



Universiteit
Leiden
The Netherlands

Maximum-entropy tools for economic fitness and complexity

Krantz, R.; Gemmetto, V.; Garlaschelli, D.

Citation

Krantz, R., Gemmetto, V., & Garlaschelli, D. (2018). Maximum-entropy tools for economic fitness and complexity. *Entropy*, 20(10), 743. doi:10.3390/e20100743

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/70722>

Note: To cite this publication please use the final published version (if applicable).

Article

Maximum-Entropy Tools for Economic Fitness and Complexity

Ruben Krantz ^{1,*}, Valerio Gemmetto ¹ and Diego Garlaschelli ^{1,2} 

¹ Lorentz Institute for Theoretical Physics, Leiden University, 2333 RA Leiden, The Netherlands; gemmetto@lorentz.leidenuniv.nl (V.G.); diego.garlaschelli@imtlucca.it (D.G.)

² IMT School for Advanced Studies, 55100 Lucca, Italy

* Correspondence: ruben_krantz@live.nl; Tel.: +31-6-30323147

Received: 31 July 2018; Accepted: 19 September 2018; Published: 28 September 2018

Abstract: The concepts of economic fitness and complexity, based on iterative and interdependent definitions of the quality of exporting countries and exported products, have led to novel insights into the dynamics of production and trade. A key step in the calculation of these quantities is the preliminary identification of statistically relevant country-product pairs. In this paper, we propose a method that could improve the current practice of filtering based on the revealed comparative advantage, by employing the maximum-entropy principle to construct an unbiased link weight probability distribution that, unlike the traditional thresholding method, allows for the statistical assessment of empirical trade volumes. The result is an adjusted geometric distribution for trade links that refines the revealed comparative advantage approach. This allows us to define the statistical significance of each trade link weight, leading to statistically supported trade link filtering decisions. Using this statistically justified filtering method, we have obtained results that are similar in nature to those that were found without this method, even though there are significant deviations in the details. In addition, the statistical information thus obtained on each trade link allows us to perform a spectral analysis of the export portfolio of individual economies.

Keywords: economic complexity; entropy; complex networks

1. Introduction

Economists have made many attempts at defining the competitiveness of economies on a national scale, starting with the “father of modern economy”, Adam Smith, in the wealth of nations. Approaches have ranged from attempting to point out the factors that make for a competitive economy, like Smith did, to an a posteriori analysis of the economic successfulness of a country, e.g., by measuring the gross domestic product (GDP). Thus far, none of these approaches have accomplished a comprehensive method of both explaining and measuring competitiveness. A recent attempt has been made by defining economic complexity, which explains why certain economies are more successful than others and gives a good estimate of the relative success they have.

1.1. Economic Complexity

Counter to standard economical theory, which states that while poorly developed countries specialize in exporting the least complex products, highly developed countries presumably produce solely the more complex products, the paper by Hidalgo and Hausmann [1] has shown that the latter have, instead, a highly diversified basket of export products, ranging from the most complex down to the simplest commodities. This called for a new approach, one that defines the competitiveness of a country’s economy on the diversity of the products it produces.

While acknowledging the pioneering work by Hidalgo and Hausmann [1], which in short states that a country’s “fitness” is simply determined by the sum of the “complexity” of its products

and vice versa, it serves us best in our current effort to focus on the approach developed later by Tacchella et al. [2–4], which has proven more robust as it reaches a stable solution whereas the original algorithm does not.

The conceptual reasoning behind this method is that a commodity (product) that is produced only by fit economies can be labeled complex, while those that are produced by a large number of economies, with high and low fitness, are marked as less complex. Inversely, economies that export only simple commodities are taken to be the least fit, while those that export a diversified range of complex and non-complex commodities are labeled as fit. This allows a ranking of different countries along the line of their fitness. As product complexity relies on countries' fitness and vice versa, this naturally is an iterative algorithm. Formally, it can be expressed as follows:

$$\begin{cases} \tilde{F}_i^{(n)} = \sum_{\alpha} M_i^{\alpha} Q^{\alpha(n-1)}, \\ \tilde{Q}^{\alpha(n)} = \frac{1}{\sum_i M_i^{\alpha} \frac{1}{F_i^{(n-1)}}}, \end{cases} \quad (1)$$

where both F_i and Q^{α} start with a value of 1 for all countries and commodities at the first iteration. In this definition, i represents the exporting country, while α denotes the exported commodity. The iteration number is represented by n . A normalization step after each iteration ensures the divergence of both fitness and complexity:

$$\begin{cases} F_i^{(n)} = \frac{\tilde{F}_i^{(n)}}{\langle \tilde{F}_i^{(n)} \rangle_i}, \\ Q_{\alpha}^{(n)} = \frac{\tilde{Q}^{\alpha(n)}}{\langle \tilde{Q}^{\alpha(n)} \rangle_{\alpha}}. \end{cases} \quad (2)$$

It is important to note the role of the binary bipartite country-commodity matrix M_i^{α} in these defining equations. This matrix represents the bipartite network of countries and the commodities they do or do not export (represented by 1 or 0, respectively), which is derived from the world trade network (WTN). The WTN is a multi-layered network with a layer (regular network) for each commodity α , showing how much each exporting country i trades in that commodity with importing country j . This (weighted) WTN is represented by the weighted adjacency matrix W with components $w_{i,j}^{\alpha}$. In current practice, this matrix is first summed ($w_i^{\alpha} = \sum_j w_{i,j}^{\alpha}$) and then filtered to retrieve the matrix M_i^{α} . In this paper, we will point out why the current approach is flawed at this point, and put forward an improved methodology to replace it.

1.2. Revealed Comparative Advantage: Current Practice and Flaws

In their criticism of the notions of fitness and complexity, Morrison et al. [5] point out the instability of the fitness and complexity algorithm. Their work shows that the addition or removal of a single product into the analysis can lead to significant changes in the resulting fitness of all countries, not only those who supposedly export it, leading them to question the usefulness of the algorithm. This emphasizes the importance of a well chosen filtering approach.

Thus far, the filtering of the weighted bipartite adjacency matrix to obtain its binary counterpart M_i^{α} has been limited to the straightforward application of the revealed comparative advantage (RCA), an economical concept first conceived by Balassa [6]. Conceptually, it is the share of a single country in

the total trade of a certain commodity divided by the country's share in the total world trade of all commodities. The mathematical definition of the RCA is rather straightforward:

$$RCA_i^\alpha = \frac{\frac{w_i^\alpha}{\sum_{i'} w_{i'}^\alpha}}{\frac{\sum_{\alpha'} w_i^{\alpha'}}{\sum_{i' \alpha'} w_{i'}^{\alpha'}}}. \quad (3)$$

Originally, this measure determines the relative importance of the trade in a certain commodity for a country, as compared to other countries or other commodities that the country in scope trades. In the current application, it serves to determine whether a country is a relevant exporter of a commodity. For each country and commodity, when the RCA is larger than or equal to 1, the corresponding trade link in the country-commodity matrix has a value of 1 and 0 otherwise. This is formalized by the filtering rule:

$$M_i^\alpha = \begin{cases} 1, & \text{if: } RCA_i^\alpha \geq 1, \\ 0, & \text{if: } RCA_i^\alpha < 1, \end{cases} \quad (4)$$

which, with a slight change of perspective, can also be regarded as a comparison between the real world values w_i^α and the expected value yielded by an RCA based null model for the network

$$M_i^\alpha = \begin{cases} 1, & \text{if: } w_i^\alpha \geq \langle w_i^\alpha \rangle, \\ 0, & \text{if: } w_i^\alpha < \langle w_i^\alpha \rangle, \end{cases} \quad (5)$$

where the null model's expectation values are created by equating $RCA_i^\alpha = 1$:

$$\langle w_i^\alpha \rangle = \frac{\sum_{\alpha'} w_i^{\alpha'} \cdot \sum_{i'} w_{i'}^\alpha}{\sum_{i' \alpha'} w_{i'}^{\alpha'}}. \quad (6)$$

The phrase null model in the context of graph theory encompasses a network that mimics the original network in some properties, but is randomized, or generalized, in all other. The null model is generally used for comparison to extract certain characteristics from the original network—in our current case, relevant trade links.

