SCIENTIFIC REPORTS

OPEN

Randomizing bipartite networks: the case of the World Trade Web

Fabio Saracco[1], Riccardo Di Clemente[1], Andrea Gabrielli[1,2,3] & Tiziano Squartini[1]

Received: 14 January 2015
Accepted: 20 April 2015
Published: 01 June 2015

Within the last fifteen years, network theory has been successfully applied both to natural sciences and to socioeconomic disciplines. In particular, bipartite networks have been recognized to provide a particularly insightful representation of many systems, ranging from mutualistic networks in ecology to trade networks in economy, whence the need of a pattern detection-oriented analysis in order to identify statistically-significant structural properties. Such an analysis rests upon the definition of suitable null models, i.e. upon the choice of the portion of network structure to be preserved while randomizing everything else. However, quite surprisingly, little work has been done so far to define null models for real bipartite networks. The aim of the present work is to fill this gap, extending a recently-proposed method to randomize monopartite networks to bipartite networks. While the proposed formalism is perfectly general, we apply our method to the binary, undirected, bipartite representation of the World Trade Web, comparing the observed values of a number of structural quantities of interest with the expected ones, calculated via our randomization procedure. Interestingly, the behavior of the World Trade Web in this new representation is strongly different from the monopartite analogue, showing highly non-trivial patterns of self-organization.

In the last fifteen years network science has exploded, revealing a world composed by interconnected systems ubiquitously found both in natural sciences and in socioeconomic disciplines[1–3]. Since the very beginning of network science, many different network representations have been adopted in order to study the particular system at hand[4]. However, the class of networks represented by bipartite networks has been recognized to provide a particularly insightful representation of many different systems[5]: ecological networks[6], trade networks[7–9], citations and collaboration networks[10,11] represent only few examples.

One could thus expect a relevant amount of work aimed at identifying the statistically-relevant patterns observed in real bipartite networks, at least comparable to the mass of results obtained so far for monopartite networks[12–21]: however, quite surprisingly, little work has been done so far to implement null models on real bipartite networks. Generally speaking, null models are statistical models used to make inference on a real system on the basis of partial information. The latter usually corresponds to some observable property of interest as the number of trade partners of a country, its exports and imports, the total exposure of a bank, etc. In particular, null models for bipartite networks being *real-data rooted* and showing the desirable features of *general applicability* and *analytical character* are currently missing. More in detail, the algorithms proposed so far show several limitations, ranging from being purely numerical (thus lacking the analytical character)[6,22,23], to assuming an *a priori* functional form either for the distribution of the quantities of interest[6] or for the model parameters (thus not being real data-rooted)[24] or, lastly, using approximate analytical models[25]. Moreover, almost all the aforementioned approaches are tailored on ecological networks, thus lacking the character of general applicability.

The lack of such models is, maybe, also due to the misconception that bipartite networks could be analysed by, firstly, projecting them on one of the layers and, secondly, analysing the projection with one of the models currently available for monopartite networks. As we will show in what follows, the

[1]Istituto dei Sistemi Complessi (ISC) - CNR, UoS Sapienza, Dipartimento di Fisica, Universitá "Sapienza" di Roma, P.le A. Moro 5, 00185 Roma (Italy). [2]IMT Institute for Advanced Studies, P.zza S. Ponziano 6, 55100 Lucca (Italy). [3]INFN - Unitá Roma1, Dipartimento di Fisica, Universitá "Sapienza" di Roma, P.le A. Moro 5, 00185 Roma (Italy). Correspondence and requests for materials should be addressed to T.S. (email: tiziano.squartini@roma1.infn.it)

monopartite and the bipartite representations enclose different kinds of information, irreducible to each other (in the most general case).

The aim of the present paper is to fill this gap, proposing a theoretical framework guaranteeing the three aforementioned properties. In order to do this, we extend a recently-proposed method to randomize monopartite networks[19] to bipartite networks. The method rests upon the sequential maximizations of Shannon entropy and the network likelihood function, a combination which has been proven to be rather effective both for detecting patterns and to reconstruct the structure of several real-world networks[20,26–30]. To the best of our knowledge, the only other paper proposing a method satisfying the three requirements above is[31]: we will comment on the differences with the one proposed here in the Discussion section.

While the proposed formalism is perfectly general, in this paper we apply our method to the binary, undirected, bipartite representation of the World Trade Web (hereafter WTW). We focused on this particular system precisely because of its popularity among network scientists, who have applied null models to all its possible representations[26,27,32–35], with the exception of the bipartite one. As we will show in what follows, representing the WTW as a bipartite network allows to gain a substantially new insight into an already deeply explored system.

The rest of the paper is organized as follows: Data section is devoted to the description of the dataset used for the present analysis, Methods section reports the detailed description of our method and Results section illustrates the results which are discussed in Discussion section, where conclusions are also drawn.

Data

The WTW can be represented in many different ways, depending on the level of information that we want to process. The most popular ones represent it via an adjacency matrix with nodes playing the role of world-countries and links indicating the presence of (any kind) of trade exchange between them. This framework has been recently extended to analyse the WTW as a multiplex, where trade exchanges corresponding to different commodities are distinguished[35,36].

Here we represent the WTW as a bipartite network, i.e. by considering the set of world-countries and the set of products as different entities and linking a given country to a given product if (and only if) the former exports the latter *above* a certain threshold (the so-called RCA[8,9]). Applying the latter rises the probability that the exported commodity is actually produced by the exporting country. In this representation, any two countries (as well as any two products) cannot be directly linked (i.e. links connecting nodes of the same set are not allowed): thus, any two nodes of the same set can be still thought as "interacting" but only indirectly, via a connection with the same nodes of the other family. This way of representing the WTW allows us to analyze the global economy from a different perspective, by making the *productivity relations* between countries explicit (i.e. *which country produces which product*).

The dataset we have considered for the present analysis is the NBER database, collecting data for the 38 years 1963-2000[37] and categorizing products according to the SITC revision 2 at four-digits level. Data have been further processed, building upon the data-mining procedure adopted in[38], to produce a dataset with 538 products across all years and a number of countries varying from 130 to 151.

Methods

The distinction between countries and products leads naturally to the definition of a biadjacency matrix, which will be indicated with $\mathbf{M}$. In the present paper we focus on the binary, undirected representation of the WTW: thus, the matrix entries will be either $m_{cp} = 1$, indicating that country c exports an amount of product p above the RCA threshold, or $m_{cp} = 0$, indicating that the production of p by country c is below the RCA threshold and, thus, has been ignored. As a consequence, each row represents the export basket of a given country, while each column represents the subset of producers of a given product. A pictorial representation of the WTW biadjacency matrix in the year 2000 is shown in Fig. 1, with the blue dotsrepresenting the ones and the white dots the zeros.

If we indicate with C the total number of countries and with P the total number of products, the total number of elements of the biadjacency matrix (i.e. its volume) is $C \cdot P$, also representing the maximum observable number of connections. In fact, unlike the usual square representation, the problems arising from the presence of self-connections are not encountered here. Moreover, the presence of two different subsets (also known as *layers*) induces a measure of "rectangularity" of our matrix $\mathbf{M}$[6], i.e. $R = \frac{|C-P|}{C+P}$, ranging in $R \in [0,1)$, with values closer to 1 indicating a large asymmetry between the number of countries and the number of products and values closer to 0 indicating equivalence between the two layers cardinality (notice that the information on the sign would be based on the arbitrary choice of the layers ordering).

The definitions of other topological quantities of interest easily follow from the usual ones, as the *number of links* (i.e. the total number of connections)

$$L(\mathbf{M}) = \sum_{c=1}^{C}\sum_{p=1}^{P} m_{cp}, \tag{1}$$

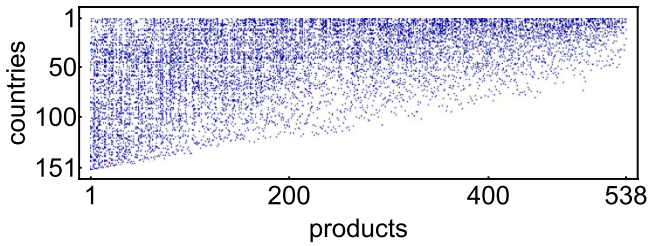

Figure 1. The binary, undirected, bipartite representation of the World Trade Web in the year 2000 [37]: countries are listed along the rows, products along the columns. Blue dots represent the ones, white dots represent the zeros. Rows and columns are reordered according to the algorithm introduced in[8,9].

and the *connectance* $c(M) = \frac{L(M)}{C \cdot P}$, measuring the percentage of observed connections. Fundamental properties are represented by the number of node-specific connections, i.e. the *degree of countries*, also named *diversification*[7–9], measuring the number of products exported by each country.

$$d_c(\mathbf{M}) = \sum_{p=1}^{P} m_{cp}, \tag{2}$$

and the *degree of products*, also named *ubiquity*[7–9], measuring the number of countries exporting each product.

$$u_p(\mathrm{M}) = \sum_{c=1}^{C} m_{cp}. \tag{3}$$

Definitions (2) and (3) induce the notions of *countries mean degree* and *products mean degree*.

$$\overline{d}(\mathbf{M}) = \frac{\sum_{c=1}^{C} d_c(\mathbf{M})}{C} = \frac{L(\mathbf{M})}{C}, \tag{4}$$

$$\overline{u}(\mathbf{M}) = \frac{\sum_{p=1}^{P} u_p(\mathbf{M})}{P} = \frac{L(\mathbf{M})}{P}. \tag{5}$$

The last passage follows from noticing that $L(M) = \sum_{c=1}^{C} d_c(M) = \sum_{c=1}^{P} u_p(M)$.

In order to make the connections between nodes of the same family explicit, a bipartite network can be projected on its layers, thus recovering two traditional, monopartite representations. This operation can be straightforwardly implemented by considering the matrix products.

$$\mathcal{C} = \mathbf{M} \cdot \mathbf{M}^{\mathbf{T}}, \ \mathcal{P} = \mathbf{M}^{\mathbf{T}} \cdot \mathbf{M} \tag{6}$$

where $\boldsymbol{M}^{\boldsymbol{T}}$ is the transpose of the biadjacency matrix $\mathbf{M}$. While the dimensions of $\mathbf{M}$ are $C \times P$, the dimensions of its transpose are $P \times C$. This implies that $\mathcal{C}$ results in a $C \times C$ matrix whose generic element $\mathcal{C}_{cc'}$, with $c \neq c'$, counts the number of patterns of length two between countries c and c'. The generic, diagonal element $\mathcal{C}_{cc}$ is precisely the degree of country c. Similarly, $\mathcal{P}$ results in a $P \times P$ matrix whose generic element $\mathcal{P}_{pp'}$, with $p \neq p'$, counts the number of patterns of length two between products P and p'. As before, the generic, diagonal element is the degree of product P. Remarkably, the entries of matrices $\mathcal{C}$ and $\mathcal{P}$ have a clear macroeconomic interpreation: while $\mathcal{C}_{cc'}$ counts the number of products shared by countries c and c', $\mathcal{P}_{pp'}$ counts the number of countries exporting both products P and p'.

