

A microscopic study of the fitness-dependent topology of the world trade network

K. Hoppe, G. J. Rodgers

Department of Mathematical Sciences, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom

Abstract

Previous studies have suggested that the world-trade network belongs to the class of static hidden variable models. In this article we investigate the microscopic structure of the world trade network, that is the hidden variable correlation matrix of the network. The hidden variable is defined as a rank ordering of gross domestic products. This choice significantly reduces the noise in the statistical analysis found in previous studies. The hidden variable correlation matrix, that expresses the probability that a trade relationship between two countries of given fitness exists, suggests an attachment kernel that at least partially favours trading pairs or dissimilar fitness rather than the purely multiplicative one found previously. Additionally, we provide an in-depth look at the data source and reveal that first-order results, such as the degree distribution, exhibit significant qualitative differences depending on the data provider. Furthermore, we shed light on the intertemporal activity of international trade and point out that fluctuations occur mostly between countries with strong dissimilarities of fitness and connectivity.

Keywords: World trade network, Fitness model, Complex networks

1. Introduction

The world trade network (WTN) has received a considerable amount of attention in the last decade. A detailed understanding of the network structure of international trade facilitates a better comprehension of crises that spread across a country's frontiers. It is generally agreed that a static hidden variable model [1, 2] best describes the topology of the WTN. The static model is most suitable because there are only a finite number of distinct countries in the world. Depending on the scope of the analysis, previous contributions understand the network as growing since the number of vertices increased from 86 in the year 1950 to 190 in the year 2000 [3]. However, since there are –depending on the definition– approximately 195 countries in the world to date, the growth of the network cannot be sustained. A dynamic perspective on the world trade network, by investigating the duration and strength of individual connections and also the relationship between the development status of a country and its ability to establish long-lasting trade relationships has been discussed in [4, 5]. It has been shown in [6, 7], that the static fitness model is an adequate model to describe the structure of the world trade network for several years, not just on the level of node degree distributions, but up to higher order statistics, such as the average nearest neighbor degree. In this article, we look into the microscopic structure of the WTN. The main quantity of interest is the edge density, that is the probability matrix for any two nodes of fitness x and y to share an edge. The fitness model is found to be an excellent fit for the WTN. However, the previously found multiplicative attachment kernel varies significantly from the kernel that we find by investigating the edge density.

A review of the literature on the network of international trade reveals that many results are not without controversy. Even first order results, such as the form of the degree distribution, differ between various publications. One reason for these discrepancies is the choice of data source. Every trade relationship between two countries should be reported twice, once as an export and once as an import. However, this is not always the case [8]. As a result, different data providers report different trading data. Another issue is the way these disparities are corrected. Some [9, 10] simply use only import data, which is deemed more reliable than export data, others [11, 12, 13] interpolate between reports of exports and imports.

The data set we use for the present article is provided by the National Bureau of Economic Research (NBER) [8] and forms an excellent starting point for this investigation. The special feature of this data set is that one part of it has been adjusted for errors, while another part has not. The adjustments have been made by specialised macroeconomists and can therefore be deemed reliable.

The world trade network is treated as directed and binary in this article. That means that directionality of trade is taken into account, but the traded amounts are ignored. Others [7, 14, 12, 15] have argued that directionality could be ignored because of the high reciprocity, i.e. export links are usually reciprocated with import links. However, since the world trade network is directed by nature, the analysis in the present article is laid out in terms of a directed network. A weighted representation of the WTN has been investigated for example in Refs. [12, 16, 13]. Curiously, the additional information of edge weight is not always of greater explanatory power than binary network analysis [10, 17]. In [10, 17] a network randomisation technique was used to show that by knowing the degree sequence of the binary WTN, higher order statistics like the average neighbor degree and the clustering coefficient can be obtained. However, using a weighted network approach and fixing the sequence of interaction strengths, these higher order statistics cannot be found anymore. This suggests that binary network analysis used in economics is a powerful tool, because all the necessary information is contained in the first order statistics.

In principle, the WTN is a spatially embedded network. However it has been illustrated that distances between countries do not add a significant amount of information to a binary analysis of international trade [18]. Similarly, also in the weighted regime, the importance of geographic distance for understanding international trade is declining over time [19]. The results of [19] are in contrast to the standard notion in economics, that the intensity of trade between pairs of countries is strongly related to their distance, see for instance [20]. Another aspect that will not feed into the analysis in this article is the multi-layered architecture of international trade. Every reported trade flow is an aggregation over different product categories. Refs. [10, 17, 9, 21] investigate these different layers separately.

The investigation of the topology of the world trade is not conducted as an end in itself. Its aim is to understand how trade can affect economic welfare. Many aspects are yet to be understood. The theory of complex systems is just one of the building blocks towards a good understanding of those effects [22, 23]. Network-theoretic measures have been shown to explain parts of nation's income. In [24] it is shown that an improvement of the degree centrality ranking by ten units increases the average GDP per capita by 0.27%. Others [11, 6, 25, 3] investigate correlation structures of income, connectivity and interaction strength in weighted networks and illustrate that an involvement in international trade has a direct impact on income and vice versa. Additionally, it has been shown that network properties have good explanatory power to detect vulnerable economies in the WTN [26, 27].

This paper is organised as follows. In section 2, the static fitness model is reviewed and the central quantities for this study derived. In section 3, the data set is introduced and reasons for its choice are discussed. In section 4, the definition of fitness is elucidated. In section 5, the inter-temporal structure of the WTN is clarified and in section 6, the static structure is investigated. Section 7 closes the article with concluding remarks.

2. The static fitness model

The empirical analysis in this article is based on the static fitness model, as it was introduced in Refs. [1, 28, 2]. The investigation that is presented later in the text relies on results that are reviewed in the following.

The static fitness model is a network model with N nodes and M directed edges. Each node inside the network is endowed with a fitness value x , that is drawn from a probability density function $\rho(x)$. The probability that a node with fitness x originates a link toward a node with fitness y is proportional to the attachment kernel $f(x, y)$. Although the network is static, the edge deployment process can be understood as dynamic. The probability that an edge is added to a pair of nodes $i \rightarrow j$ is $f(x_i, x_j) / \sum_{k,l} f(x_k, x_l)$. The reader should be reminded at this point that this formalism ignores the possible problem of edge duplications. Edge duplications are not an issue in sparse networks, i.e. $M \ll N^2$, however they will lead to bias in dense networks, such as the WTN. However, correcting for the impossibility of duplications leads to non-linearities in the formalism, so that closed form solutions can no longer be found.

