The double role of GDP in shaping the structure of the International Trade Network

Assaf Almog*

Instituut-Lorentz for Theoretical Physics,
Leiden Institute of Physics,
University of Leiden,
Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
Email: almog@lorentz.leidenuniv.nl
*Corresponding author

Tiziano Squartini

IMT School for Advanced Studies Lucca,
P.zza S. Ponziano 6,
55100 Lucca, Italy
Email: squartini@libero.it
Email: tiziano.squartini@imtlucca.it

Diego Garlaschelli

Instituut-Lorentz for Theoretical Physics,
Leiden Institute of Physics,
University of Leiden,
Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
Email: garlaschelli@lorentz.leidenuniv.nl

Abstract: The International Trade Network (ITN) is the network formed by trade relationships between world countries. The complex structure of the ITN impacts important economic processes such as globalisation, competitiveness, and the propagation of instabilities. Modelling the structure of the ITN in terms of simple macroeconomic quantities is therefore of paramount importance. While traditional macroeconomics has mainly used the gravity model to characterise the magnitude of trade volumes, modern network theory has predominantly focused on modelling the topology of the ITN. Combining these two complementary approaches is still an open problem. Here we review these approaches and emphasise the double role played by gross domestic product (GDP) in empirically determining both the existence and the volume of trade linkages. Moreover, we discuss a unified model that exploits these patterns and uses only the GDP as the relevant macroeconomic factor for reproducing both the topology and the link weights of the ITN.

Keywords: network theory; econophysics; exponential random graph model; fitness model; GDP; gross domestic product.

Reference to this paper should be made as follows: Almog, A., Squartini, T. and Garlaschelli, D. (2017) 'The double role of GDP in shaping the structure of the International Trade Network', *Int. J. Computational Economics and Econometrics*, Vol. 7, No. 4, pp.381–398.

Biographical notes: Assaf Almog is PhD candidate at the Lorentz Institute for Theoretical Physics, Leiden University (NL), since 2013. He graduated in 2012 with a Master's degree in Applied Physics from Delft Technical University (NL). His research interests focus on network theory and economic complexity.

Tiziano Squartini is Assistant Professor at the IMT School for Advanced Studies Lucca. He defended his PhD thesis in Siena ('Information-theoretic approach to the analysis of complex networks') in 2011. In 2012–2013, he was Postdoctoral Researcher at the Lorentz-Institute for Theoretical Physics (Leiden, NL) under the supervision of Diego Garlaschelli. In 2014–2015, he was Postodoctoral Researcher at the Institute for Complex Systems in Rome, under the supervision of Luciano Pietronero. His research activity focuses on the application of network theory to financial and economic systems.

Diego Garlaschelli is an Associate Professor at the Lorentz Institute for Theoretical Physics, Leiden University, and Associate Fellow of the Saïd Business School, Oxford University. Since 2011, he leads a research group with strongly interdisciplinary interests, including network theory, economic complexity, social dynamics, statistical physics and graph theory. He teaches courses in complex networks and econophysics at the Faculty of Science in Leiden.

This paper is a revised and expanded version of a paper entitled 'The double role of GDP in shaping the structure of the International Trade Network' presented at *IWcee15. International Workshop on Computational Economics and Econometrics*, Rome, 28–29 May, 2015.

1 Introduction

The bilateral trade relationships existing between world countries form a complex network known as the International Trade Network (ITN). The observed complex structure of the network is at the same time the outcome and the determinant of a variety of underlying economic processes, including economic growth, integration and globalisation. Moreover, recent events such as the financial crisis clearly pointed out that the interdependencies between financial markets can lead to cascading effects which, in turn, can severely affect the real economy. International trade plays a major role among the possible channels of interaction among countries (Kali et al., 2007, 2010; Schiavo et al., 2010; Saracco et al., 2015), thereby possibly further propagating these cascading effects worldwide and adding one more layer of contagion. Characterising the networked worldwide economy is therefore an important open problem and modelling the ITN represents a crucial step of this challenge (Serrano et al., 2003, 2007; Fagiolo et al., 2008; Barigozzi et al., 2010; De Benedictis and Tajoli, 2011; Cristelli et al., 2013; Sinha et al., 2010).

Historically, macroeconomic models have mainly focused on modelling the trade volumes between countries. The gravity model, which was introduced in the early 1960s by Tinbergen (1962) (see discussion in Squartini et al., 2014), serves as powerful empirical model that aims at predicting the bilateral trade flow between any two (trading) countries based on the knowledge of their gross domestic product (GDP) and mutual geographic distance. Although the model has been upgraded, over the years, to include other possible factors of macroeconomic relevance, like common language and trade agreements, GDP and distance remain the two factors with largest explanatory power.

The gravity model can reproduce the observed trade volumes between countries satisfactorily. However, at least in its simplest and most popular implementation, the model cannot account for zero volumes, therefore predicting a fully-connected trade network. This outcome is totally inconsistent with the observed, heterogeneous, topology of the ITN, which represents the backbone on which trade is made. Subsequent refinements of the gravity model allowing for zero trade flows succeeded only in reproducing the total number of missing links, not their position in the trade network, thereby producing sparser but still non-realistic topologies (Fagiolo et al., 2009; Fagiolo, 2010).

In conjunction with the traditional macroeconomic approach, recent years have witnessed an approach to modelling the ITN using tools from network theory (Garlaschelli et al., 2004, 2005, 2007; Fronczak et al., 2012; Bhattacharya et al., 2008), among which maximum-entropy techniques (Squartini et al., 2011, 2013, 2015) have been proven to be particularly successful.Maximum-entropy models aim at reproducing higher-order structural properties of real-world networks using lower-order information (more precisely, node-specific), which is constrained to be reproduced (Wells, 2004; Bargigli and Gallegati, 2011; Musmeci et al., 2013; Caldarelli et al., 2013). Important examples of local properties that can be chosen as constraints are the *degree*, i.e., the number of links of a node (in the ITN case, this is the number of trade partners of a country) and the *strength*, i.e., the total weight of the links of a node (in the ITN case, this is the total trade volume of a country). Examples of higher-order properties that the method aims at reproducing are the *clustering coefficient*, which refers to the fraction of realised triangles around nodes, and the *degree correlations*.

These studies have focused on both binary and weighted representations of the ITN, i.e., the two representations defined by the *existence* and by the *magnitude* of trade exchanges among countries, respectively. In principle, depending on which local properties are chosen as constraints, maximum-entropy models can either fail or succeed in replicating the higher-order properties of the ITN. As an example, it has been shown that inferring the network topology only from purely weighted properties such as the strength of all nodes (i.e., the trade volumes of all countries) results in a trivial, uniform structure (almost fully connected and, thus, unrealistic) (Squartini et al., 2011). This limitation is similar to the one discussed above for the gravity model, which aims at reproducing the pairwise traded volumes exclusively, while completely ignoring the underlying network topology. By contrast, the knowledge of purely topological properties such as the degrees of all nodes (i.e., the number of trade partners of all countries), which are usually neglected in traditional macroeconomic models, turns out to be essential for reproducing the heterogeneous topology observed in the ITN (Squartini et al., 2011). A combination of weighed and topological local properties allows to reconstruct the higher-order properties of the ITN with extremely high accuracy (Mastrandrea et al., 2014).