Since nodes of the same layer cannot be directly linked, it is enough that a path of length two (i.e. the minimum allowed length) connects any two nodes of the same family to directly link them in the corresponding monopartite projection. Thus, by first applying the Heaviside step-function $\Theta[\ldots]$ to matrices $\mathcal{C}$ and $\mathcal{P}$ element-wise (i.e. $\Theta[\mathcal{C}] = \{\Theta[\mathcal{C}_{cc'}]\}_{c,c'=1}^{C}$, where $\Theta[\mathcal{C}_{cc'}]$ can be 0 or 1, if $\mathcal{C}_{cc'} = 0$ and $\mathcal{C}_{cc'} > 0$ respectively - and similarly for $\mathcal{P}$) and then subtracting the diagonal elements, the binary, adjacency matrices describing the two monopartite projections are recovered, i.e.

$$\mathbf{C} = \Theta[\mathcal{C}] - \mathbf{I}_C, \ \mathbf{P} = \Theta[\mathcal{P}] - \mathbf{I}_P \tag{7}$$

where $\mathbf{I}_C$ and $\mathbf{I}_P$ are the identity matrices having dimensions $C \times C$ and $P \times P$ respectively.

Topological measures for binary, undirected, bipartite networks. Several quantities have already been proposed to analyse bipartite networks[6]. However, here we define different measures by

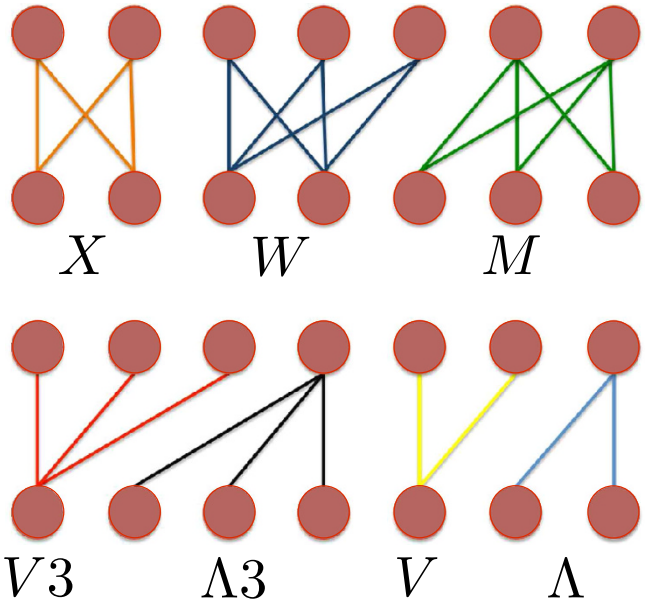

Figure 2. Motifs for bipartite networks. Countries are reported in the upper layer, products in the bottom layer. The bottom panel shows motifs belonging to the Vn and Λn families, with $n = 2,3$.

extending some of the most used indicators in network theory, better capturing, in our opinion, the particular features of a given bipartite network's structure.

a. Assortativity. The traditional definition of assortativity is intended to quantify the degrees correlations, by distinguishing the assortative behavior (signalling positive degrees correlations) from the disassortative behavior (signalling negative degrees correlations). When dealing with bipartite networks, we can measure such correlations both with respect to countries and with respect to products, by respectively defining the *average nearest products ubiquity* (or ANPU)

$$u_c^{nn}(\mathbf{M}) = \frac{\sum_{p=1}^{P} m_{cp} u_p}{d_c} \tag{8}$$

and the *average nearest countries diversification* (or ANCD) as

$$d_p^{nn}(\mathbf{M}) = \frac{\sum_{c=1}^{C} m_{cp} d_c}{u_p}. \tag{9}$$

As in the monopartite case, assortativity is quantified by respectively scattering the ANPU and ANCD values versus the degree sequences $\{d_c\}_{c=1}^{C}$ and $\{u_p\}_{p=1}^{P}$.

b. Complexity and fitness. As recently pointed out[8,9], countries and products can be assigned two purely network-based quantities, known as *fitness*, F_c (to be assigned to countries), and *complexity*, Q_p (to be assigned to products), playing the role of non-monetary indicators of the economy development and providing a highly non-trivial way to rank the world-countries economic health (see also the Supplementary Information).

c. Motifs. The usual clustering coefficient, measuring the hierarchical structure of a monopartite network, cannot be defined for bipartite networks: in fact, since no odd cycles of any length can be observed in bipartite networks (precisely because links within the same layer are forbidden) triangles cannot be observed as well; similarly, the usual triangular motifs cannot be defined[3,39].

However, higher-order correlations between nodes can still be captured by defining a completely new class of motifs. The first examples we provide are the V-*motifs* and the Λ-*motifs* (see Fig. 2). The former count how many couples of countries export the same products, quantifying the productivities' similarity; the latter count how many couples of products are in the basket of the same producer, providing a measure of products correlation. Remembering that $\mathcal{C}_{cc'}$, with $c \neq c'$, counts the number of products exported by both c and c', the total number of V-motifs connecting any pair of countries is

$$N_V(\mathbf{M}) = \sum_{c=1}^{C}\sum_{c'=c+1}^{C} \mathcal{C}_{cc'} = \sum_{c=1}^{C}\sum_{c'=c+1}^{C}\sum_{p=1}^{P} m_{cp}m_{c'p} = \sum_{p=1}^{P}\binom{u_p}{2} \tag{10}$$

and, remembering the analogous role of , the total number of Λ-motifs connecting any pair of products is.

$$N_\Lambda(\mathbf{M}) = \sum_{p=1}^{P}\sum_{p'=p+1}^{P} \mathcal{P}_{pp'} = \sum_{p=1}\sum_{p'=p+1}\sum_{c=1} m_{cp}m_{cp'} = \sum_{c=1}^{C}\binom{d_c}{2}. \tag{11}$$

The last passages follow from noticing that each V-motif (Λ-motif) is constituted by a pair of links having the same product (country) as a common vertex. The number of countries competing on the same product, as well as the number of products in the same basket, can be further risen, leading to the following generalizations (with $V2 \equiv V$ and $\Lambda 2 \equiv \Lambda$):

$$N_{Vn}(\mathbf{M}) = \sum_{p=1}^{P}\binom{u_p}{n}, \;\; N_{\Lambda n}(\mathbf{M}) = \sum_{c=1}^{C}\binom{d_c}{n}; \tag{12}$$

Figure 2 shows an example of V3-motifs and Λ3-motifs. From defintions (12) it follows that $V1 = \Lambda 1 = L$.

Higher-order correlations can be captured by allowing for a higher number of connected nodes in the same layers (see X-*motifs*, M-*motifs* and W-*motifs* in the Supplementary Information). Remarkably, all the defined kinds of motifs:

- can be compactly expressed in terms of products of biadjacency matrix entries;
- can be defined for specific subsets of countries and products, thus allowing for a finer analysis of the production dynamics. For example, a measure of correlation of countries *a* and *b* production is given by the motif $N_{V^{a,b}} = C_{ab} = \sum_{p=1}^{P} m_{ap}m_{bp}$;
- may have an application also in the analysis of ecological networks, especially mutualistic networks (e.g. impollinators-flowers): in fact, measures of co-occurrence can be directly applied to ecosystems to quantify the species' competitiveness for the available resources.

In what follows we will focus on the Vn and Λn families (a more detailed discussion about all motifs is provided in the Supplementary Information).

d. Assortativity coefficient. Beside our definitons, we have also considered the assortativity measure proposed in[40] and called r. The latter ranges in the domain $r \in [-1,1]$, with $r = 1$ indicating the tendency of links to connect nodes with similar degrees and $r = -1$ indicating the tendency of links to connect nodes with different degrees.

e. Nestedness. On the basis of the two aforementioned measures F_c and Q_p, one can reorder the matrix rows and columns (i.e. countries and products) by, respectively, decreasing the fitness along rows (from top to bottom) and increasing the complexity along columns (from left to right), thus obtaining the triangular structure shown in Fig. 7. In order to quantify the shape of such a matrix, several measures have been recently proposed[41–44], under the common name of *nestedness*. Here we adopt the one proposed in[41] (called NODF - see also the Supplementary Information). Notice that the measure of nestedness adopted here doesn't depend on the rows and columns ordering criterion (in what follows we will adopt the one based on F_c and Q_p measures)[8,9].

Randomizing bipartite networks. In order to implement suitable null models to detect the statistically-relevant patterns of real bipartite networks, the lines of the method proposed in[19] can be followed. In particular, an ensemble $\mathcal{G}$ of binary, undirected, bipartite networks must be considered, in order to maximize Shannon entropy

$$S = -\sum_{\mathbf{M}\in\mathcal{G}} P(\mathbf{M}) \ln P(\mathbf{M}) \tag{13}$$

under a given set of constraints $\overrightarrow{C}(M)$[16,19]. Notice that the probability coefficient $P(\boldsymbol{M})$ is assigned to every adjacency matrices in the esemble and the constraints are defined in terms of the entries of $\boldsymbol{M}$. The result is the well-known exponential distribution:

$$P\left(\mathbf{M}\middle|\overrightarrow{\theta}\right) = \frac{e^{-H\left(\mathbf{M},\,\overrightarrow{\theta}\right)}}{Z\left(\overrightarrow{\theta}\right)} \tag{14}$$