To begin with, the probability that two nodes with fitness x_i and x_j are connected in a network with M edges and N nodes is derived. Denote this probability with $C_{M,N}(x_i, x_j)$ and further define $\bar{C}_{M,N}(x_i, x_j) = 1 - C_{M,N}(x_i, x_j)$. We will refer to $C_{M,N}$ as the microscopic structure of this network because it contains all information on the network and

thus all topological properties of the network can be derived from it. Assuming a sequential update of the network, $C_{M,N}$ can be derived from first principles using

$$\bar{C}_{M+1,N}(x_i, x_j) = \bar{C}_{M,N}(x_i, x_j) \left(1 - \frac{f(x_i, x_j)}{\sum_{k,l} f(x_k, x_l)} \right). \quad (1)$$

That is, the probability that there exists no edge $i \rightarrow j$ in a network with $M + 1$ edges is equal to the probability that there was no such edge in a network with M edges and no edge has been added to the pair $i \rightarrow j$. Assuming now sufficiently large N , such that fitness can be approximated as continuous variable leads to

$$\bar{C}_{M+1,N}(x, y) = \bar{C}_{M,N}(x, y) \left(1 - \frac{f(x, y)}{N^2 \int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta} \right). \quad (2)$$

Using the boundary condition $\bar{C}_{0,N}(x, y) = 1$ and a continuous time approximation, the solution of Eq. (2) is given by

$$\bar{C}_{M,N}(x, y) = \exp \left\{ -\frac{M}{N^2} \frac{f(x, y)}{\int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta} \right\}. \quad (3)$$

This expression can be approximated, assuming a sufficiently sparse network, such that the probability for the existence of an edge is given by

$$C_{M,N}(x, y) \cong \frac{M}{N^2} \frac{f(x, y)}{\int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta}. \quad (4)$$

Another quantity of interest in this article is the expected in-degree of a node with a given fitness value. This quantity can also be derived from first principles. Define the probability that a node with fitness x increases its in-degree in an edge-addition step with

$$\lambda(x, N) = \frac{1}{N} \frac{\int_0^1 f(y, x) \rho(y) dy}{\int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta}. \quad (5)$$

Since the size of the network N is a factor of $\lambda(x, N)$, also define $\lambda(x) = N\lambda(x, N)$. Furthermore, $p_{M,N}(k|x)$ is the probability that a node with fitness x has in-degree k in a network with M edges and N nodes. $p_{M,N}(k|x)$ obeys

$$p_{M+1,N}(k|x) = p_{M,N}(k|x)(1 - \lambda(x, N)) + p_{M,N}(k-1|x)\lambda(x, N). \quad (6)$$

This equation can be solved using a generating function approach. The resulting conditional in-degree distribution is

$$p_{M,N}(k|x) = \frac{e^{-M/N \lambda(x)} \left(\frac{M}{N} \lambda(x) \right)^k}{\Gamma(k+1)}. \quad (7)$$

Where $\Gamma(x)$ is the standard Gamma function. Notice that the out-degree distribution can be described in a very similar way. The only part that changes is the attractor $\lambda(x, N)$. The probability that a node increases its out-degree by one in an edge addition step is

$$\mu(x, N) = \frac{1}{N} \frac{\int_0^1 f(x, y) \rho(y) dy}{\int_0^1 \int_0^1 f(\xi, \eta) \rho(\xi) \rho(\eta) d\xi d\eta}. \quad (8)$$

Thus if $f(x, y)$ is a symmetric function, then the conditional out-degree- and in-degree distributions are identical. The fitness conditional expectation of the in-degree of a randomly chosen node is

$$\bar{k}(x) = \mathbb{E}[k|x] = \sum_{k \geq 0} k p_{M,N}(k|x) = \frac{M}{N} \lambda(x). \quad (9)$$

Thus, the form of the conditional expectation is solely determined by the form of λ . The quantities $C_{M,N}(x, y)$ and $\bar{k}(x)$ are closely related, more specifically:

$$\int_0^1 C_{M,N}(x, y)\rho(x)dx = \frac{M}{N^2} \frac{\int_0^1 f(x, y)\rho(x)dx}{\int_0^1 \int_0^1 f(\xi, \eta)\rho(\xi)\rho(\eta)d\xi d\eta} \quad (10)$$

$$= \frac{M}{N^2} \lambda(y) \quad (11)$$

$$= \frac{1}{N} \bar{k}(y). \quad (12)$$

This relationship will be used in a later section to assess the validity of the fitness model for the WTN.

3. The choice of a data source

The choice of the data set is crucial for the validity of the derived results. Data for the WTN can be found in various publications [29, 30, 8]. The concern of data validity is raised in various places [13, 12, 10, 9, 8]. Data on world trade should always be reported in two statements. Every flow of goods or services occurs in one country as an export and in another country as import. However, it occurs that a significant amount of trade is only reported on one side. This leads to inconsistencies in the data analysis. Some suggest methods to approximate the true value of money flow by averaging over the existing data [13, 12], while others [10, 9] suggest to replace all export data with import data, because import data is believed to be more reliable [8].

A technique to correct for the inconsistencies is not suggested here, merely the severity of the problem is pointed out. The network data for the present study is obtained from the National Bureau of Economic Research (NBER). A detailed description is given in [8]. The primary data source for [8] is two-fold. Data for the years 1962-1983 is taken from UN data, collected at various times and has not been revised during later stages. Data from 1984-2000 is also taken from UN Comtrade data, but has undergone significant revisions and corrections that are explained in detail in [8]. The corrections can be considered reliable, as experts in the field have searched for data and corrected it manually. The corrections were conducted at component level, i.e. each commodity group was considered in isolation.

The effect of this data adjustment is illustrated in Fig. 1. The figure shows density estimators for the in- and out-degree distribution of the WTN over the years 1962-2000. The two colors indicate the separation into data that was curated and data that has undergone only little adjustment, if any. It is evident that the year 1984 marks a change of regime, which is most likely not inherent in the network formation process, but in the way the data is collected and revised, which has changed in that year. Comprehension of stochastic processes that take place on networks is often formulated in terms of the degree distribution [31]. It is therefore of high importance for further analysis of the network to have a correct representation of this quantity.

It is anticipated that a different data collection scheme will lead to slightly different empirical results, but Fig. 1 presents an entire regime shift of first-order results that is solely initiated by a change of the data source. All further results of this article rely on the relatively short, but well adjusted data that is discussed in Ref. [8] for the years 1984-2000.

Next to the data on international trade, also data to resemble the hidden variables was collected. The gross-domestic-product (GDP) was obtained from UN-data ¹. The countries that are covered in the present report are displayed in Tab. A.1.

4. Definition of fitness

It is undisputed that the world trade network falls into the class of hidden variable models [6, 7, 3]. However, it is not clear how to define fitness. There is general agreement that fitness is defined over some constraint interval, usually [0, 1]. Two different definitions of fitness are investigated in the following.

