

A GDP-driven model for the binary and weighted structure of the International Trade Network

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 New J. Phys. 17 013009

(<http://iopscience.iop.org/1367-2630/17/1/013009>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 90.147.23.94

This content was downloaded on 05/11/2015 at 13:50

Please note that [terms and conditions apply](#).



OPEN ACCESS

RECEIVED

29 September 2014

ACCEPTED FOR PUBLICATION

24 November 2014

PUBLISHED

9 January 2015

Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).

Any further distribution of
this work must maintain
attribution to the author
(s) and the title of the
work, journal citation and
DOI.



PAPER

A GDP-driven model for the binary and weighted structure of the International Trade Network

Assaf Almog¹, Tiziano Squartini^{1,2} and Diego Garlaschelli¹¹ Instituut-Lorentz for Theoretical Physics, Leiden Institute of Physics, University of Leiden, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands² 'Sapienza' University of Rome, P.le Aldo Moro 5, 00185 Rome, ItalyE-mail: almog@lorentz.leidenuniv.nl**Keywords:** complex networks, econophysics, maximum entropy models

Abstract

Recent events such as the global financial crisis have renewed the interest in the topic of economic networks. One of the main channels of shock propagation among countries is the International Trade Network (ITN). Two important models for the ITN structure, the classical gravity model of trade (more popular among economists) and the fitness model (more popular among networks scientists), are both limited to the characterization of only one representation of the ITN. The gravity model satisfactorily predicts the volume of trade between connected countries, but cannot reproduce the missing links (i.e. the topology). On the other hand, the fitness model can successfully replicate the topology of the ITN, but cannot predict the volumes. This paper tries to make an important step forward in the unification of those two frameworks, by proposing a new gross domestic product (GDP) driven model which can simultaneously reproduce the binary and the weighted properties of the ITN. Specifically, we adopt a maximum-entropy approach where both the degree and the strength of each node are preserved. We then identify strong nonlinear relationships between the GDP and the parameters of the model. This ultimately results in a weighted generalization of the fitness model of trade, where the GDP plays the role of a 'macroeconomic fitness' shaping the binary and the weighted structure of the ITN simultaneously. Our model mathematically explains an important asymmetry in the role of binary and weighted network properties, namely the fact that binary properties can be inferred without the knowledge of weighted ones, while the opposite is not true.

1. Introduction

After the 2008 financial crisis, it has become clear that a better understanding of the mechanisms and dynamics underlying the networked worldwide economy is vital [1]. Among the possible channels of interaction among countries, international trade plays a major role [2–4]. Combined together, the worldwide trade relations can be interpreted as the connections of a complex network, the International Trade Network (ITN) [5–21], whose understanding and modeling is one of the traditional goals of macroeconomics. The standard model of non-zero trade flows, inferring the volume of bilateral trade between any two countries from the knowledge of their gross domestic product (GDP) and mutual geographic distance (D), is the so-called 'gravity model' of trade [22–26]. In its simplest form, the gravity model predicts that the volume of trade between countries i and j is

$$F_{ij} = \alpha \frac{\text{GDP}_i^\beta \cdot \text{GDP}_j^\beta}{D_{ij}^\gamma}. \quad (1)$$

Note that the above expression, as well as the rest of this paper, focuses on the undirected version of the network for simplicity. In this representation, the trade from country i to country j and the trade from country j to country i are combined together. Given the highly symmetric structure of the ITN, this simplification retains all the basic network properties of the system [7, 15, 16, 27–29].

In its simplest form, the gravity model is fitted on the non-zero weights observed between all pairs of *connected* countries. This means that the model can predict the pair-specific volume of trade only *after* the presence of the trade relation itself has been established [25]. This intrinsic limitation is alarming, since almost half of the links in the ITN are missing [6, 15–17]. Although several improvements and generalizations of the standard gravity model have been proposed to overcome this problem (see [25, 26] for excellent reviews), so far none of them succeeded in reproducing the observed complex topology and the observed volumes simultaneously. Moreover, the various attempts have not been conceived under a unique theoretical framework and are therefore based on the combination of different mechanisms (e.g. one for establishing the presence of a trade relation, and one for establishing its intensity). In general, the challenge of successfully predicting, via only one mechanism, both trade probabilities and trade volumes remains an open problem.

Over the past years, the problem of replicating the observed structure of the ITN has been extensively approached using network models [6, 7, 9, 18, 30] and, more indirectly, maximum-entropy techniques to reconstruct networks from partial information [15–17, 31–34]. These studies have focused both on the purely binary architecture (defined solely by the existence of trade exchanges between world countries) [6, 7, 9, 15, 33] and on the weighted structure (when also the magnitude of these interactions is taken into account) [16–18, 30]. What clearly emerges is that both topological and weighted properties of the network are deeply connected with purely macroeconomic properties (in particular the GDP) governing bilateral trade volumes [2, 3, 5–7, 9–14, 18, 21]. However, it has also been clarified that, while the knowledge of the degree sequence (i.e. the number of trade partners for each country) allows to infer the entire binary structure of the network with great accuracy [15, 17], the knowledge of the strength sequence (i.e. the total volume of trade for each country) gives a very poor prediction of all network properties [16, 17]. Indeed, the network inferred only from the strength sequence has a trivial topology, being much denser (if integer link weights are assumed [16]) or even fully connected (if continuous weights are assumed [18]), and in any case much more homogeneous than the empirical one. This limitation leads back to the main drawback of the gravity model. Indeed, it has been shown that a simplified version of the gravity model (with $\beta = 1$ and $\gamma = 0$) can be recovered as a particular case of a maximum-entropy model with given strength sequence (and continuous link weights) [18].

Combined together, the high informativeness of the degree sequence for the binary representation of the ITN and the low informativeness of the strength sequence for its weighted representation contradict the naive expectation that, once aggregated at the country level, weighted structural properties (the strengths) are *per se* more informative than purely binary properties (the degrees). This empirical puzzle still calls for a theoretical explanation and has generated further interest around the challenge of finding a unique mechanism predicting link probabilities and link weights simultaneously. In this paper, we will propose a model that successfully implements such mechanism. The model can reproduce the observed properties of the ITN and finally highlights a clear mathematical explanation for the observed binary/weighted asymmetry.

Our approach builds on previous theoretical results. Recently, an improved reconstruction approach [35], based on an analytical maximum-likelihood estimation method [36], has been proposed in order to define more sophisticated maximum-entropy ensembles of weighted networks. This approach exploits previous mathematical results [37] characterizing a network ensemble where both the degree and the strength sequences are constrained. The graph probability is the so-called generalized Bose–Fermi distribution [37], and the resulting network model goes under the name of enhanced configuration model (ECM) [35]. When used to reconstruct the properties of several empirical networks, the ECM shows a significant improvement with respect to the case where either only the degree sequence (binary configuration model, BCM for short) or only the strength sequence (weighted configuration model, WCM for short) is constrained. One therefore expects that combining the knowledge of strengths and degrees is precisely the ingredient required in order to successfully reproduce the ITN from purely local information. Indeed, a more recent study has shown that, when applied to international trade data (both aggregated and commodity-specific), the method successfully reproduces the key properties of the ITN, across different years and for different levels of aggregation (i.e. for different commodity-specific layers) [38].