Despite the ability of maximum-entropy models to provide a better agreement with the data with respect to gravity models, in principle the former do not provide any hint on the underlying (macro)economic factors shaping the structure of the network under consideration. These models, in fact, assign 'hidden variables' or 'fitness parameters' to each country. These quantities arise as Lagrange multipliers involved in the constrained maximisation of the entropy and control for the probability that a link is established and/or has a given weight. Even if, a priori, these parameters have no economic interpretation, here we propose a macroeconomic identification for the underlying variables defining the maximum-entropy models. This interpretation is supported by previous studies showing that both topological and weighted properties of the ITN are strongly connected with purely macroeconomic quantities, in particular the GDP (Almog et al., 2015).

In this paper we first focus on various empirical relations existing between the GDP and a range of country-specific properties. These properties convey basic but important local information from a network perspective. We also show that these relations are robust and very stable throughout different decades. We then illustrate to which extent the GDP affects the binary and weighted representations of the ITN. These results suggest a justification for the use of GDP as an empirical fitness to be used in maximum-entropy models, thus providing a macroeconomic interpretation for the abstract mathematical parameters defining the model themselves. Reversing the perspective, this result enables us to introduce a novel GDP-driven model (Almog et al., 2015) that successfully reproduces the binary *and* the weighted properties of the ITN simultaneously.

The different structural mechanisms (binary and weighted) in the model can be mapped to existing economic literature on extensive and intensive margins of trade, hence providing an elegant bridge between the two approaches. More specifically, our formalism allows us to introduce the (static) concept of extensive and intensive bias, definable as a the tendency of nodes to either prefer the formation of extra links or the reinforcement of existing link weights. Our results show that while the topology of the ITN can be successfully inferred without any information about the weighted properties, the ITN weighted structure cannot be inferred without any topological information: this is the origin of the limitations of the traditional Gravity Model which, disregarding the network topology, is unable to faithfully reproduce the ITN inter-linkages structure. The mathematical structure of the model translates this puzzling asymmetry into the informativeness of binary and weighted constraints (degree and strength) (Almog et al., 2015). These results represent a promising step forward in the formulation of a unified model for modelling the structure of the ITN.

2 Data

In this study, we have used data from the Gleditsch database which spans the years 1950–2000 (Gleditsch et al., 2002), focusing only on the first year of each decade, i.e., six years in total. The datasets are available in the form of weighted matrices of bilateral trade flows w_{ij}, the associated adjacency matrices a_{ij} and vectors of GDPs. There are approximately 200 countries in the dataset covering the considered 51 years; the GDP is measured in US dollars.

We have analysed this dataset precisely because it has been the subject of many studies so far, focusing both on the binary and on the weighted representation of the ITN. This will allow us to compare the performance of our GDP-driven (two-steps) method with other reconstruction algorithms already present in the literature (Mastrandrea et al., 2014).

3 Empirical evidence

Trade exchanges between countries play a crucial role in many macroeconomic phenomena. As a consequence, it is fundamental to be able to characterise the observed structure of the

ITN and its properties. More specifically, the ITN can be represented in two different ways, depending on the kind of information used to analyse the system: the first one concerns only the existence of trade relations and gives origin to the ITN *binary representation*; the second one also takes into account the volume of the trade exchanges and gives origin to the ITN *weighted representation*. While the binary representation describes the skeleton of the ITN, relating exclusively to the presence of trade relations, the weighted representation also accounts for the volume of trade occurring 'over' the links, i.e., the weight of the link once it is formed. The two representations convey very important information regarding the 'trade patterns' of each country and, most importantly, correspond to different trade mechanisms.

Traditionally, macroeconomic models have mainly focused on the weighted representation, because economic theory perceives the latter as being genuinely more informative than the purely binary representation: such models make use of countries gross domestic product (GDP), their geographic distance and any other possible quantity of (supposed) macroeconomic relevance to infer trading volumes between countries. The GDP is the most popular measure in the economic literature. Although it is generally used as a proxy to infer the evolution of many macroeconomic properties describing the weighted representation of the ITN (as the countries trade exchanges), here we will show that the GDP plays a key role not only to explain the ITN weighted structure, but also the emergence of its binary structure.

Let us start with an empirical analysis of the GDP. We first define new rescaled quantities of the GDP: g_i and $\tilde{g}_i$

$$g_i \equiv \frac{\text{GDP}_i}{\sum_j \text{GDP}_j},\ \forall\, i \qquad \tilde{g}_i \equiv \frac{\text{GDP}_i}{\text{GDP}_{mean}},\ \forall\, i, \tag{1}$$

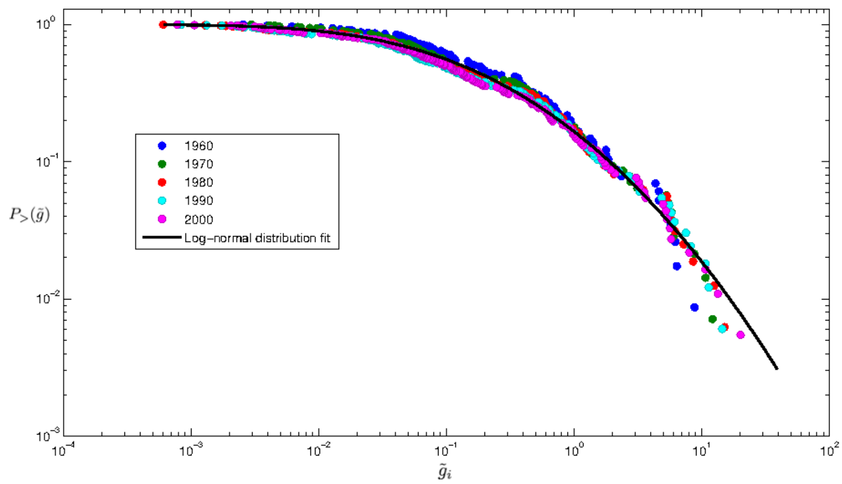
where $\text{GDP}_{\text{mean}} \equiv \frac{\sum_i^N \text{GDP}_i}{N}$ is the average GDP for an observed year. The two quantities adjust the values of the countries GDPs for both the size of the network and the growth, and are a connected by a simple relation $\tilde{g}_i = N \cdot g_i$. We use the two quantities of the rescaled GDP throughout our analysis, mainly using g_i for the reason that the quantity is bounded $0 \leq g_i \leq 1$ which coincides with our model.

In Figure 1, we plot the cumulative distribution of the rescaled GDP $\tilde{g}_i$ with i indexing the countries for the different decades collected into our dataset. What emerges is that the distributions of the rescaled GDPs can be described by log-normal distribution characterised by similar values of the parameters. The log-normal curve is fitted to all the values (from the different decades). This suggests that the rescaled GDPs are quantities which do not vary much with the evolution of the system, thus potentially representing the (constant) hidden macroeconomic fitness ruling the entire evolution of the system itself. This, in turn, implies understanding the functional dependence of the key topological quantities on the countries rescaled GDP.