¹<http://www.data.un.org/>

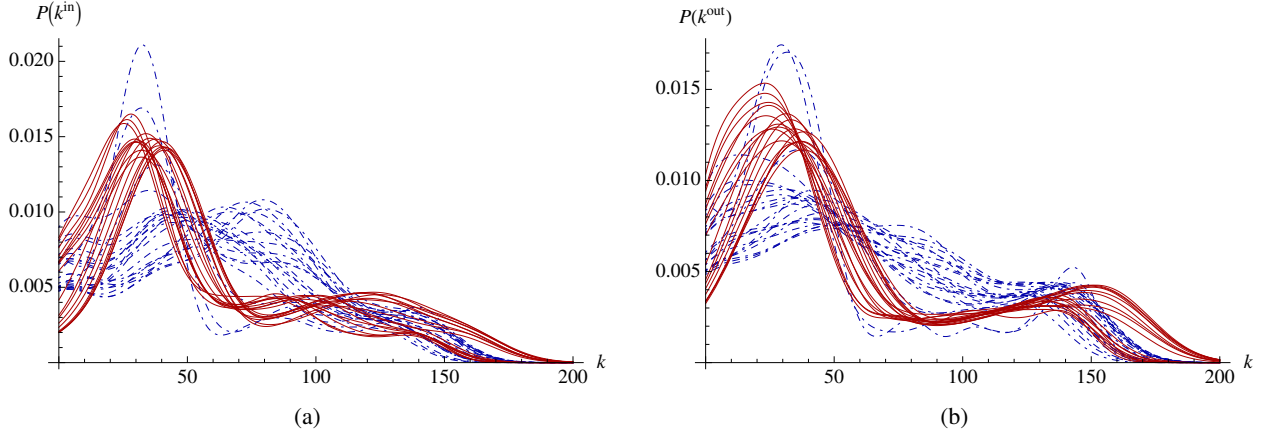


Figure 1: Density estimators for the (a) in-degree and (b) out-degree distribution in the world trade network for the years 1962-2000. The data before the year 1984 is indicated with dot-dashed lines (blue), the distributions for the years 1984 and after is plotted solid red.

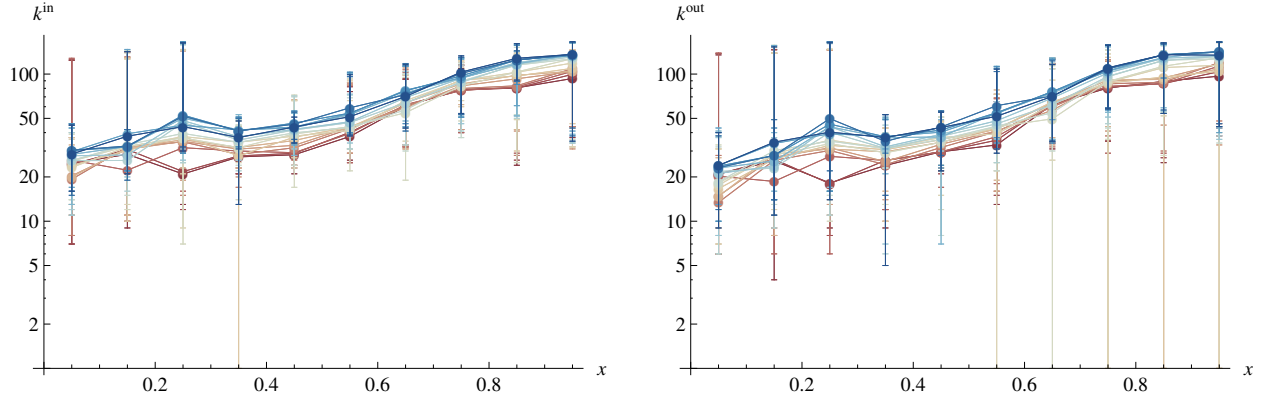


Figure 2: Relationship between the in-degree (a), respectively the out-degree (b) of a node with its fitness, when fitness is defined as in Eq. (13). The log-linear plots show an almost linear dependence which suggests that $\mathbb{E}[k|x]$ has an exponential form. The data is binned into intervals of length 0.1, the averages of these intervals are indicated with bulletpoints. The maximum and minimum of the intervals are shown as errorbars. The different colors represent different years starting from 1962 (red) until 2000 (blue).

One possibility is to impose a ranking on the node's GDP and normalise this ranking to the unit interval. Formally, denote the income of country i as w_i . Then the corresponding fitness of node i is defined as

$$x_i = \vartheta(w_i) = \frac{1}{N} \sum_{j=1}^N H(w_i - w_j) \quad (13)$$

Where $H(x)$ is the Heaviside step function. Fig. 2 illustrates the relationship between fitness and expected node-degree. The almost linear form in the log-linear scaled plot implies an approximately exponential form of the fitness conditional expected node-degree $\mathbb{E}[k|x]$. This empirical finding will be compared later with analytical results from Section 2. Notice that the particular choice of fitness in Eq. (13) simplifies the expression of $\lambda(x)$ significantly. Suppose that the national income X is distributed with density $f_X(x)$. Consider fitness as a transformed variable, in

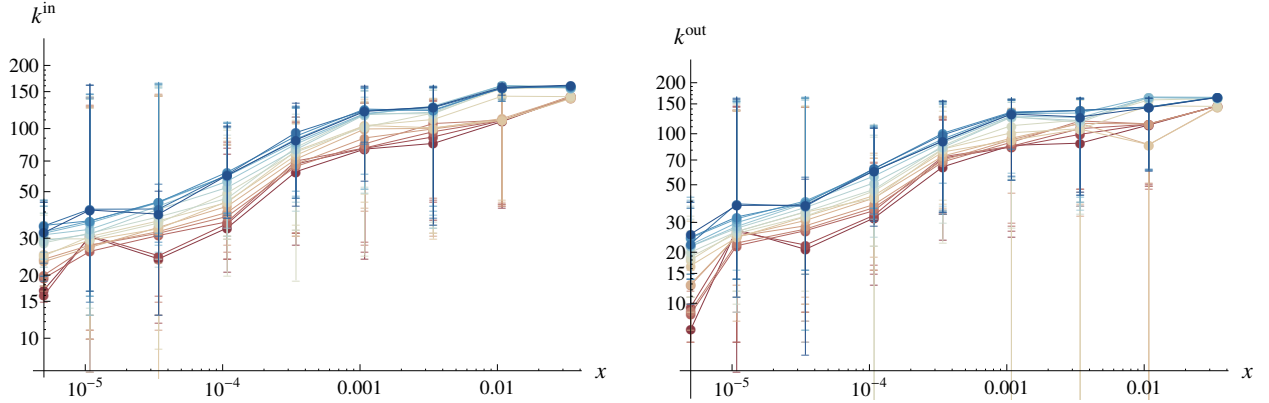


Figure 3: Relationship between the in-degree (a), respectively the out-degree (b) of a node with its fitness, when fitness is defined as in Eq. (18). The similar shape to the graphs in Fig. 2 shows that the expected degree is more stretched here, since the abscissa is logarithmically scaled, as opposed to the linear scaling in Fig. 2. The data is binned into logarithmic intervals, the averages of these intervals are indicated with bulletpoints. The maximum and minimum of the intervals are shown as errorbars. The different colors represent different years starting from 1962 (red) until 2000 (blue).

the sense that $Y = \vartheta(X)$. The cumulative distribution of fitness can be written as

$$R_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}[\vartheta(X) \leq y] \quad (14)$$

$$= \mathbb{P}[X \leq \vartheta^{-1}(y)] \quad (15)$$

$$= \int_0^{\vartheta^{-1}(y)} f_X(x) dx. \quad (16)$$

The definition of $\vartheta(x)$ is simply a discrete formulation of the cumulative distribution function of income X . Thus $\int_0^x f(\xi) d\xi = \vartheta(x)$ and therefore it follows directly from Eq. (16) that

$$R_Y(y) = y \quad \text{thus} \quad \rho(y) = 1. \quad (17)$$

The fact that the fitness density becomes a constant eliminates one degree of freedom in the definition of $\lambda(x)$.