However, in itself the ECM is a network reconstruction method, rather than a genuine model of network formation. To turn it into a proper network model for the ITN structure, it would be necessary to find a macroeconomic interpretation for the underlying variables involved in the method. This operation would correspond to what has already been separately performed at a purely binary level (by identifying a strong relationship between the GDP and the variable controlling the degree of a country in the BCM [6, 39, 40]) and at a purely weighted level (by finding a relationship between the GDP and the variable controlling the strength of a country in the WCM [18], in the same spirit of the gravity model). Generalizing the above results to the combination of strengths and degrees is not obvious, given the different mathematical expressions characterizing the ECM, the BCM and the WCM.

In this paper we show that, indeed, the variables of the ECM are all strongly correlated with the GDP. This result gives a macroeconomic interpretation of the parameters' values satisfying the ITN constraints. Reversing

the perspective, this result enables us to introduce the first GDP-driven model that successfully reproduces the binary and the weighted properties of the ITN simultaneously. Finally, we show that the ECM model can be replaced by a simpler, two-step (TS) model that reconciles the binary projection of the ECM model with the topology predicted by the BCM. These results represent a promising step forward in the formulation of a unified model for the structure of the ITN. It is the mathematical structure of the TS model that finally explains the puzzling asymmetry in the informativeness of weighted and binary properties. This result has a general applicability to the analysis of weighted networks, and is therefore not restricted to our study of the ITN.

2. Maximum-entropy approaches to the ITN

Since our results are a generalization of previous maximum-entropy approaches to the characterization of the ITN, in this section we first briefly review the main results of those approaches, while our new findings are presented in the next section. In so doing, we gradually introduce the mathematical building blocks of our analysis and illustrate our main motivations. Moreover, since previous studies have used different data sets, we also recalculate the quantities of interest on the same data set that we will use later for our own investigation. This allows us to align the results of previous approaches and properly compare them with our new findings.

2.1. Data

We use yearly bilateral data on exports and imports from the United Nations Commodity Trade Database (UN COMTRADE) [41] from year 1992 to 2004. The sample refers to 13 years, 1992–2004, represented in current US dollars, and disaggregated over 97 commodity classes. In this paper we analyze the aggregated level, which results in 13 yearly temporal snapshots of undirected total trade flows. Our network consists of $N = 173$ countries, present in the data throughout the considered temporal interval.

This data set was the subject of many studies exploring both purely the binary representation, and its full weighted representation [15, 16, 35, 38]. Another data set which is widely used to represent the ITN network is the trade data collected by Gleditsch [42]. The data contain the detailed list of bilateral import and export volumes, for each country in the period 1950–2000.

2.2. Binary structure

If one focuses solely on the binary undirected projection of the ITN, then the BCM represents a very successful maximum-entropy model. In the BCM, the local knowledge of the number of trade partners of each country, i.e. the degree sequence, is specified. It has been shown that higher-order properties of the ITN can be simply traced back to the knowledge of the degree sequence [15]. This result adds considerable information to the standard results of traditional macroeconomic analyzes of international trade. In particular, it suggests that the degree sequence, which is a purely topological property, needs to be considered as an important target quantity that international trade models, in contrast with the mainstream approaches in economics, should aim at reproducing [15, 17].

Let us first represent the observed structure of the ITN as a weighted undirected network specified by the square matrix \mathbf{W}^* , where the specific entry w_{ij}^* represents the weight of the link between country i and country j .

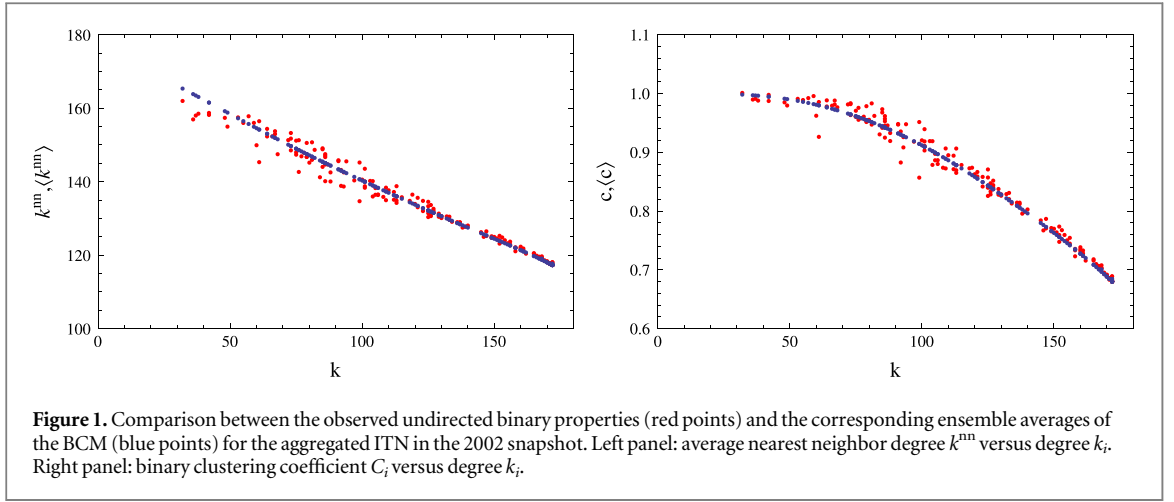
Then, let us represent the binary projection of the network in terms of the binary adjacency matrix \mathbf{A}^* , with entries defined as $a_{ij}^* = \Theta[w_{ij}^*]$, $\forall i, j$, where Θ is a Heaviside step function. A maximum-entropy ensemble of networks is a collection of graphs where each graph is assigned a probability of occurrence determined by the choice of some constraints. The BCM is a maximum-entropy ensemble of binary graphs, each denoted by a generic matrix \mathbf{A} , where the chosen constraint is the degree sequence. In the canonical formalism [36], the latter can be constrained by writing the following Hamiltonian:

$$H(\mathbf{A}) = \sum_{i=1}^N \theta_i k_i(\mathbf{A}), \quad (2)$$

where the degree sequence is defined as $k_i(\mathbf{A}) = \sum_{j \neq i}^N a_{ij} = \sum_{j \neq i}^N \Theta[w_{ij}^*]$, $\forall i$, and θ_i are the free parameters (Lagrange multipliers) [36]. As a result of the constrained maximization of the entropy [36], the probability of a given configuration \mathbf{A} can be written as

$$P(\mathbf{A}) = \frac{e^{-H(\mathbf{A})}}{\sum_{\mathbf{A}'} e^{-H(\mathbf{A}')}} = \prod_{i < j} \left[\frac{z_i z_j}{1 + z_i z_j} \right]^{a_{ij}} \left[\frac{1}{1 + z_i z_j} \right]^{1-a_{ij}}, \quad (3)$$

where $z_i \equiv e^{-\theta_i}$ and $p_{ij} \equiv \frac{z_i z_j}{1 + z_i z_j}$. The latter represents the probability of forming a link between nodes i and j , which is also the expected value



$$\langle a_{ij} \rangle = \frac{z_i z_j}{1 + z_i z_j} = p_{ij}. \quad (4)$$

According to the maximum-likelihood method proposed in [36], the vector of unknowns \vec{z} can be numerically found by solving the system of N coupled equations

$$\langle k_i \rangle = \sum_{j \neq i}^N p_{ij} = k_i(A^*) \quad \forall i, \quad (5)$$

where the expected value of each degree k_i is matched to the observed value $k_i(A^*)$ in the real network A^* . The (unique) solution will be indicated as \vec{z}^* . When inserted back into equation (4), this solution allows us to analytically describe the binary ensemble matching the observed constraints. Being the result of the maximization of the entropy, this ensemble represents the least biased estimate of the network structure, based only on the knowledge of the empirical degree sequence.