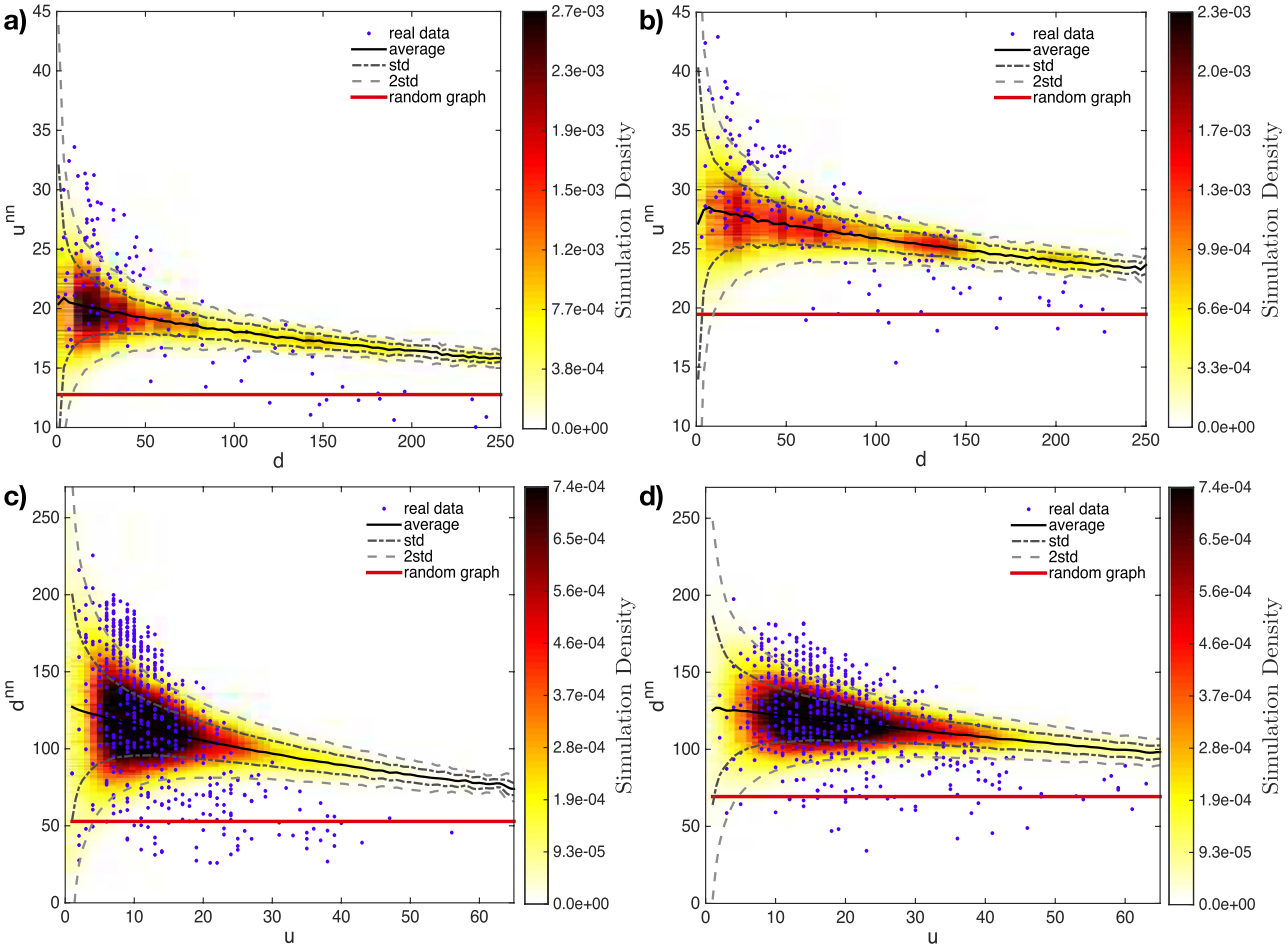

Figure 3. Application of our method to the binary, undirected, bipartite World Trade Web in the year 1963 (left column) and 2000 (right column). Panels report u_c^{nn} VS d_c (a, b) and d_p^{nn} VS u_p (c, d). Observed points are in blue; the black, solid curves are CM-induced ensemble averages; the red, solid lines are RG-induced ensemble averages; the gray, dashed curves indicate the ± 1 standard deviation region; the gray, dash-dotted curves indicate the ± 2 standard deviations region. Colored areas represent the ensemble density of expected points (sampling 5000 matrices). Although the BiCM captures the disassortative trend of the WTW, its striking similarity with the BiRG predictions proves that the explanatory power of the degree sequence is far more limited in the bipartite representation than in the monopartite one[26].

with the hamiltonian $H(\mathrm{M}, \vec{\theta}) = \vec{\theta} \cdot \vec{C}(\mathrm{M})$ compactly expressing the imposed set of constraints, being the vector of Lagrange multipliers associated to the vector of constraints and $Z(\vec{\theta}) = \sum_{\mathrm{M}\in G} e^{-H(\mathrm{M},\vec{\theta})}$ being the normalization.

In the monopartite case, one of the most insightful null models has been proven to be the so-called Configuration Model (CM)[12,14]. Let us now implement the bipartite extension of the CM (BiCM, in what follows), by constraining the degree sequence of the binary, undirected, bipartite WTW and analyzing the system beyond the information contained into it. Since now we have two different layers of nodes, the hamiltonian reads.

$$H(\mathrm{M}, \vec{\theta}) = \vec{\alpha} \cdot \vec{d}(\mathbf{M}) + \vec{\beta} \cdot \vec{u}(\mathbf{M}). \tag{15}$$

Now we can calculate the probability coefficient (14), associating a probability to each network in the ensemble on the basis of the specific degree sequences $\vec{d}(M)$ and $\vec{u}(M)$:

$$P(\mathbf{M}|\vec{\theta}) = \frac{e^{-\vec{\alpha}\cdot\vec{d}(\mathbf{M})-\vec{\beta}\cdot\vec{u}(\mathbf{M})}}{\sum_{\mathbf{M}} e^{-\vec{\alpha}\cdot\vec{d}(\mathbf{M})-\vec{\beta}\cdot\vec{u}(\mathbf{M})}} = \prod_{c,p} p_{cp}^{m_{cp}} \left(1 - p_{cp}\right)^{1-m_{cp}} \tag{16}$$

the notation $\Pi_{c,p}$ being equivalent to $\prod_{c=1}^{C}\prod_{p=1}^{P}$ (see the Supplementary Information for the detailed calculations). The coefficient $p_{cp} = \frac{x_c y_p}{1+x_c y_p}$, with $e^{-\alpha_c} = x_c$ and $e^{-\beta_p} = y_p$, is the ensemble probability of having a link between country c and product p, as $\langle m_{cp} \rangle = \sum_{M\in\mathcal{G}} m_{cp}(M) P(M|\vec{\theta}) = p_{cp} \equiv \frac{x_c y_p}{1+x_c y_p}$.

Our null model provides the analytical expression of a network probability as a product over all the accessible $C \times P$ pairs of nodes. In other words, the BiCM interprets the links as independent random variables, thus defining a grandcanonical probability measure where links correlations are discarded. Notice also that no probability coefficients controlling for the presence of links between nodes in the same layer appear in the expression (16). This is a consequence of having considered an *ensemble of bipartite networks* as the support of our probability distribution: in so doing, the forbidden intra-layer links are automatically excluded by the choice of the allowable configurations volume.

The probability distribution in (16) depends on $C+P$ unknown parameters (i.e. the Lagrange multipliers), also called *hidden variables*[13,24]. The recipe provided by statistical mechanics to estimate the hidden variables is summed up by the equations

$$-\frac{\partial \ln Z}{\partial \alpha_c} = \langle d_c \rangle, \ \forall \ c; \ -\frac{\partial \ln Z}{\partial \beta_p} = \langle u_p \rangle, \ \forall \ p. \tag{17}$$

However, no indication about the numerical value to be assigned to the ensemble average of constraints is provided. Thus, in order to estimate the hidden variables from data, let us first note that $P(M|\vec{\theta})$ can be rewritten solely in terms of the observed constraints value, i.e. $P(\mathbf{M}|\vec{\theta}) = \prod_c x_c^{d_c(\mathbf{M})} \prod_p y_p^{u_p(\mathbf{M})} \prod_{c,p} \left(1 + x_c y_p\right)^{-1}$[19]. Then, let us consider the log-likelihood function $\mathcal{L}(\vec{x}, \vec{y}) = \ln P(\mathbf{M}|\vec{x}, \vec{y})$:

$$\mathcal{L}(\vec{x}, \vec{y}) = \sum_{c=1}^{C} d_c(\mathbf{M}) \ln x_c + \sum_{p=1}^{P} u_p(\mathbf{M}) \ln y_p - \sum_{c=1}^{C}\sum_{p=1}^{P} \ln\left(1 + x_c y_p\right). \tag{18}$$

The recipe provided by statistics to estimate the unknown parameters of a given probability distribution prescribes to maximize $\mathcal{L}$[19]. This means solving the system $\vec{\nabla}\mathcal{L}(\vec{x}, \vec{y}) = \vec{0}$ of $C+P$ equations in $C+P$ unknowns[19]:

$$\begin{cases} d_c(\mathbf{M}) = \sum_{p=1}^{P} \frac{x_c y_p}{1 + x_c y_p}, \ c = 1 \dots C, \\ u_p(\mathbf{M}) = \sum_{c=1}^{C} \frac{x_c y_p}{1 + x_c y_p}, \ p = 1 \dots P. \end{cases} \tag{19}$$

In what follows the vector of solutions satisfying the system (19), for given $\vec{d}(\mathbf{M})$ and $\vec{u}(\mathbf{M})$ as degree mean values, will be indicated as $(\vec{x}^*, \vec{y}^*)$. Notice that the coefficients appearing at the second member of the system equations have the same functional form both for countries and products. This is a consequence of assigning only one Lagrange multiplier to each node but in such a way to distinguish the nodes in the first layer from the nodes in the second layer.

Expected topological measures for binary, undirected, bipartite networks. In the previous subsections several quantities of interest to be measured on binary, undirected, bipartite networks have been listed. In this subsection we will show how our method can be implemented to calculate their expected value (to be compared with the observed one) and the relative errors (to quantify the discrepancies) in order to assess up to what level our null model is able to explain the higher-order structure of the network.

Our method allows us to proceed in a two-fold way. The first one is analytical. Using the link-specific probability coefficients p_{cp} and the passages sketched in[19], we are able to analytically calculate both the expected value and the standard deviation of the (analytically-definable) quantities of the previous subsections. However, because of the impossibility to perform analytical evaluation of the average for some key quantities, we have adopted a different strategy: we have sampled the grancanonical ensemble of binary, undirected, bipartite networks induced by the BiCM according to the probability coefficients $P(\mathbf{M}|\vec{x}^*, \vec{y}^*)$, measured the aforementioned properties on our sample $\widetilde{\mathcal{G}}$ and calculated the statistical moments, as average and standard deviation, of the generic quantity X as.

$$\langle X \rangle \simeq \widetilde{X} = \sum_{\mathbf{M}\in\widetilde{\mathcal{G}}} X(\mathbf{M}) \widetilde{P}(\mathbf{M}), \tag{20}$$

$$\sigma_X \simeq \sigma_{\widetilde{X}} = \sum_{\mathbf{M}\in\widetilde{\mathcal{G}}} (X(\mathbf{M}) - \widetilde{X})^2 \widetilde{P}(\mathbf{M}) \tag{21}$$

i.e. as sampling moments according to the sampling frequencies $\widetilde{P}(\mathbf{M}) = \frac{N_{\mathbf{M}}}{|\widetilde{\mathcal{G}}|}$ ($N_{\boldsymbol{m}}$ being the number of networks in the ensemble having biadjacency matrix equal to $\mathbf{m}$). Since our method is unbiased[19,21]], numerically sampling $\mathcal{G}$ provides a faithful representation of the whole ensemble. We have also calculated the probability distribution (induced by $\widetilde{P}(\mathbf{M})$) of some of the properties of interest, in order to quantify the statistical significance of their observed value (via the z-score, for example).