However, this RCA “null model” is chosen implicitly; moreover, in our opinion, this choice is not well-motivated and leads to flaws that can have implications all the way to the results of the complexity algorithm itself. Our critique is summarized by these three arguments:

- The RCA as a null model represents a fully connected or very dense network, as by the definition in Equation (6) it has a non-zero value for each i and α that have a non-zero total trade. In practice, this is the case for almost all links. In contrast, the world trade network is quite sparse with only 2–4% of all potential links realized throughout the analyzed years.
- The current definition of the RCA only applies to the bipartite network of countries and commodities, while the original world trade network contains another dimension of information, being the receiving importing country. Keeping in mind that a null model should mimic the original network, this importer dimension should also be represented in any appropriate null model—especially so, because the trade weight that the RCA null model would expect does not depend at all on the receiving country, while in reality this is of course of major importance (one would expect more trade to a country with a lot of incoming trade).
- Most importantly, the current methodology does not take into account the statistical significance of the filtered values. An RCA of over 1 could signify an important export product of a country, but could just as well be due to a statistical fluctuation through the years. This flaw is something

that Tacchella et al. also partially realized (see supplement of [7]), leading them to develop a hidden Markov model approach to binarize the country-commodity matrix to reduce this noise. We choose a different path, keeping to the original data and performing a statistical analysis to keep noise at bay.

The innovation of this paper is that we replace the current method with a null model that extends the RCA to three dimensions (exporter, importer and commodity), mimics the original network in its sparsity by controlling the probability that a link exists, and includes a probability distribution (with the expected weight and variance thereof) for each link weight in order to make statistically justified filtering choices.

2. Methodology

Given the considerations in the introduction, we aim at stepping back from the country-commodity matrix to the underlying multiplex trade layers (a multiplex network is a network consisting of multiple layers with the same nodes but different links between these nodes in each layer). Hence, we will first extend the RCA to a layer-specific version dependent on the importer, as well as the exporter and the commodity, like in the original, in Section 2.1. Thereby, we introduce a multiplex null model, for which we will develop an unbiased weight distribution around the RCA expected value. Our path towards that goal heavily relies on the maximum likelihood method as described in [8,9]. As in other recent research in complex networks concerning economics and innovation [10,11], entropy plays an important role in this approach.

In an approach analogous to statistical mechanics, the idea of this method is to use Shannon–Gibbs entropy and the Lagrange multiplier technique to establish link probabilities. Given an ensemble of graphs that satisfy a set of topological constraints linked to the original graph, this approach allows us to establish the graph in the ensemble with the highest entropy, where the concept of maximum entropy in the network context means a graph with the least possible amount of information or graph-specific patterns [8]. The maximum likelihood method is applied to find the values for the Lagrange multipliers that ensure the constrained topological features are most likely to align with the real world graph. Using the Lagrange multipliers, we can define a link weight probability distribution for every link in the WTN, allowing a statistical filtering on the weights of links, providing an improved input for the country-commodity matrix used in the fitness and complexity algorithm.

However, the world trade network requires a specific approach as it is both a multiplex and a weighted network. Therefore, we will first develop a multiplex framework building on previous work in Section 2.2, before moving on to the core of our improvements in subsequent subsections.

2.1. The Extended RCA

Our criticism that the RCA only functions as a null model for the bipartite country-commodity network can be countered by a straightforward extension to the importer dimension. When one regards the denominator in Equation (6) as the normalization that ensures that $\sum_i \langle w_i^\alpha \rangle = \sum_i w_i^\alpha$ and $\sum_\alpha \langle w_i^\alpha \rangle = \sum_\alpha w_i^\alpha$, the obvious expression for the expectation value $\langle w_{i,j}^\alpha \rangle$ becomes

$$\langle w_{i,j}^\alpha \rangle = \frac{\sum_{j',\alpha'} w_{i,j'}^{\alpha'} \cdot \sum_{i',\alpha'} w_{i',j}^{\alpha'} \cdot \sum_{i',j'} w_{i',j'}^\alpha}{(\sum_{i',j',\alpha'} w_{i',j'}^{\alpha'})^2}. \quad (7)$$

This extended RCA can function as a null model for the weights of the world trade network in its full detail.

2.2. Extension to the Multiplex Network

The maximum likelihood method builds upon the concept of maximum entropy. This requires us to extend the graph's Shannon–Gibbs entropy and with that, the graph probability, from a regular

single layer network to a multiplex network. This was previously developed and described in [12]. We will express the probability that a multiplex network with the set of adjacency matrices $\{W^\alpha\}$ exists ($\mathcal{P}(\{W^\alpha\})$) in terms of the single layer network probability $P^\alpha(W^\alpha)$, starting off by extending the Hamiltonian:

$$\mathcal{H}(\{W^\alpha\}) = \sum_{\alpha} H^\alpha(W^\alpha), \tag{8}$$

$$\mathcal{S} = - \sum_{\{W^\alpha\}} P(\{W^\alpha\}) \ln P(\{W^\alpha\}). \tag{9}$$

Now, if the probability on a single layer in the multiplex is:

$$P^\alpha(W^\alpha) = \frac{e^{-\mathcal{H}(\{W^\alpha\})}}{Z} \tag{10}$$

(where Z is the partition function: $Z = \sum_{\{W^\alpha\}} e^{-\mathcal{H}(\{W^\alpha\})}$). Then, the full multiplex graph probability is:

$$\begin{aligned} \mathcal{P}(\{W^\alpha\}) &= \prod_{\alpha} P^\alpha(W^\alpha) \\ &= \frac{\prod_{\alpha} e^{-\mathcal{H}(\{W^\alpha\})}}{Z}. \end{aligned} \tag{11}$$

2.3. Link Weight Probability Distribution

With the formalities out of the way, we can move on to our goal of expressing a link weight probability distribution $q_{i,j}^\alpha(w_{i,j}^\alpha)$ that makes up the components of the (multiplex) graph probability

$$\mathcal{P}(\{W^\alpha\}) = \prod_{i,j,\alpha} q_{i,j}^\alpha(w_{i,j}^\alpha). \tag{12}$$

Considering the fact that the weight of a trade link can theoretically range from zero to infinity, the parallel with Fermi statistics or a geometric probability distribution immediately comes to mind. We believe that this would be oversimplifying the practical situation, as it is hard to defend the viewpoint that the first traded weight has the same probability linked to it as all the following. We therefore follow [13] in their generalization of Bose and Fermi statistics, leading us to a modified geometric distribution that takes into account the difficulty of establishing the first trade link:

$$q_{i,j}^\alpha(w_{i,j}^\alpha) = \begin{cases} p_{i,j}^\alpha (r_{i,j}^\alpha)^{w_{i,j}^\alpha - 1} \cdot (1 - r_{i,j}^\alpha), & \text{if: } w_{i,j}^\alpha > 0, \\ 1 - p_{i,j}^\alpha, & \text{if: } w_{i,j}^\alpha = 0. \end{cases} \tag{13}$$

In this distribution, $p_{i,j}^\alpha$ is the probability of establishing the link in the first place and $r_{i,j}^\alpha$ is the probability of adding a unit of weight.

This leaves us with the task of finding expressions for $p_{i,j}^\alpha$ and $r_{i,j}^\alpha$ in order to complete the link weight probability distribution (and be able to find the statistical significance of each actual link). We have developed three distinct approaches to finding a reasonable link existence probability ($p_{i,j}^\alpha$):

- Directed binary configuration model (DBCM),
- Multiplex directed binary configuration model (MDBCM),
- Strength-replaced MDBCM.

2.4. The Directed Binary Configuration Model

For all of these approaches, we follow the same basic rules of the maximum likelihood method described in [8], in order to arrive at the directed binary configuration model (DBCM) expression for

$p_{i,j}^\alpha$ first. The rather straightforward extension to the MDBCM and the strength-replaced MDBCM will be discussed in the following subsections.

