Inferring monopartite projections of bipartite networks: an entropy-based approach

You may also be interested in:

Unbiased sampling of network ensembles
Tiziano Squartini, Rossana Mastrandrea and Diego Garlaschelli

Enhanced reconstruction of weighted networks from strengths and degrees
Rossana Mastrandrea, Tiziano Squartini, Giorgio Fagiolo et al.

Fast and accurate determination of modularity and its effect size
Santiago Treviño III, Amy Nyberg, Charo I Del Genio et al.

Evaluating accuracy of community detection using the relative normalized mutual information
Pan Zhang

Individual nodes contribution to the mesoscale of complex networks
Florian Klimm, Javier Borge-Holthoefer, Niels Wessel et al.

Identification of clusters of investors from their real trading activity in a financial market
Michele Tumminello, Fabrizio Lillo, Jyrki Piilo et al.

Extracting weights from edge directions to find communities in directed networks
Darong Lai, Hongtao Lu and Christine Nardini

Network-based association of hypoxia-responsive genes with cardiovascular diseases
Rui-Sheng Wang, William M Oldham and Joseph Loscalzo

Eigenspaces of networks reveal the overlapping and hierarchical community structure moreprecisely
Xiaoke Ma, Lin Gao and Xuerong Yong

New Journal of Physics

The open access journal at the forefront of physics

CrossMark

OPEN ACCESS

PAPER

Inferring monopartite projections of bipartite networks: an entropy-based approach

Fabio Saracco[1,2], **Mika J Straka**[1], **Riccardo Di Clemente**[3], **Andrea Gabrielli**[1,2], **Guido Caldarelli**[1,2] and **Tiziano Squartini**[1,2]

1 IMT School for Advanced Studies, P.zza S. Ponziano 6, I-55100 Lucca, Italy
2 Institute for Complex Systems (ISC-CNR) UOS Sapienza, 'Sapienza' University of Rome, P.le A. Moro 5, I-00185 Rome, Italy
3 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Massachusetts Avenue 77, Cambridge, MA 02139, United States of America

E-mail: tiziano.squartini@imtlucca.it

Abstract

Bipartite networks are currently regarded as providing a major insight into the organization of many real-world systems, unveiling the mechanisms driving the interactions occurring between distinct groups of nodes. One of the most important issues encountered when modeling bipartite networks is devising a way to obtain a (monopartite) projection on the layer of interest, which preserves as much as possible the information encoded into the original bipartite structure. In the present paper we propose an algorithm to obtain statistically-validated projections of bipartite networks, according to which any two nodes sharing a statistically-significant number of neighbors are linked. Since assessing the statistical significance of nodes similarity requires a proper statistical benchmark, here we consider a set of four null models, defined within the exponential random graph framework. Our algorithm outputs a matrix of link-specific *p*-values, from which a validated projection is straightforwardly obtainable, upon running a multiple hypothesis testing procedure. Finally, we test our method on an economic network (i.e. the countries-products World Trade Web representation) and a social network (i.e. MovieLens, collecting the users' ratings of a list of movies). In both cases non-trivial communities are detected: while projecting the World Trade Web on the countries layer reveals modules of similarly-industrialized nations, projecting it on the products layer allows communities characterized by an increasing level of complexity to be detected; in the second case, projecting MovieLens on the films layer allows clusters of movies whose affinity cannot be fully accounted for by genre similarity to be individuated.

1. Introduction

Many real-world systems, ranging from biological to socio-economic ones, are bipartite in nature, being defined by interactions occurring between pairs of distinct groups of nodes (be they authorships, attendances, affiliations, etc) [1, 2]. This is the reason why bipartite networks are ubiquitous tools, employed in many different research areas to gain insight into the mechanisms driving the organization of the aforementioned complex systems.

One of the issues encountered when modeling bipartite networks is obtaining a (monopartite) projection over the layer of interest while preserving as much as possible the information encoded into the original bipartite structure. This problem becomes particularly relevant when, e.g. a direct measurement of the relationships occurring between nodes belonging to the same layer is impractical (as gathering data on friendship within social networks [3]).