However, the particular definition of fitness is arbitrary. Another form that was suggested in Ref. [6] is

$$\vartheta'(w_i) = \frac{w_i}{\sum_{j=1}^N w_j}, \quad (18)$$

or alternatively in Ref. [7]

$$\vartheta''(w_i) = \frac{w_i}{N \sum_{j=1}^N w_j} = \frac{\vartheta'(w_i)}{N}. \quad (19)$$

The corresponding relationship between the degree and the fitness of a node, using the fitness definition in Eq. (18) is illustrated in Fig. 3. This sub-linear form in a log-log-scaled plot is the same as found in Ref. [6]. Notice that the empirical $\lambda(x)$ is slightly more noisy in the ϑ' regime than when fitness is defined by ϑ . However, the averages do not represent good approximations for the bins, the data in the bins is distributed over a wide range of k . The original derivation of various topological properties in Ref. [2] uses the fact that $\mathbb{E}[k|x]$ is invertible. That implies that the distribution of k inside the bins must be relatively narrow. Both fitness definitions seem to violate this assumption, however the results using ϑ appear narrower. For both definitions, there seems to be a more explicit relationship between fitness and the in-degree than between fitness and the out-degree.

5. Intertemporal structure of the WTN

The structure of the world trade network is permanently changing, new trade relations are established and existing ones are terminated. Fig. 4 illustrates the macroscopic behaviour of the WTN over time. The number of nodes is

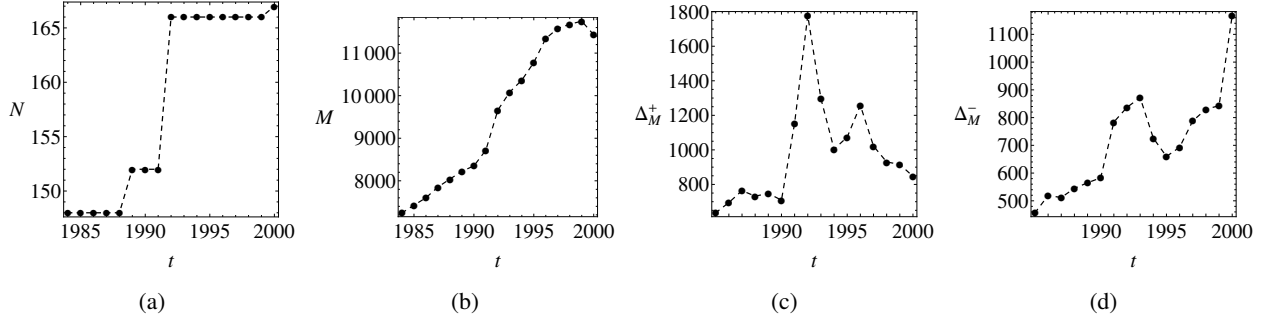


Figure 4: Macroscopic development of the WTN over time. (a) Temporal development of the number of nodes N . (b) Temporal development of the number of edges M . (c) Temporal development of the number of added edges as defined in Eq. (20). (d) Temporal development of the number of removed edges as defined in Eq. (20)

almost static apart from the sudden increase of trading countries when the Soviet Union collapsed. Since this event is not system inherent, the network can be regarded as static with respect to the number of nodes. The number of edges is almost constantly increasing. In order to understand the change of the number of edges better, we define

$$\Delta_M^\pm(t) = \sum_{i,j=1}^{N(t)} \delta[A_{ij}(t) - A_{ij}(t-1), \pm 1]. \quad (20)$$

Where $\delta[k, k']$ is the Kronecker Delta and $A_{ij}(t)$ is the adjacency matrix of the directed network at time t . Figs. 4 (c-d) reveal that the steady influx of new edges is composed by both edge deletion and dominating edge addition. Notice that the link density, defined as $M/(N(N-1))$ is steadily increasing over the reported time, here from 0.33... up to 0.41.... These figures remain the same in the case of the undirected network. This finding is in contrast to Ref. [13], where a roughly constant link density 0.52 over 53 years from 1948 to 2000 using an undirected network is reported. This difference illustrates very clearly the importance of a carefully chosen data source, as link density is a very coarse measure.

Apart from the macroscopic observation that edges constantly emerge and disappear, it is important to understand whether fitness is the main driver of these processes. Recently, activity driven networks have been investigated [32, 33]. Every node in this class of networks is endowed with an activity value that determines the probability that a particular node originates a new link inside the network. In order to identify a coupling between activity in new trade links and fitness, we define

$$e_{i,\pm}^{\text{out}}(t) = \sum_{j=1}^{N(t)} \delta[A_{ij}(t) - A_{ij}(t-1), \pm 1] \quad \text{and} \quad (21)$$

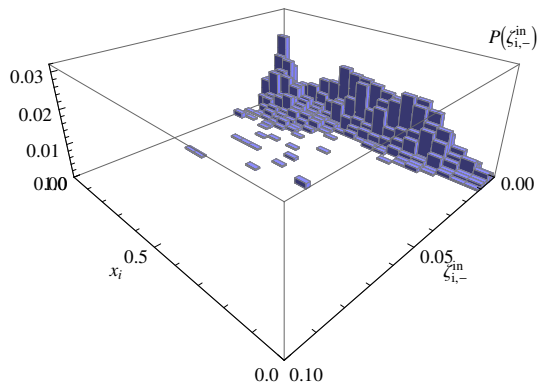
$$e_{i,\pm}^{\text{in}}(t) = \sum_{j=1}^{N(t)} \delta[A_{ji}(t) - A_{ji}(t-1), \pm 1]. \quad (22)$$

$e_{i,\pm}^{\text{out}}(t)$ and $e_{i,\pm}^{\text{in}}(t)$ measure the number of edges that are added to node i either as outward or inward links. Activity levels can be measured as fractions of overall activity. Thus, we define additionally

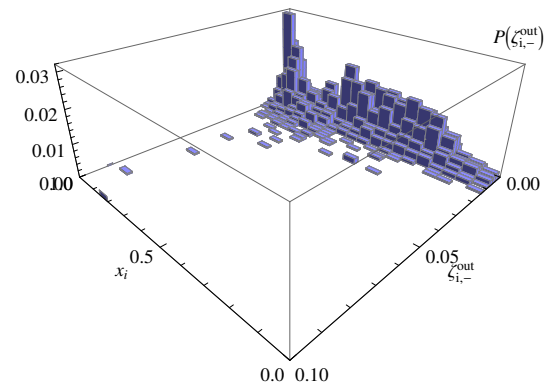
$$\zeta_{i,\pm}^{\text{out}}(t) = \frac{e_{i,\pm}^{\text{out}}(t)}{\sum_{j=1}^N e_{j,\pm}^{\text{out}}(t)} \quad \text{and} \quad (23)$$

$$\zeta_{i,\pm}^{\text{in}}(t) = \frac{e_{i,\pm}^{\text{in}}(t)}{\sum_{j=1}^N e_{j,\pm}^{\text{in}}(t)}. \quad (24)$$