The maximum likelihood method is applied to construct an unbiased ensemble of all possible graphs $\{\mathbf{G}\}$ that resemble the original \mathbf{G}^* in a predefined topological property. As constructing a micro-canonical ensemble, where all the topological constraints are met exactly, can only be done numerically and not analytically, we opt for the computationally faster constructed canonical ensemble, where the expectation values of the topological constraints meet the real world originals. Thus, instead of the strict $\vec{C}(\{\mathbf{G}\}) = \vec{C}(\mathbf{G}^*)$ (i.e., requiring that the topological properties \vec{C} of all graphs in ensemble \mathbf{G} are equal to those in the original graph), we require only that:

$$\langle \vec{C}(\{\mathbf{G}\}) \rangle = \vec{C}(\mathbf{G}^*). \tag{14}$$

In this subsection, we will develop the DBCM, which can be defined as the canonical ensemble constructed using the the in- and out-degree—the simplest first order topological property—of all nodes of only a single layer of the full network as constraint. This means that, in the current application of the WTN, the DBCM will be applied layer for layer. For our current purposes, we limit ourselves to a binary representation of a graph, constraining degrees instead of strengths. Thus, the sparsity of the original graph will be conserved:

$$\langle \vec{k}(\{\mathbf{G}\}) \rangle = \vec{k}(\mathbf{G}^*) \tag{15}$$

or more specifically, for directed graphs like the WTN:

$$\langle \vec{k}_{in/out}(\{\mathbf{G}\}) \rangle = \vec{k}_{in/out}(\mathbf{G}^*), \tag{16}$$

where \vec{k} is the vector containing degrees k_i of all nodes i .

To ensure a completely unbiased randomization of the canonical ensemble, we require that the probability $\mathcal{P}(\mathbf{G})$ that a graph \mathbf{G} exists maximizes the Shannon–Gibbs entropy, subject to the constraint defined above (Equation (16)) and enforcing a normalized probability distribution:

$$S \equiv - \sum_{\mathbf{G}} \mathcal{P}(\mathbf{G}) \ln \mathcal{P}(\mathbf{G}), \quad \text{where:} \tag{17}$$

$$\sum_{\mathbf{G}} \mathcal{P}(\mathbf{G}) = 1.$$

These requirements can be met by introducing into the graph probability expression a set of Lagrange multipliers $\vec{\theta} = \{\theta_a\}$ and $\vec{\phi} = \{\phi_a\}$ that enforce the ensemble constraints (Equation (16)). The general expression for the graph probability then becomes

$$\mathcal{P}(\mathbf{G}|\vec{\theta}, \vec{\phi}) = \frac{e^{-H(\mathbf{G}|\vec{\theta}, \vec{\phi})}}{Z(\vec{\theta}, \vec{\phi})}, \tag{18}$$

where the Hamiltonian is defined as the product of the Lagrange multipliers and the constraints

$$H(\mathbf{G}|\vec{\theta}, \vec{\phi}) \equiv \sum_{i,j} \theta_i k_{i,out}(\mathbf{G}) + \phi_j k_{j,in}(\mathbf{G}) = \vec{\theta} \cdot \vec{k}_{out}(\mathbf{G}) + \vec{\phi} \cdot \vec{k}_{in} \tag{19}$$

and the partition function $Z(\vec{\theta})$ normalizes the probability distribution

$$Z(\vec{\theta}, \vec{\phi}) = \sum_{\mathbf{G}} e^{-H(\mathbf{G}|\vec{\theta}, \vec{\phi})} \tag{20}$$

and, finally, a topological property's expectation value can be expressed using the graph probability

$$\langle X \rangle_{\vec{\theta}, \vec{\phi}} \equiv \sum_{\mathbf{G}} X(\mathbf{G}) \mathcal{P}(\mathbf{G} | \vec{\theta}, \vec{\phi}). \tag{21}$$

The key step to fit the model that is defined by Equation (18) to the real world graph \mathbf{G}^* is to tune the Lagrange multipliers so that the likelihood of retrieving that original graph is maximized. Following [14], this is achieved by setting

$$\langle \vec{C}(\{\mathbf{G}\}) \rangle_{\vec{\theta}^*, \vec{\phi}^*} = \sum_{\mathbf{G}} \vec{C}(\mathbf{G}) \mathcal{P}(\mathbf{G} | \vec{\theta}, \vec{\phi}) = \vec{C}(\mathbf{G}^*), \tag{22}$$

which, in our configuration model approach, leads to fixing:

$$\begin{cases} \sum_{\mathbf{G}} \vec{k}_{out}(\mathbf{G}) \mathcal{P}(\mathbf{G} | \vec{\theta}, \vec{\phi}) = \vec{k}_{out}(\mathbf{G}^*), \\ \sum_{\mathbf{G}} \vec{k}_{in}(\mathbf{G}) \mathcal{P}(\mathbf{G} | \vec{\theta}, \vec{\phi}) = \vec{k}_{in}(\mathbf{G}^*). \end{cases} \tag{23}$$

The key to establishing the Lagrange multipliers using the above expression is factorizing the graph probability into local components $\mathcal{P}(\mathbf{G} | \vec{\theta}, \vec{\phi}) = \prod_{i,j} P_{ij}(g_{i,j} | \vec{\theta}, \vec{\phi})$, by rewriting the Hamiltonian:

$$\begin{aligned} H(\mathbf{G} | \vec{\theta}, \vec{\phi}) &= \sum_{i,j} \theta_i k_{i,out}(\mathbf{G}) + \phi_j k_{j,in}(\mathbf{G}) \\ &= \sum_{i,j} (\theta_i + \phi_j) g_{i,j}, \end{aligned} \tag{24}$$

which leads to:

$$\begin{aligned} \mathcal{P}(\mathbf{G} | \vec{\theta}, \vec{\phi}) &= \frac{e^{-\sum_{i,j} (\theta_i + \phi_j) g_{i,j}}}{\sum_{\mathbf{G}} e^{-\sum_{i,j} (\theta_i + \phi_j) g_{i,j}}} \\ &= \frac{\prod_{i,j} (e^{-\theta_i - \phi_j})^{g_{i,j}}}{\prod_{i,j} (1 + e^{-\theta_i - \phi_j})} \end{aligned}$$

(as $g_{i,j}$ is either 0 or 1 in a binary graph). Then, substituting the Lagrange multiplier exponentials with what we would call "hidden variables" $x_i = e^{-\theta_i}$ and $y_j = e^{-\phi_j}$

$$\begin{aligned} &= \prod_{i,j} \left(\frac{(x_i y_j)^{g_{i,j}}}{1 + x_i y_j} \right) \\ &= \prod_{i,j} \left(\frac{(x_i y_j)^{g_{i,j}}}{1 + x_i y_j} \cdot \left(\frac{1 + x_i y_j}{1 + x_i y_j} \right)^{(g_{i,j}-1)} \right) \\ &= \prod_{i,j} \left(\left(\frac{x_i y_j}{1 + x_i y_j} \right)^{g_{i,j}} \cdot \left(1 - \frac{x_i y_j}{1 + x_i y_j} \right)^{(1-g_{i,j})} \right) \\ &= \prod_{i,j} p_{i,j}^{g_{i,j}} (1 - p_{i,j})^{(1-g_{i,j})} \end{aligned} \tag{25}$$

$$= \prod_{i,j} P_{ij}(g_{i,j} | \vec{\theta}, \vec{\phi}), \tag{26}$$

where $p_{i,j} = \frac{x_i y_j}{1 + x_i y_j}$ represents the probability that a link from i to j is present and therefore the success probability in the Bernoulli probability function $P_{ij}(g_{i,j} | \vec{\theta}, \vec{\phi})$.