Nevertheless, the analytical expressions of the expected value and standard deviation of the quantities explicitly defined in the previous subsections has been derived in the Supplementary Information.

Results

Let us first show our results on the temporal snapshot of the WTW corresponding to the year 2000. The number of nodes is $C_{2000} = 151$ and $P_{2000} = 538$, causing the R index to be $R0 \simeq 0.56$ (see section Methods). The high asymmetry of our network is also pointed out by the different mean degrees, $\overline{d} \simeq 70$ and $\overline{u} \simeq 20$, indicating that countries are, on average, almost three times more connected than products. However, the connectance is $c_{2000} \simeq 0.13$: thus, our bipartite WTW is much sparser than its monopartite counterpart[26]. Notice that our null model, constraining (on average) the degree sequence, exactly reproduces any network's connectance by definition, spanning the domain of applicability of both the sparse and the dense network reconstruction algorithms.

Assortativity

Figure 3 shows the comparison between observed and expected values of our coefficients of assortativity. Having plotted u_c^{nn} VS d_c and d_p^{nn} VS u_p, we firstly observe that the bipartite WTW shows a *disassortative* behavior, signalled by a globally decreasing trend of our measures. More detailedly, two distinct behaviors seem to characterize u_c^{nn} as a function of d_c: while countries with *low diversification* are preferentially linked to products with *high ubiquity* (left side of panels 3a and 3b), countries with *high diversification* are linked to *almost all products* (right side of panels 3a and 3b). This is also reflected in the triangular structure of the matrix (see Fig. 1). For products, this distinction is less sharp (panels 3c and 3d): in fact, while *high-ubiquity* products are linked to *almost all countries*, *low-ubiquity* products can be found connected to *both high-* and *low-diversification countries*.

As can be seen from Fig. 3, the BiCM captures the disassortative behavior of both u_c^{nn} and d_p^{nn}; however, only part of the observed points lies within the ± 2 standard deviations region. This means that the mechanism shaping the disassortative behavior of the WTW is not completely explained by our null model, signalling a non-trivial origin of the WTW degree correlations. What is strikingly surprising is the prediction based on the Random Graph model (BiRG): the corresponding trend is closer to the BiCM prediction than in the monopartite representation of the WTW[26]. Moreover, since disassortativity is more pronounced in real data, our results indicate that the BiCM performs better than BiRG for small values of d_c and u_p, while the BiRG correctly capture their flat behavior at large d_c and u_p (i.e. for competitive countries and ubiquitous products, for which $\left\langle d_p^{nn} \right\rangle_{BiRG} \simeq L/C$, $\langle u_c^{nn} \rangle_{BiRG} \simeq L/P$). This seems to indicate that the explanatory power of the degree sequence is far more limited in the bipartite representation than in the monopartite one and that additional information is required to improve the agreement between observations and predictions (even at the simplest level of binary, undirected networks).

Figure 4 extends our assortativity analysis to the entire dataset. In order to condensate the information of 38 scatter plots, we have computed the barycenter and sparseness of both the observed and expected clouds of points. In particular, we have calculated the arithmetic mean of both the observed values $\{u_c^{nn}\}_{c=1}^{C}$.

$$\overline{u^{nn}} = \frac{\sum_{c=1}^{C} u_c^{nn}}{C} = \frac{1}{C}\sum_{c=1}^{C}\sum_{c'=1}^{C}\frac{C_{cc'}}{d_c} \tag{22}$$

and $\left\{d_p^{nn}\right\}_{p=1}^{P}$, the expected values $\{\langle u_c^{nn} \rangle\}_{c=1}^{C}$ and $\left\{\left\langle d_p^{nn} \right\rangle\right\}_{p=1}^{P}$ and the corresponding confidence intervals (CI) at 95% level. As for the motifs, also $\overline{u^{nn}}$ and $\overline{d^{nn}}$ can be interpreted in macroeconomic terms. In fact, $\sum_{c'=1}^{C} C_{cc'}/d_c$ measures the country-specific number of competitions, thus quantifying the (average) presence of a country on the global market. Further averaging over all countries provides a measure of the integration of world-countries production.

What emerges is that the evolution of expected points closely follows the evolution of the observed ones, pointing out that the BiCM correctly describes the temporal trend of the assortativity measures. Notice that, even if observed points are systematically more concentrated on higher levels (as shown in panels 4a and 4c), the confidence intervals are still close enough to let us interpret the BiCM predictions

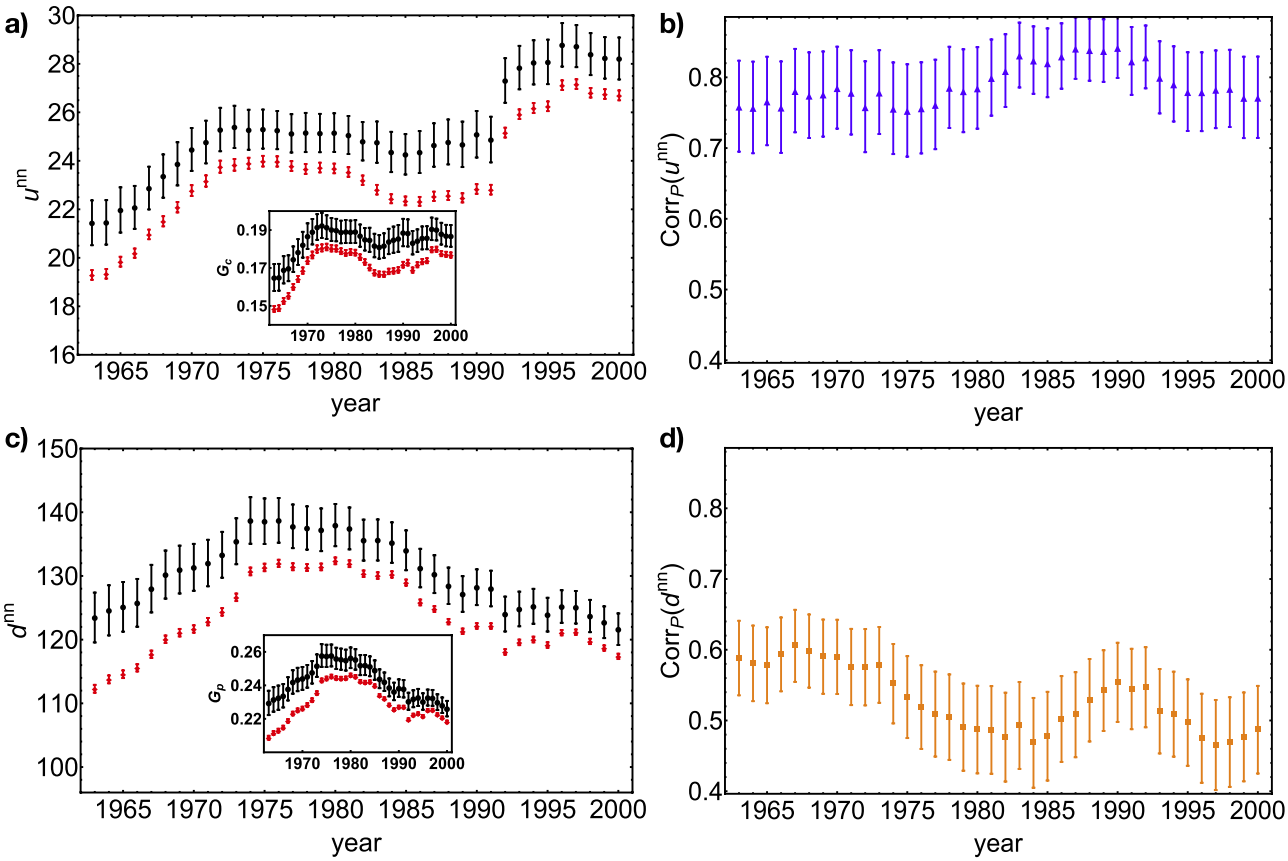

Figure 4. Temporal evolution of the arithmetic mean of the observed $\{u_c^{nn}\}_{c=1}^{C}$ (•) and expected $\{\langle u_c^{nn}\rangle\}_{c=1}^{C}$ (•), together with the 95% CI (panel a); temporal evolution of the arithmetic mean of the observed $\{d_p^{nn}\}_{p=1}^{P}$ (•) and expected $\{\langle d_p^{nn}\rangle\}_{p=1}^{P}$ (•), together with the 95% CI (panel c); temporal evolution of the Pearson correlation coefficient between $\{u_c^{nn}\}_{c=1}^{C}$ and $\{\langle u_c^{nn}\rangle\}_{c=1}^{C}$ (▲) together with the 95% CI (panel b) and between $\{d_p^{nn}\}_{p=1}^{P}$ and $\{\langle d_p^{nn}\rangle\}_{p=1}^{P}$ (■) together with the 95% CI (panel d). The evolution of expected points closely follows the evolution of the observed ones, pointing out that the BiCM correctly describes the temporal

as correct. Moreover, the constancy of the amplitude of the confidence intervals for both observed and expected ANPU values indicates that the corresponding clouds of points maintain the same sparseness across our 38 years dataset; on the other hand, the amplitude of the observed ANCD confidence intervals slightly reduces, indicating a shrinkage of the corresponding cloud of points (compare panels 3b and 3d).

The temporal trends of $\overline{u^{nn}}$ and $\overline{d^{nn}}$ show interesting differences. In fact, while $\overline{u^{nn}}$ keeps increasing across the whole dataset, $\overline{d^{nn}}$ does not (and from 1975 starts decreasing). Since the countries mean degree keeps rising as well ($\overline{d}_{1963} \simeq 48$ and $\overline{d}_{2000} \simeq 70$), the increasing trend is probably due to the birth of new links, indicating that while existing countries have enlarged their production, new-born countries have started theirs. The results seem also to be compatible with the picture of several "appealing" products behaving as hubs and attracting links, including the ones of the new-born countries which in turn, having a low degree, reduce the value of the d_p^{nn}.

Since $\overline{u^{nn}}$ ranges in the interval $[0,C]$, the effect due to the varying number of countries can be washed away by further dividing it by C, $G_C = \overline{u^{nn}}/C$ (and thus normalizing it to the interval [0,1]). Remarkably, our index G_C can be now interpreted a "genuine" measure of globalization, not affected by any spurious effect. Very interestingly, the temporal trend of G_C after 1970 becomes now almost flat. This means that the WTW evolution does not actually affect the value of countries integration which organize in such a way to maintain the same value of G_C, irrespectively of the rising number of countries, their higher diversification, etc. This seems to confirm the stationary evolution of such network, recently pointed out[45]. A similar reasoning leads us to interpret $G_P = \overline{d^{nn}}/P$ as a measure of products homogeneity.