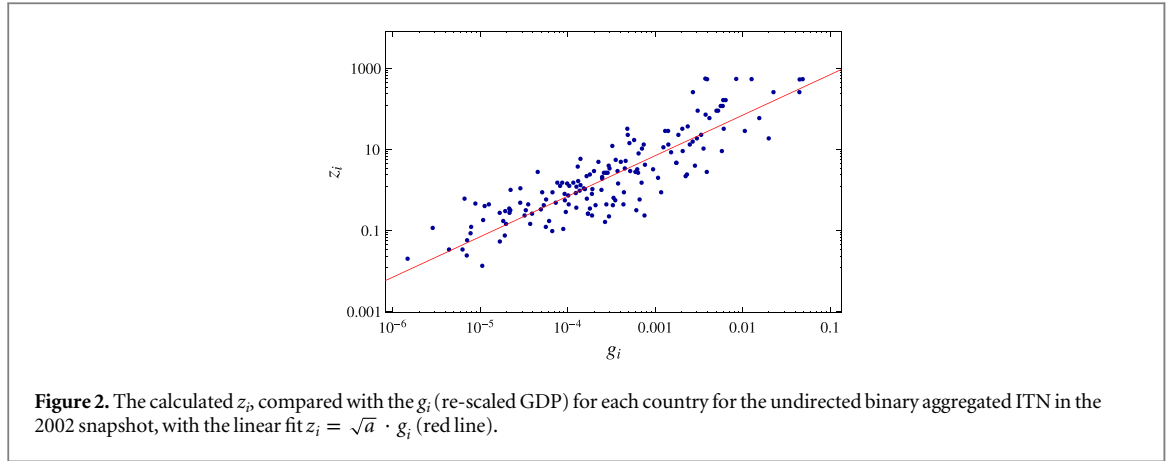
In figure 1 we plot some higher-order topological properties of the ITN as a function of the degree of nodes, for the 2002 snapshot. These properties are the so-called average nearest neighbor degree and the clustering coefficient. For both quantities, we plot the observed values (red points) and the corresponding expected values predicted by the BCM (blue points). The exact expressions for both empirical and expected quantities are provided in the appendix. We see that the expected values are in very close agreement with the observed properties. These results replicate recent findings [15, 17] based on the same UN COMTRADE data. They show that at a binary level, the degree correlations (disassortativity) and clustering structure of the ITN are excellently reproduced by the BCM. As we also confirmed in the present analysis, these results were found to be very robust, as they hold true over time and for various resolutions (i.e., for different levels of aggregation of traded commodities) [15, 17].

2.2.1. Relation with the fitness model

It should be noted that equation (4) can be thought of as a particular case of the so-called *Fitness Model* [43], which is a popular model of binary networks where the connection probability p_{ij} is assumed to be a function of the values of some ‘fitness’ characterizing each vertex. Indeed, the variables \vec{z}^* can be treated as fitness parameters [6, 39] which control the probability of forming a link. A very interesting correlation between a fitness parameter of a country (assigned by the model) and the GDP of the same country has been found [39]. This relation is replicated here in figure 2, where the rescaled GDP of each country ($g_i \equiv \frac{\text{GDP}_i}{\sum_i \text{GDP}_i}$) is compared to the value of the fitness parameter z_i^* obtained by solving equation (5). The red line is a linear fit of the type

$$z_i = \sqrt{a} \cdot g_i. \quad (6)$$

This leads to a more economic interpretation where the fitness parameters can be replaced (up to a proportionality constant) with the GDP of countries, and used to reproduce the properties of the network. This procedure, first adopted in [6], can give predictions for the network based only on macroeconomic properties of countries, and reveals the importance of the GDP to the binary structure of the ITN. Importantly, this observation was the first empirical evidence in favor of the fitness model as a powerful network model [43]. Likewise, other studies have shown that the observed topological properties turn out to be important in explaining macroeconomics dynamics [2, 3].



2.3. Weighted structure

Despite the importance of the topology, the latter is only the backbone over which goods are traded, and the knowledge of the volume of such trade is extremely important. To be able to give predictions about the weight of connections, one needs to switch from an ensemble of binary graphs to one of weighted graphs.

The simplest weighted counterpart of the BCM is the WCM, which is a maximum-entropy ensemble of weighted networks where the constraint is the strength sequence, i.e. the total trade of each country in the case of the ITN. Recent studies have shown that the higher-order binary quantities predicted by the WCM, as well as the corresponding weighted quantities, are very different from the observed counterparts [16, 17]. More specifically, the main limitation of the model is that of predicting a largely homogeneous and very dense (sometimes fully connected) topology. Roughly speaking, the model excessively ‘dilutes’ the total trade of each country by distributing it to almost all other countries. This failure in correctly replicating the purely topological projection of the real network is the root of the bad agreement between expected and observed higher-order properties.

2.3.1. Relation with the Gravity Model

Just like the BCM has been related to the Fitness Model [6], a variant of the WCM has been related to the Gravity Model [18]. The variant is actually a continuous version of the WCM, where the strength sequence is constrained and the weights are real numbers instead of integers. When applied to the ITN, the model gives the following expectation for the weight of the links:

$$\langle w_{ij} \rangle = T \cdot g_i g_j \quad \forall i, j, \quad (7)$$

where T is the total strength in the network, and g_i is the re-scaled GDP as before [18]. In essence, the above expression identifies again a relationship between the GDP and the hidden variable (analogous to the fitness in the binary case) specifying the strength of a node.

Equation (7) coincides with equation (1) where $\beta = 1$ and $\gamma = 0$. The model therefore corresponds to a particularly simple version of the Gravity Model. Indeed, the model reproduces reasonably well the observed non-zero weights of the ITN [18]. However, just like the Gravity Model, the model predicts a complete graph where $a_{ij} = 1 \quad \forall i, j$, and dramatically fails in reproducing the binary architecture of the network. This can be easily shown by realizing that the continuous nature of edge weights, which can take non-negative real values in the model, implies that there is a zero probability of generating zero weights (i.e. missing links). We will show the prediction of this model in comparison with our results later on in the paper.

3. A GDP-driven model of the ITN

Motivated by the challenge to satisfactorily model both the topology and the weights of the ITN, the ECM has been recently proposed as an improved model of this network [38]. The ECM focuses on weighted networks, and can enforce the degree and strength sequence simultaneously [35]. It builds on the so-called generalized Bose–Fermi distribution that was first introduced as a null model of networks with coupled binary and weighted constraints [37].