The simplest way of inferring the presence of otherwise unaccessible connections is linking any two nodes, belonging to the same layer, as long as they share at least one neighbor: however, this often results in a very dense

network whose topological structure is almost trivial. A solution which has been proposed prescribes to retain the information on the number of common neighbors, i.e. to project a bipartite network into a *weighted* monopartite network [3]. This prescription, however, causes the nodes with larger degree in the original bipartite network to have, in turn, larger strengths in the projection, thus masking the genuine statistical relevance of the induced connections. Moreover, such a prescription lets spurious clusters of nodes emerge (e.g. cliques induced by the presence of—even—a single node connected to all nodes on the opposite layer).

In order to face this problem, algorithms to retain only the significant weights have been proposed [3]. Many of them are based on a thresholding procedure, a major drawback of which lies in the arbitrariness of the chosen threshold [4–6]. A more statistically-grounded algorithm prescribes to calculate the statistical significance of the projected weights according to a properly-defined null model [7]; the latter, however, encodes relatively little information on the original bipartite structure, thus being more suited to analyze natively monopartite networks. A similar-in-spirit approach aims at extracting the backbone of a weighted, monopartite projection by calculating its minimum spanning tree and provides a recipe for community detection by calculating the minimum spanning forest [8, 9]. However, the lack of a comparison with a benchmark makes it difficult to asses the statistical relevance of its outcome.

The approaches discussed so far represents attempts to validate a projection *a posteriori*. A different class of methods, on the other hand, focuses on *projecting* a statistically validated network by estimating the tendency of any two nodes belonging to the same layer to share a given portion of neighbors. All approaches define a similarity measure which either ranges between 0 and 1 [10, 11] or follows a probability distribution on which a *p*-value can be computed [12–14]. While in the first case the application of an arbitrary threshold is still unavoidable, in the second case prescriptions rooted in traditional statistics can be applied.

In order to overcome the limitations of currently-available algorithms, we propose a general method which rests upon the very intuitive idea that any two nodes belonging to the same layer of a bipartite network should be linked in the corresponding monopartite projection if, and only if, significantly similar. To stress that our benchmark is defined by constraints which are satisfied *on average*, we will refer to our method as to a *grand canonical* algorithm for obtaining a statistically-validated projection of any binary, undirected, bipartite network. A *microcanonical* projection method has been defined as well [15] which, however, suffers from a number of limitations imputable to its nature of purely numerical algorithm [3].

The rest of the paper is organized as follows. In the methods section, our approach is described: first, we introduce a quantity to measure the similarity of any two nodes belonging to the same layer; then, we derive the probability distribution of this quantity according to four bipartite null models, defined within the exponential random graph (ERG) formalism [16]. Subsequently, for any two nodes, we quantify the statistical significance of their similarity and, upon running a multiple hypothesis test, we link them if recognized as significantly similar. In the results section we employ our method to obtain a projection of two different data sets: the countries-products World Trade Web and the users-movies MovieLens network. Finally, in the discussions section we comment our results.
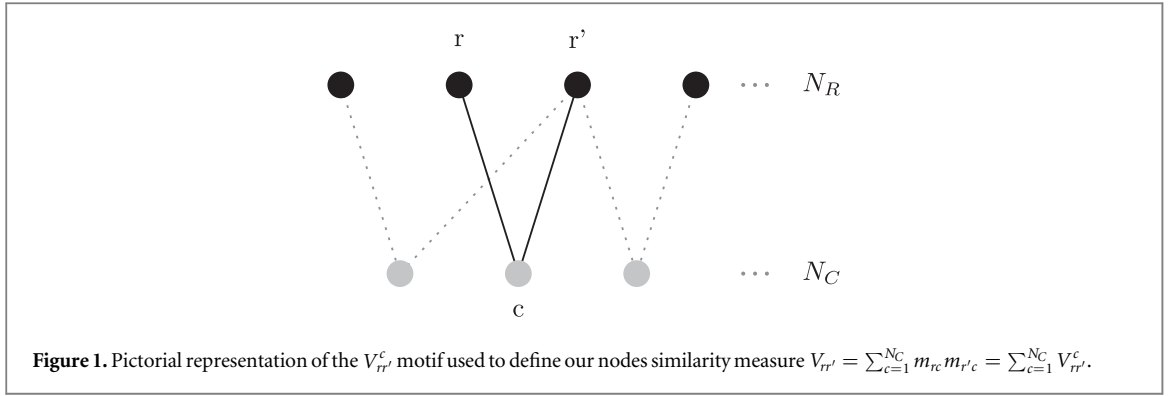
2. Methods