We can now rewrite Equation (23) using the factorized version of $\mathcal{P}(\mathbf{G}|\vec{\theta}, \vec{\phi})$ to get to expressions that we can use to numerically determine x_i and y_j :

$$\langle g_{i,j} \rangle = \sum_g g_{i,j} p_{i,j} = p_{i,j}, \tag{27}$$

leading to:

$$\begin{cases} \langle \vec{k}_{i,out}(\mathbf{G}) \rangle = \sum_j p_{i,j} = \vec{k}_{i,out}(\mathbf{G}^*), \\ \langle \vec{k}_{j,in}(\mathbf{G}) \rangle = \sum_i p_{i,j} = \vec{k}_{j,in}(\mathbf{G}^*). \end{cases} \tag{28}$$

This is the mechanism behind the maximum likelihood method to build a statistically unbiased null model based on a single real world example, for a single layer, binary, directed network. However, the goal of this section was to find an expression for the link presence probability $p_{i,j}^\alpha$, for the link i, j in each layer α of the multiplex WTN, which was required in Equation (13). The key lies in simply solving the equivalent of Equation (28) generalized for the multiplex network numerically:

$$\begin{cases} \sum_j p_{i,j}^\alpha = \vec{k}_{i,out}^\alpha(\mathbf{G}^*), \\ \sum_i p_{i,j}^\alpha = \vec{k}_{j,in}^\alpha(\mathbf{G}^*), \end{cases} \quad \text{where: } p_{i,j}^\alpha = \frac{x_i^\alpha y_j^\alpha}{1 + x_i^\alpha y_j^\alpha}. \tag{29}$$

2.5. Multiplex Directed Binary Configuration Model

By construction, the DBCM is applied to each individual layer of the multiplex WTN separately. This implies that the null model that is thus created relies—as far as the expression for $p_{i,j}^\alpha$ goes—only on characteristics of that layer. This implies that only the trade activities of each country within a certain commodity are considered, while any general, inter-layer characteristics of a country in general are ignored. On top of that, solving the set of Equations (29) for all layers means finding $2n \cdot l$ hidden variables (where n = number of nodes and l = number of layers), whereas this can be limited to $2n + l$ in the MDBCMM.

With the goal of treating the WTN as a whole, instead of a large number of single layer networks, we can also construct a null model that takes cross-layer properties into account. This is achieved in the multiplex directed binary configuration model (MDBCMM) by setting a whole new set of constraints to the configuration model, this time including the total amount of links in each layer (\vec{L}^α , referred to as the “layer degree”). Also note that the constrained out- and in-degrees are now replaced by their totals over all layers α :

$$\begin{cases} \langle \vec{k}_{i,out}^{tot}(\mathbf{G}) \rangle = \vec{k}_{i,out}^{tot}(\mathbf{G}^*) = \sum_{j,\alpha} g_{i,j}^\alpha, \\ \langle \vec{k}_{j,in}^{tot}(\mathbf{G}) \rangle = \vec{k}_{j,in}^{tot}(\mathbf{G}^*) = \sum_{i,\alpha} g_{i,j}^\alpha, \\ \langle \vec{L}^\alpha(\mathbf{G}) \rangle = \vec{L}^\alpha(\mathbf{G}^*) = \sum_{i,j} g_{i,j}^\alpha. \end{cases} \tag{30}$$

In general, we can follow the steps described in the previous section for the DBCM with just minor adjustments in the Hamiltonian, which will eventually lead to a new set of defining expressions for $p_{i,j}^\alpha$

$$H(\mathbf{G}|\vec{\theta}, \vec{\phi}, \vec{\zeta}) = \sum_{i,j,\alpha} \theta_i k_{i,out}^{tot}(\mathbf{G}) + \phi_j k_{j,in}^{tot}(\mathbf{G}) + \zeta^\alpha L^\alpha(\mathbf{G}) \tag{31}$$

$$= \sum_{i,j,\alpha} (\theta_i + \phi_j + \zeta^\alpha) g_{i,j}^\alpha. \tag{32}$$

Following the same steps as in the above derivation for the DBCM and substituting the Lagrange multipliers with the hidden variables $x_i = e^{-\theta_i}$, $y_j = e^{-\phi_j}$ and $z^\alpha = e^{-\zeta^\alpha}$ as before, we find:

$$p_{i,j}^\alpha = \frac{x_i y_j z^\alpha}{1 + x_i y_j z^\alpha}. \tag{33}$$

The construction of the MDBCM is then limited to the simple operation of numerically solving the following set of defining equations, in parallel to Equation (29):

$$\begin{cases} \sum_{j,\alpha} p_{i,j}^\alpha = \vec{k}_{i,out}^{tot}(\mathbf{G}^*), \\ \sum_{i,\alpha} p_{i,j}^\alpha = \vec{k}_{j,in}^{tot}(\mathbf{G}^*), \\ \sum_{i,j} p_{i,j}^\alpha = \vec{L}^\alpha(\mathbf{G}^*), \end{cases} \tag{34}$$

which clearly indicates the dependence of $p_{i,j}^\alpha$ on the total out- and in-degrees as well as the layer degrees. The former two, as required at the outset of the MDBCM, represent the network wide characteristics of individual nodes, while the latter represents inter-layer variations.

2.6. Strength-Replaced MDBCM

The final implementation of the maximum likelihood method that we have developed builds upon the MDBCM, instead of starting of with a new set of constraints. As the name suggests, we will replace two of the hidden variables by strengths. This is common practice in the use of the configuration model, where hidden variables are often replaced by some sort of “fitness” of the nodes. (Note that in this sense fitness means a measure of the performance of a node from a network theory point of view. This use of the term fitness is inherited from previous work [15,16]).

In our case, we will replace the Lagrange multiplier x_i by the total out-strength of the corresponding node i , and y_j by the total in-strength of node j

$$x_i \rightarrow s_{i,out}^{tot}(\mathbf{G}^*) \equiv \sum_{j,\alpha} w_{i,j}^\alpha(\mathbf{G}^*), \tag{35}$$

$$y_j \rightarrow s_{j,in}^{tot}(\mathbf{G}^*) \equiv \sum_{i,\alpha} w_{i,j}^\alpha(\mathbf{G}^*), \tag{36}$$

$$p_{i,j}^\alpha = \frac{s_{i,out}^{tot} s_{j,in}^{tot} z^\alpha}{1 + s_{i,out}^{tot} s_{j,in}^{tot} z^\alpha}. \tag{37}$$

Besides this replacement, there are no changes with respect to the previously discussed MDBCM. An advantage of this method is that it would be easier for laymen to understand and that it is simpler and theoretically faster to solve numerically, with only one set of hidden variables. For the scope of this paper, we will focus on the more rigorous MDBCM.

2.7. The Weight Unit Probability

So far, we have covered one out of two missing components of the geometric link weight probability distribution (Equation (13)), while $r_{i,j}^\alpha$ remains to be defined. We will exploit the choice of the extended RCA as the weight expectation value (Equation (7)) to solve Equation (13) for $r_{i,j}^\alpha$, while keeping $p_{i,j}^\alpha$ generic as we have several possible definitions for it:

$$\begin{aligned}
\langle w_{i,j}^\alpha \rangle &= \sum_{w \geq 0} q_{i,j}^\alpha(w_{i,j}^\alpha) \cdot w_{i,j}^\alpha & (38) \\
&= p_{i,j}^\alpha \cdot \sum_{w \geq 0} (r_{i,j}^\alpha)^{w_{i,j}^\alpha - 1} \cdot (1 - r_{i,j}^\alpha) \cdot w_{i,j}^\alpha \\
&= \frac{p_{i,j}^\alpha \cdot (1 - r_{i,j}^\alpha)}{r_{i,j}^\alpha} \sum_{w > 0} (r_{i,j}^\alpha)^{w_{i,j}^\alpha} \cdot w_{i,j}^\alpha \\
&= \frac{p_{i,j}^\alpha \cdot (1 - r_{i,j}^\alpha)}{r_{i,j}^\alpha} \cdot \frac{r_{i,j}^\alpha}{(r_{i,j}^\alpha - 1)^2} \\
&= \frac{p_{i,j}^\alpha}{1 - r_{i,j}^\alpha}
\end{aligned}$$

and therefore:

$$r_{i,j}^\alpha = 1 - \frac{p_{i,j}^\alpha}{\langle w_{i,j}^\alpha \rangle}. \quad (39)$$

2.8. Statistical Significance

Now, we can return to our initial goal of replacing the crude RCA filtering method with one that gives a statistical justification. To that aim, we need the statistical significance of each real world link, as compared with the link weight probability distribution derived above. We will express this in z-scores: the number of standard deviations σ a weight deviates from the expected weight. Instead of filtering on $RCA \geq 1$, we can from then on apply the filter $z \geq \tau$, where τ is a chosen threshold. The values of the country-commodity matrix remain 1 for successful filtering and 0, otherwise, as was the case with the RCA.