As already indicated by a number of results (Squartini et al., 2013), the topological quantities which play a major role in determining the ITN structure are the countries degrees (i.e., the number of their trading partners) and the countries strengths (i.e., the total volume of their trading activity). Thus, the first step to understand the role of the rescaled GDP in shaping the ITN structure is quantifying the dependence of degrees and strengths on it. Since we will now analyse each snapshot at a time (correction for size is not needed), here we will use the bounded rescaled GDP g_i. Moreover, this form of the rescaled GDP coincides with a bounded macroeconomic fitness value, which is consistent with the models presented in the next sections.

To this aim, let us explicitly plot k_i vs. g_i and s_i vs. g_i for a particular decade, as shown in Figure 2. The red points represent the relations between the two pairs of observed quantities for the 2000 snapshot. Interestingly, the rescaled GDP is directly proportional to the strength (in a log-log scale), thus indicating that *the wealth of countries is strongly correlated to the total volume of trade they participate in*. Such an evidence provides the empirical basis for the definition of the gravity model, stating that *the trade between any two countries is directly proportional to the (product of the) countries GDP*.

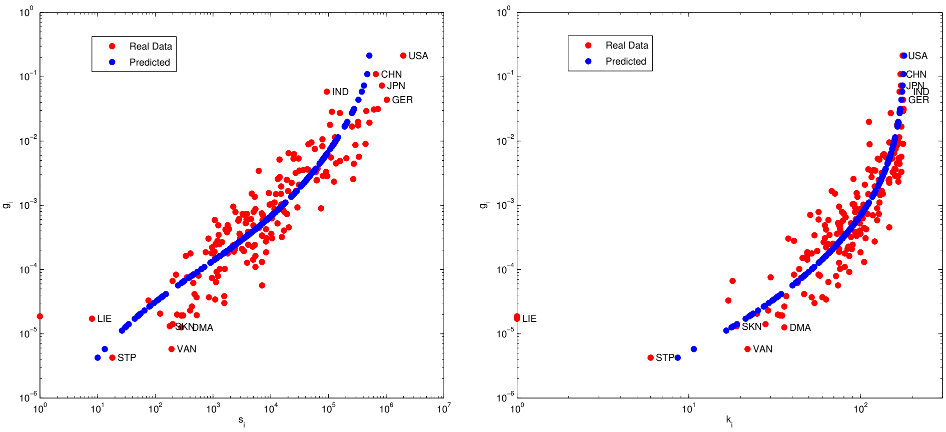
Figure 1 Empirical cumulative distributions $P_>(\tilde{g})$ of the GDP rescaled to the mean, for different years. The curve is a log-normal distribution fitted to the data (see online version for colours)



On the other hand, the functional dependence of the degrees on the g_i values is less simple to decipher. Generally speaking, the relation is monotonically increasing and this means that countries with high GDP have also an high degree, i.e., are strongly connected with the others; coherently, countries characterised by a low value of the GDP have also a low degree, i.e., are less connected to the rest of the world. Moreover, while for low values of the GDP there seems to exist a linear relation (in a log-log scale) between k_i and g_i, as the latter rises a saturation effect is observed (in correspondence of the value $k_{\max} = N - 1$), due to the finite size of the network under analysis. Roughly speaking, richest countries lie on the vertical trait of the plot, while poorest countries lie on the linear trait of the same plot: in other words, *the degree of countries represents a purely topological indicator of the countries wealth*.

To sum up, Figure 2 shows that countries GDP plays a double role in shaping the ITN structure: first, it controls for the number of trading channels each country establishes; second, it controls for the volume of trade each country participates in, via the established connections. The blue points in Figure 2, instead, represent the relation between $\langle k_i \rangle$ vs. g_i and $\langle s_i \rangle$ vs. g_i, where the quantities in brackets are the predicted values for degrees and strengths generated by our model, which we will discuss later.

Figure 2 Comparison between observed (red points) and expected (blue points) degrees and strengths for the aggregated ITN in the 2000 snapshot. Right panel: degree k_i vs. normalised GDP g_i and expected degree $\langle k_i \rangle$ vs. normalised GDP g_i. Left panel: strength s_i vs. normalised GDP g_i and expected strength $\langle s_i \rangle$ vs. normalised GDP g_i (see online version for colours)



4 Null models

In order to formalise the evidence highlighted in the previous section, a theoretical framework is needed. To this aim, we can make use of the *exponential random graph* formalism (ERG in what follows). Under this formalism, one 'generates' a ensemble of random networks by maximising the entropy of the ensemble. However, the maximisation is done under certain 'constraints' which enforce certain properties of the random ensemble (expectations) to be equal specific observables that are measured in the real system. Different maximum-entropy models enforce different constraints, different properties of the real network, and this corresponds to different probabilities and expectations of the models.

Here, we use the formulas defining the so-called *enhanced configuration model* (ECM in what follows) which has been recently proposed as an improved model for the ITN reconstruction (Mastrandrea et al., 2014). The ECM aims at reconstructing weighted networks, by enforcing the degree and the strength sequences simultaneously (i.e., the number of neighbours of each node and the total sum of the weights attached to each node's connection) (Mastrandrea et al., 2014). In words, this amounts at choosing only some structural quantities of a given network $\mathbf{W}$ and treating them as constraints of a Shannon entropy maximisation process. In particular, degrees and strengths, respectively defined as $k_i(\mathbf{W}) = \sum_{j \neq i}^{N} a_{ij} = \sum_{j \neq i}^{N} \Theta[w_{ij}]$, $\forall\, i$ and $s_i(\mathbf{W}) = \sum_{j \neq i}^{N} w_{ij}$, $\forall\, i$, can be simultaneously constrained within the ERG framework (Mastrandrea et al., 2014). Such a recipe amounts at choosing a network $\mathbf{W}$ with a probability coefficient given by $P(\mathbf{W}) \propto e^{-\sum_{i=1}^{N} \alpha_i k_i(\mathbf{W}) + \beta_i s_i(\mathbf{W})}$: the higher the effectiveness of the chosen constraints in reproducing the network structure, the higher the probability of picking it as the result of the drawing process. From the perspective of network theory, specifying the countries degrees amounts to reproduce the binary structure of the ITN or, as previously said, its skeleton; on the other hand, specifying the countries strengths amounts to reconstruct the weight of each link. In economic terms, this amounts to retain two different kinds of information: the number of trading partners of each country *and* the total volume of trade of each country.

Notice that previous attempts to infer the binary structure of the ITN from the information encoded into the strength sequence alone have led to the prediction of a largely homogeneous and very dense (sometimes fully connected) network, not compatible with the observed one. In other words, predicting the number of partners of a given country from the total volume of its trade leads to 'dilute' the total trade of each country by distributing it to almost all other countries, dramatically overestimating the number of trading partners (Squartini et al., 2013). This failure in correctly replicating the purely topological projection of the real network is at the root of the bad agreement between expected and observed higher-order properties and makes it necessary to explicitly constrain the degree of each country. This evidence should lead us to reconsider the quantities traditionally used in economic models and the actual role played by them in explaining a given network structure. Particularly, one must add additional information regarding the topology of the network in order to reproduce the complex structure of the ITN.