A bipartite, undirected, binary network is completely defined by its biadjacency matrix, i.e. a rectangular matrix **M** whose dimensions will be indicated as $N_R \times N_C$, with N_R being the number of nodes in the top layer (i.e. the number of rows of **M**) and N_C being the number of nodes in the bottom layer (i.e. the number of columns of **M**). **M** sums up the structure of the corresponding bipartite matrix: $m_{rc} = 1$ if node r (belonging to the top layer) and node c (belonging to the bottom layer) are linked, otherwise $m_{rc} = 0$. Links connecting nodes belonging to the same layer are not allowed.

In order to obtain a (layer-specific) monopartite projection of a given bipartite network, a criterion for linking the considered pairs of nodes is needed. Schematically, our grand canonical algorithm works as follows:

A. choose a specific pair of nodes belonging to the layer of interest, say r and r', and measure their similarity;

B. quantify the statistical significance of the measured similarity with respect to a properly-defined null model, by computing the corresponding *p*-value;

C. link nodes r and r' if, and only if, the related *p*-value is statistically significant;

∗ repeat the steps above for every pair of nodes.

We will now describe each step of our algorithm in detail.

Figure 1. Pictorial representation of the $V_{rr'}^c$ motif used to define our nodes similarity measure $V_{rr'} = \sum_{c=1}^{N_C} m_{rc} m_{r'c} = \sum_{c=1}^{N_C} V_{rr'}^c$.

2.1. Measuring nodes similarity

The first step of our algorithm prescribes to measure the degree of similarity of nodes r and r'. A straightforward approach is counting the number of common neighbors $V_{rr'}$ shared by nodes r and r'. By adopting the formalism proposed in [16], our measure of similarity is provided by the number of bi-cliques $K_{1,2}$ [17], also known as V-motifs [16]:

$$V_{rr'} = \sum_{c=1}^{N_C} m_{rc} m_{r'c} = \sum_{c=1}^{N_C} V_{rr'}^c, \tag{2.1}$$

where we have adopted the definition $V_{rr'}^c \equiv m_{rc} m_{r'c}$ for the single V-motif defined by nodes r and r' and node c belonging to the opposite layer (see figure 1 for a pictorial representation). From the definition, it is apparent that $V_{rr'}^c = 1$ if, and only if, both r and r' share the (common) neighbor c.

Notice that naïvely projecting a bipartite network corresponds to considering the monopartite matrix defined as $\mathbf{V}_{rr'}^{\text{naive}} = V_{rr'}$ whose densely connected structure, described by $\mathbf{R}_{rr'}^{\text{naive}} = \Theta[V_{rr'}]$, is characterized by an almost trivial topology.

2.2. Quantifying the statistical significance of nodes similarity

The second step of our algorithm prescribes to quantify the statistical significance of the similarity of our nodes r and r'. To this aim, a benchmark is needed: a natural choice leads to adopt the ERG class of null-models [16, 18–22].

Within the ERG framework, the generic bipartite network $\mathbf{M}$ is assigned an exponential probability $P(\mathbf{M}) = \frac{e^{-H(\vec{\theta}, \vec{C}(\mathbf{M}))}}{Z(\vec{\theta})}$, whose value is determined by the vector $\vec{C}(\mathbf{M})$ of topological constraints [18]. In order to determine the unknown parameters $\vec{\theta}$, the likelihood-maximization recipe can be adopted: given an observed biadjacency matrix $\mathbf{M}^*$, it translates into solving the system of equations $\langle \vec{C} \rangle(\vec{\theta}) = \sum_{\mathbf{M}} P(\mathbf{M}) \vec{C}(\mathbf{M}) = \vec{C}(\mathbf{M}^*)$ which prescribes to equate the ensemble averages $\langle \vec{C} \rangle(\vec{\theta})$ to their observed counterparts, $\vec{C}(\mathbf{M}^*)$ [19].