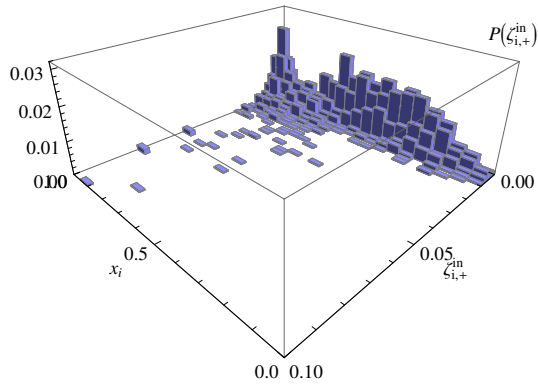
Fig. 5 illustrates the dependency of fitness on $\zeta_{i,\pm}^{\text{in,out}}$ in a two-dimensional histogram. It is clear from Fig. 5 that fitness is not a good descriptor for activity levels, there is no clearly visible correlation between activity and fitness. Since no



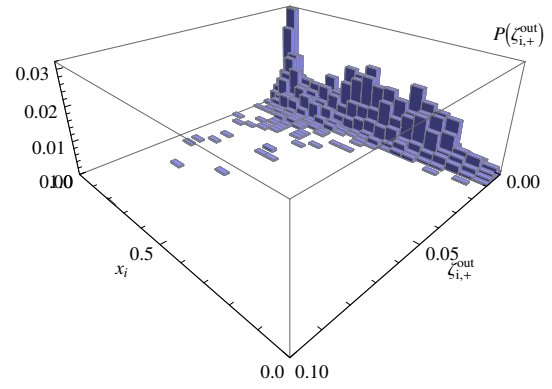
(a)



(b)



(c)



(d)

Figure 5: 2D-Histograms of activity levels as defined in Eqs. (23-24). (a) $\zeta_{i,-}^{\text{in}}$ vs. fitness. (b) $\zeta_{i,-}^{\text{out}}$ vs. fitness. (c) $\zeta_{i,+}^{\text{in}}$ vs. fitness. (d) $\zeta_{i,+}^{\text{out}}$ vs. fitness.

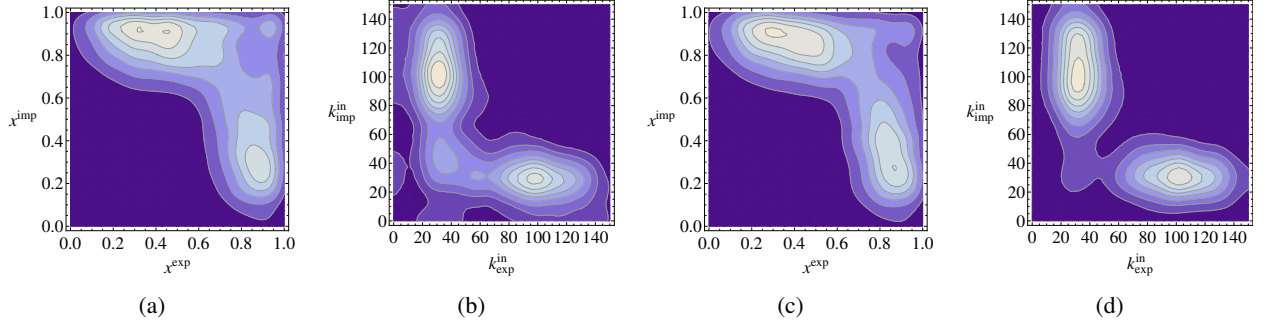


Figure 6: Contour-plots of two dimensional histograms of in-degrees and fitness adjacent to removed and added edges. Dark areas indicate regions of low probabilities, white areas indicate high probabilities. (a) Fitness of nodes adjacent to newly deployed edges. (b) In-degrees of nodes adjacent to newly deployed edges. (c) Fitness of nodes adjacent to removed edges (d) In-degrees of nodes adjacent to removed edges.

particular pattern emerges in Fig. 5, it is clear that a unilateral mechanism is not driving the intertemporal dynamics of the WTN as defined in Eqs. (23-24). Fig. 6 shows contour plots of the fitness and degrees of nodes adjacent to newly deployed and removed edges. Interestingly, two distinct patterns emerge. This shows that in contrast to the previously discussed activity histograms, a bilateral mechanism can explain the addition and removal well. Fig. 6 shows that new links emerge and disappear preferably between countries with dissimilar fitness and degree. The fitness contour plots (a), (c) for addition and removal of edges are qualitatively similar. Both have a high concentration around pairs of high fitness and further high densities for pairs of dissimilar fitness. Since this is the case not only for addition, but also for removals, it suggests that fitness does not only account for attractiveness, but also for activity in general. Curiously, the contour plots of node degrees adjacent to newly deployed and removed edges show an almost inverted pattern, compared to the fitness based contour plots. The fitness dependent plots show high activity mostly between high fitness nodes and nodes of differing fitness, while the degree based plots show most of the activity with at least one low-degree node involved. This is surprising as it is pointed out earlier in this text that there is a strong positive relationship between fitness and degree. Another aspect that can be inferred from Fig. 6 is that new trade relationships emerge equally from low-connected to highly connected ones and vice versa. In other words, the attachment mechanism can be assumed to be symmetric.

6. Static structure of the WTN

In order to get a more complete picture of the internal mechanism of the WTN, it is useful to look also at the mere existence of links, rather than at their emergence. We will restrict ourselves here to the years 1984 and 2000. Fig. 7 pictures the fitness and in-degrees of nodes that are adjacent to edges in the years 1984 and 2000. The results are similar to the findings in the previous section. Most of the trade flows between dissimilar countries, low connected countries trade with highly connected ones and this is true for trade in both directions, the plots are almost symmetrical for import and export. This disassortativity has as well been described in Refs. [12, 14, 16]. Fig. 7 also shows that trade patterns have hardly changed between the years 1984 and 2000. Merely existing patterns have become more apparent. The fitness-fitness histograms in Figs. 7a and 7c show a slightly different pattern compared to the histogram in Fig. 6a. While edges are added most preferentially between countries of different fitness, the static view reveals that most of the trade is taking place between high fitness countries at the top of the ranking scale. Fig. 8 shows some example contours for different functional forms. It is clear that, if fitness is defined as in Eq. (13), the microscopic structure of trade in Fig. 7d cannot be explained with a multiplicative attachment kernel that has been suggested in [6] and is illustrated in Fig. 8c. The form of the fitness-fitness correlations in Fig. 7d is more similar to the functional forms in Figs. 8a and 8b.

