equations, or considering dynamic specifications of the gravity equation in line with, e.g., Baier and Bergstrand (2007).

Appendix A

See Table 4.

Table 4 List of countries in the balanced panel (1981–2000)

ID	Acro	Name	ID	Acro	Name	ID	Acro	Name
2	USA	United States	355	BUL	Bulgaria	600	MOR	Morocco
20	CAN	Canada	360	RUM	Rumania	615	ALG	Algeria
31	BHM	Bahamas	365	RUS	Russia	616	TUN	Tunisia
40	CUB	Cuba	375	FIN	Finland	620	LIB	Libya
41	HAI	Haiti	380	SWD	Sweden	625	SUD	Sudan
42	DOM	Dominican Rep.	385	NOR	Norway	630	IRN	Iran
51	JAM	Jamaica	390	DEN	Denmark	640	TUR	Turkey
52	TRI	Trinidad/Tobago	395	ICE	Iceland	645	IRQ	Iraq
53	BAR	Barbados	402	CAP	Cape Verde	651	EGY	Egypt
54	DMA	Dominica	403	STP	Sao Tome	652	SYR	Syria
55	GRN	Grenada	404	GNB	Guinea-Bissau	660	LEB	Lebanon
56	SLU	Saint Lucia	411	EQG	Eq. Guinea	663	JOR	Jordan

Table 4 continued

ID	Acro	Name	ID	Acro	Name	ID	Acro	Name
57	SVG	St. Vincent	420	GAM	Gambia	666	ISR	Israel
58	AAB	Antigua	432	MLI	Mali	670	SAU	Saudi Arabia
70	MEX	Mexico	433	SEN	Senegal	678	YEM	Yemen
80	BLZ	Belize	434	BEN	Benin	690	KUW	Kuwait
90	GUA	Guatemala	435	MAA	Mauritania	692	BAH	Bahrain
91	HON	Honduras	436	NIR	Niger	694	QAT	Qatar
92	SAL	El Salvador	437	CDI	Cote Dć6Ivoire	696	UAE	Arab Emirates
93	NIC	Nicaragua	438	GUI	Guinea	698	OMA	Oman
94	COS	Costa Rica	439	BFO	Burkina Faso	700	AFG	Afghanistan
95	PAN	Panama	450	LBR	Liberia	710	CHN	China
100	COL	Colombia	451	SIE	Sierra Leone	712	MON	Mongolia
101	VEN	Venezuela	452	GHA	Ghana	713	TAW	Taiwan
110	GUY	Guyana	461	TOG	Togo	731	PRK	North Korea
115	SUR	Surinam	471	CAO	Cameroon	732	ROK	South Korea
130	ECU	Ecuador	475	NIG	Nigeria	740	JPN	Japan
135	PER	Peru	481	GAB	Gabon	750	IND	India
140	BRA	Brazil	482	CEN	Centr African Rep.	760	BHU	Bhutan
145	BOL	Bolivia	483	CHA	Chad	770	PAK	Pakistan
150	PAR	Paraguay	484	CON	Congo	771	BNG	Bangladesh
155	CHL	Chile	490	DRC	Congo (Zaire)	775	MYA	Myanmar
160	ARG	Argentina	500	UGA	Uganda	780	SRI	Sri Lanka
165	URU	Uruguay	501	KEN	Kenya	781	MAD	Maldives
200	UKG	United Kingdom	510	TAZ	Tanzania	790	NEP	Nepal
205	IRE	Ireland	516	BUI	Burundi	800	THI	Thailand
210	NTH	Netherlands	517	RWA	Rwanda	811	CAM	Cambodia
211	BEL	Belgium	520	SOM	Somalia	812	LAO	Laos
212	LUX	Luxembourg	522	DJI	Djibouti	816	DRV	Vietnam
220	FRN	France	530	ETH	Ethiopia	820	MAL	Malaysia
225	SWZ	Switzerland	540	ANG	Angola	830	SIN	Singapore
230	SPN	Spain	541	MZM	Mozambique	840	PHI	Philippines
235	POR	Portugal	551	ZAM	Zambia	850	INS	Indonesia
260	GFR	Germany	552	ZIM	Zimbabwe	900	AUL	Australia
290	POL	Poland	553	MAW	Malawi	910	PNG	Papua
305	AUS	Austria	560	SAF	South Africa	920	NEW	New Zealand
310	HUN	Hungary	570	LES	Lesotho	935	VAN	Vanuatu
325	ITA	Italy	571	BOT	Botswana	940	SOL	Solomon's
338	MLT	Malta	572	SWA	Swaziland	950	FJI	Fiji
339	ALB	Albania	580	MAG	Madagascar	970	KBI	Kiribati
345	YUG	Yugoslavia	581	COM	Comoros	971	NAU	Nauru
350	GRC	Greece	590	MAS	Mauritius	972	TON	Tonga
352	CYP	Cyprus	591	SEY	Seychelles	973	TUV	Tuvalu

Appendix B

See Table 5.