As a result of constraining both degrees and strengths, the ECM predicts that a trade relation between countries i and j exists with a probability p_{ij} equal to

$$\langle a_{ij}\rangle(\mathbf{x},\mathbf{y}) \equiv p_{ij}(\mathbf{x},\mathbf{y}) = \frac{x_i x_j y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j} \tag{2}$$

(with $x_i = e^{-\alpha_i}$ and $y_i = e^{-\beta_i}$) and involves an expected volume of trade amounting to

$$\langle w_{ij}\rangle(\mathbf{x},\mathbf{y}) = \frac{p_{ij}(\mathbf{x},\mathbf{y})}{1 - y_i y_j} = \frac{x_i x_j y_i y_j}{(1 - y_i y_j + x_i x_j y_i y_j)(1 - y_i y_j)}. \tag{3}$$

The unknown vectors $\mathbf{x}$ and $\mathbf{y}$ can be estimated according to the maximum-of-the-likelihood prescription (Mastrandrea et al., 2014), by solving the system of $2N$ coupled equations

$$k_i(\mathbf{W}^*) = \sum_{j\neq i}^{N} p_{ij}(\mathbf{x}^*,\mathbf{y}^*), \forall\, i \quad \text{and} \quad s_i(\mathbf{W}^*) = \sum_{j\neq i}^{N} \langle w_{ij}\rangle(\mathbf{x}^*,\mathbf{y}^*), \forall\, i, \tag{4}$$

where $\mathbf{W}^*$ indicates the particular weighted network under analysis and $\mathbf{x}^*$ and $\mathbf{y}^*$ indicate the values of the Lagrange multipliers satisfying equations (4). These parameters can be treated as fitness parameters, respectively controlling for the probability that a link exists and that its expected weight assumes a given value.

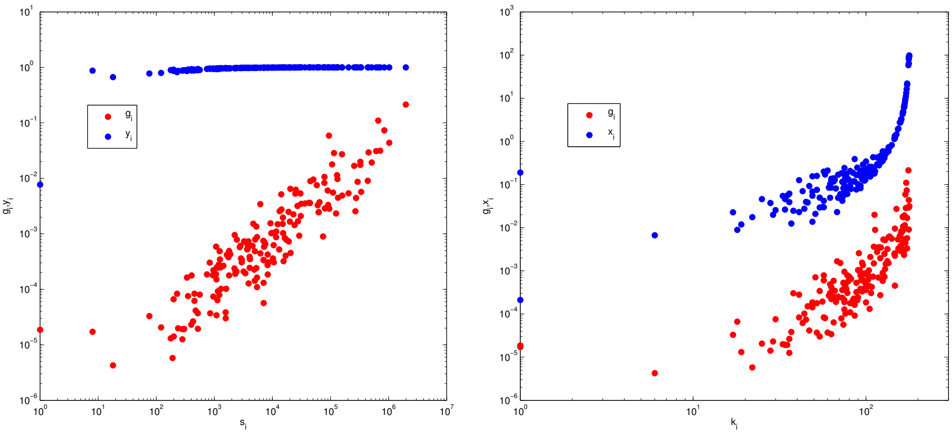
The application of the ECM to various real-world networks shows that the model can accurately reproduce the higher-order empirical properties of these networks (Mastrandrea et al., 2014). When applied to the ITN in particular, the ECM replicates both binary and weighted empirical properties, for different levels of disaggregation, and for several years (Mastrandrea et al., 2014).

5 A GDP-driven model of the ITN

Let us now make a step forward and check whether the hidden variables x_i and y_i, which effectively reproduce the observed ITN (Mastrandrea et al., 2014), can be thought of as parameters having a clear (macro)economic interpretation. Let us start our analysis by first inspecting the relationship between the ECM statistics k_i and s_i and the hidden variables extracted from the model.

As Figure 3 shows, nodes degrees k_i seems to be related to the quantities x_i and g_i through a very similar relationship; on the other hand, the functional relation between s_i and y_i appears to be less straightforward, showing a saturation effect in correspondence of the value $y = 1$. In order to discover the mathematical form of these relations, let us repeat the analysis which led to Figure 3, by plotting x_i and y_i vs. g_i.

Figure 3 Comparison between observed (red points) and expected (blue points) degrees and strengths for the aggregated ITN in the 2000 snapshot. Right panel: degree k_i vs. normalised GDP g_i and fitness parameter x_i (calculated by the model). Left panel: strength s_i vs. normalised GDP g_i and fitness parameter y_i (calculated by the model) (see online version for colours)



In Figure 4, we show the relationship between the two ECM parameters x_i and y_i and the rescaled GDP for each country of the ITN in the 2000 snapshot. Such quantities are strongly correlated, confirming the linear dependence between x_i and g_i and $y_i/(1 - y_i)$ and g_i respectively. The latter, in particular, is the simplest functional form guaranteeing the presence of the vertical asymptote emerging from the plot as s_i vs. y_i.

5.1 The GDP as a macroeconomic fitness

Figure 4 seems to suggest that the fitness parameter x_i satisfies a approximately linear relation with the relative GDP g_i, fitted by the curve

$$x_i = \sqrt{a} \cdot g_i \tag{5}$$

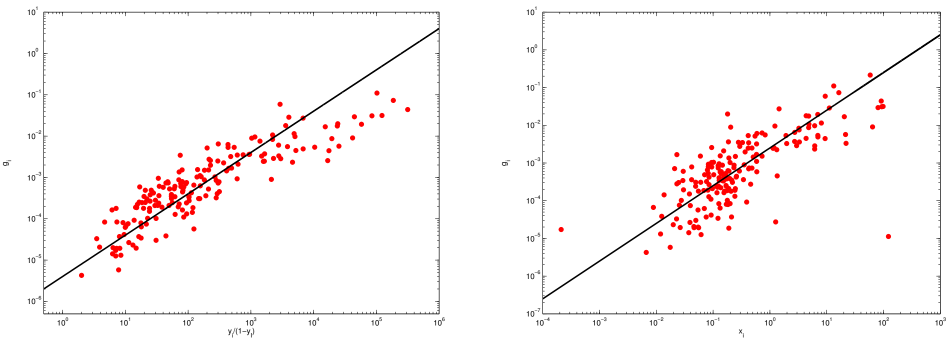
where $\sqrt{a}$ is a parameter and $g_i = \frac{\text{GDP}_i}{\sum_i \text{GDP}_i}$.

By contrast, since the GDP is an unbounded quantity, while the fitness parameter y_i is bounded between 0 and 1 (this is a mathematical property of the model (Mastrandrea et al., 2014; Garlaschelli and Loffredo, 2009)), the relation between y_i and g_i must be necessarily non-linear. A simple functional form for such a relationship is given by

$$y_i = \frac{b \cdot g_i^c}{1 + b \cdot g_i^c}. \tag{6}$$

Indeed, Figure 4 confirms that the above expression provides a very good fit to the data.