The histogram of fitness in Fig. 7c can be used as an estimator for the true $C_{M,N}(x, y)$. This edge density combined with the connection of $\lambda(x)$ and $C_{M,N}(x, y)$ in Eq. (7) can be utilized to estimate the resulting $\bar{k}(x)$ from the data and compare it with the actual $\bar{k}(x)$. This comparison can be found in Fig. 9a. Fig. 9 validates the hypothesis that the WTN is well described by a fitness model. The fitness conditional degree distribution, that is found in real data, is over a

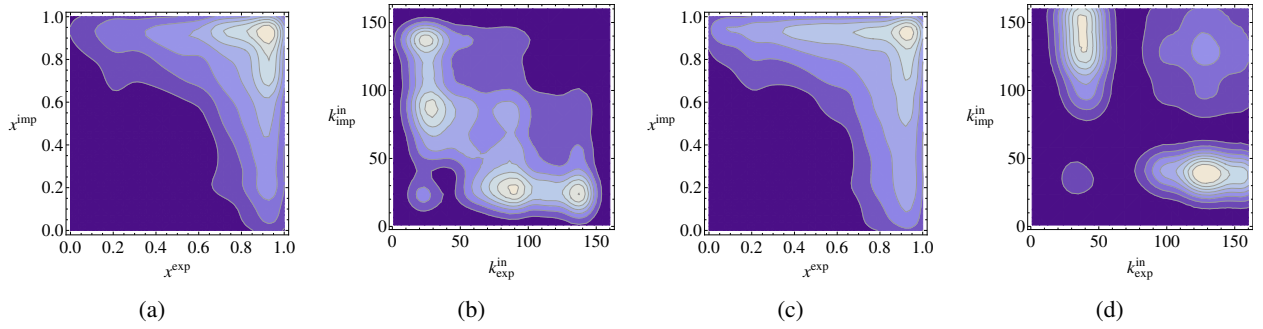


Figure 7: Contour plots of histograms of node properties adjacent to edges that are present in the years 1984 and 2000. (a) Fitness of nodes adjacent to edges in the year 1984. (b) In-degrees of nodes adjacent to edges in the year 1984. (c) Fitness of nodes adjacent to edges in the year 2000. (d) In-degrees of nodes adjacent to edges in the year 2000.

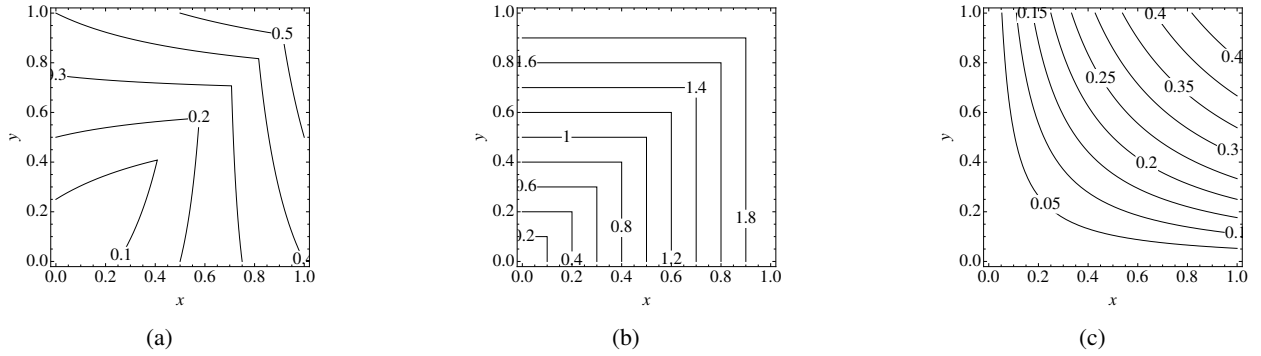


Figure 8: Contour plots of (a) $g(x, y) = 0.4|x - y| + 0.6xy$. (b) $g(x, y) = |x - y| + x + y$, (c) $g(x, y) = xy/(1 + xy)$

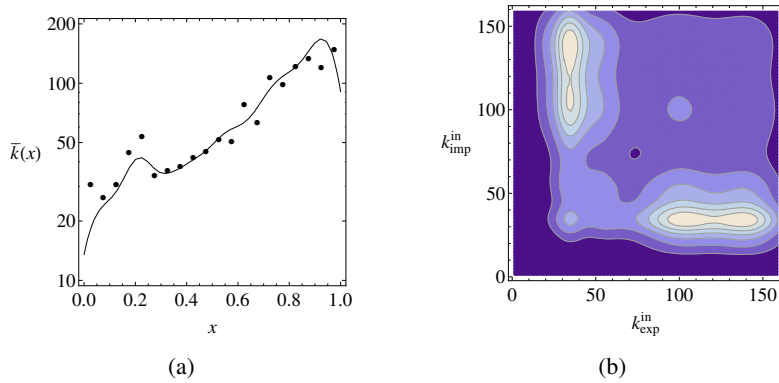


Figure 9: Theoretical results for the expected in-degree of a node and degree-degree correlations using empirical data as an estimator for the true $f(x, y)$. (a) Comparison of theoretical to analytical results of the fitness conditional expected in-degree of a randomly chosen node. The bullet points indicate the observed values. The solid line represents the estimator using the fitness-fitness histogram of nodes adjacent to edges in the year 2000 as an estimator for the true $C_{M,N}(x, y)$ that can be substituted in the expression for the average degree as in Eq. (12). (b) Contour plot of theoretical results on the degree-degree correlations $P(k, k')$, using Eq. (25) and the histogram of node's fitness adjacent to existing edges in the year 2000 as estimator for $C_{M,N}(x, y)$.

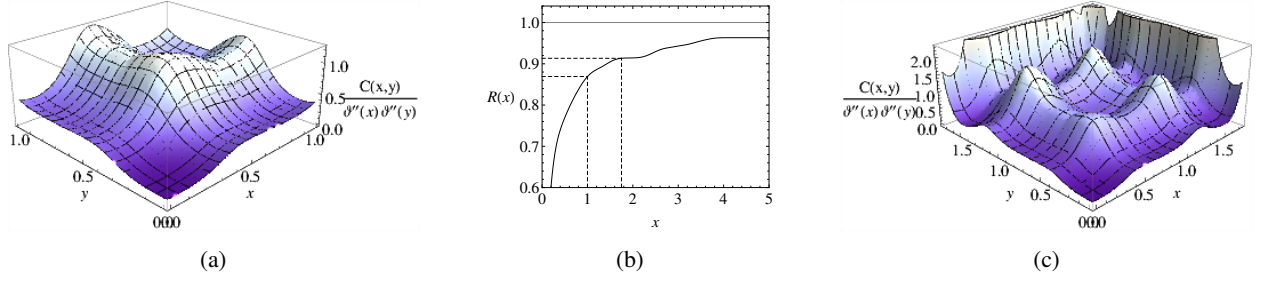


Figure 10: (a) Normalised connection density of the WTN in the year 2000 over the unit-square, using the fitness definition in Eq. (19). (b) Cumulative distribution function of fitness in the WTN in the year 2000 with fitness defined as in Eq. (19). The dashed lines mark the upper domain of the density plots in (a) and (c). (c) Normalised connection density of the WTN in the year 2000 over $[0, 1.75]$, using the fitness definition in Eq. (19)

broad range of values confirmed by the values arising from the theory on static fitness models, using the empirical $\tilde{C}_{M,N}(x, y)$ as an estimator for $\lambda(x)$.

Not only on the aggregated level of the conditional expectation of in-degrees, good agreement of degree correlation structure of the network can also be found. The probability that there exists a link from a node with in-degree k toward a node with in-degree k' is defined as

$$P(k, k') = \int_0^1 \int_0^1 p(k|x) C_{M,N}(x, x') p(k'|x') dx dx'. \quad (25)$$

Fig. 9b shows the $P(k, k')$ contour that can be computed using Eqs. (7), (25) and the observed edge density $\tilde{C}(x, y)$. The comparison of Figs. 7d and 9b shows an excellent agreement between the theory and the empirical findings. The areas of high trading intensities are almost congruent.