In the ECM, the degree and strength sequence can be constrained by writing the following Hamiltonian:

$$H(\mathbf{W}) = \sum_{i=1}^N [\alpha_i k_i(\mathbf{W}) + \beta_i s_i(\mathbf{W})], \quad (8)$$

where the strength sequence is defined as $s_i(\mathbf{W}) = \sum_{j \neq i}^N w_{ij}$, $\forall i$ and the degree sequence as

$k_i(\mathbf{W}) = \sum_{j \neq i}^N a_{ij} = \sum_{j \neq i}^N \Theta[w_{ij}]$, $\forall i$. As a result, the probability of a given configuration \mathbf{W} can be written as

$$P(\mathbf{W}) = \frac{e^{-H(\mathbf{W})}}{\sum_{\mathbf{W}'} e^{-H(\mathbf{W}')}} = \prod_{i < j} \frac{(x_i x_j)^{a_{ij}} (y_i y_j)^{w_{ij}} (1 - y_i y_j)}{1 - y_i y_j + x_i x_j y_i y_j} \quad (9)$$

with $x_i \equiv e^{-\alpha_i}$ and $y_i \equiv e^{-\beta_i}$. The ECM gives the following predictions about the probability of a link ($\langle a_{ij} \rangle$) and the expected weight of the link ($\langle w_{ij} \rangle$):

$$\langle a_{ij} \rangle = \frac{x_i x_j y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j} = p_{ij} \quad (10)$$

$$\langle w_{ij} \rangle = \frac{x_i x_j y_i y_j}{(1 - y_i y_j + x_i x_j y_i y_j)(1 - y_i y_j)} = \frac{p_{ij}}{1 - y_i y_j}. \quad (11)$$

According to the maximum-likelihood method proposed in [35], the vectors of unknowns \vec{x} and \vec{y} can be numerically found by solving the system of $2N$ coupled equations

$$\langle k_i(\mathbf{W}) \rangle = \sum_{j \neq i}^N p_{ij} = k_i(\mathbf{W}^*) \quad \forall i \quad (12)$$

$$\langle s_i(\mathbf{W}) \rangle = \sum_{j \neq i}^N \langle w_{ij} \rangle = s_i(\mathbf{W}^*) \quad \forall i \quad (13)$$

and will be indicated as \vec{x}^* and \vec{y}^* . These unknown parameters can be treated as fitness parameters which control the probability of forming a link and the expected weight of that link simultaneously.

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 [35]. 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 (temporal snapshots) [38]. Indeed, in figure 3 we show the higher-order binary quantities (average nearest neighbor degree and clustering coefficient) as well as their weighted ones (average nearest neighbor strength and weighted clustering coefficient) for the 2002 snapshot of the ITN. We compare the observed values (red points) and the corresponding quantities predicted by the ECM (blue points). The mathematical expressions for all these quantities are provided in the appendix. We find a very good agreement between data and model, confirming the recent results in [38] for the data set we are using here. We also confirmed that these results are robust for several temporal snapshots [38].

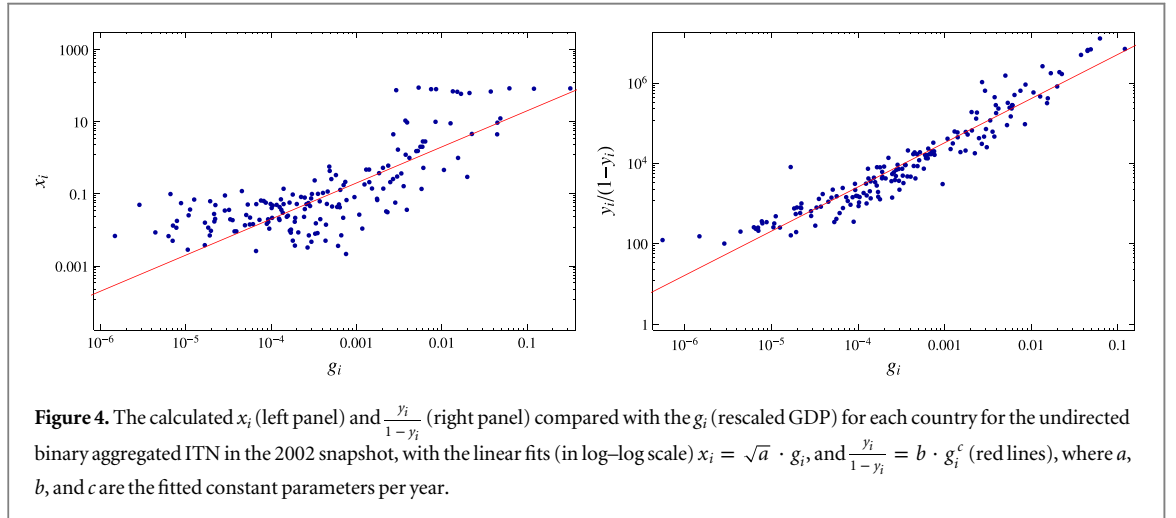
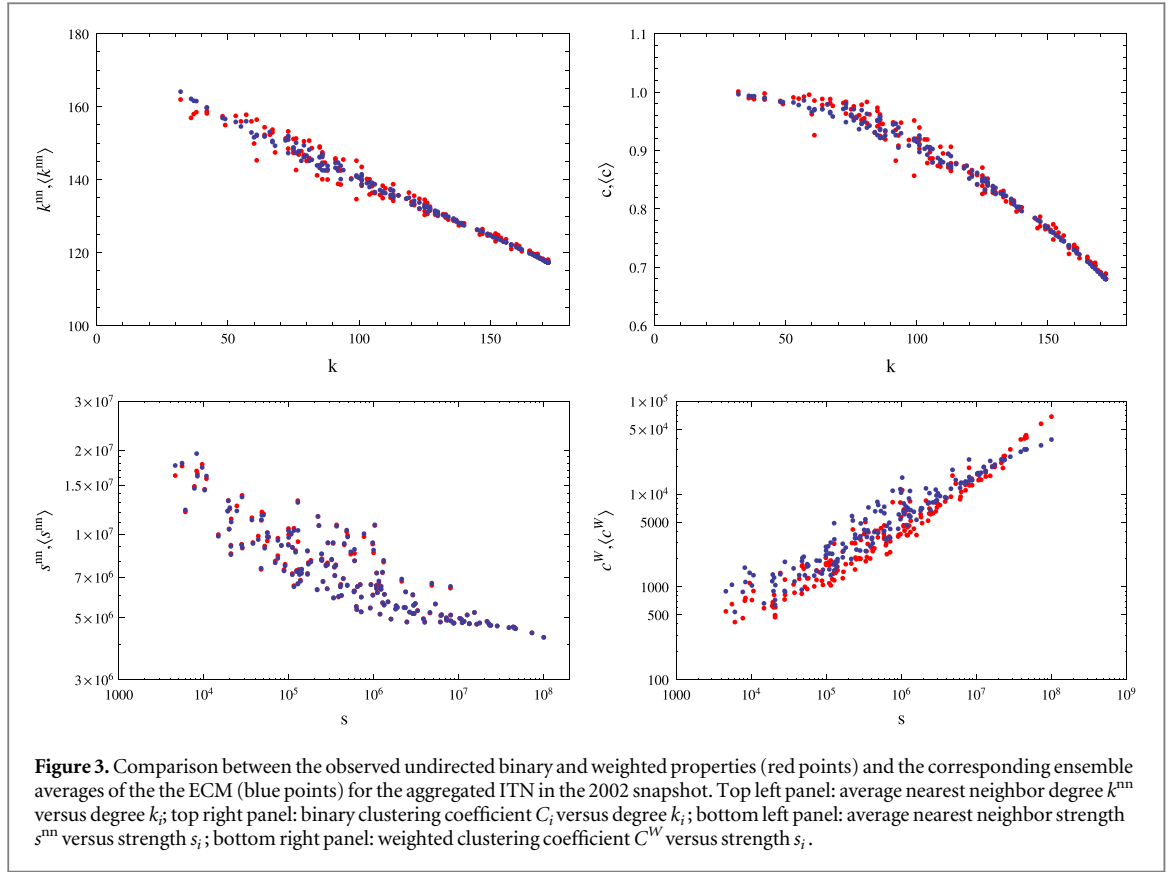
3.1. A weighted fitness model of trade

Considering the promising results of the ECM, we now make a step forward and check whether the hidden variables x_i and y_i , which effectively reproduce the observed ITN, can be thought of as ‘fitness’ parameters having a clear economic interpretation. This amounts to checking whether the relation shown previously in figure 2 for the purely binary case can be generalized in order to find a macroeconomic interpretation to the abstract fitness parameters in the general weighted case as well.