Table 5 List of link- or country-related additional variables employed in gravity-equation exercises

Label	Related to	Explanation	Source
GDP	Country	Gross-domestic product	Gleditsch (2002)
AREA	Country	Country area in km^2	CEPII (http://www.cepii.fr/)
POP	Country	Country population	Gleditsch (2002)
LL	Country	Dummy variable equal to 1 for landlocked countries	CEPII (http://www.cepii.fr/)
CONT	Country	Dummy variables recording the continent to which the country belongs	CEPII (http://www.cepii.fr/)
RM	Country	Country remoteness index defined as the weighted average of the distances of a country to all other countries, with weights equal to the country share of world GDP, see Bhavnani et al. (2002)	Our computations on CEPII and Gleditsch (2002) data
CPI	Country	Consumer price index	International Monetary Fund (www.imf.org)
DIST	Link	Geodesic geographical distance between two countries, calculated with the great circle formula	CEPII (http://www.cepii.fr/)
CTG	Link	Contiguity dummy equal to 1 if two countries share a common border	CEPII (http://www.cepii.fr/)
COMC	Link	Dummy equal to 1 if two countries use the same currency	Our computation on Andrew Rose dataset (http://faculty.haas.berkeley.edu/arose), see also Glick and Rose (2001)
COML	Link	Dummy equal to 1 if the official language (or mother tongue or second language) of the two co untries is the same	CEPII (http://www.cepii.fr/)
COL	Link	Dummy equal to 1 if the two countries share a substantial colonizer-colonized relationship	C EPII (http://www.cepii.fr/)
TA	Link	Dummy variable equal to 1 for countries involved in regional, bilateral or preferential trade agree ments in year 1995 still in place in year 2000	WTO (http://www.wto.org/)
EXC	Link	Nominal exchange rates	International Monetary Fund (www.imf.org)
COMR	Link	Dummy equal to 1 if the two countries share a common religion	Our computations on CIA F actbook dataset (https://www.cia.gov/library/publications/the-world-factbook/fields/2122.html)

Appendix C

Network statistics

Given a $N \times N$ symmetric weight matrix $W = \{w_{ij}\}$, with $0 \leq w_{ij} \leq 1$, define the associated symmetric adjacency matrix as $A = \{a_{ij}\}$, where $a_{ij} = 1$ iff $w_{ij} > 0$ and zero otherwise. In the paper, we make use of the following statistics:

– *Node degree* (Albert and Barabási 2002), defined as $\text{ND}_i = A_{(i)}\mathbf{1}$, where $A_{(i)}$ is the i-th row of A and $\mathbf{1}$ is a unary vector. ND is a measure of binary connectivity, as it counts the number of trade partners of any given node. Although we mainly focus here on a weighted-network approach, we study ND because of its natural interpretation in terms of number of trade partnerships and bilateral trade agreements.
– *Node strength* (DeMontis et al. 2005), defined as $\text{NS}_i = W_{(i)}\mathbf{1}$, where $W_{(i)}$ is the i-th row of W. NS is a measure of weighted connectivity, as it gives us an idea of how intense existing trade relationships of country i are.
– *Node average nearest-neighbor strength* (DeMontis et al. 2005), that is $\text{ANNS}_i = (A_{(i)}W\mathbf{1})/(A_{(i)}\mathbf{1})$. ANNS measures how intense are trade relationships maintained by the partners of a given node. Therefore, the correlation between ANNS and NS is a measure of network assortativity (if positive) or disassortativity (if negative).
– *Weighted clustering coefficient* (Saramaki et al. 2007; Fagiolo 2007), defined as

$$WCC_i = \frac{\left(W^{\left[\frac{1}{3}\right]}\right)^3_{ii}}{ND_i(ND_i - 1)}.$$

Here Z^3_{ii} is the i-th entry on the main diagonal of $Z \cdot Z \cdot Z$ and $Z^{[m]}$ stands for the matrix obtained from Z after raising each entry to m. WCC measures how much clustered a node i is from a weighted perspective, i.e. how much intense are the linkages of trade triangles having country i as a vertex. Replacing W with A, one obtains the standard binary clustering coefficient (BCC), which counts the fraction of triangles existing in the neighborhood of any give node (Watts and Strogatz 1998).
– *Random-walk betweenness centrality* (Newman 2005; Fisher and Vega-Redondo 2006), which is a measure of how much a given country is globally-central in the ITN. A node has a higher random-walk betweenness centrality (RWBC) the more it has a position of strategic significance in the overall structure of the network. In other words, RWBC is the extension of node betweenness centrality to weighted networks and measures the probability that a random signal can find its way through the network and reach the target node where the links to follow are chosen with a probability proportional to their weights.