The empirical analysis, that has been laid out in this paper confirms the general understanding that the network of international trade is governed by a vertex intrinsic fitness model. However, our investigation into the microscopic detail –the fitness correlation matrix – of the network revealed that the attachment kernel $f(x, y)$ is less trivial than, for example, assumed in Refs. [6, 7]. The topology in a fitness driven network is determined by two variables, the attachment kernel $f(x, y)$ and the fitness density $\rho(x)$. For the sake of robustness, we have chosen to define fitness such that one degree of freedom is eliminated and the model depends solely on the form of $f(x, y)$.

We will now turn to the attachment kernel, that has been proposed in [6, 7]. We have shown so far that the WTN is well described using an attachment kernel that at least partially favours links between dissimilar countries when fitness is defined as in Eq. (13). This does not imply that the attachment kernel $f(x, y; a) = axy/(1 + axy)$, which was proposed in [6, 7] together with fitness definitions ϑ' and ϑ'' is incorrect. Moreover, it has been shown that this kernel yields correct predictions for derived quantities, such as the average nearest neighbour-degree and the fitness conditional degree distribution [6, 7].

Fig. 10 shows empirical results of the microscopic structure of the WTN using the alternative fitness definition ϑ'' , defined in Eq. (19). Fig. 10c illustrates that the results become very noisy for larger values of fitness. The fractions of the sample that are included in the three-dimensional histograms are shown in the cumulative distribution of fitness in Fig. 10b. The data for the first $\sim 85\%$ is relatively noise free. Fig. 10a illustrates that the attachment kernel is symmetric in its arguments and favours connections between high-fitness nodes. The emerging pattern is very different from the one that was found using fitness definition ϑ . Notably, $\tilde{C}_{M,N}(x, y)$ is non-zero at the boundaries $(x, 0)$ and $(0, y)$. This is an indicator that $C_{M,N}(x, y)$ must be at least partially additive. Moreover, apart from being a technicality, a pure multiplicative kernel implies zero trade with with lowest fitness countries. This would represent a system in which trade advantages that arise from different income levels would not be monetarised. A more detailed analysis of the exact form of the edge density is not possible. Fig. 10c shows that the entire analysis with fitness defined as in Eq. (19) is not very robust. The necessary normalisation of the edge density $C_{M,N}(x, y)$ introduces a considerable amount of noise for larger fitness values. This shows that the introduction of fitness defined as in Eq. (13) is a significant advancement over previous definitions.

7. Conclusion

In this article, we have discussed empirical aspects of international trade. The microscopic structure of the world trade network, that we have illustrated by computing the edge density matrix, exhibits a pattern which is significantly different from previously assumed ones. We have found that the income dependency of trade between two countries is not purely multiplicative, but also at least partially additive with a preference for trade between dissimilar countries. It has also been shown here that the static hidden variable model gives an excellent description of the world trade network on the microscopic level. So far, only the aggregate level had been studied. For future analysis, we have proposed a novel approach to define fitness. By introducing a rank ordering of income, one degree of freedom, namely the fitness density, is eliminated. This reduces the noise in the statistical analysis significantly. Additionally, we have demonstrated that empirical results on the world trade network depend largely on the chosen data source. Even first order results, such as the degree distribution can differ depending on the data provider. The magnitude of these disparities is larger than expected and should be considered in future investigations.

Appendix A. Countries covered in this study

Afghanistan	Cameroon	Estonia	Indonesia	Mali	Philippines	Syria
Albania	Canada	Ethiopia	Iran	Malta	Poland	Taiwan
Algeria	Cent.Afr.Rep	Fiji	Iraq	Mauritania	Portugal	Tajikistan
Angola	Chad	Finland	Ireland	Mauritius	Qatar	Thailand
Argentina	Chile	Fm USSR	Israel	Mexico	Rep Moldova	Togo
Armenia	China	Fm Yemen AR	Italy	Mongolia	Romania	Trinidad Tbg
Australia	China HK SAR	Fm Yemen Dm	Jamaica	Morocco	Russian Fed	Tunisia
Austria	China MC SAR	Fm Yugoslav	Japan	Mozambique	Rwanda	Turkey
Azerbaijan	Colombia	France,Monac	Jordan	Myanmar	Samoa	Turkmenistan
Bahamas	Congo	Gabon	Kazakhstan	Nepal	Saudi Arabia	Uganda
Bahrain	Costa Rica	Gambia	Kenya	Neth.Ant.Aru	Senegal	UK
Bangladesh	Cote Divoire	Georgia	Kiribati	Netherlands	Seychelles	Ukraine
Barbados	Croatia	Germany	Korea D P Rp	New Calednia	Sierra Leone	Untd Arab Em
Belarus	Cuba	Ghana	Korea Rep.	New Zealand	Singapore	Uruguay
Belgium-Lux	Cyprus	Greece	Kuwait	Nicaragua	Slovakia	USA
Belize	Czech Rep	Greenland	Kyrgyzstan	Niger	Slovenia	Uzbekistan
Benin	Czechoslovak	Guatemala	Lao P.Dem.R	Nigeria	Somalia	Venezuela
Bermuda	Dem.Rp.Congo	Guinea	Latvia	Norway	South Africa	Viet Nam
Bolivia	Denmark	GuineaBissau	Lebanon	Occ.Pal.Terr	Spain	Yemen
Bosnia Herzg	Djibouti	Guyana	Liberia	Oman	Sri Lanka	Yugoslavia
Brazil	Dominican Rp	Haiti	Libya	Pakistan	St.Kt-Nev-An	Zambia
Bulgaria	Ecuador	Honduras	Lithuania	Panama	Sudan	Zimbabwe
Burkina Faso	Egypt	Hungary	Madagascar	Papua N.Guin	Suriname	Fm Yemen Ar
Burundi	El Salvador	Iceland	Malawi	Paraguay	Sweden	
Cambodia	Eq.Guinea	India	Malaysia	Peru	Switz.Liecht	

Table A.1: List of 173 countries that are included in the study. The naming is directly taken from the UN-data archive, which leads to multiple entries for single countries, that were used in different years.

- [1] G. Caldarelli, A. Capocci, P. De Los Rios, M. A. Muñoz, Scale-free networks from varying vertex intrinsic fitness, *Phys. Rev. Lett.* 89 (2002) 258702. doi:10.1103/PhysRevLett.89.258702.
URL <http://link.aps.org/doi/10.1103/PhysRevLett.89.258702>
- [2] V. D. P. Servedio, G. Caldarelli, P. Buttà, Vertex intrinsic fitness: How to produce arbitrary scale-free networks, *Phys. Rev. E* 70 (2004) 056126. doi:10.1103/PhysRevE.70.056126.
URL <http://link.aps.org/doi/10.1103/PhysRevE.70.056126>
- [3] D. Garlaschelli, T. Di Matteo, T. Aste, G. Caldarelli, M. I. Loffredo, Interplay between topology and dynamics in the world trade web, *The European Physical Journal B* 57 (2) (2007) 159–164.
- [4] T. Besedeš, T. J. Prusa, The role of extensive and intensive margins and export growth, *Journal of Development Economics* 96 (2) (2011) 371–379.