Figure 4 Comparison between the calculated x_i and the rescaled GDP g_i (right panel) and for the calculated $y_i/(1 - y_i)$ and the relative GDP g_i (left panel), for the aggregated ITN in the 2000 snapshot, together with a linear fit (black line) (see online version for colours)



These findings have two important consequences: first, they confirm that the GDP of world countries plays a double role, contributing to determine both the topological structure of the ITN and the amount of trade exchanges; second, since the relationships summed up by equations (5) and (6) hold true for each snapshot of the ITN in our dataset, for each year we can insert equations (5) and (6) into equations (2) and (3) to obtain a GDP-driven model of the ITN structure for that year. While this was already expected on the basis of the results obtained by implementing simpler null models – constraining either the degree sequence alone (the *binary configuration model*, or BCM (Squartini et al., 2013)) or the strength sequence alone (the *weighted configuration model*, or WCM (Squartini et al., 2013)) – finding the appropriate way to explicitly combine these results into a unified description of the ITN has remained impossible so far.

5.2 Reformulating the ECM as a 'two-step' model

It should be noted that equations (5) and (6) can be thought of as a particular case of a popular model among physicists, the so-called *fitness model* (Caldarelli et al., 2002), which prescribes to write the connection probability p_{ij} between any two nodes i and j as a function of some intrinsic 'fitness' characterising each vertex. This observation leads to the identification of the fitness parameter with the GDP of countries, thus suggesting that, from a purely economic point of view, GDP is the only relevant quantity that must be taken into account in order to explain the observed structural patterns. Such a procedure, first adopted in (Garlaschelli et al., 2004) to study the purely binary structure of the ITN[1] – where a very good agreement between the hidden variables z_i, controlling solely for the degree of node i, and the rescaled $\tilde{g}_i$ has been shown – not only allows one to make predictions of the quantities of interest based on purely (country-specific) macroeconomic properties but also provides an algorithm to test the effectiveness of the chosen quantities in reproducing such observations. In fact, equations (5) and (6) could be, in principle, refined by further inserting any supposedly relevant macroeconomic quantity (as the geographic distances);

however, their actual (macro)economic relevance would then be tested upon quantifying the actual fitting improvement.

At this point, it should be noted that we are arrived at two seemingly conflicting results. In fact we have explicitly stated that *both* the BCM *and* the ECM give a very good prediction for the binary topology of the ITN; however, the equations specifying the connection probability p_{ij} in the two models are significantly different. This finding makes us expect that, despite the different mathematical expressions, the numerical values of the probability coefficients in these two models do not differ too much: the comparison between the two probability matrices $\{p_{ij}^{\mathrm{BCM}}\}$ and $\{p_{ij}^{\mathrm{ECM}}\}$ shows that they are, indeed, very similar (Almog et al., 2015). This in turn, enables us to greatly simplify the equations defining the ECM, by replacing the expression for the p_{ij} coefficients provided by the ECM with the ones provided by the BCM. If we denote the new probability coefficients with p_{ij}^{ts}, 'ts' standing for 'two-step' (the reason will be clear in a moment), equations (2) and (3) can be naturally rewritten as

$$\langle a_{ij}\rangle^{ts}(\mathbf{z}) \equiv p_{ij}^{ts}(\mathbf{z}) = \frac{z_i z_j}{1 + z_i z_j}, \tag{7}$$

$$\langle w_{ij}\rangle^{ts}(\mathbf{z}, \mathbf{y}) = \frac{p_{ij}^{ts}(\mathbf{z})}{1 - y_i y_j}, \tag{8}$$

where, now, the unknown vector $\mathbf{z}$, and therefore the p_{ij}^{ts} coefficients, can be determined by solving a system of equations formally analogue to the one defining the BCM, i.e., $k_i(\mathbf{W}^*) = \sum_{j \neq i}^{N} p_{ij}^{ts}(\mathbf{z}^*), \forall\, i$. In this simplified model the connection probabilities no longer depend on the strengths as in the original ECM, while the weights still do. In other words, we have decoupled the structural part of the system of equations defining the ECM from the remaining one, providing a simpler set of equations to solve. This, in turn, implies that we can specify the model via a 'two-step' procedure according to which

- we first solve the N equations determining the p_{ij}^{ts}, upon constraining the nodes degrees only
- then evaluate the remaining parameters determining $\langle w_{ij}\rangle^{ts}$ through the ECM.

For this reason, we denote the model as the 'two-step' model (TS hereafter).

The TS model inherits the functional form of the link-specific distribution of weights from the ECM:

$$q_{ij}^{ts}(w_{ij}) = \frac{(z_i z_j)^{a_{ij}} (y_i y_j)^{w_{ij} - a_{ij}} (1 - y_i y_j)^{a_{ij}}}{1 + z_i z_j}. \tag{9}$$

It is instructive to rewrite equation (9) as a product of two different factors, i.e., as $q_{ij}^{ts}(w_{ij}) = \left[\frac{(z_i z_j)^{a_{ij}}}{1+z_i z_j}\right] \cdot (y_i y_j)^{w_{ij} - a_{ij}} (1 - y_i y_j)^{a_{ij}}$ to better highlight the two random processes behind the formation of each link. As a first step, one implements a Bernoulli trial with probability p_{ij}^{ts} in order to determine whether a link connecting i and j is created or not. The second part of our algorithm can be interpreted as a drawing from a geometric distribution, with parameter $y_i y_j$: if a link (or, equivalently, a unitary weight) is indeed established, a second random process determines whether the weight of the same link is increased by another unit (with probability $y_i y_j$) or whether the process stops (with probability $1 - y_i y_j$). Iterating this procedure to determine the probability of obtaining larger weights leads precisely to equation (9). As a consistency check, one can explicitly calculate the expected weight $\langle w_{ij}\rangle^{ts}$ for the nodes pair i-j through the formula $\sum_{w=0}^{+\infty} w \cdot q_{ij}^{ts}(w)$, which correctly leads to equation (8).

In more economic terms, the analysis of the ITN clearly proves that a substantial difference exists between establishing a new trade relation and reinforcing an existing one (by rising the exchanged amount of goods of e.g., ‘one unit’ of trade). These two processes are described, respectively, by the coefficients p_{ij}^{ts} and $y_i y_j$. In order to understand which one is more probable, we can study the behaviour of the ratio

$$\frac{p_{ij}^{ts}}{y_i y_j} = f_{ij}(g_i, g_j | a, b, c) \tag{10}$$

for each pair of countries. In fact, whenever $p_{ij}^{ts}/(y_i y_j) > 1$ countries i and j would probably establish a new trade relation quite easily, however experiencing a certain resistance to reinforce it. On the other hand, whenever $p_{ij}^{ts}/(y_i y_j) < 1$ countries i and j would experience a certain resistance to start trading; however, in the case such a relation were established, it would represent a channel with relatively low ‘friction’, inducing the involved partners to strengthen it.

Before analysing the case $p_{ij}^{ts}/(y_i y_j) = 1$ let us rewrite it as $\frac{z_i z_j}{y_i y_j}(1 - y_i y_j) = 1$. The expression at the first member also appears in equation (9) which, in fact, can be restated in the following way: $q_{ij}^{ts}(w_{ij}) = \left[\frac{z_i z_j}{y_i y_j}(1 - y_i y_j)\right]^{a_{ij}} \frac{(y_i y_j)^{w_{ij}}}{1+z_i z_j}$. Imposing the first factor to be equal to 1 implies reducing equation (9) to $q_{ij}^{ts}(w_{ij}) = (y_i y_j)^{w_{ij}}(1 - y_i y_j)$, i.e., to the WCM probability distribution. This model does not discriminate between the first link and the subsequent ones, reducing *tout court* $q_{ij}^{ts}(w_{ij})$ to a simple geometric distribution: thus, the failure of the WCM in reproducing the observed properties of the ITN lies precisely in its incapability to give the right importance to the very first link, treating it as a simple unit of weight and not as the channel making the trade exchanges possible.