References

Abeysinghe T, Forbes K (2005) Trade linkages and output-multiplier effects: a structural var approach with a focus on asia. Rev Int Econ 13:356–375
Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Mod Phys 74:47–97
Anderson JE (1979) A theoretical foundation for the gravity equation. Am Econ Rev 69:106–116

Anderson JE, van Wincoop E (2003) Gravity with gravitas: a solution to the border puzzle. Am Econ Rev 93:170–192

Artis M, Galvão A-B, Marcellino M (2003) The transmission mechanism in a changing world. Economics working papers ECO2003/18. European University Institute

Baier SL, Bergstrand JH (2007) Do free trade agreements actually increase members' international trade?. J Int Econ 71:72–95

Baldwin R, Taglioni D (2006) Gravity for dummies and dummies for gravity equations. NBER working papers 12516. National Bureau of Economic Research

Barrat A, Barthélemy R, Pastor-Satorras R, Vespignani A (2004) The architecture of complex weighted networks. Proc Natl Acad Sci 101:3747–3752

Barthélemy M, Barrat A, Pastor-Satorras R, Vespignani A (2005) Characterization and modeling of complex weighted networks. Physica A 346:34–43

Bergstrand JH (1985) The gravity equation in international trade: some microeconomic foundations and empirical evidence. Rev Econ Stat 67:474–481

Bhattacharya K, Mukherjee G, Manna S (2007) The international trade network. In: Chatterjee A, Chakrabarti B (eds) Econophysics of markets and business networks. Springer, Milan

Bhattacharya K, Mukherjee G, Särämaki J, Kaski K, Manna S (2008) The international trade network: weighted network analysis and modeling. J Stat Mech Theory Exp A 2:P02002

Bhavnani R, Coe DT, Subramanian A, Tamirisa NT (2002) The missing globalization puzzle. IMF working papers 02/171. International Monetary Fund

Burger M, Oort Fv, Linders G (2009) On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. Research Paper ERS-2009-003-ORG Revision. Erasmus Research Institute of Management (ERIM)

Cormen TH, Leiserson CE, Rivest RL, Stein C (2001) Introduction to algorithms. MIT Press and McGraw-Hill, Cambridge and New York

Deardorff A (1998) Determinants of bilateral trade: does gravity work in a neoclassical world? In: The regionalization of the world economy. NBER Chapters. National Bureau of Economic Research, pp 7–32

DeMontis A, Barthélemy M, Chessa A, Vespignani A (2005) The structure and evolution of inter-urban traffic: a weighted network analysis. Environ Plan B Plan Design 34:905–924

Dorogovtsev S, Mendes J (2003) Evolution of networks: from biological nets to the internet and WWW. Oxford University Press, Oxford

Eichengreen B, Irwin DA (1996) The role of history in bilateral trade flows. NBER Working Papers 5565. National Bureau of Economic Research

Fagiolo G (2007) Clustering in complex directed networks. Phys Rev E 76:026107

Fagiolo G, Schiavo S, Reyes J (2008) On the topological properties of the world trade web: a weighted network analysis. Physica A 387:3868–3873

Fagiolo G, Schiavo S, Reyes J (2009a) The ecolution of the world trade web: a weighted-network approach. J Evol Econ (forthcoming)

Fagiolo G, Schiavo S, Reyes J (2009b) World-trade web: Topological properties, dynamics, and evolution. Physical Review E 79:036115

Feenstra RC, Markusen JR, Rose AK (2001) Using the gravity equation to differentiate among alternative theories of trade. Can J Econ 34:430–447

Fisher E, Vega-Redondo F (2006) The linchpins of a modern economy. Working paper. Cal Poly

Forbes K (2002) Are trade linkages important determinants of country vulnerability to crises? In: Sebastian E, Jeffrey F (eds) Preventing currency crises in emerging markets. University of Chicago Press, Chicago

Frankel J (1997) Regional trading blocs in the world economic system. Institute for International Economics, Washington DC

Fratianni M (2009) The gravity model in international trade. In: Rugman AM (ed) The Oxford handbook of international business. Oxford University Press, Oxford

Garlaschelli D, Di Matteo T, Aste T, Caldarelli G, Loffredo M (2007) Interplay between topology and dynamics in the world trade web. Eur Phys J B 57:1434–6028

Garlaschelli D, Loffredo M (2004) Fitness-dependent topological properties of the world trade web. Phys Rev Lett 93:188701