As we intend to replace the bipartite country-commodity RCA as a filter applied before the fitness and complexity algorithm, we will limit ourselves to the bipartite z-score instead of finding the full three-dimensional version—which is trivial after the following derivation of the bipartite one:

$$\begin{aligned}
z(w_i^\alpha) &= \frac{w_i^\alpha - \langle w_i^\alpha \rangle}{\sigma(w_i^\alpha)} \\
&= \frac{\sum_j w_{i,j}^\alpha - \sum_j \langle w_{i,j}^\alpha \rangle}{\sqrt{\sum_j \sigma^2(w_{i,j}^\alpha)}}, & (40)
\end{aligned}$$

requiring us to find an expression for the variance:

$$\begin{aligned}
\sigma^2(w_{i,j}^\alpha) &= \langle (w_{i,j}^\alpha)^2 \rangle - \langle w_{i,j}^\alpha \rangle^2, & (41) \\
\langle (w_{i,j}^\alpha)^2 \rangle &= \sum_{w \geq 0} q_{i,j}^\alpha(w_{i,j}^\alpha) \cdot (w_{i,j}^\alpha)^2 \\
&= \frac{p_{i,j}^\alpha \cdot (1 - r_{i,j}^\alpha)}{r_{i,j}^\alpha} \sum_{w > 0} (r_{i,j}^\alpha)^{w_{i,j}^\alpha} \cdot (w_{i,j}^\alpha)^2 \\
&= \frac{p_{i,j}^\alpha \cdot (1 - r_{i,j}^\alpha)}{r_{i,j}^\alpha} \cdot \frac{r_{i,j}^\alpha (r_{i,j}^\alpha - 1)}{(r_{i,j}^\alpha - 1)^3} \\
&= \langle w_{i,j}^\alpha \rangle \cdot \frac{1 + r_{i,j}^\alpha}{1 - r_{i,j}^\alpha}
\end{aligned}$$

$$= \frac{2\langle w_{i,j}^\alpha \rangle^2}{p_{i,j}^\alpha} - \langle w_{i,j}^\alpha \rangle, \quad (42)$$

thus:

$$\sigma^2(w_{i,j}^\alpha) = \frac{2 - p_{i,j}^\alpha}{p_{i,j}^\alpha} \langle w_{i,j}^\alpha \rangle^2 - \langle w_{i,j}^\alpha \rangle. \quad (43)$$

2.9. Practical Implementation

Now that we have formally derived all the necessary components of the null model: the extended version of the RCA, the link weight probability distribution and the expression for the z-score that follows from it, we will briefly describe the practical implementation of the complete method in a comprehensive list:

1. Calculate $\langle w_{i,j}^\alpha \rangle$ for each link using the extended RCA as defined in Equation (7).
2. Find the hidden variables—and with that, $p_{i,j}^\alpha$ —applying either the DBCM or the (regular or strength-replaced) MDBCM, by solving Equations (29) or (34), respectively.
3. Combine $\langle w_{i,j}^\alpha \rangle$ and $p_{i,j}^\alpha$ in Equation (43) to find the variance on each link i, j, α .
4. Use Equation (40) to find the z-score of each country-commodity pair i, α and filter all links in the bipartite network using a threshold on the z-score (typically $z \geq 1$, $z \geq 2$ or similar).
5. Apply the fitness and complexity algorithm as developed Tacchella et al. in [2].

3. Results

In this section, we will show that even though we propose a change to current practice that requires theoretical and methodological adjustments, there are no fundamental implications for the results. However, while they remain similar in nature, there are obvious discrepancies between the results of the original and our statistically justified approach—which tells us that it was, indeed, necessary to develop this more rigorous method. Furthermore, we will show a new set of results that are a direct outcome of our new approach, which can tell us more about the performance of each individual economy on a detailed level. Note that more details about the datasets used in this research are listed in Section 5. An important detail to mention here is that we have opted to apply the MDBCM method for finding $p_{i,j}^\alpha$, as this has proven to be faster than the DBCM and takes cross-layer patterns into account. We have chosen it over its strength replaced counterpart for the present, as it is theoretically the most profound of the two, leaving a thorough comparison for later concern.

3.1. Comparison with Previous Results

We have two goals in this comparison with the results that were achieved in [2–4] among others: firstly, we will shortly show that we have been able to reproduce their results (by applying a $z \geq 0$ filter, which is equal to the original RCA filtering method) and secondly we will show that, with the introduction of larger z-score threshold, the complexity algorithm still yields results of the same nature—while, naturally, we will also highlight the discrepancies.

3.1.1. Evolution of Fitness and Complexity

An indication of the improved working of the algorithm proposed by Tacchella et al. [2–4] is the evolution of the fitness and complexity indicators throughout the iterations of the algorithm. The graph in Figure 1 shows a clear divergence of fitness in the earlier iterations, while it converges to a fixed point for all countries in the end. Similar behavior is observed for complexity. This indicates the correct reproduction of their algorithm.

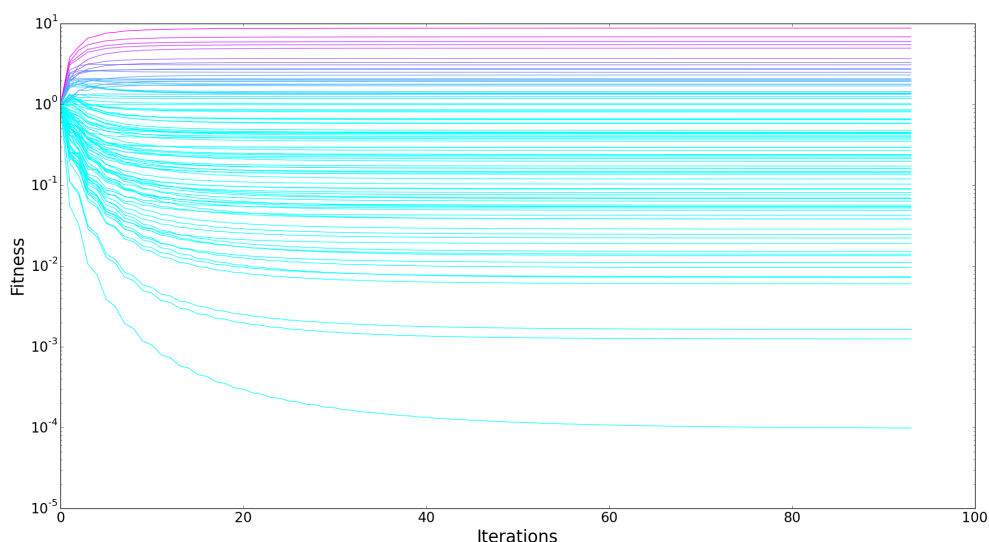


Figure 1. Evolution of country fitnesses throughout the iterations of the fitness and complexity algorithm both clearly show a convergence to a fixed point (using $z \geq 0$ for exact replication of revealed comparative advantage (RCA) filtering).

3.1.2. Ranking Countries by Fitness

One of the main results presented in [2–4] is the ranking of countries according to their fitness. Note that the absolute fitness is of no real interest due to the normalization in each iteration of the algorithm. We have compared a top 10 ranking as found earlier using data from 2010 and the ranking we have produced using the 2012 data, as shown in the table below (Table 1). The discrepancies can largely be attributed to two differences: obviously the different years in scope, and the higher granularity of the data we have used. This table shows that, even though discrepancies are expected, the results are rather similar in nature.

Table 1. A comparison of the fitness ranking results [2–4] for the year 2010 and our replication for 2012. Note the differences in the bottom half of the top 10, which can be attributed to the different years that are analyzed and the different datasets.