5.3 A GDP-driven model of the ITN

Equations (7) and (8) provide the expressions into which we can input the vector of fitness parameters g_i, $\forall\, i$, according to the prescriptions of equations (5) and (6). As a result, we obtain the following formulas that mathematically characterise our GDP-driven specification of the TS model:

$$\langle a_{ij}\rangle^{ts}(a) \equiv p_{ij}^{ts}(a) = \frac{a \cdot g_i g_j}{1 + a \cdot g_i g_j}, \tag{11}$$

$$\langle w_{ij}\rangle^{ts}(a, b, c) = p_{ij}^{ts} \frac{(1 + b \cdot g_i^c)(1 + b \cdot g_j^c)}{(1 + b \cdot g_i^c + b \cdot g_j^c)}. \tag{12}$$

Equations (12) can be used to reverse the approach used so far: rather than determining the $2N$ free parameters either of the ECM ($\mathbf{x}$ and $\mathbf{y}$) or of the TS model ($\mathbf{z}$ and $\mathbf{y}$), upon constraining degrees and strengths to their observed values, we can now use the knowledge of the GDP of all countries to obtain a model that only depends on the three parameters a, b, c. Since the model consists of two subsequent steps, we can first assign a value to the parameter a and, only once a is set, fit the remaining parameters b and c.

Parameter a can be determined quite easily. In fact, following (Garlaschelli et al., 2004; Garlaschelli and Loffredo, 2008), the value of a can be chosen as the one ensuring that the density of connections is reproduced, i.e.,

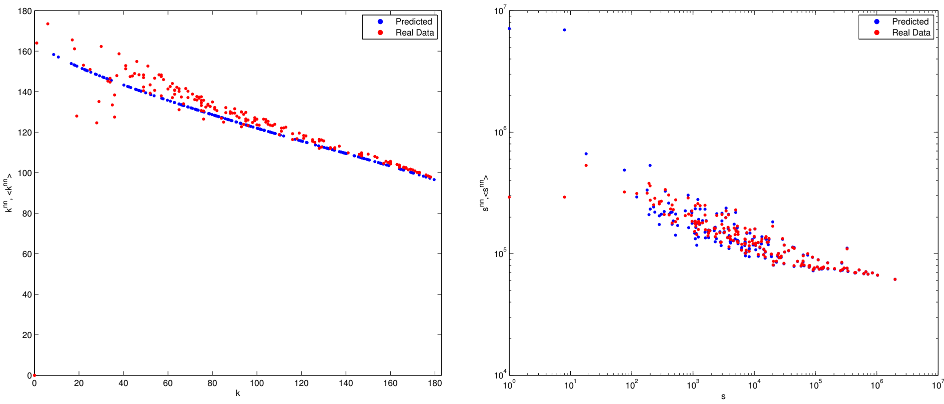
$$L = \sum_{i}^{N} \sum_{j \neq i}^{N} \frac{a \cdot g_i g_j}{1 + a \cdot g_i g_j}; \tag{13}$$

such a prescription overcomes the limitation of econometric models (as the gravity model) in failing to predict the right density of connections, allowing us to fix it from the very beginning. Notice that satisfying equation (13) is equivalent to maximising the likelihood function of the fitness model, as shown in (Garlaschelli et al., 2005).

Fixing the values of b and c is slightly more complicated. In fact, we could imagine to impose a similar condition, as constraining the total weight W of the network. However, since the TS model uses approximate expressions, rather than those of the ECM, maximising the likelihood function in the second step of the model no longer coincides with the desired condition $\langle W \rangle = W$. Similarly, extracting the parameters from the fit shown in Figure 3 does not preserve the total weight of the network. However, in absence of any *a priori* preference, we chose the latter procedure, due to its relative numerical simplicity with respect to the former one.

In Figure 5, we show a comparison between the higher-order observed properties of the ITN in 2000 and their expected counterparts predicted by the GDP-driven TS model (the mathematical expressions of these properties are provided in Appendix). As a baseline comparison, we also show the predictions of the GDP-driven WCM model with continuous weights proposed in (Fronczak et al., 2012), which coincides with a simplified version of the gravity model.

Figure 5 Comparison between the observed properties (red points), the corresponding ensemble averages of the GDP-driven 'two-step' model (blue points) of the aggregated ITN in the 2000 snapshot. Left panel: average nearest neighbours degree k_i^{nn} vs. degree k_i. Right panel: average nearest neighbours strength s_i^{nn} vs. strength s_i (see online version for colours)



Naturally, as expected, the predictions in Figure 4 are more noisy than the ECM predicted values (the TS model makes use of three parameters only, while the ECM is defined by

$2N$ parameters): this is due to the fact that equations (5) (and the corresponding BCM equation) and equation (6) describe fitting curves rather than exact relationships. However, as a general comment, the GDP-driven TS model reproduces the empirical trends very well; most importantly, our model performs significantly better than the GDP-driven WCM in replicating both binary and weighted properties. Again, the drawback of these models lies in the fact that they predict a fully connected topology and a relatively homogeneous network. More specifically, the plot of the average nearest neighbour strengths, s^{nn}, predicted by our model is slightly shifted with respect to the observed points. This effect is due to the fact that, as we mentioned, the total weight of the network W (hence the average trend of s^{nn}) is only approximately reproduced by our model, as a consequence of the simplification leading from the ECM to the TS model. Our findings are also robust over the entire time span of our dataset. We can therefore conclude that the ECM model, as well as its simplified TS variant, can be successfully turned into a fully GDP-driven model that simultaneously reproduces both the topology and the weighted structure of the ITN.

6 Conclusion

In this paper we have demonstrated the capabilities of a novel GDP-driven model which successfully reproduces both the binary and weighted properties of the ITN. The model uses the GDP of world countries as a sort of macroeconomic fitness that in turns determine the probabilities for the formation of the network links. The use of the GDP as a macroeconomic fitness parameter is motivated in the first section, where we show the extent to which this quantity is entangled with the first order, country-specific, properties of the network. The model also represent an improvement in the reconstruction ability of a network, by extending it to both the binary ant the weighted representations.

Furthermore, the model dependence on the parameters a, b, c suggests an interpretation of the latter as traditional macroeconomic elasticities. As an example, the stationary character of the rescaled GDP would allow for a temporal analysis of bilateral trade exchanges simply in terms of the functions $a(t)$, $b(t)$ and $c(t)$: for example, an increasing trend of $a(t)$ would suggest an increasing tendency of countries to establish new trade relationships. Thus, while our model shares the same mathematical features of traditional macroeconomic models (which are defined in terms of data-driven parameters, whose change reflects a deeper change in the system macroeconomic organisation), it overcomes their traditional limitations, restoring the importance of purely structural quantities beside the traditionally inspected ones.