In figure 4 we show the relationship between the two parameters x_i and y_i and the rescaled GDP (g_i) for each country of the ITN in the 2002 snapshot. We find strong correlations between these quantities. The fitness parameter x_i turns out to be in a roughly linear relation with the rescaled GDP g_i , fitted by the curve

$$x_i = \sqrt{a} \cdot g_i, \quad (14)$$

where \sqrt{a} is the fitted constant, and $g_i = \frac{\text{GDP}_i}{\sum_i \text{GDP}_i}$ (all the GDPs are relative to that specific year). It should be noted that this relation is similar to that found between z_i and g_i in the BCM and shown previously in figure 2, but less accurate. This observation will be useful later. 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 [35, 37]), the relation between y_i and g_i is necessarily highly nonlinear. A simple functional form for such a relationship is given by



$$y_i = \frac{b \cdot g_i^c}{1 + b \cdot g_i^c}. \quad (15)$$

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

We checked that the above results hold systematically over time, for each snapshot of the ITN in our data set. This implies that, in a given year, we can insert equations (14) and (15) into equations (10) and (11) to obtain a GDP-driven model of the ITN structure for that year. Such a model highlights that the GDP has a crucial role in shaping both the binary and the weighted properties of the ITN. While this was already expected on the basis of the aforementioned results obtained using the BCM and the WCM (or the corresponding simplified gravity model) separately, finding the appropriate way to explicitly combine these results into a unified description of the ITN has remained impossible so far. Rather than exploring in more detail the predictions of the GDP-driven model in the form described above, we first make some considerations leading to a simplification of the model itself.

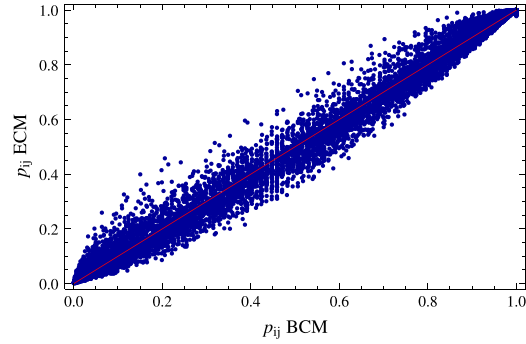


Figure 5. The probability of forming a link in the ECM p_{ij} ECM compared to the probability in the BCM p_{ij} BCM for the undirected binary aggregated ITN in the 2002 snapshot. The red line describes the identity line.

3.2. Reduced TS model

At this point, it should be noted that we arrived at two seemingly conflicting results. We showed that both the BCM and ECM give a very good prediction for the binary topology of the network. However, equations (4) and (10), which specify the connection probability p_{ij} in the two models, are significantly different. The comparable performance of the BCM and the ECM at a binary level (see figures 1 and 3) makes us expect that, when the specific values \bar{z}^* and \bar{x}^* are inserted into equations (4) and (10) respectively, the values of the connection probability become comparable in the two models, despite the different mathematical expressions.

In figure 5 we compare the two probabilities for the ITN in the 2002 snapshot. Note that each point refers to the probability of creating a link between a pair of countries, which results in $\frac{N(N-1)}{2}$ points. Indeed, we can see that the values are scattered along the identity line, confirming the expectation that the connection probability has similar value in the two models.

The above result allows us to make a remarkable simplification. In equations (10) and (11), we can replace the expression for p_{ij} provided by the ECM with that provided by the BCM in equation (4). To avoid confusion, we denote the new probability with p_{ij}^{ts} , where ts stands for ‘two-step’, for a reason that will be clear immediately. This results in the following equations for the expected network properties:

$$\langle a_{ij} \rangle^{\text{ts}} = \frac{z_i z_j}{1 + z_i z_j} \equiv p_{ij}^{\text{ts}}, \quad (16)$$

$$\langle w_{ij} \rangle^{\text{ts}} = \frac{p_{ij}^{\text{ts}}}{1 - y_i y_j}, \quad (17)$$

where the z_i ’s, and therefore the p_{ij}^{ts} ’s, depend only on the degrees through equation (5), while the y_i ’s and the $\langle w_{ij} \rangle^{\text{ts}}$ ’s depend on both strengths and degrees through equations (12) and (13).

In this simplified model the connection probability, which fully specifies the topology of the ensemble of networks, no longer depends on the strengths as in the ECM, while the weights still do. This implies that we can specify the model via a TS procedure where we first solve the N equations determining p_{ij}^{ts} via the degrees, and then find the remaining variables determining $\langle w_{ij} \rangle^{\text{ts}}$ through the ECM. For this reason, we denote the model as the TS model.

The probability of a configuration \mathbf{W} reads

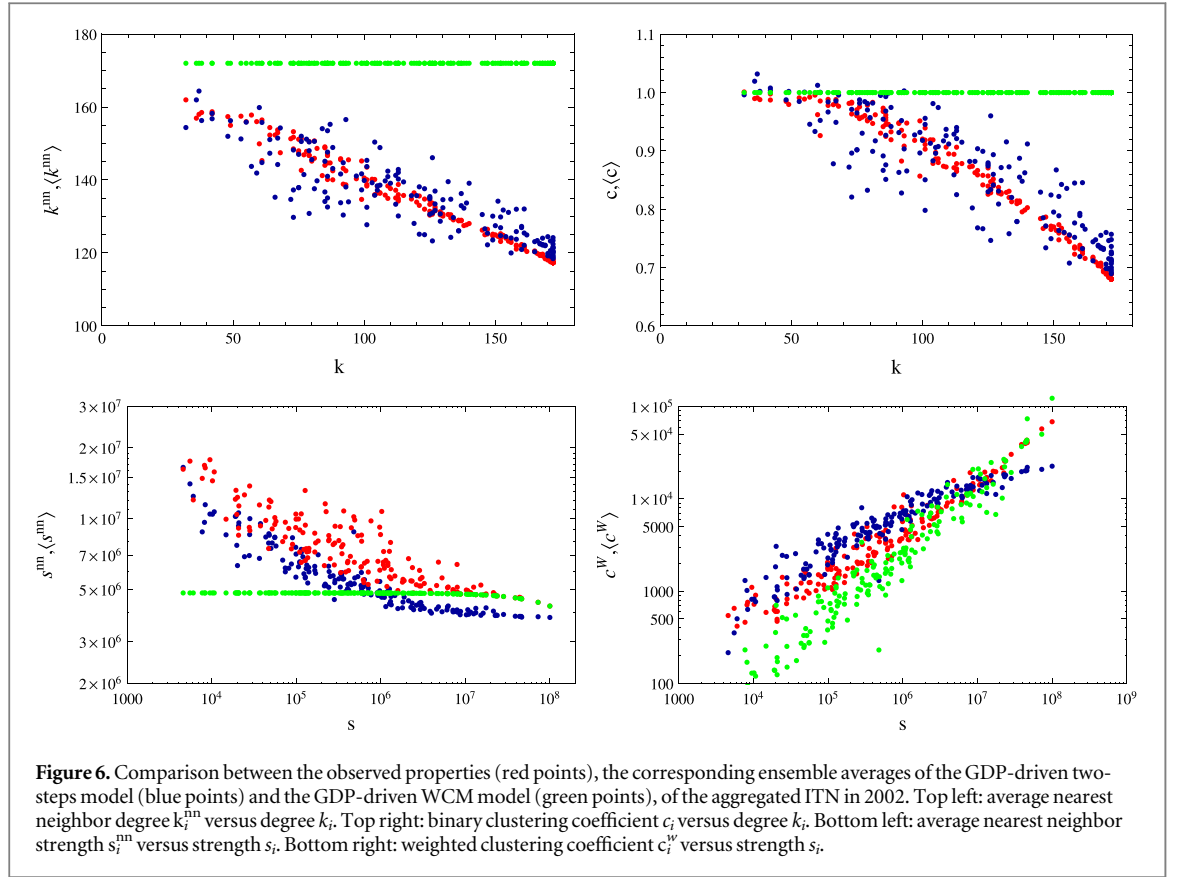
$$P^{\text{ts}}(\mathbf{W}) = \prod_{i < j} q_{ij}^{\text{ts}}(w_{ij}), \quad (18)$$