Tacchella et al. (2010)	Replication (2012)
Germany	Germany
China	China
Italy	Italy
Japan	Japan
USA	USA
France	Belgium
UK	France
Austria	Netherlands
Spain	India
Belgium	UK

3.1.3. Correlation with GDP per Capita

To conclude the comparison with the original results, we show that we have reproduced the global correlation that was found between the fitness of each country and its GDP per capita in Figure 2. We will not go into the supposed significance of this correlation, which in the first place was disputed

in [5] and could be further questioned after a comparison with similar plots using higher z-score filters, as we will show in the following section.

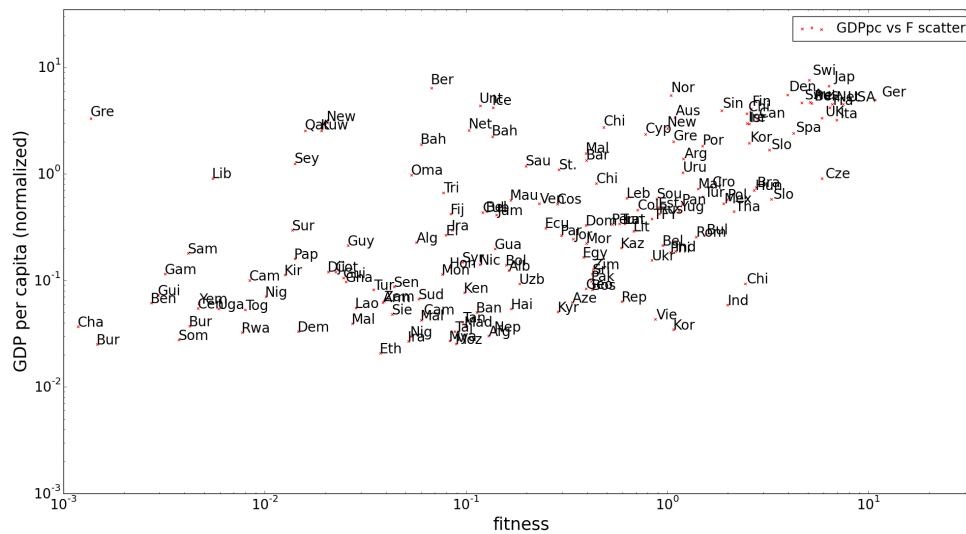


Figure 2. The normalized gross domestic product (GDP) per capita and fitness in 1995 are correlated with a correlation coefficient of 0.64 in the reproduction of the original (RCA filtered) results, with a standard error of 0.052 (which is substantial on this scale).

3.2. Results with Filtering on Higher Statistical Significance

We will cover the new results we obtained using a higher statistical significance threshold in the filtering procedure ($z \geq a$ where a is an integer larger than 0) in this subsection, emphasizing that although some discrepancies between the original results and these statistically justified ones do occur, they remain similar in nature. We would like to point out some practical implications of a stricter filtering procedure first, as these will help to clarify any different results. Naturally, when we take a higher filtering threshold, say $z \geq 1$ instead of $z \geq 0$, more of the links of the original network will be left out of the resulting matrix M_i^a , which will become even sparser. This is can be seen in Figure 3, showing that there is a sharper drop in the out-degrees of the exporters as a smaller fraction of the commodities they export are passed through the filter.

The impact of the increased threshold is more practically illustrated by looking into the products that are allowed by one threshold value, but denied by another. For example, in 2004, the Netherlands had a z-score greater than zero (i.e., RCA filtering) for the products in Table 2 (and many others), but these did not pass a $z \geq 1$ filter. Some of these products would have been an important part of Dutch export several decades ago, but are clearly no longer as relevant (e.g., fish and electric lamps). Others are harder to interpret.

Table 2. Some examples of commodities with $0 \leq z \leq 1$ for the Netherlands in 2004. These are allowed to pass through the original revealed comparative advantage (RCA) filter, but are denied by any filter with $z \geq 1$.

Fish (fresh or chilled)
Peas
Rubber inner tyre tubes
Sacks and bags of jute
Compacting machinery
Resistance welding machines
Electric lamps and light fittings

We will see that the impact of this sparsifying of M_i^x varies per country. Based on trivial statistics, one could expect that eliminating a certain fraction of all links for countries with a small original amount of trade links would lead to more extreme variations than eliminating that same fraction from countries highly connected countries. This intuition proves right, as we see that, for different z-score thresholds, the resulting top 10 countries in fitness ranking is relatively stable, while larger variations occur while going down the ranking. This is clearly shown in Table 3. Countries that perform steadily while we raise the z-score threshold thereby prove to have a more robust economy. This “robustness” can be visualized, something we will cover in Section 3.3.

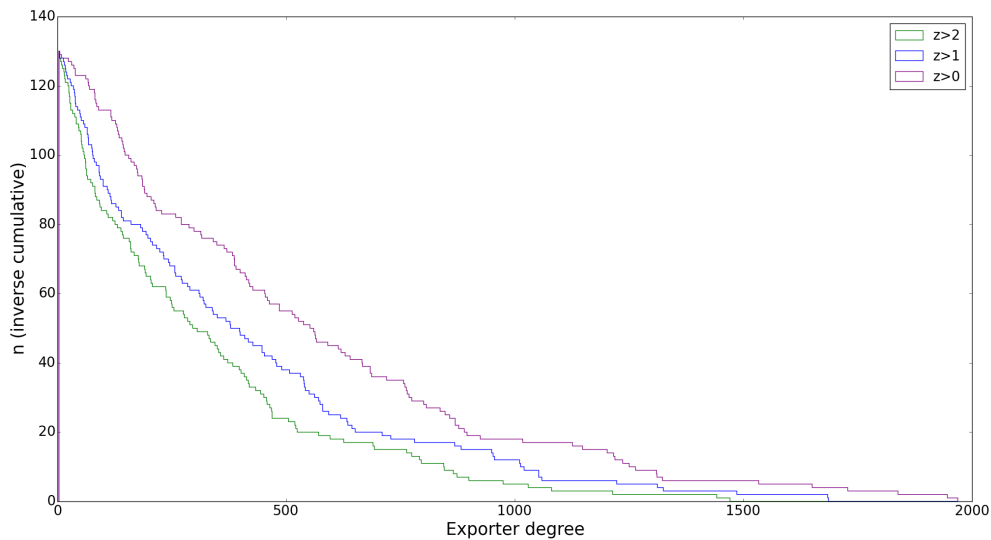


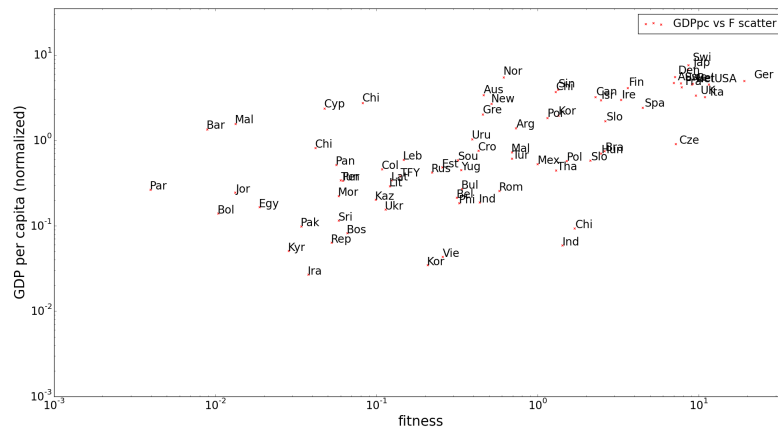
Figure 3. The inverse cumulative degree distribution of exporting countries in 2012 after filtering with different z-score thresholds clearly exposes the increased sparseness of the network after filtering.

Table 3. Ranking of countries according to their fitness, in 2012.