- [5] M. Riccaboni, S. Schiavo, Stochastic trade networks, IMT Lucca Eic Working Paper Series 1.
- [6] D. Garlaschelli, M. I. Loffredo, Fitness-dependent topological properties of the world trade web, *Phys. Rev. Lett.* 93 (2004) 188701. doi:10.1103/PhysRevLett.93.188701. URL <http://link.aps.org/doi/10.1103/PhysRevLett.93.188701>
- [7] D. Garlaschelli, M. I. Loffredo, Structure and evolution of the world trade network, *Physica A: Statistical Mechanics and its Applications* 355 (1) (2005) 138 – 144. doi:<http://dx.doi.org/10.1016/j.physa.2005.02.075>. URL <http://www.sciencedirect.com/science/article/pii/S0378437105002852>
- [8] R. C. Feenstra, R. E. Lipsey, H. Deng, A. C. Ma, H. Mo, World trade flows: 1962-2000, Tech. rep., National Bureau of Economic Research (2005).
- [9] M. Barigozzi, G. Fagiolo, D. Garlaschelli, Multinetwork of international trade: A commodity-specific analysis, *Phys. Rev. E* 81 (2010) 046104. doi:10.1103/PhysRevE.81.046104. URL <http://link.aps.org/doi/10.1103/PhysRevE.81.046104>
- [10] T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. i. a binary network analysis, *Phys. Rev. E* 84 (2011) 046117. doi:10.1103/PhysRevE.84.046117. URL <http://link.aps.org/doi/10.1103/PhysRevE.84.046117>
- [11] A. Fronczak, P. Fronczak, Statistical mechanics of the international trade network, *Phys. Rev. E* 85 (2012) 056113. doi:10.1103/PhysRevE.85.056113. URL <http://link.aps.org/doi/10.1103/PhysRevE.85.056113>
- [12] G. Fagiolo, J. Reyes, S. Schiavo, The evolution of the world trade web: a weighted-network analysis, *Journal of Evolutionary Economics* 20 (4) (2010) 479–514.
- [13] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, S. S. Manna, The international trade network: weighted network analysis and modelling, *Journal of Statistical Mechanics: Theory and Experiment* 2008 (02) (2008) P02002.
- [14] M. A. Serrano, M. Boguñá, Topology of the world trade web, *Phys. Rev. E* 68 (2003) 015101. doi:10.1103/PhysRevE.68.015101. URL <http://link.aps.org/doi/10.1103/PhysRevE.68.015101>
- [15] G. Fagiolo, Directed or undirected? a new index to check for directionality of relations in socio-economic networks, *Economics Bulletin* 3 (34) (2006) 1–12.
- [16] G. Fagiolo, J. Reyes, S. Schiavo, World-trade web: Topological properties, dynamics, and evolution, *Phys. Rev. E* 79 (2009) 036115. doi:10.1103/PhysRevE.79.036115. URL <http://link.aps.org/doi/10.1103/PhysRevE.79.036115>
- [17] T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. ii. a weighted network analysis, *Phys. Rev. E* 84 (2011) 046118. doi:10.1103/PhysRevE.84.046118. URL <http://link.aps.org/doi/10.1103/PhysRevE.84.046118>
- [18] F. Picciolo, T. Squartini, F. Ruzzenenti, R. Basosi, D. Garlaschelli, The role of distances in the world trade web, in: *The 8th International Conference on Signal Image Technology and Internet Based Systems (SITIS 2012)*, IEEE, Conference Publishing Services, Sorrento, Italy, 2012, pp. 784–792.
- [19] R. Chiarucci, F. Ruzzenenti, M. I. Loffredo, Detecting spatial homogeneity in the world trade web with detrended fluctuation analysis, *Physica A: Statistical Mechanics and its Applications* 401 (0) (2014) 1 – 7. doi:<http://dx.doi.org/10.1016/j.physa.2014.01.019>. URL <http://www.sciencedirect.com/science/article/pii/S0378437114000235>
- [20] M. Lawless, Deconstructing gravity: trade costs and extensive and intensive margins, *Canadian Journal of Economics/Revue canadienne d'économique* 43 (4) (2010) 1149–1172.
- [21] M. Barigozzi, G. Fagiolo, G. Mangioni, Identifying the community structure of the international-trade multi-network, *Physica A: Statistical Mechanics and its Applications* 390 (11) (2011) 2051 – 2066. doi:<http://dx.doi.org/10.1016/j.physa.2011.02.004>. URL <http://www.sciencedirect.com/science/article/pii/S0378437111001129>
- [22] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D. R. White, Economic networks: The new challenges, *Science* 325 (5939) (2009) 422–425. doi:10.1126/science.1173644.
- [23] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, D. R. White, Economic networks: What do we know and what do we need to know?, *Advances in Complex Systems* 12 (04n05) (2009) 407–422.
- [24] R. Kali, J. Reyes, The architecture of globalization: a network approach to international economic integration, *Journal of International Business Studies* 38 (4) (2007) 595–620.
- [25] D. Garlaschelli, M. I. Loffredo, Effects of network topology on wealth distributions, *Journal of Physics A: Mathematical and Theoretical* 41 (22) (2008) 224018. URL <http://stacks.iop.org/1751-8121/41/i=22/a=224018>
- [26] R. Kali, J. Reyes, Financial contagion on the international trade network, *Economic Inquiry* 48 (4) (2010) 1072–1101. doi:10.1111/j.1465-7295.2009.00249.x. URL <http://dx.doi.org/10.1111/j.1465-7295.2009.00249.x>
- [27] A. Garas, P. Argyrakis, C. Rozenblat, M. Tomassini, S. Havlin, Worldwide spreading of economic crisis, *New Journal of Physics* 12 (11) (2010) 113043. URL <http://stacks.iop.org/1367-2630/12/i=11/a=113043>
- [28] M. Boguñá, R. Pastor-Satorras, Class of correlated random networks with hidden variables, *Phys. Rev. E* 68 (2003) 036112. doi:10.1103/PhysRevE.68.036112. URL <http://link.aps.org/doi/10.1103/PhysRevE.68.036112>
- [29] Organisation for Economic Cooperation and Development, International trade by commodities statistics, Tech. rep., Mimas, University of Manchester, dOI: <http://dx.doi.org/10.5257/oced/itcs/2007> (2010).
- [30] K. S. Gleditsch, Expanded trade and gdp data, *Journal of Conflict Resolution* 46 (5) (2002) 712–724.
- [31] A. Barrat, M. Barthlemy, A. Vespignani, *Dynamical processes on complex networks*, Cambridge University Press, 2008.

- [32] N. Perra, B. Gonçalves, R. Pastor-Satorras, A. Vespignani, Activity driven modeling of time varying networks, Scientific reports 2.
- [33] K. Hoppe, G. J. Rodgers, Mutual selection in time-varying networks, Phys. Rev. E 88 (2013) 042804. doi:10.1103/PhysRevE.88.042804.
URL <http://link.aps.org/doi/10.1103/PhysRevE.88.042804>