where

$$q_{ij}^{\text{ts}}(w_{ij}) = \frac{(z_i z_j)^{a_{ij}} (y_i y_j)^{w_{ij} - a_{ij}} (1 - y_i y_j)^{a_{ij}}}{1 + z_i z_j} \quad (19)$$

is the probability that a link of weight w_{ij} connects the nodes (countries) i and j . The above probability has the same general expression as in the original ECM [35], but here z_i comes from the estimation of the simpler BCM. It is instructive to rewrite (19) as

$$q_{ij}^{\text{ts}}(0) = \frac{1}{1 + z_i z_j} = (1 - p_{ij}^{\text{ts}}); \quad (20)$$



$$q_{ij}^{\text{ts}}(w) = p_{ij}^{\text{ts}} \left(\gamma_i \gamma_j \right)^{w-1} \left(1 - \gamma_i \gamma_j \right), \forall w > 0 \quad (21)$$

to highlight the random processes creating each link. As a first step, one determines whether a link is created or not with a probability p_{ij}^{ts} . If a link (of unit weight) is indeed established, a second attempt determines whether the weight of the same link is increased by another unit (with probability $\gamma_i \gamma_j$) or whether the process stops (with probability $1 - \gamma_i \gamma_j$). Iterating this procedure, the probability that an edge with weight w is established between nodes i and j is given precisely by $q_{ij}^{\text{ts}}(w)$ in equation (21). The expected weight $\langle w_{ij} \rangle^{\text{ts}}$ is then correctly retrieved as $\sum_{w=1}^{+\infty} w \cdot q_{ij}^{\text{ts}}(w)$.

Using the relations found in equations (6) and (15), we can input the g_i as the fitness parameters into equations (16) and (17) to get the following expressions that mathematically characterize our GDP-driven specification of the TS model:

$$\langle a_{ij} \rangle^{\text{ts}} = \frac{a g_i g_j}{1 + a g_i g_j} \equiv p_{ij}^{\text{ts}} \quad (22)$$

$$\langle w_{ij} \rangle^{\text{ts}} = p_{ij}^{\text{ts}} \frac{(1 + b g_i^c)(1 + b g_j^c)}{(1 + b g_i^c + b g_j^c)}. \quad (23)$$

The above equations can be used to reverse the approach used so far: rather than using the $2N$ free parameters of the ECM (\vec{x} and \vec{y}) or of the TS model (\vec{z} and \vec{y}) to fit the models on the observed values of the degrees and strengths, 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 . Assigning values to these parameters can be done using two techniques: maximization of the likelihood function and nonlinear curve fitting. Since the model is a TS one, we can first assign a value to the parameter a , and only in the second step (once a is set) we fit the parameters b and c .

We chose to fix a by maximizing the likelihood function [39], which results in constraining the expected number of links to the observed number ($\langle L \rangle = L$), as in [6]. Fixing the values of b and c is slightly more complicated. Since the model uses the approximated expressions of the TS model, rather than those of the original ECM model, maximizing the likelihood function in the second step no longer yields the desired condition $\langle T \rangle = T$, where T is the total strength in the network. Similarly, extracting the parameters from the fit

as shown in figure 4 does not maintain the total strength in the network. 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 6 we show a comparison between the higher-order observed properties of the ITN in 2002 and their expected counterparts predicted by the GDP-driven TS model. Again, the mathematical expressions of these properties are provided in the appendix. As a baseline comparison, we also show the predictions of the GDP-driven WCM model with continuous weights described by equation (7) [18], which coincides with a simplified version of the gravity model as we mentioned.

We see that the GDP-driven TS model reproduces the empirical trends very well. Of course, as expected, the predictions in figure 6 (which use only three free parameters) are more noisy than those in figure 3 (which use $2N$ free parameters). This is due to the fact that equations (6) and (15) describe fitting curves rather than exact relationships. Importantly, our model performs significantly better than the WCM/gravity model 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.

It should also be noted that the plot of average nearest neighbor strength (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 strength T (hence the average trend of the s^{nn}) is only approximately reproduced by our model, as a result of the simplification from the ECM to the TS model.

As for all the other results in this paper, we checked that our findings are robust over the entire time span of our data set. We 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 weights of the ITN.

The success of the TS model has an important interpretation. Looking back at equations (16) and (22), we recall that the effect of the TS approximation is the fact that the connection probability p_{ij}^{ts} can be estimated separately from the weights $\langle w_{ij} \rangle^{ts}$, using only the knowledge of the degree sequence if equation (16) is used, or the GDP and total number of links if equation (22) is used, while discarding that of the strengths. By contrast, the estimation of the expected weights $\langle w_{ij} \rangle^{ts}$ cannot be carried out separately, as it requires that the connection probability p_{ij}^{ts} appearing in equations (17) and (23) is estimated first. This asymmetry of the model means 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 topological information. The expressions defining the TS model provide a mathematical explanation for this otherwise puzzling effect that has already been documented in previous analyzes of the ITN [15–17, 38].

4. Conclusions

In this paper we have introduced a novel GDP-driven model which successfully reproduces both the binary and weighted properties of the ITN. The model uses the GDP of countries as a sort of macroeconomic fitness, and reveals the existence of strong relations between the GDP and the model parameters controlling the formation and the volume of trade relations. In light of the limitations of the existing models (most notably the binary-only nature of the fitness model and the weighted-only nature of the gravity model), we believe that our results represent a promising step forward in the development of a unified model of the ITN structure. We have also shown that the full ECM model can be effectively reduced to the simpler TS model. The success of the latter provides a mathematical explanation of an otherwise puzzling asymmetry, namely the fact that purely topological properties can be successfully predicted without knowing the weights, while weighted network properties can only be predicted if the topological ones are preliminary estimated. Future work should explore how to further improve these results and possibly expand them by introducing additional macroeconomic parameters like geographic distance, thus fully bridging the gap between network-based and gravity-based approaches to the structure of international trade.

Acknowledgments

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 project MULTIPLEX (contract 317532) and the Netherlands Organization for Scientific Research (NWO/OCW).

Appendix A. Higher-order properties

In this appendix we give a summarized description of the binary and weighted network quantities which are studied in this paper. Specifically, we first show how the properties are measured over a real network, and then how the expected values under the ECM and the TS model are constructed.

A.1. Observed properties

Let us note a weighted undirected network as a square matrix W , where the specific entry w_{ij} represents the edge weight between country i and country j . The binary representation of the network is noted by a binary matrix A , where the entries are $a_{ij} = \Theta[w_{ij}]$, $\forall i, j$.