Garlaschelli D, Loffredo M (2005) Structure and evolution of the world trade network. Physica A 355:138–144

Gleditsch K (2002) Expanded trade and GDP data. J Conflict Resolut 46:712–724

Glick R, Rose AK (2001) Does a currency union affect trade? The time series evidence. NBER working papers 8396. National Bureau of Economic Research

Gower JC, Ross GJS (1969) Minimum spanning trees and single linkage cluster analysis. Appl Stat 18:54–64

Hayfield T, Racine JS (2008) Nonparametric econometrics: the np package. J Stat Softw 27

Helliwell JF, Padmore T (1985) Empirical studies of macroeconomic interdependence. In: Jones R, Kenen P (eds) Hand book of international economies. Elsevier, Amsterdam, The Netherlands

Hurvich C, Simonoff J, Tsai C (1998) Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. J Royal Stat Soc Series B 60:271–293

Jackson MO, Demange G, Goyal S, Nouwel AVD (2003) A survey of models of network formation: stability and efficiency. In: In group formation in economics: networks, clubs and coalitions. Cambridge University Press, Cambridge

Kali R, Méndez F, Reyes J (2007) Trade structure and economic growth. J Int Trade Econ Dev 16:245–269

Kali R, Reyes J (2007) The architecture of globalization: a network approach to international economic integration. J Int Bus Stud 38:595–620

Krempel L, Pluemper T (2003) Exploring the dynamics of international trade by combining the comparative advantages of multivariate statistics and network visualizations. J Soc Struct 4:1–22

Krugman P (1995) Growing world trade: causes and consequences. Brookings Pap Econ Act 26:327–377

Leamer EE, Levinsohn J (1995) International trade theory: the evidence (chapter 26). In: Grossman GM, Rogoff K (eds) Handbook of international economics, vol 3 of handbook of international economics. Elsevier. pp 1339–1394

Li Q, Racine J (2004) Cross-validated local linear nonparametric regression. Statistica Sinica 14:485–512

Li X, Jin YY, Chen G (2003) Complexity and synchronization of the world trade web. Physica A Stat Mech Appl 328:287–296

Linders G-JM, de Groot HL (2006) Estimation of the gravity equation in the presence of zero flows. Tinbergen Institute discussion papers 06-072/3. Tinbergen Institute

Mantegna R (1999) Hierarchical structure in financial markets. Eur Phys J B 11:193–197

Mitzenmacher M (2004) A brief history of generative models for power law and lognormal distributions. Internet Math 1:226–251

Newman M (2005) A measure of betweenness centrality based on random walks. Soc Networks 27:39–54

Overman HG, Redding S, Venables AJ (2003) The economic geography of trade, production, and income: a survey of empirics. In: Harrigan J, Choi K (eds) Handbook of international trade. Blackwell Publishers, Oxford pp 353–387

Reichardt J, White D (2007) Role models for complex networks. Eur Phys J B 60:217–224

Reyes J, Schiavo S, Fagiolo G (2008) Assessing the evolution of international economic integration using random-walk betweenness centrality: the cases of East Asia and Latin America. Adv Complex Syst 11:685–702

Rose AK (2000) One money, one market: the effects of common currency on trade. Econ Policy 30:9–45

Rose AK, Spiegel MM (2002) A gravity model of sovereign lending: trade, default and credit. Working papers in applied economic theory 2002–09. Federal Reserve Bank of San Francisco

Rosser JB (2008) Econophysics and economic complexity. Adv Complex Syst 11:745–760

Santos Silva JMC, Tenreyro S (2006) The log of gravity. Rev Econ Stat 88:641–658

Saramaki J, Kivelä M, Onnela J, Kaski K, Kertész J (2007) Generalizations of the clustering coefficient to weighted complex networks. Phys Rev E 75:027105

Serrano A, Boguñá M (2003) Topology of the world trade web. Phys Rev E 68:015101(R)

Serrano A, Boguñá M, Vespignani A (2007) Patterns of dominant flows in the world trade web. J Econ Interact Coord 2:111–124

Stanley MHR, Buldyrev SV, Havlin S, Mantegna RN, Salinger MA, Eugene Stanley H (1995) Zipf plots and the size distribution of firms. Econ Lett 49:453–457

StataCorp (2007) Stata statistical software: release 10. StataCorp LP, College Station

Sutton J (1997) Gibrat's legacy. J Econ Lit 35:40–59

Tzekina I, Danthi K, Rockmore D (2008) Evolution of community structure in the world trade web. Eur Phys J B Condens Matter 63:541–545

Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57:307–333

Watts D, Strogatz S (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442

Wooldridge JM (2001) Econometric analysis of cross section and panel data. The MIT Press, Boston