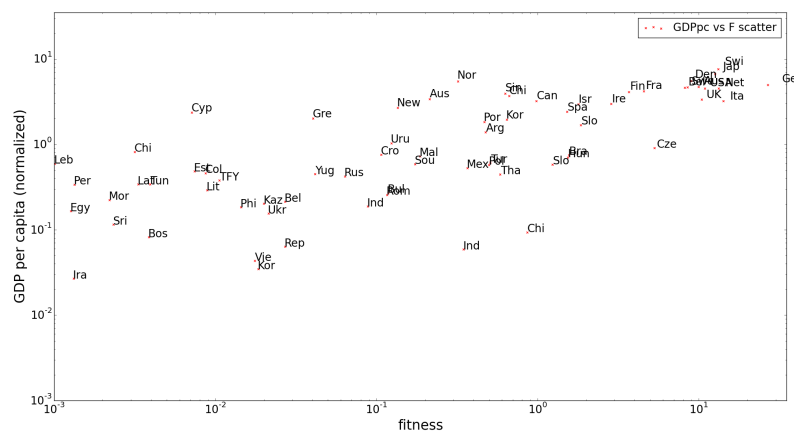
$z \geq 0$	$z \geq 1$	$z \geq 2$
China	China	China
Germany	Germany	Germany
USA	USA	Italy
Japan	Japan	Japan
Italy	Italy	USA
India	Belgium	Belgium
Belgium	India	India
France	France	France
Netherlands	Netherlands	Netherlands
Spain	Spain	UK
UK	UK	Spain
Hong Kong	Hong Kong	Switzerland
Switzerland	Switzerland	Hong Kong
Czech Republic	Austria	Austria
Austria	Czech Republic	Czech Republic
South Korea	South Korea	South Korea
Sweden	Sweden	Sweden
Poland	Turkey	Thailand
Turkey	Thailand	Denmark
Denmark	Malaysia	Turkey
Thailand	Denmark	Singapore
Malaysia	Poland	Malaysia

These effects are also visible in the plots showing the correlation between GDP per capita and fitness of countries. Again, the top-ranked countries remain more or less stable, while countries with low initial fitness at $z \geq 0$ filtering show major variations for $z \geq 1$ and $z \geq 2$ in Figure 4. These plots point out once again the instability of the fitness and complexity algorithm as reported in [5].

A remarkable change is the much lower standard error of the correlation between GDP per capita and fitness, which can be interpreted as a justification of our approach. This also means that in an attempt to improve GDP growth forecasts using fitness, as in [7], applying the methods described in this paper could improve results.



(a) $z \geq 1$ filtered, correlation coefficient: 0.66, standard error: 0.034.



(b) $z \geq 2$ filtered, correlation coefficient: 0.64, standard error: 0.028.

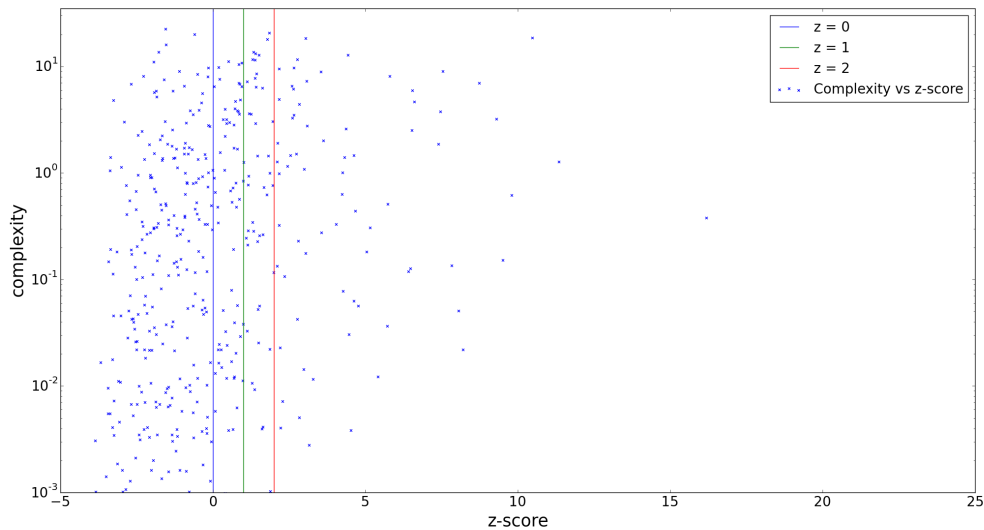
Figure 4. Normalized GDP per capita versus fitness (1995) plots show a similar correlation for higher z -score filtering thresholds. The standard error of the correlation decreases remarkably with higher thresholds (compared to 0.052 for $z \geq 0$). (a) $z \geq 1$ filtered, correlation coefficient: 0.66, standard error: 0.034; (b) $z \geq 2$ filtered, correlation coefficient: 0.64, standard error: 0.028.

3.3. Additional Results: z -Score Spectra

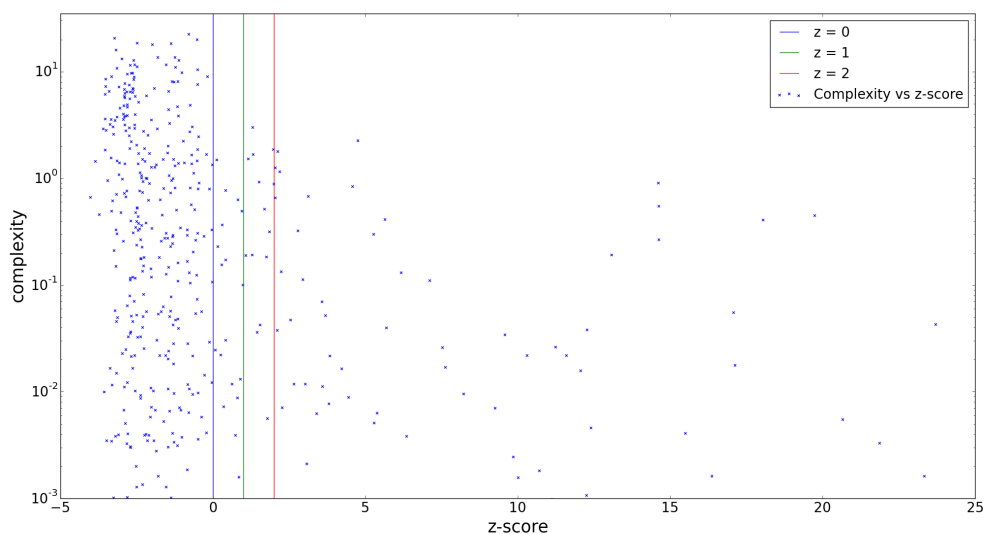
As a by-product of the methodology proposed in this paper, we have obtained a new type of results during our research. In a more detailed analysis of the performance of a single economy, it can be enlightening to plot the “spectrum” of the countries commodities, with their z -scores and complexities on the x - and y -axes, respectively.

These plots immediately yield a lot of information about the diversity of a country’s export portfolio, as can be seen in Figure 5. A diverse portfolio in this graph is, rather intuitively, represented

by large numbers of complex and less complex commodities on the high z-score end of the spectrum, like the USA in 1995, while China in that same year shows a strongly left-leaning spectrum with only simpler products at the right end. From these spectra, it is obvious that the USA's economy is more "robust" in the sense that it would be impacted less by a raised z-score threshold than China would.



(a) USA in 1995, with a relatively large number of high complexity commodities with $z \geq 2$



(b) China in 1995, with a most high complexity commodities just under $z = 0$

Figure 5. Commodity complexity spectra, showing the export product baskets of the USA and China in terms of their complexity and z-score. (a) USA in 1995, with a relatively large number of high complexity commodities with $z \geq 2$; (b) China in 1995, with a most high complexity commodities just under $z = 0$.

One of the most interesting features of these spectra is that they can convey a country's potential. When one finds a great number of complex commodities in a spectrum just under the usual z-score threshold, this could indicate that, within a reasonable time, many of these will be added to the significant export basket of this country. The plot of China's export portfolio in Figure 5 is a good

example of such a spectrum. Potentially, an analysis of the temporal evolution of such spectra could be very informative. Moreover, these spectra could point countries to what products they should invest in to improve their economical fitness (those with high complexity and a low z-score)—something, as was suggested by [4] that could in turn lead to an improvement in GDP.