We compute the average nearest neighbor degree as:

$$k_i^{\text{nn}}(W) = \sum_{j \neq i} \frac{a_{ij} k_j}{k_i} = \frac{\sum_{j \neq i} \sum_{k \neq j} a_{ij} a_{jk}}{\sum_{j \neq i} a_{ij}}. \quad (\text{A.1})$$

It is calculated as the arithmetic mean of the degrees of the neighbors of a specific node, which is a measure of correlation between the degrees of adjacent nodes.

The binary clustering coefficient has the following expression:

$$c_i(W) = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ij} a_{jk} a_{ki}}{\sum_{j \neq i} \sum_{k \neq i, j} a_{ij} a_{ki}}. \quad (\text{A.2})$$

It is a measure of the tendency to which nodes in a graph form cluster together. More specifically, it counts how many closed triangles are attached to each node with respect to all the possible triangles.

The corresponding weighted properties are the average nearest neighbor strength and the weighted clustering coefficient. The average nearest neighbor strength, defined as:

$$s_i^{\text{nn}}(W) = \sum_{j \neq i} \frac{a_{ij} s_j}{k_i} = \frac{\sum_{j \neq i} \sum_{k \neq j} a_{ij} w_{jk}}{\sum_{j \neq i} a_{ij}}, \quad (\text{A.3})$$

where $s_i = \sum_j w_{ij}$ is the strength (total flow) of a country. The s_i^{nn} measure the average strength of the neighbors for a specific node i . Like its binary counterpart, it gives the magnitude of activity of a specific node neighbors (weighted activity).

The weighted clustering coefficient [44] is defined as:

$$c_i^W(W) = \frac{\sum_{j \neq i} \sum_{k \neq i, j} (w_{ij} w_{jk} w_{ki})^{\frac{1}{3}}}{\sum_{j \neq i} \sum_{k \neq i, j} a_{ij} a_{ki}}. \quad (\text{A.4})$$

The $c_i^W(W)$ is a measure of the weight density in the neighborhood of a node. It classifies the tendency of a specific node to cluster in a triangle taking into account also the edge-values.

Now, the measured properties of the real network need to be compared with the reproduced properties of the different models. These reproduced properties are the expected values of the maximum entropy ensemble that each model is generating, and can be calculated analytically. The expected values can be obtained by simply replacing a_{ij} with the probability p_{ij} for the different models (p_{ij} is different to each model). This next step is what we will discuss in the next sections.

A.2. Expected values in the BCM and ECM

Since the BCM model is only dealing with the binary representation, we will have expected values just for the two binary higher-order properties. While the ECM gives expectations for the weighted counterparts of the binary properties.

For the binary higher-order properties, we replace a_{ij} with p_{ij} which is the probability of creating a link, and also the expected value of the edge $p_{ij} = \langle a_{ij} \rangle$. This simple procedure yields the analytic formula of the expected value for the properties. We compute the expected average nearest neighbor degree as:

$$\langle k_i^{\text{nn}} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq j} p_{ij} p_{jk}}{\sum_{j \neq i} p_{ij}} \quad (\text{A.5})$$

and the expected binary clustering coefficient as:

$$\langle c_i \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij} p_{jk} p_{ki}}{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij} p_{ki}}, \quad (\text{A.6})$$

where for the BCM model we input $p_{ij} = \frac{z_i z_j}{1 + z_i z_j}$, and for the ECM the more complex term $p_{ij} = \frac{x_i x_j y_i y_j}{1 - y_i y_j + x_i x_j y_i y_j}$.

In the weighted case (weighted higher-order properties), we are left only with the ECM. The expected average nearest neighbor strength is calculated as:

$$\langle s_i^{\text{nn}} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij} \langle w_{jk} \rangle}{\sum_{j \neq i} p_{ij}}, \quad (\text{A.7})$$

where $\langle w_{jk} \rangle = \frac{x_j x_k y_j y_k}{(1 - y_j y_k + x_j x_k y_j y_k)(1 - y_j y_k)}$ and we input the p_{ij} of the ECM model as before.

In the expected value of the c^W we should be more careful, since it is necessary to calculate the expected product of (powers of) distinct matrix entries

$$\langle c_i^W \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i, j} \langle (w_{ij} w_{jk} w_{ki})^{\frac{1}{3}} \rangle}{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij} p_{ki}}. \quad (\text{A.8})$$

We know that

$$\left\langle \sum_{i \neq j \neq k} w_{ij}^\alpha \cdot w_{jk}^\beta \cdot \dots \right\rangle = \sum_{i \neq j \neq k} \langle w_{ij}^\alpha \rangle \cdot \langle w_{jk}^\beta \rangle \cdot \langle \dots \rangle \quad (\text{A.9})$$

with the generic term for the ECM case

$$\langle w_{ij}^\gamma \rangle = \sum_0^\infty w^\gamma q_{ij} (w | \vec{x}, \vec{y}) = \frac{x_i x_j (1 - y_i y_j) Li_\gamma(y_i y_j)}{1 - y_i y_j + x_i x_j y_i y_j} \quad (\text{A.10})$$

where $Li_n(R) = \sum_{l=1}^\infty \frac{R^l}{l^n}$ is the n th polylogarithm of R . For a more comprehensive description please refer to [35].

A.3. Expected values in the TS model

Here again we use the known expressions for the properties and replacing the terms p_{ij}^{ts} and w_{ij}^{ts} with the expected values $\langle a_{ij} \rangle$ and $\langle w_{ij} \rangle$ correspondingly. However, here the expected values are a function of the GDP of the countries, or more specifically the re-scaled GDP g_i . The expressions for the higher-order binary properties are as before:

$$\langle k_i^{\text{nn}} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij}^{\text{ts}} p_{jk}^{\text{ts}}}{\sum_{j \neq i} p_{ij}^{\text{ts}}} \quad (\text{A.11})$$

and

$$\langle c_i \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij}^{\text{ts}} p_{jk}^{\text{ts}} p_{ki}^{\text{ts}}}{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij}^{\text{ts}} p_{ki}^{\text{ts}}}, \quad (\text{A.12})$$

where $p_{ij}^{\text{ts}} = \frac{a_i g_j}{1 + a_i g_j}$.

In the weighted case, the expected average nearest neighbor strength is calculated as:

$$\langle s_i^{\text{nn}} \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i, j} p_{ij}^{\text{ts}} \langle w_{jk}^{\text{ts}} \rangle}{\sum_{j \neq i} p_{ij}^{\text{ts}}}, \quad (\text{A.13})$$

where $\langle w_{ij}^{\text{ts}} \rangle = \frac{a_i g_j}{1 + a_i g_j} \cdot \frac{(1 + b g_i^c)(1 + b g_j^c)}{(1 + b g_i^c + b g_j^c)}$ and we input the p_{ij}^{ts} of the TS model as before.