We have also calculated the Pearson correlation coefficient between the vectors $\{u_c^{nn}\}_{c=1}^{C}$ and $\{\langle u_c^{nn}\rangle\}_{c=1}^{C}$ (panel 4b) and between the vectors $\{d_p^{nn}\}_{p=1}^{P}$ and $\{\langle d_p^{nn}\rangle\}_{p=1}^{P}$ (panel 4d), in order to quantify the agree-

ment on the "shape" of the clouds of points. The correlation of the latter is lower than the correlation of the former: this is due to the shape of the empirical cloud of ANCD which is less linear than the empirical ANPU, thus worsening the agreement with the corresponding expectations (which show an almost perfectly linear trend).

Complexity and fitness

Complexity and fitness can be obtained only numerically, as the result of the convergence of the algorithm proposed in[8,9,46]. Panels 5a and 5b show the comparison between observed and expected complexity (plotted VS ubiquity) for the years 1963 and 2000; panels 5c and 5d show the comparison between observed and expected fitness (plotted VS diversification) for the same years. Our null model capture both trends with a larger accuracy than in the measure of assortativity: notice how the expected trend under the BiCM reproduces the "beak" of the observed complexity in real data and the vast majority of the observed cloud lies within the ± 2 standard deviations region.

Similarly, the expected trend of reconstructed fitness captures the different growth regimes of the observed fitness in the WTW data, showing few sparse points outside the same error region (clearly visible in the log-log plots of Fig. 5). The regime with lower slope (left side of panels 5e and 5f) represents the so-called "poverty trap"[8,9], i.e. the area populated by the group of countries with lowest fitness: notice how all such countries lie within the ± 2 standard deviation region (or immediately outside). Similar considerations hold for all the remaining years, indicating a constant performance of our method across our 38-years dataset.

The average trends in Fig. 5 are computed differently from those in Fig. 3: while the latter represent the node-specific, ensemble averages $\{\langle d_c^{nn}\rangle\}_{c=1}^{C}$ and $\left\{\left\langle u_p^{nn}\right\rangle\right\}_{p=1}^{P}$, the former represent averages taken over ranked nodes, ordered according to their complexity - panels a and b - and fitness - panels c and d. Generally speaking, ordering nodes on the basis of such procedure will produce a different ranking for different bipartite networks of the ensemble. Moreover, the ranking operation guarantees neither that the identity of ranked nodes remains the same (e.g. two different countries can be ranked first for two different networks), nor that the corresponding complexity and fitness maintain their value across our sample (i.e. the nodes ranked first will, in general, have different values of F_c and Q_p): this in turn implies that each ranked node degree may change as well (i.e. the nodes ranked first for different networks will, in general, have different degrees). From these considerations, the need of quantifying 1) the variation of any country diversification as a function of its fitness and 2) the variation of any product ubiquity as a function of its complexity follows. This is in line with the spirit of the research in[8,9]: trying to establish a biunivocal relation both between ubiquity and complexity and between fitness and diversification, in order to unambiguously rank countries and products. This kind of analysis represents a highly non-trivial test bench of our model which appear to perform very well.

Motifs. The motifs analysis has been carried on by calculating two different quantities. The first one has been defined as

$$s_m = \frac{N_m(\mathbf{M}) - \langle N_m\rangle}{\langle N_m\rangle} \tag{23}$$

and named *similarity*: it quantifies the goodness of our prediction, measuring the difference between the observed and expected abundances. Beside similarity, we have also considered the traditional z-scores[3,28,39], defined as the ratio of the difference between the observed and expected abundances and the corresponding standard deviation

$$z_m = \frac{N_m(\mathbf{M}) - \langle N_m\rangle}{\sigma_m} \tag{24}$$

with $\sigma_m = \sqrt{\langle N_m^2\rangle - \langle N_m\rangle^2}$ and m indicating the particular motif considered. Even if z-scores have been recognized to be dependent on the network size [47] (at least for monopartite networks), our dataset collects matrices with very similar volume ($R \in [0.56, 0.61]$): thus, we can imagine this effect to be very small.

Notice that similarity and z-scores provide complementary information: in particular, the latter measures the statistical significance of the agreement found by the former, accounting for the role of higher-order correlations not included in our constraints. Moreover, their ratio $s_m/z_m = \sigma_m/\langle N_m\rangle$ coincides with the motif-specific coefficient of variation, quantifying to what extent the average sums up the relevant information encoded into the corresponding ensemble distribution. Naturally, as for the observed abundances, both s_m and z_m can be defined for specific subsets of nodes as well.

Figure 6 shows the analysis of the Vn and Λn motifs. First, we have sampled the V-motifs and Λ-motifs abundance on the ensemble, in order to verify their distribution (see the Supplementary Information): both follow a gaussian very closely. Since all our motifs are sums of (neither independent nor identically distributed) random variables, this may be seen as a consequence of the generalized Central Limit

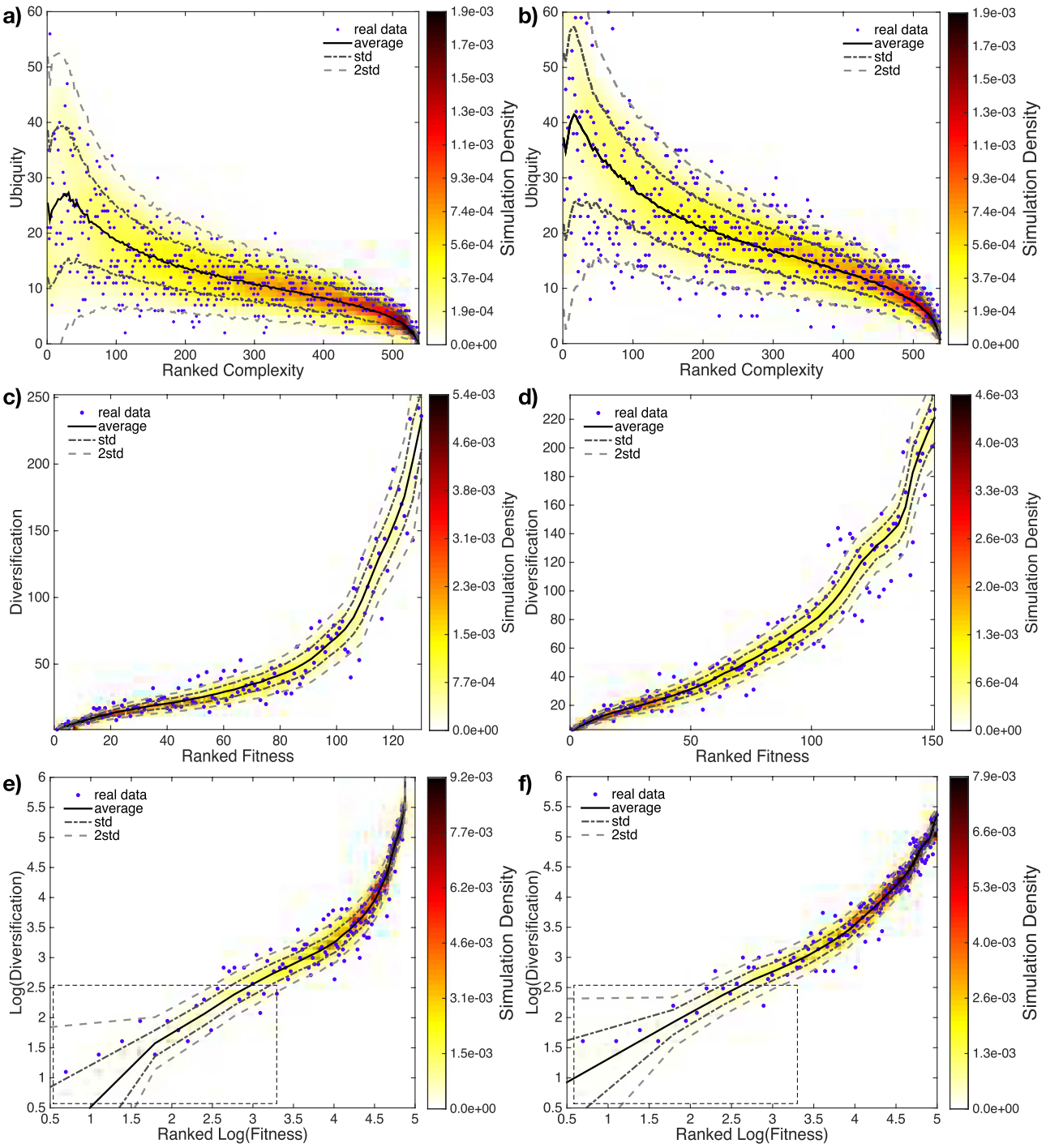

Figure 5. Application of our method to the binary, undirected, bipartite World Trade Web in the year 1963 (left column) and 2000 (right column). Panels report u_p VS Q_p (a, b) and d_c VS F_c (c, d). Observed points are in blue; the black solid curves are BiCM-induced ensemble averages; the gray dashed curves indicate the ± 1 standard deviation region; the gray dash-dotted curves indicate the ± 2 standard deviations region. Colored areas represent the ensemble density of expected points (sampling 5000 matrices). Our null model seems to satisfactorily capture both trends. Panels (e, f) show the so-called "poverty trap", i.e. the group of countries with lowest fitness[8,9]. Notice how all such countries lie within the ± 2 standard deviation region (or immediately outside).

Theorem. z-scores can be thus attributed the correct probabilistic meaning of (gaussian) standardized variables[39],[47] and choosing a threshold z_0 for z allows the identification of significantly deviating patterns. In what follows we will choose $z_0 = \pm 1.65$ as threshold values for the aggregated Vn and Λn families and $z_0 = \pm 2$ for the subsets-specific corresponding ones (see the Supplementary Information for a justification of such values). Naturally, if the observations were exactly reproduced by our null model, the z-scores would be zero[49].

The evolution of both similarity and z-scores across the years in our database point out that the Λn family is better reproduced than the Vn family (showing a similarity and a z-score closer to zero - see panels 6a and 6b). In particular, Vn z-scores lie outside the boundary of the significance region, showing values lower than -1.65. This indicates that for the binary, bipartite representation of the WTW, the degree sequence is far more effective in reproducing the products correlations than the correlations between countries. In other words, we correctly capture the countries tendency to expand their production, which seems to co-exist with a certain superposition of the countries baskets of products (see M-motifs in the Supplementary Information). However, the BiCM overestimates the resemblance of the different baskets: as z-scores indicate, world-countries tend to form less V-motifs than expected under our null model (further confirmed by the trend of X-motifs and W-motifs - see the Supplementary Information). Summing up, world countries show a clear tendency to diversify their production, at the same time avoiding to directly compete on the same products.