The misunderstanding concerning the preminent role of weighted properties has its roots in the observation that the presence of a non-zero trade relation, $w_{ij} > 0$, implies the existence of the relation itself (i.e., $a_{ij} > 0$ *if and only if* $w_{ij} > 0$). When applied to a given country i, this line of reasoning leads one to conclude that knowing the trade exchanges with all its partners (i.e., the strength of country i) implies knowing also its degree. However, this is valid only from a merely *empirical* point of view, at which level the dependence of degrees on strengths, $k_i(s_i)$, is well known. In fact, strengths fail when requested to *predict* a given binary structure: as highlighted by other works (Mastrandrea et al., 2014), $\langle k_i \rangle(s_i) \simeq N - 1$ dramatically differs from $k_i(s_i)$.

The success of the TS model has an important interpretation. We recall that the effect of the approximation leading from the ECM to the TS model lies in the fact that the connection probability p_{ij}^{ts} can be estimated separately from the weights $\langle w_{ij} \rangle^{ts}$, using either the knowledge of the degree sequence – if equation (7) is used – or that of the GDPs and total number of links – if equation (12) is used. By contrast, the estimation of the expected weights cannot be carried out separately, as it requires the evaluation of the connection probability p_{ij}^{ts}. This asymmetry implies that the topology of the ITN can be successfully inferred without any information about the weighted properties, while the weighted structure cannot be inferred without any topological information.

This effect is thus the origin of the limitation of 'purely weighted' models, such as the gravity model, which focus on trade volumes while disregarding the connectivity of countries. The TS model provides a mathematical explanation for this otherwise puzzling effect observed in the ITN.

Acknowledgements

AA and DG acknowledge support from the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands) with funds from beneficiaries of Duyfken Trading Knowledge BV, Amsterdam, the Netherlands. This work was also supported by the EU projects MULTIPLEX (FP7-ICT, grant n. 317532), SIMPOL (FP7-ICT, grant n. 610704) and DOLFINS (H2020-EU.1.2.2., grant n. 640772).

References

Almog, A., Squartini, T. and Garlaschelli, D. (2015) 'A GDP-driven model for the binary and weighted structure of the International Trade Network', *New J. Phys.*, Vol. 17, p.013009.

Bargigli, L. and Gallegati, M. (2011) 'Random digraphs with given expected degree sequences: a model for economic networks', *J. Econ. Behav. and Organ.*, Vol. 78, p.396.

Barigozzi, M., Fagiolo, G. and Garlaschelli, D. (2010) 'Multinetwork of international trade: a commodity-specific analysis', *Phys. Rev. E*, Vol. 81, p.046104.

Bhattacharya, K., Mukherjee, G., Saramaki, J., Kaski, K. and Manna, S.S. (2008) 'The International Trade Network: weighted network analysis and modeling', *J. Stat. Mech.*, p.P02002.

Caldarelli, G., Capocci, A., De Los Rios, P. and Muñoz, M.A. (2002) 'Scale-free networks from varying vertex intrinsic fitness', *Phys. Rev. Lett.*, Vol. 89, p.258702.

Caldarelli, G., Chessa, A., Pammolli, F., Gabrielli, A. and Puliga, M. (2013) 'Reconstructing a credit network', *Nat. Phys.*, Vol. 9, p.125.

Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. and Pietronero, L. (2013) 'Measuring the intangibles: a metrics for the economic complexity of Countries and products', *PLoS ONE*, Vol. 8, p.0070726.

De Benedictis, L. and Tajoli, L. (2011) 'The world trade network', *The World Economy*, Vol. 34, p.1417.

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

Fagiolo, G., Reyes, J. and Schiavo, S. (2009) 'World-trade web: topological properties, dynamics, and evolution', *Phys. Rev. E*, Vol. 79, p.036115.

Fagiolo, G. (2010) 'The international-trade network: gravity equations and topological properties', *J. Econ. Interact. Coord.*, Vol. 5, No. 5, pp.1–25.

Fagiolo, G., Squartini, T. and Garlaschelli, D. (2013) 'Null models of economic networks: the case of the world trade web', *J. Econ. Interac. Coord.*, Vol. 8, No. 1, p.75.

Fronczak, A., Fronczak, P. and Holyst, J.A. (2012) 'Statistical mechanics of the international trade network', *Phys. Rev. E*, Vol. 85, p.056113.

Garlaschelli, D. and Loffredo, M.I. (2004) 'Fitness-dependent topological properties of the world trade web', *Phys. Rev. Lett.*, Vol. 355, p.188701.

Garlaschelli, D. and Loffredo, M.I. (2005) 'Structure and evolution of the world trade network', *Physica A*, Vol. 10, p.138.

Garlaschelli, D., Di Matteo, T., Aste, T., Caldarelli, G. and Loffredo, M. (2007) 'Interplay between topology and dynamics in the world trade web', *Eur. Phys. J. B*, Vol. 57, p.1434.

Garlaschelli, D. and Loffredo, M.I. (2008) 'Maximum likelihood: Extracting unbiased information from complex networks', *Phys. Rev. E*, Vol. 78, No. 1, p.015101.

Garlaschelli, D. and Loffredo, M.I. (2009) 'Generalized Bose-Fermi statistics and structural correlations in weighted Networks', *Phys. Rev. Lett.*, Vol. 102, p.038701.

Gleditsch, K.S. (2002) 'Expanded trade and GDP data', *Journal of Conflict Resolution*, Vol. 46, p.712.

Kali, R. and Reyes, J. (2007) 'The architecture of globalization: a network approach to international economic integration', *J. Int. Bus. Stud.*, Vol. 38, p.595.

Kali, R. and Reyes, J. (2010) 'Financial contagion on the international trade network', *Economic Inquiry*, Vol. 48, p.1072.

Mastrandrea, R., Squartini, T., Fagiolo, G. and Garlaschelli, D. (2014) 'Enhanced reconstruction of weighted networks from strengths and degrees', *New J. Phys.*, Vol. 16, p.043022.

Mastrandrea, R., Squartini, T., Fagiolo, G. and Garlaschelli, D. (2014) 'Reconstructing the world trade multiplex: the role of intensive and extensive biases', *Phys. Rev. E*, Vol. 90, p.062804.

Musmeci, N., Battiston, S., Caldarelli, G., Puliga, M. and Gabrielli, A. (2013) 'Bootstrapping topological properties and systemic risk of complex networks using the fitness model', *J. Stat. Mech.*, Vol. 151, p.720.

Saracco, F., Di Clemente, R., Gabrielli, A. and Squartini, T. (2015) 'Detecting early signs of the 2007-2008 crisis in the world trade', *Sci. Rep.*, Vol. 6, No. 30286.

Schiavo, S., Reyes, J. and Fagiolo, G. (2010) 'International trade and financial integration: a weighted network analysis', *Quantitative Finance*, Vol. 10, p.389.

Serrano, A. and Boguna, M. (2003) 'Topology of the world trade web', *Phys. Rev. Lett.*, Vol. 68, p.015101.

Serrano, A., Boguna, M. and Vespignani, A. (2007) 'Patterns of dominant flows in the world trade web', *J. Econ. Interact. Coord.*, Vol. 2, p.111.

Sinha, S., Chatterjee, A., Chakraborti, A. and Chakrabarti, B.K. (2010) *Econophysics: An Introduction*, Wiley-VCH, Weinheim.