Two of the null models we have considered in the present paper are known as the bipartite random graph (BiRG) model and the bipartite configuration model (BiCM) [16]; the other ones are the two 'partial' configuration models BiPCM_r and BiPCM_c: the four null models are defined, respectively, by constraining the total number of links, the degrees of nodes belonging to both layers and the degrees of nodes belonging to one layer only (see appendix for the analytical definitions).

The use of linear constraints allows us to write $P(\mathbf{M})$ in a factorized form, i.e. as the product of pair-specific probability coefficients

$$P(\mathbf{M}) = \prod_{r=1}^{N_R} \prod_{c=1}^{N_C} p_{rc}^{m_{rc}} (1 - p_{rc})^{1 - m_{rc}} \tag{2.2}$$

the numerical value of the generic coefficient p_{rc} being determined by the likelihood-maximization condition (see appendix). As an example, in the case of BiRG, $p_{rc} = p_{\text{BiRG}} = \frac{L}{N_R \cdot N_C}$, $\forall r, c$ with L being the total number of links in the actual bipartite network.

Since ERG models with linear constraints treat links as independent random variables, the presence of each $V_{rr'}^c$ can be regarded as the outcome of a Bernoulli trial:

$$f_{\text{Ber}}(V_{rr'}^c = 1) = p_{rc} p_{r'c}, \tag{2.3}$$

$$f_{\text{Ber}}(V_{rr'}^c = 0) = 1 - p_{rc} p_{r'c}. \tag{2.4}$$

It follows that, once r and r' are chosen, the events describing the presence of the N_C single $V_{rr'}^c$ motifs are independent random experiments: this, in turn, implies that each $V_{rr'}$ is nothing else than a sum of independent Bernoulli trials, each one described by a different probability coefficient.

The distribution describing the behavior of each $V_{rr'}$ turns out to be the so-called Poisson–Binomial [23, 24]. More explicitly, the probability of observing zero V-motifs between r and r' (or, equivalently, the probability for nodes r and r' of sharing zero neighbors) reads

$$f_{\mathrm{PB}}(V_{rr'} = 0) = \prod_{c=1}^{N_C} (1 - p_{rc}p_{r'c}), \tag{2.5}$$

the probability of observing only one V-motif reads

$$f_{\mathrm{PB}}(V_{rr'} = 1) = \sum_{c=1}^{N_C} \left[p_{rc}p_{r'c} \prod_{\substack{c'=1 \\ c' \neq c}}^{N_C} (1 - p_{rc'}p_{r'c'}) \right], \tag{2.6}$$

etc. In general, the probability of observing n V-motifs can be expressed as a sum of $\binom{N_C}{n}$ addenda, running on the n-tuples of considered nodes (in this particular case, the ones belonging to the bottom layer). Upon indicating with C_n all possible nodes n-tuples, this probability reads

$$f_{\mathrm{PB}}(V_{rr'} = n) = \sum_{C_n} \left[\prod_{c_i \in C_n} p_{rc_i}p_{r'c_i} \prod_{c_i' \notin C_n} (1 - p_{rc_i'}p_{r'c_i'}) \right] \tag{2.7}$$

(notice that the second product runs over the complement set of C_n).

Measuring the statistical significance of the similarity of nodes r and r' thus translates into calculating a p-value on the aforementioned Poisson–Binomial distribution, i.e. the probability of observing a number of V-motifs greater than, or equal to, the observed one (which will be indicated as $V_{rr'}^*$):

$$p\text{-value}(V_{rr'}^*) = \sum_{V_{rr'} \geqslant V_{rr'}^*} f_{\mathrm{PB}}(V_{rr'}). \tag{2.8}$$

Upon repeating such a procedure for each pair of nodes, we obtain an $N_R \times N_R$ matrix of p-values (see also the appendix). In order to speed up the numerical computation of p-values, a Python code has been made publicly available by the authors[4].