The comparison between similarity and z-score clarifies the role of average in characterizing the ensemble distribution of Vn and Λn families: the ratio $s_m/z_m \leq 0.1$ justifies our interest in their ensemble average alone.

However, z-scores of Vn and Λn families result in almost flat trends which allow us to draw only general conclusions on the WTW as a whole. The reason lies in the "aggregated" character of such motifs, not distinguishing between different subsets of countries or products. To be more precise, let us consider the temporal evolution of our motifs on specific subsets of nodes (see panels 6c and 6d): in particular, the Asian Tigers (South Korea, Singapore, Taiwan, Hong Kong), the BRICS countries (Brazil, USSR/Russia, India, China, South Africa), the european countries belonging to G7 (France, Italy, Germany, United Kingdom) and a number of eastern-european countries (Hungary, Romania, Bulgaria, Poland, USSR/Russia) and let us calculate the temporal evolution of V4 and V5 motifs restricted to them. The european countries show a z-score almost constantly equal to 4, indicating a significant affinity which is maintained over time. An even stronger internal affinity is shown by the Asian Tigers to which China should be added (in fact, its addition to the group rises the z-score). On the other hand, BRICS countries show a very limited affinity[8,9,48]: their trend becomes more and more consistent with the null model, to become negative in the recent years. The last two examples point out the limitations of the traditional economic classification (usually distinguishing China from Asian Tigers and gathering BRICS together), not capturing any actual economic likeness.

Eastern-european countries, on the other hand, show a strong correlation before 1989, gradually declining as this topical year approaches. Interestingly enough, after 1989 such correlation doesn't disappear, remaining statistically significant (and stabilizing around $z \simeq 2$): this seems to indicate a significant connection still persisting, having Russia replaced USSR as "reference" country. An additional test is provided by the random choice of four countries (Ghana, China, Mozambique, Austria): although close to zero, their trend is constantly negative. In fact, being Ghana and Mozambique low-diversification countries, they will be linked only to high-ubiquity products, common to all countries (see Fig. 3): thus, their basket will be far more limited than China's and Austria's, limiting in turn their possibility to compete. The constantly negative sign indicates, in this case, the impossibility to compete.

This kind of analysis can be repeated for Λn motifs as well, allowing us to gain a substantial insight into the products correlations. Panels 6e and 6f show some examples. While the food sector we have considered shows a constantly high value of z, indicating the common origin of the chosen dairy products, the pink trend signals a non-trivial positive correlation between the sectors represented by worked aluminium artifacts, tractors and fruit. A possible explanation may rest upon the consideration that tractors are constituted by parts in aluminium to be, in turn, used to transport the picked fruit. Consistently, the last group of products (cheese, rods and locomotives) is characterized by the value $z \simeq 0$.

Notice that while for some groups of nodes the first moment encloses great part of the relevant information ($s_m/z_m \leq 0.5$), for other groups higher-order moments could provide additional, useful information ($s_m/z_m \simeq 1$), e.g. the distribution asymmetry. Interestingly, these circumstance are mostly encountered for countries and products, respectively.

Assortativity coefficient and nestedness. As for the Vn and Λn motifs, the assortativity coefficient has a gaussian ensemble distribution (see the Supplementary Information). Both the observed value r and its z-score signal that we are globally overestimating the network assortativity: more exactly, since our expected coefficient $\langle r \rangle$ is still negative, we are predicting a less disassortative network than observed (see Fig. 7). This is a consequence of our randomization procedure, distributing links between nodes

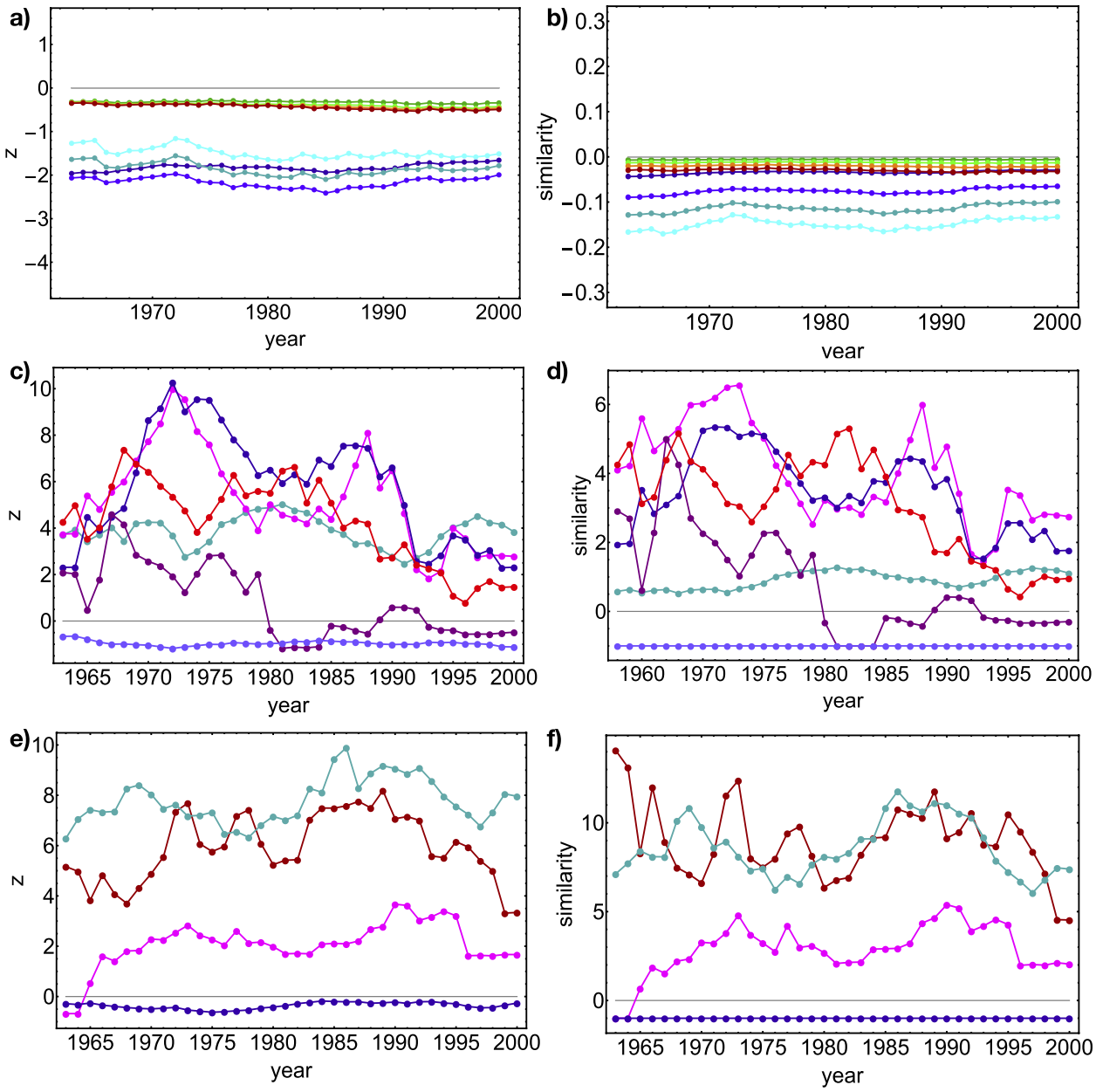

Figure 6. Analysis of motifs. Top panels: z-scores (panel a) and similarity (panel b) evolution across our database years of N_V (•), N_{V3} (•), N_{V4} (•), N_{V5} (•), N_Λ (•), $N_{\Lambda 3}$ (•), $N_{\Lambda 4}$ (•), $N_{\Lambda 5}$ (•). Middle panels: z-scores (panel c) and similarity (panel d) evolution of Vn-motifs, restricted to subsets of countries - Asian Tigers (•), Asian Tigers plus China (•), EU countries in G7 (•), BRICS (•), eastern countries (•), four randomly chosen countries (•). Bottom panels: z-scores (panel e) and similarity (panel f) evolution of Λn-motifs, restricted to subsets of products - "fruit and parts of plants", "aluminium and aluminium alloys", "road tractors" (•), "milk and cream", "butter", "cheese" (•), four randomly chosen products (•). Right column, panel f: similarity evolution across our database years of the same motifs. Our method correctly captures the countries tendency to expand their production (Λn-motifs), even if the resemblance of the different baskets of products is overestimated (Vn-motifs). Moreover, our method identifies statistically significative correlations among subsets of countries and products.

more homogeneously (recall that, consistently, our predicted $\left\{d_p^{nn}\right\}_{p=1}^{P}$ and $\{u_c^{nn}\}_{c=1}^{C}$ show less steeply decreasing trends than the observed ones - see Fig. 3).

In order to better understand the concept of nestedness, let us explicitly draw a matrix from the BiCM-induced grandcanonical ensemble, ranking its rows and columns according to the F_c and Q_p measures[8,9]. The result is shown in Fig. 7. Notice that nestedness cannot be simply reduced to the concept of

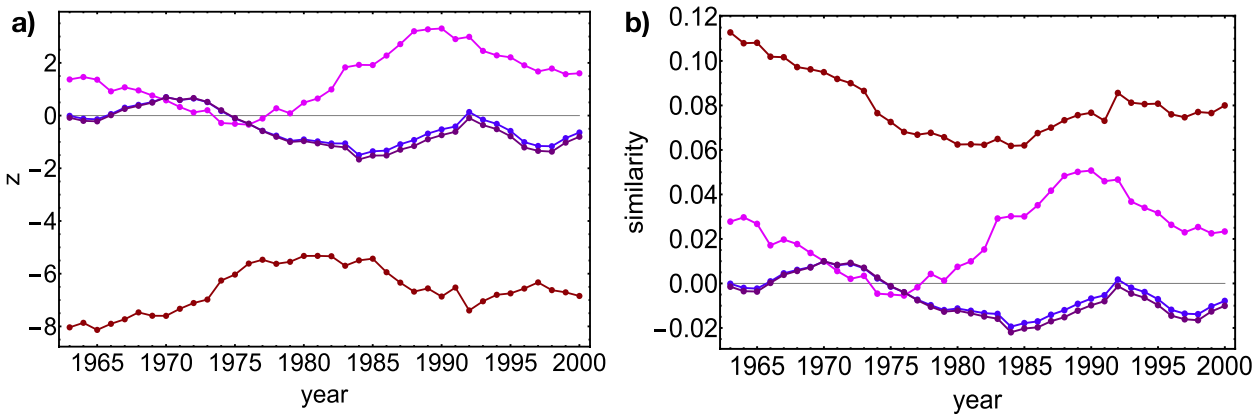

Figure 7. Analysis of the assortativity coefficient and nestedness. z-scores (panel a) and similarity (panel b) evolution across our database years of r (•), NODF (•), nestedness along rows (•) and columns (•). While we are predicting a less disassortative network than observed, our method correctly reproduces the matrix nestedness.