For convenience reasons we will write the expression for the weighted clustering coefficient c^W first as a function of the fitness parameters z_i and y_j , and later replaced them with the corresponding GDP terms. The expected value of the c^W in the TS case is:

$$\langle c_i^W \rangle = \frac{\sum_{j \neq i} \sum_{k \neq i,j} \left(w_{ij} w_{jk} w_{ki} \right)^{\frac{1}{3}}}{\sum_{j \neq i} \sum_{k \neq i,j} p_{ij}^{\text{ts}} p_{ki}^{\text{ts}}} \quad (\text{A.14})$$

As before we observe that

$$\left\langle \sum_{i \neq j \neq k} w_{ij}^\alpha \cdot w_{jk}^\beta \cdot \dots \right\rangle = \sum_{i \neq j \neq k} \langle w_{ij}^\alpha \rangle \cdot \langle w_{jk}^\beta \rangle \cdot \langle \dots \rangle \quad (\text{A.15})$$

with the generic term for the TS model

$$\langle w_{ij}^\chi \rangle = \sum_0^\infty w^\chi q_{ij}(w | \vec{z}, \vec{\gamma}) = \frac{z_i z_j (1 - y_i y_j) Li_\gamma(y_i y_j)}{(1 + z_i z_j) y_i y_j}, \quad (\text{A.16})$$

where $Li_n(R) = \sum_{l=1}^\infty \frac{R^l}{l^n}$ is the n th polylogarithm of R .

Once we input the expressions of z_i and y_i

$$\begin{aligned} z_i &= \sqrt{a} \cdot g_i, \\ y_i &= \frac{b \cdot g_i^c}{1 + b \cdot g_i^c} \end{aligned} \quad (\text{A.17})$$

equation (A.16) yields

$$\langle w_{ij}^\chi \rangle = \frac{a g_i g_j}{1 + a g_i g_j} \cdot \frac{1 + c g_i^c + b g_j^c}{b^2 g_i^c g_j^c} Li_\gamma \left(\frac{b^2 g_i^c g_j^c}{(1 + b g_i^c)(1 + b g_j^c)} \right) \quad (\text{A.18})$$

References

- [1] Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A and White D R 2009 Economic networks: the new challenges *Science* **325** 422
- [2] Kali R and Reyes J 2007 The architecture of globalization: a network approach to international economic integration *J. Int. Bus. Stud.* **38** 595
- [3] Kali R and Reyes J 2010 Financial contagion on the international trade network *Econ. Inquiry* **48** 1072
- [4] Schiavo S, Reyes J and Fagiolo G 2010 International trade and financial integration: a weighted network analysis *Quant. Finance* **10** 389–99
- [5] Serrano A and Boguna M 2003 Topology of the world trade web *Phys. Rev. E* **68** 015101(R)
- [6] Garlaschelli D and Loffredo M I 2004 Fitness-dependent topological properties of the world trade web *Phys. Rev. Lett.* **93** 188701
- [7] Garlaschelli D and Loffredo M 2005 Structure and evolution of the world trade network *Physica A* **355** 138
- [8] Serrano A, Boguna M and Vespignani A 2007 Patterns of dominant flows in the world trade web *J. Econ. Interact. Coord.* **2** 111
- [9] 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* **57** 1434
- [10] Fagiolo G, Reyes J and Schiavo S 2008 On the topological properties of the world trade web: A weighted network analysis *Physica A* **387** 3868–73
- [11] Fagiolo G, Reyes J and Schiavo S 2009 World-trade web: topological properties, dynamics, and evolution *Phys. Rev. E* **79** 036115
- [12] Fagiolo G, Reyes J and Schiavo S 2010 The evolution of the world trade web: a weighted-network analysis *J. Evolutionary Econ.* **20** 479–514
- [13] Barigozzi M, Fagiolo G and Garlaschelli D 2010 Multinetwork of international trade: a commodity-specific analysis *Phys. Rev. E* **81** 046104
- [14] De Benedictis L and Tajoli L 2011 The world trade network *World Economy* **34** 1417–54
- [15] Squartini T, Fagiolo G and Garlaschelli D 2011 Randomizing world trade: I. A binary network analysis *Phys. Rev. E* **84** 046117
- [16] Squartini T, Fagiolo G and Garlaschelli D 2011 Randomizing world trade: II. A weighted network analysis *Phys. Rev. E* **84** 046118
- [17] Fagiolo G, Squartini T and Garlaschelli D 2013 Null models of economic networks: the case of the world trade web *J. Econ. Interact. Coord.* **8** 75–107
- [18] Fronczak A, Fronczak P and Holyst J A 2012 Statistical mechanics of the international trade network *Phys. Rev. E* **85** 056113
- [19] Squartini T and Garlaschelli D 2012 *In Self-Organizing Systems (Lecture Notes in Computer Science vol 7166)* (Berlin: Springer) p 2435
- [20] 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* **8** 0070726
- [21] Sinha S, Chatterjee A, Chakraborti A and Chakraborti B K 2010 *Econophysics: An Introduction* (Weinheim: Wiley)
- [22] Tinbergen J 1962 *Shaping the World Economy: Suggestions for an International Economic Policy* (New York: The Twentieth Century Fund)
- [23] 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* (Berlin: Springer) pp 161–86
- [24] Fagiolo G 2010 The international-trade network: gravity equations and topological properties *J. Econ. Interact. Coord.* **5** 1–25
- [25] Duenas M and Fagiolo G 2013 Modeling the international-trade network: a gravity approach *J. Econ. Interact. Coord.* **8** 155–78
- [26] de Benedictis L and Taglioni D 2011 The gravity model in international trade *The Trade Impact of European Union Preferential Policies* (Berlin: Springer) pp 55–89

- [27] Garlaschelli D and Loffredo M I 2004 Patterns of link reciprocity in directed networks *Phys. Rev. Lett.* **93** 268701
- [28] Fagiolo G 2006 Directed or undirected? a new index to check for directionality of relations in socio-economic networks *Econ. Bull.* **3** 1–12
- [29] Squartini T, Picciolo F, Ruzzenenti F and Garlaschelli D 2013 Reciprocity of weighted networks *Sci. Rep.* **3** 2729
- [30] 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.* **P02002**
- [31] 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
- [32] Bargigli L and Gallegati M 2011 Random digraphs with given expected degree sequences: a model for economic networks *J. Econ. Behav. Organ.* **78** 396–411
- [33] 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. Phys.* **151** 720–34
- [34] Caldarelli G, Chessa A, Pammolli F, Gabrielli A and Puliga M 2013 Reconstructing a credit network *Nat. Phys.* **9** 125–6
- [35] Mastrandrea R, Squartini T, Fagiolo G and Garlaschelli D 2014 Enhanced reconstruction of weighted networks from strengths and degrees *New J. Phys.* **16** 043022
- [36] Squartini T and Garlaschelli D 2011 Analytical maximum-likelihood method to detect patterns in real networks *New J. Phys.* **13** 083001
- [37] Garlaschelli D and Loffredo M I 2009 Generalized Bose–Fermi statistics and structural correlations in weighted networks *Phys. Rev. Lett.* **102** 038701
- [38] Mastrandrea R, Squartini T, Fagiolo G and Garlaschelli D 2014 Reconstructing the world trade multiplex: the role of intensive and extensive biases (arXiv:1402.4171)
- [39] Garlaschelli D and Loffredo M I 2008 Maximum likelihood: extracting unbiased information from complex networks *Phys. Rev. E* **78** 015101
- [40] Picciolo F, Squartini T, Ruzzenenti F, Basosi R and Garlaschelli D 2012 The role of distances in the world trade web *Proc. 8th Int. Conf. on Signal-Image Technology and Internet-Based Systems* pp 784–92
- [41] <http://comtrade.un.org/>.
- [42] Gleditsch K S 2002 Expanded trade and GDP data *J. Conflict Resolution* **46** 712–24
- [43] Caldarelli G, Capocci A, de Los Rios P and Muñoz M A 2002 Scale-free networks from varying vertex intrinsic fitness *Phys. Rev. Lett.* **89** 258702
- [44] Fagiolo G 2007 Clustering in complex directed networks *Phys. Rev. E* **76** 026107