Squartini, T. and Garlaschelli, D. (2011) 'Analytical maximum-likelihood method to detect patterns in real networks', *New J. Phys.*, Vol. 13, p.083001

Squartini, T., Fagiolo, G. and Garlaschelli, D. (2011) 'Randomizing world trade. I. A binary network analysis', *Phys. Rev.*, Vol. 84, p.046117.

Squartini, T., Fagiolo, G. and Garlaschelli, D. (2011) 'Randomizing world trade. II. A weighted network analysis', *Phys. Rev.*, Vol. 84, p.046118.

Squartini, T. and Garlaschelli, D. (2014) 'Jan Tinbergen's legacy for economic networks: from the gravity model to quantum statistics', *Econophysics of Agent-Based Models*, Springer, pp.161–186.

Squartini, T., Mastrandrea, R. and Garlaschelli, D. (2015) 'Unbiased sampling of network ensembles', *New J. Phys.*, Vol. 17, p.023052.

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

Wells, S. (2004) *Financial Interlinkages in the United Kingdom's Interbank Market and the Risk of Contagion*, Bank of England Working Paper, No. 230/2004.

Note

[1]In the BCM, the probability that any two nodes i and j are connected has the expression $p_{ij}^{BCM} = \frac{z_i z_j}{1+z_i z_j}$. The unknown parameters $\mathbf{z}$ can be numerically evaluated upon solving the system of N equations $k_i(\mathbf{W}^*) = \sum_{j\neq i}^{N} p_{ij}^{BCM}(\mathbf{z}^*), \forall\, i$.

Appendix: higher-order properties of the undirected representation of the ITN

Table A1 gives a summarised description of the binary and weighted network quantities analysed in this paper. Specifically, it both shows their analytical definition and the corresponding expected value under the ECM and the GDP-driven TS model.

Table A1 Mathematical expressions for the empirical and expected properties of the undirected representation of the ITN

Empirical properties	*Expected properties under the ECM*	*Expected properties under the TS*
a_{ij}	$\langle a_{ij}\rangle = p_{ij} = \frac{x_i x_j y_i y_j}{1-y_i y_j + x_i x_j y_i y_j}$	$\langle a_{ij}\rangle = p_{ij}^{ts} = \frac{z_i z_j}{1+z_i z_j}$
$k_i = \sum_{j\neq i} a_{ij}$	$\langle k_i\rangle = \sum_{j\neq i} p_{ij} = k_i$	$\langle k_i\rangle^{ts} = \sum_{j\neq i} p_{ij}^{ts}$
$k_i^{nn} = \frac{\sum_{j\neq i} a_{ij} k_j}{k_i}$	$\langle k_i^{nn}\rangle = \frac{\sum_{j\neq i} p_{ij} k_j}{k_i}$	$\langle k_i^{nn}\rangle^{ts} = \frac{\sum_{j\neq i} p_{ij}^{ts} \langle k_j\rangle^{ts}}{\langle k_i\rangle^{ts}}$
$c_i = \frac{\sum_{j\neq i}\sum_{k\neq i,j} a_{ij} a_{jk} a_{ki}}{k_i(k_i-1)}$	$\langle c_i\rangle = \frac{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij} p_{jk} p_{ki}}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij} p_{ik}}$	$\langle c_i\rangle^{ts} = \frac{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^{ts} p_{jk}^{ts} p_{ki}^{ts}}{\sum_{j\neq i}\sum_{k\neq i,j} p_{ij}^{ts} p_{ik}^{ts}}$
w_{ij}	$\langle w_{ij}\rangle = \frac{p_{ij}}{1-y_i y_j}$	$\langle w_{ij}\rangle^{ts} = \frac{p_{ij}^{ts}}{1-y_i y_j}$
$s_i = \sum_{j\neq i} w_{ij}$	$\langle s_i\rangle = \sum_{j\neq i} \langle w_{ij}\rangle$	$\langle s_i\rangle^{ts} = \sum_{j\neq i} \langle w_{ij}\rangle^{ts}$
$s_i^{nn} = \frac{\sum_{j\neq i} a_{ij} s_j}{k_i}$	$\langle s_i^{nn}\rangle = \frac{\sum_{j\neq i} p_{ij} s_j}{k_i}$	$\langle s_i^{nn}\rangle^{ts} = \frac{\sum_{j\neq i} p_{ij}^{ts} \langle s_j\rangle^{ts}}{\langle k_i\rangle^{ts}}$

Let us recall that a weighted undirected network can be represented through a square matrix $\mathbf{W}$, where the specific entry w_{ij} represents the edge weight between country i and country j. The binary representation of the network, encoded into the matrix $\mathbf{A}$, is straightforwardly obtained upon defining $a_{ij} \equiv \Theta[w_{ij}]$.

The *degree* and the *strength* of a given node, respectively defined as $k_i(\mathbf{W}) = \sum_{j\neq i}^{N} a_{ij} = \sum_{j\neq i}^{N} \Theta[w_{ij}], \forall\, i$ and $s_i(\mathbf{W}) = \sum_{j\neq i}^{N} w_{ij}, \forall\, i$, are first-order properties, describing the neighbourhood of the node itself and, specifically, the number of its first neighbours (i.e., the other nodes sharing a direct connection with it) and its total volume.

Exploring the topological properties of more distant nodes (i.e., the neighbours of the neighbours) implies considering longer pathways starting from node i. The simpler second-order properties that can be defined are the *average nearest neighbours degree*, k_i^{nn}, i.e., the arithmetic mean of the degrees of the neighbours of node i and the *average nearest neighbours strength*, s_i^{nn}, i.e., the arithmetic mean of the strengths of the neighbours of node i. Once plotted vs. the corresponding node degree (strength), the k_i^{nn} (s_i^{nn}) provides information on the tendency of nodes degrees (strengths) to be either positively or negatively correlated. In economic terms, the k^{nn} quantifies the tendency of strongly connected countries to trade with strongly connected partners as well.

Another important feature of complex networks concerns the tendency of nodes to cluster together. It can be quantified through the clustering coefficient, c_i, which measures the percentage of closed triangles node i is part of. In economic terms, the clustering coefficient quantifies the tendency of countries to form small communities and, at a more general level, the hierarchical character of the ITN structure.

The measured properties of the real network need to be compared with the different models predictions. The expected values can be obtained by simply replacing a_{ij} with the probability coefficients $\langle a_{ij} \rangle$ predicted by the different models (e.g., $\langle a_{ij} \rangle = \frac{z_i z_j}{1+z_i z_j} = p_{ij}^{ts}$ for the TS, $\langle a_{ij} \rangle = \frac{x_i x_j y_i y_j}{1-y_i y_j+x_i x_j y_i y_j}$ for the ECM, etc.) and w_{ij} with $\langle w_{ij} \rangle$ (e.g., $\langle w_{ij} \rangle = \frac{p_{ij}^{ts}}{1-y_i y_j}$ for the TS, etc.). Whenever considering the GDP-driven TS model, the mathematical expressions for $\langle a_{ij} \rangle$ and $\langle w_{ij} \rangle$ are the ones illustrated by equations (9) and (12).