As a final remark, notice that this approach describes a one-tail statistical test, where nodes are considered as significantly similar if, and only if, the observed number of shared neighbors is 'sufficiently large'. In principle, our algorithm can be also used to carry out the reverse validation, linking any two nodes if the observed number of shared neighbors is 'sufficiently small': this second type of validation can be performed whenever interested in highlighting the 'dissimilarity' between nodes.

2.3. Validating the projection

In order to understand which p-values are significant, it is necessary to adopt a statistical procedure accounting for testing multiple hypotheses at a time.

In the present paper we apply the so-called false discovery rate (FDR) procedure [25]. Whenever M different hypotheses, $H_1 \ldots H_M$, characterized by M different p-values, must be tested at a time, FDR prescribes to, first, sort the M p-values in increasing order, $p\text{-value}_1 \leqslant \ldots \leqslant p\text{-value}_M$ and, then, to identify the largest integer $\hat{i}$ satisfying the condition

$$p\text{-value}_{\hat{i}} \leqslant \frac{\hat{i}t}{M} \tag{2.9}$$

with t representing the usual single-test significance level (e.g. $t = 0.05$ or $t = 0.01$). The third step of the FDR procedure prescribes to reject all the hypotheses whose p-value is less than, or equal to, $p\text{-value}_{\hat{i}}$, i.e. $p\text{-value}_1 \leqslant \ldots \leqslant p\text{-value}_{\hat{i}}$. Notably, FDR allows one to control for the expected number of false 'discoveries' (i.e. incorrectly-rejected null hypotheses), irrespectively of the independence of the hypotheses tested (our hypotheses, for example, are not independent, since each observed link affects the similarity of several pairs of nodes).

In our case, the FDR prescription translates into adopting the threshold $\hat{i}t / \binom{N_R}{2}$ which corresponds to the largest $p\text{-value}_{\hat{i}}$ satisfying the condition

---

[4] Python code for computing p-values under the null models discussed in the paper: https://github.com/tsakim/bicm.

$$p\text{-value}_{\hat{i}} \leqslant \frac{\hat{i}\,t}{\binom{N_R}{2}} \tag{2.10}$$

(with i indexing the sorted $\binom{N_R}{2}$ p-value ($V_{rr'}$) coefficients) and considering as significantly similar only those pairs of nodes r, r' whose $p\text{-value}(V^*_{rr'}) \leqslant p\text{-value}_{\hat{i}}$. In other words, every couple of nodes whose corresponding p-value is validated by the FDR is joined by a binary, undirected link in our projection. In what follows, we have used a single-test significance level of $t = 0.01$.

Summing up, the recipe for obtaining a statistically-validated projection of the bipartite network **M** by running the FDR criterion requires that $\mathbf{R}^{nm}_{rr'} = 1$ if, and only if, $p\text{-value}(V_{rr'}) \leqslant p\text{-value}_{\hat{i}}$, according to null model nm used. Notice that the validation process naturally circumvents the problem of spurious clustering (see also appendix).

The aforementioned approaches providing an algorithm to project a validated network differ in the way the issue of comparing multiple hypotheses is dealt with. While in some approaches this step is simply missing and each test is carried out independently from the other ones [3, 14], in others the Bonferroni correction is employed [12, 13]. Both solutions are affected by drawbacks.

The former algorithms, in fact, overestimate the number of *incorrectly rejected* null hypotheses (i.e. of incorrectly validated links). A simple argument can, indeed, be provided: the probability that, by chance, at least one, out of M hypotheses, is incorrectly rejected (i.e. that at least one link is incorrectly validated) is $\text{FWER} = 1 - (1 - t)^M$ which is $\text{FWER} \simeq 1$ for just $M = 100$ tests conducted at the significance level of $t = 0.05$.