"triangularity" of a matrix. In fact, even if the drawn matrix shows a more curved boundary than the observed one, both the nestedness ensemble distribution (see the Supplementary Information) and its z-score (Fig. 7) signal that our method reproduces it correctly.

We have also measured the nestedness along rows and the nestedness along columns separately (according to the definitions in[41]. While the latter is reproduced and closely follows the trend of the global one, the former is, for a few years, significantly underestimated. This is non-trivially related to the way our null model redistributes V-motifs and Λ-motifs. However, as the bottom panel in Fig. 8 suggests, a role seems to be played by the asymmetry of our bipartite matrix as well: in other words, the higher cardinality of the products layer seems to induce a preferential filling of the rows, making them more homogenenous and lowering their expected nestedness.

It should be also noted that the ensemble coefficient of variation for both r and NODF show such a small value ($s_m/z_m \simeq 10^{-2}$ for both, across our temporal dataset) that the ensemble average can be considered as the only moment carrying relevant information.

Discussion

In this paper we have both proposed a method to randomize binary, undirected, bipartite networks, by constraining essential network features as the total number of links and the nodes connectivity, and tested it on a real system as the World Trade Web. While, on the one hand, specifying the degree sequence allows highly non-trivial properties like countries fitness, products complexity and the matrix nestedness to become reproduced across our whole dataset, on the other quantities like assortativity and motifs still elude a satisfactorily explanation.

This is even more surprising, when considering the high level of accuracy achieved by the CM predictions in the analysis of the monopartite representation of the WTW. Our findings suggest that analysing different representations of the same network can indeed convey additional information, as proved by the agreement between the observed assortativity and the expected one (see Fig. 3), lower than in the corresponding monopartite WTW[26]. In words, the correlations between countries induced by their productivity relations, clearly displayed by the bipartite representation of the WTW, are only partially explained by the degree sequence, calling for a higher amount of information to achieve the same level of accuracy obtained for the monopartite representation (and analogously for products). Otherwise stated, representing the same system via different network models (even belonging to the same class of binary, undirected configurations) may strongly affect the effectiveness of the corresponding piece of information (as the nodes connectivity) in reproducing the observed structure.

Assortativity provides again the clearest example: as previously pointed out, the bipartite Configuration Model predicts trends quite similar to those expected under the bipartite Random Graph. To better quantify this difference, we have calculated the Shannon entropy (normalized to the total number of nodes pairs, i.e. the network volume) of the probability distributions induced by the BiRG and the BiCM:

$$S = \frac{-\sum_{c=1}^{C}\sum_{p=1}^{P}\left[p_{cp}\ \ln\ p_{cp} + \left(1 - p_{cp}\right)\ln\left(1 - p_{cp}\right)\right]}{C \cdot P} \tag{25}$$

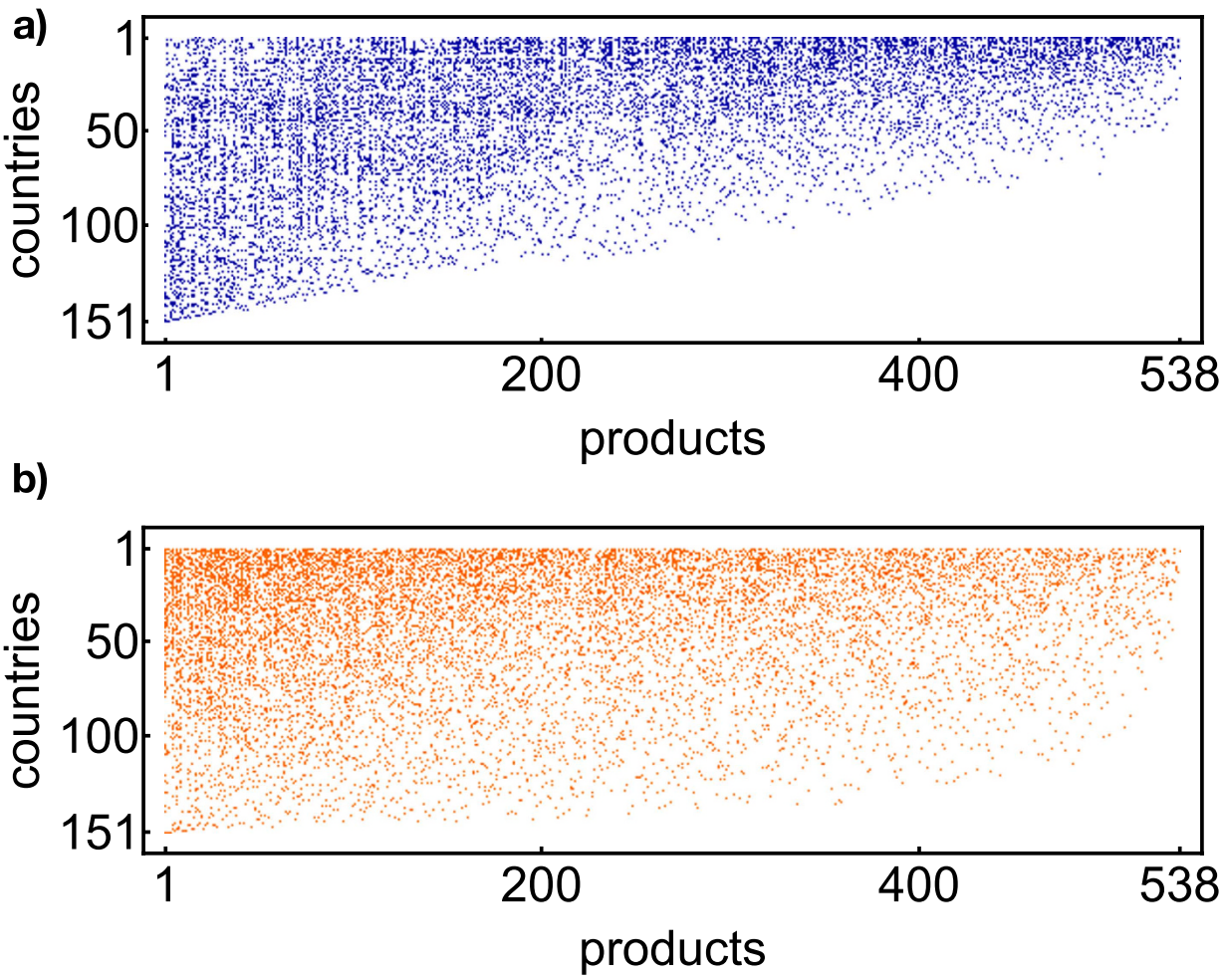

Figure 8. Upper panel: the real World Trade Web matrix in the year 2000, with rows and columns in increasing order of fitness and complexity [8,9]. Lower panel: matrix drawn from the BiCM-induced grandcanonical ensemble for the same year and ordered according to the same criterion.

where $p_{cp} = \frac{x_c x_p}{1 + x_c x_p}$ for the BiCM and $p_{cp} = \frac{x}{1+x}$ for the BiRG (see Supplementary Information). Results are shown in the bottom panel of Fig. 9. As evident from the trends, while specifying the total number of links strongly reduces the uncertainty (as signalled by the low value of the connectance, reducing the ensemble entropy to half its maximum value), further specifying the degree sequence produces a less relevant effect one could expect on the basis of the well known, monopartite results[26]. Comparing the analyses of degree correlations for the bipartite and the projected WTW (on both countries and products layers - top and middle panels of Fig. 9 for the year 2000), what emerges is quite impressive: while the CM prediction correctly overlaps to the observed trend, the RG predicts a flat trend completely missing the observed cloud of points (in line with the results already obtained for the monopartite representation[26]. In terms of Shannon entropy, when passing from the RG to the CM the reduction of uncertainty on the observed, projected WTW amounts to 41%; for the bipartite WTW, this percentage reduces to only 16% (see Fig. 9). This findings clearly indicate a future extension of our work: constraining those quantities having a significant impact on nodes correlations, as V-motifs, Λ-motifs or nestedness, in order to define a more effective null model.

However, as the analysis of motifs reveals, the BiCM provides the right benchmark to highlight meaningful correlations between countries and products, representing a purely topological alternative to the traditional economic classification, whose limitations have been already pointed out[8,9,36]. Remarkably, this kind of analysis can be repeated for different years, in order to monitor our system over time and detect significant temporal trends of the world economies co-evolution.

We stress that our approach is grandcanonical and possible extensions of the method move in the same direction. The paper in[31], on the other hand, implements the microcanonical version of a mono-layer

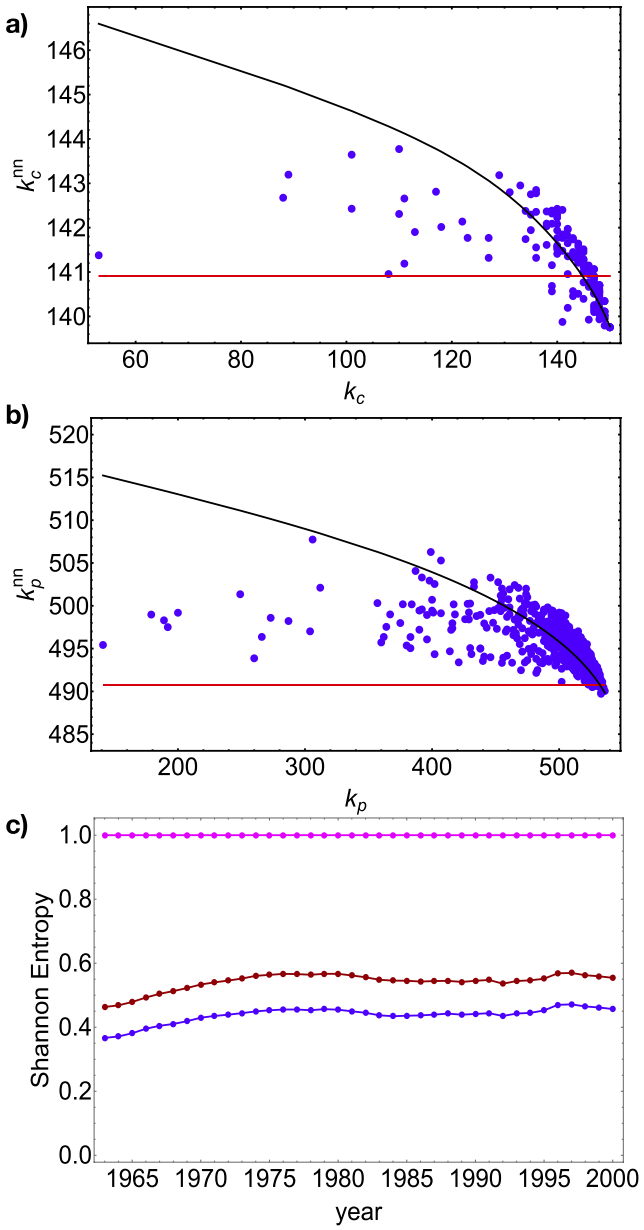

Figure 9. Top panel: analysis of the degrees correlations on the projected WTW, in the year 2000, on the countries layer (blue: observed trend; black: prediction under the CM; red: prediction under the RG). Middle panel: analysis of the degrees correlations on the projected WTW, in the year 2000, on the products layer (blue: observed trend; black: prediction under the CM; red: prediction under the RG). Bottom panel: Shannon entropy of the uniform distribution (•), of the bipartite Random Graph model (•) and of the bipartite Configuration Model (•) over the grandcanonical ensemble of binary, undirected, bipartite networks.

regular random graph: as for monopartite networks, comparing the performance of the two available approaches represents a challenging, future research direction.