4. Discussion

The above leaves a couple of items open to discuss, including a number of caveats and some potential research tangents. One is the nature of the data that we have used (for a detailed description of the data, please refer to Section 5). Firstly, the datasets only convey information on the traded commodities that an economy exports and therefore leave out all internally consumed products by definition. This is an important consideration, although one could argue that when a country really is a significant producer of a commodity, it will in most cases also export it.

A second consideration regarding the use of world trade data is that only physical products are taken into account. The whole contribution of the services sector to the global economy (around 64%) is neglected in this analysis. We will leave it to economists to estimate the importance of this neglect.

Continuing to the fitness and complexity algorithm, we have one consideration that we deem worth sharing. It could be argued that there is a slight circular reasoning to first measure the algorithm's performance by comparison of the resulting fitness with GDP and then investigating the correlation between the both of them.

Another problem arises when choosing the configuration model to construct $p_{i,j}^{\alpha}$ with (DBCM, MDBCM or its strength-replaced counterpart). So far, we have not been able to identify a single measure to tell which model is most successful and should therefore be applied. Comparing results with GDP is not necessarily a good approach, as we are not necessarily trying to imitate it. We have opted for the (regular) MDBCM, as it was computationally faster and conceptually more appealing than the DBCM, but even more options than the three mentioned in this paper could be considered.

Lastly, one can imagine applying the fitness and complexity algorithm to importers instead of exporters, now that the new methodology offers the opportunity to construct a importer-commodity matrix as input. This was out of the scope of the current paper, but we would argue that if the complexity of the products a country exports tells us something about that country's economy, the complexity of the products it imports would as well. Potentially, it could yield information on the wealth of a nation, so a comparison between each country's exporting and importing fitness might prove to be an interesting next step of investigation.

5. Materials and Methods

5.1. Data

All the data that we have used are originally collected by the UN statistics division and combined in the UN Comtrade database citecomtrade. We have used the available data from the years 2004 and 2012 directly from this database [17].

We also employed parts of a longer range of data that was retrieved from Comtrade and cleaned by Robert Feenstra and Robert Lipsey from the University of California, Davis [18]. This dataset covers the years 1962 up and until 2000. This data range has been most useful for an analysis of temporal evolution of economies.

These datasets contain information of the following kind: exporter i trades commodity α with importer j , with a total value of x dollar. Throughout this paper, we have treated this data as a multiplex network, consisting of many individual networks per commodity combined as "layers". These individual layers are regular weighted, directed networks. Note that, despite all efforts by Comtrade and Feenstra and Lipsey, neither of these datasets is complete, due to practical issues rather than scientific ones.

The first, more recent pair of datasets has six-digit commodity specification codes, whereas the second range of datasets consists of four-digit specifications for commodities. This means that the first database is far more detailed (including circa 5000 commodities) than the second (under 1000 commodities). This in itself has led to differences in (the nature of) our results, both within our own research and when reproducing other's results.

This multiplex, weighted, directed network is represented by an adjacency matrix that we have called W . Links in this network, or rather, the value of their weights, are referred to by $w_{i,j}^\alpha$, with the subscripts i and j referring to the exporting and importing countries respectively, while the superscript is reserved for the layer, or commodity of the trade link.

5.2. Numerical Methods

All numerical efforts have been executed in Python code. For the numerical solving of the equations involving the hidden variables in $p_{i,j}^\alpha$, we have applied the `scipy.optimize` package. All other code is straightforward Python and Numpy. The code developed for our research purposes will be made publicly available to simplify reproduction of our results.

Author Contributions: D.G., V.G. and R.K. worked out the new methodology, R.K. performed the computations, D.G., V.G. and R.K. analyzed the results, R.K. wrote the paper while D.G. revised it.

Funding: D.G. acknowledges 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 Netherlands Organization for Scientific Research (NWO/OCW).

Conflicts of Interest: The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

WTN	World Trade Network
RCA	Revealed Comparative Advantage
DBCM	Directed Binary Configuration Model
MDBCM	Multiplex Directed Binary Configuration Model

References

- Hidalgo, C.; Hausmann, R. The building blocks of economic complexity. *Proc. Natl. Acad. Sci. USA* **2009**, *26*, 10570–10575. [[CrossRef](#)] [[PubMed](#)]
- Tacchella, A.; Cristelli, M.; Caldarelli, G.; Gabrielli, A.; Pietronero, L. A new metrics for countries' fitness and products' complexity. *Sci. Rep.* **2012**, *2*, 723. [[CrossRef](#)] [[PubMed](#)]
- Cristelli, M.; Gabrielli, A.; Tacchella, A.; Caldarelli, G.; Pietronero, L. Measuring the Intangibles: A metrics for the economic complexity of countries and products. *PLoS ONE* **2013**, *8*, e70726. [[CrossRef](#)] [[PubMed](#)]
- Cristelli, M.; Tacchella, A.; Pietronero, L. The heterogeneous dynamics of economic complexity. *PLoS ONE* **2015**, *10*, e0117174. [[CrossRef](#)] [[PubMed](#)]
- Morrison, G.; Buldyrev, S.; Imbruno, M.; Doria Arrieta, O.; Rungi, A.; Riccaboni, M.; Pammolli, F. On economic complexity and the fitness of nations. *Sci. Rep.* **2017**, *7*, 15332 [[CrossRef](#)] [[PubMed](#)]
- Balassa, B. Trade liberalisation and revealed comparative advantage. *Manch. Sch.* **1965**, *33*, 99–123. [[CrossRef](#)]
- Tacchella, A.; Mazzilli, D.; Pietronero, L. A dynamical systems approach to gross domestic product forecasting. *Nat. Phys.* **2018**, *14*, 861–865 [[CrossRef](#)]
- Squartini, T.; Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. *New J. Phys.* **2011**, *13*, 083001. [[CrossRef](#)]
- Squartini, T.; Mastrandrea, R.; Garlaschelli, D. Unbiased sampling of network ensembles. *New J. Phys.* **2015**, *17*, 023052. [[CrossRef](#)]

10. Napolitano, L.; Evangelou, E.; Pugliese, E.; Zeppini, P.; Room, G. Technology networks: The autocatalytic origins of innovations. *R. Soc. Open Sci.* **2018**, *5*. [[CrossRef](#)] [[PubMed](#)]
11. Pugliese, E.; Cimini, G.; Patelli, A.; Zaccaria, A.; Pietronero, L.; Gabrielli, A. Unfolding the innovation system for the development of countries: Co-evolution of Science, Technology and Production. *arXiv* **2017**, arXiv:1707.05146.
12. Gemmetto, V.; Squartini, T.; Picciolo, F.; Ruzzenenti, F.; Garlaschelli, D. Multiplexity and multireciprocity in directed multiplexes. *Phys. Rev. E* **2016**, *94*, 042316. [[CrossRef](#)] [[PubMed](#)]
13. Garlaschelli, D.; Loffredo, M. Generalized Bose-Fermi statistics and structural correlations in weighted networks. *Phys. Rev. Lett.* **2009**, *102*, 038701. [[CrossRef](#)] [[PubMed](#)]
14. Garlaschelli, D.; Loffredo, M. Maximum likelihood: Extracting unbiased information from complex networks. *Phys. Rev. E* **2008**, *78*, 015101. [[CrossRef](#)] [[PubMed](#)]
15. Caldarelli, G.; Capocci, A.; Rios, P.D.L.; Mun, M.A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **2002**, *89*, 258702. [[CrossRef](#)] [[PubMed](#)]
16. Garlaschelli, D.; Loffredo, M.I. Fitness-dependent topological properties of the world trade web. *Phys. Rev. Lett.* **2004**, *18*, 2–5 [[CrossRef](#)] [[PubMed](#)]
17. United Nations Comtrade Database. Available online: <http://comtrade.un.org/data/> (accessed on 29 September 2015).
18. Feenstra, R.; Lipsey, R. Long Range Comtrade Data. Available online: <http://cid.econ.ucdavis.edu/nberus.html> (accessed on 29 September 2015).



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).