The latter algorithms, on the other hand, adopt a criterion deemed as severely overestimating the number of *incorrectly retained* null hypotheses (i.e. of incorrectly discarded links) [25]. Indeed, if the stricter condition $\text{FWER} = 0.05$ is now imposed, the threshold p-value can be derived as $p\text{-value}_{\text{th}} = t \simeq 0.05/M$ which rapidly vanishes as M grows. As a consequence, very sparse (if not empty) projections are often obtained.

Naturally, deciding which test is more suited for the problem at hand depends on the importance assigned to false positive and false negatives. As a rule of thumb, the Bonferroni correction can be deemed as appropriate when *few* tests, out of a *small* number of multiple comparisons, are expected to be significant (i.e. when even a *single* false positive would be problematic). On the contrary, when *many* tests, out of a *large* number of multiple comparisons, are expected to be significant (as in the case of socio-economic networks), using the Bonferroni correction may, in turn, produce a too large number of false negatives, an undesired consequence of which may be the impairment of, e.g. a recommendation system.

As a final remark, we stress that an *a priori* selection of the number of validated links is not necessarily compatible with the existence of a level t of statistical significance ensuring that the FDR procedure still holds. As an example, let us suppose we retain only the first k p-values; the FDR would then require the following inequalities to be satisfied: $p\text{-value}_k \leqslant kt/M$ and $p\text{-value}_{k+1} > (k + 1)t/M$. This, in turn, would imply $p\text{-value}_k/k < p\text{-value}_{k+1}/(k + 1)$. The aforementioned condition, however, can be easily violated by imaging a pair of subsequent p-values close enough to each other (e.g. $p\text{-value}_3 = 0.039$ and $p\text{-value}_4 = 0.040$).

2.4. Testing the projection algorithm

2.4.1. Community detection

In order to test the performance of our method, the Louvain algorithm has been run on the validated projections of the real networks considered for the present analysis [26]. Since Louvain algorithm is known to be order-dependent [27, 28], we considered N outcomes of the former, each one obtained by randomly reshuffling the order of nodes taken as input (N being the network size), and chose the one providing the maximum value of the modularity. This procedure can be shown to enhance the detection of partitions characterized by a higher value of the modularity itself (a parallelized Python version of the reshuffled Louvain method is available at the public repository[5]).

3. Results

3.1. World trade web

Let us now test our validation procedure on the first data set considered for the present analysis: the World Trade Web. In the present paper we consider the COMTRADE database (using the HS 2007 code revision), spanning the years 1995–2010[6]. After a data-cleaning procedure operated by BACI [29] and a thresholding procedure induced by the RCA (for more details, see [30]), we end up with a bipartite network characterized by $N_R = 146$