Future work moves towards the direction of extending the present framework to directed, as well as weighted, networks models, to test the robustness of our findings also for configurations beyond the binary, undirected ones.

References

1. Albert R. & Barabasi. A.-L. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.* **74,** 47–96 (2002).
2. Newman M. E. J. The structure and function of complex networks. *SIAM Rev.* **45,** 167 (2003).
3. Caldarelli G. in Scale-free Networks. Complex Webs in Nature and Technology (Oxford University Press 2007).

4. Boccaletti S., Latora V., Moreno Y., Chavez M. & Hwang D.-U. Complex Networks: Structure and Dynamics. *Phys. Rep*. **424,** 175–308 (2006).
5. Guillaume J.-L. & Latapy M. Bipartite structure of all complex networks. *Inform. Proc. Lett.* **90,** 215–221 (2004).
6. Dormann C. F., Fründ J., Blüthgen N., Gruber B. Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks. *Open Ecol. J*. **2,** 7–24 (2009).
7. Hidalgo C. & Hausmann R. The building blocks of economic complexity *Proc. Nat. Acad. Sci*. **26,** 10570–10575 (2009).
8. Tacchella A., Cristelli M., Caldarelli G., Gabrielli A. & Pietronero L. A New Metrics for Countries' Fitness and Products' Complexity. *Sci. Rep.* **2,** 723, doi:10.1038/srep00723 (2012) (Date of access: 13/03/2015).
9. Tacchella A., Cristelli M., Caldarelli G., Gabrielli A. & Pietronero L. Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products. *PLoS ONE* **8,** doi:10.1371/journal.pone.0070726 (2013) (Date of access: 13/03/2015).
10. Cimini G., Gabrielli A. & Sylos Labini F. The Scientific Competitiveness of Nations. *PLoS ONE* **9,** doi: 10.1371/journal.pone.0113470 (2014) (Date of access: 13/03/2015).
11. Peltomäki M. & Alava M. Correlations in Bipartite Collaboration Networks. *J. Stat. Mech*. **2006,** P01010 (2006).
12. Chung F. & Lu L. Connected Components in Random Graphs with Given Expected Degree Sequences. *Ann. Comb.* **6,** 125–145 (2002).
13. Caldarelli G., Capocci A., De Los Rios P. & Muñoz M. Scale-Free Networks from Varying Vertex Intrinsic Fitness. *Phys. Rev. Lett.* **89,** 258702 (2002).
14. Park J. & Newman M. E. J. The statistical mechanics of networks. *Phys. Rev. E* **70,** 066117 (2004).
15. Serrano M. A., Boguna M. & Pastor-Satorras R. Correlations in weighted networks. *Phys. Rev. E* **74,** 055101(R) (2006).
16. Garlaschelli D. & Loffredo M. I. Maximum likelihood: extracting unbiased information from complex networks. *Phys. Rev. E* **78,** 015101(R) (2008).
17. Bianconi G. The entropy of network ensembles. *Phys. Rev. E* **79,** 036114 (2009).
18. Fronczak A. in Encyclopedia of Social Network Analysis and Mining (Springer-Verlag 2014).
19. Squartini T. & Garlaschelli D. Analytical maximum-likelihood method to detect patterns in real networks, *New. J. Phys.* **13,** 083001 (2011).
20. Mastrandrea R., Squartini T., Fagiolo G. & Garlaschelli D. Enhanced reconstruction of weighted networks from strengths and degrees. *New J. Phys.* **16,** 043022 (2014).
21. Squartini T., Mastrandrea R. & Garlaschelli D. Unbiased sampling of network ensembles. *New J. Phys.* **17,** 023052 (2015).
22. Dormann C. F., Gruber B. & Fründ J. Introducing the bipartite Package: Analysing Ecological Networks. *R News* **8,** 8–11 (2008).
23. Strona G., Nappo D., Boccacci F., Fattorini S. & San-Miguel-Ayanz J. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nat. Comm.* **5,** doi:10.1038/ncomms5114 (2014) (Date of access: 13/03/2015).
24. Kitsak M. & Krioukov D. Hidden Variables in Bipartite Networks. *Phys. Rev. E* **82,** 026114 (2011).
25. Dormann C. F., Strauss R. A method for detectingmodules in quantitative bipartite networks. *Methods Ecol. Evol*. **5,** 90–98 (2014).
26. Squartini T., Fagiolo G. & Garlaschelli D. Randomizing world trade. I. A binary network analysis. *Phys. Rev. E* **84,** 046117 (2011).
27. Squartini T., Fagiolo G. & Garlaschelli D. Randomizing world trade. II. A weighted network analysis. *Phys. Rev. E* **84,** 046118 (2011).
28. Squartini T. & Garlaschelli D. Triadic Motifs and Dyadic Self-Organization in the World Trade Network. *Lec. Notes Comp. Sci*. **7166,** 24–35 (2012).
29. Musmeci N., Battiston S., Puliga M. & Gabrielli A. Bootstrapping topology and systemic risk of complex network using the fitness model. *J. Stat. Phys*. **151,** 720–734 (2013).
30. Caldarelli G., Chessa A., Pammolli F., Gabrielli A. & Puliga M. Reconstructing a credit network. *Nat. Phys.* **9,** 125–126, doi:10.1038/nphys2580 (2013) (Date of access: 13/03/2015).
31. Tumminello M., Micciché S., Lillo F., Piilo J. & Mantegna R.N. Statistically Validated Networks in Bipartite Complex Systems. *PLoS ONE* **6,** doi:10.1371/journal.pone.0017994 (2011) (Date of access: 13/03/2015).
32. Fronczak A. & Fronczak P. Statistical mechanics of the international trade network. *Phys. Rev. E* **85,** 056113 (2012).
33. Serrano M. A. & Boguna M. Topology of the world trade web. *Phys. Rev. E* **68,** 015101(R) (2003).
34. Fagiolo G., Reyes J. & Schiavo S. The evolution of the world trade web: a weighted-network analysis. *J. Evol. Econ.* **20,** 479–514 (2010).
35. Barigozzi M., Fagiolo G. & Garlaschelli D. Multinetwork of international trade: A commodity-specific analysis. *Phys. Rev. E* **81,** 046104 (2010).
36. Mastrandrea R., Squartini T., Fagiolo G. & Garlaschelli D. Reconstructing the world trade multiplex: the role of intensive and extensive biases. *Phys. Rev. E* **90,** 062804 (2014).
37. National Bureau of Economic Research dataset: http://www.nber.org/data/ (Date of access: 13/03/2015).
38. Feenstra R. C., Lipsey R. E., Deng H., Ma A. C. & Mo H. World Trade Flows: 1962-2000. *National Bureau of Economic Research working paper* 11040, doi:10.3386/w11040 (2005) (Date of access: 13/03/2015).
39. Milo R., Shen-Orr S., Itzkovitz S., Kashtan N., Chklovskii D. & Alon U. Network motifs: simple building blocks of complex networks. *Science* **298,** 824–827 (2002).
40. Newman M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89,** 208701 (2002).
41. Almeida-Neto M., Guimaraes P., Guimaraes P. R. Jr., Loyola R. D. & Ulrich W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* **8,** 1227–1239 (2008).
42. Bastolla U., Fortuna M. A., Pascual-Garcia A., Ferrera A., Luque B. & Bascompte J. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* **7241,** 1018–1020 (2009).
43. Staniczenko P. P. A., Kopp J. & Allesina S. The ghost of nestedness in ecological networks. *Nat. Comm.* **4,** doi: 10.1038/ncomms2422 (2013) (Date of access: 13/03/2015).
44. Jonhson S., Dominguez-Garcia V. & Munoz M. A. Factors Determining Nestedness in Complex Networks. *PLoS ONE* **8,** doi:10.1371/journal.pone.0074025 (2013) (Date of access: 13/03/2015).
45. Squartini T. & Garlaschelli D. Stationarity, non-stationarity and early warning signals in economic networks. *Journal of Complex Networks*, doi: 10.1093/comnet/cnu012 (2014) (Date of access: 13/03/2015).
46. Pugliese E., Zaccaria A. & Pietronero L. On the convergence of the Fitness-Complexity Algorithm. *arXiv:1410.0249* (2014) (Date of access: 13/03/2015).
47. Milo R., Itzkovitz S., Kashtan N., Levitt R., Shen-Orr S., Ayzenshtat I., Sheffer M. & Alon U. Superfamilies of evolved and designed networks. *Science* **303,** 1538–1542 (2004).
48. Caldarelli G., Cristelli M., Gabrielli A., Pietronero L., Scala A. & Tacchella A. A Network Analysis of Countries' Export Flows: Firm Grounds for the Building Blocks of the Economy. *PLoS ONE* **7,** doi:10.1371/journal.pone.0047278 (2012) (Date of access: 13/03/2015).
49. Freund J. E. & Perles B. M. in Modern Elementary Statistics (Prentice-Hall College Div 1996).

Acknowledgments
This work was supported by the EU project GROWTHCOM (611272) and the Italian PNR project CRISIS-Lab. The authors thank Giulio Cimini, Matthieu Cristelli and Andrea Tacchella for useful discussions.

Author Contributions
F.S. and R.D.C. analysed the data and prepared all figures. A.G. wrote the article. T.S. planned the research and wrote the article. All authors reviewed the manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Saracco, F. *et al.* Randomizing bipartite networks: the case of the World Trade Web. *Sci. Rep.* **5**, 10595; doi: 10.1038/srep10595 (2015).