---

[5] https://github.com/tsakim/Shuffled_Louvain

[6] http://comtrade.un.org/

Figure 2. From top to bottom, pictorial representation of the validated projections of the WTW in the year 2000 (ones are indicated as black dots, zeros as white dots): naïve projection $\mathbf{R}_{rr'}^{\text{naïve}}$, BiRG-induced projection and BiCM-induced projection. Rows and columns of each matrix have been reordered according to the same criterion.
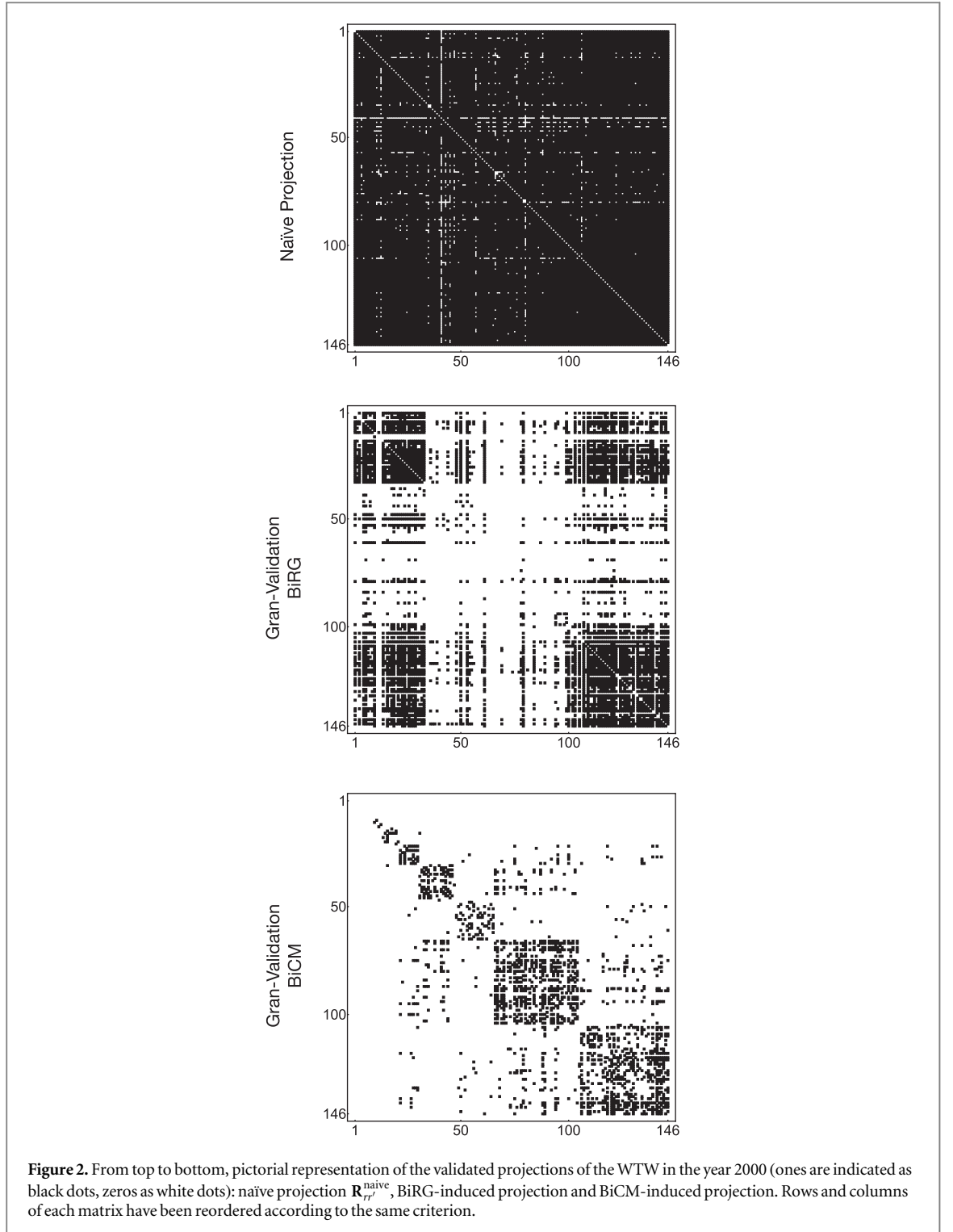
countries and $N_C = 1131$ classes of products, whose generic entry $m_{rc} = 1$ indicates that country r exports product c above the RCA threshold.

Countries layer. Figure 2 shows three different projections of the WTW. The first panel shows a pictorial representation of the WTW topology in the year 2000, upon naïvely projecting it (i.e. by joining any two nodes if at least one neighbor is shared, thus obtaining a matrix $\mathbf{R}_{rr'}^{\text{naïve}} = \Theta[V_{rr'}]$). The high density of links (which oscillates between 0.93 and 0.95 throughout the period covered by the data set) causes the network to be characterized by trivial values of structural quantities (e.g. all nodes have a clustering coefficient very close to 1).

The second panel of figure 2 represents the projected adjacency matrix using the BiRG as a null model. In this case, the only parameter defining our reference model is $p_{\text{BiRG}} = \frac{L}{N_R \cdot N_C} \simeq 0.13$. As a consequence, $p_{rc} = p_{\text{BiRG}}$ for every pair of nodes and formula (2.7) simplifies to the binomial
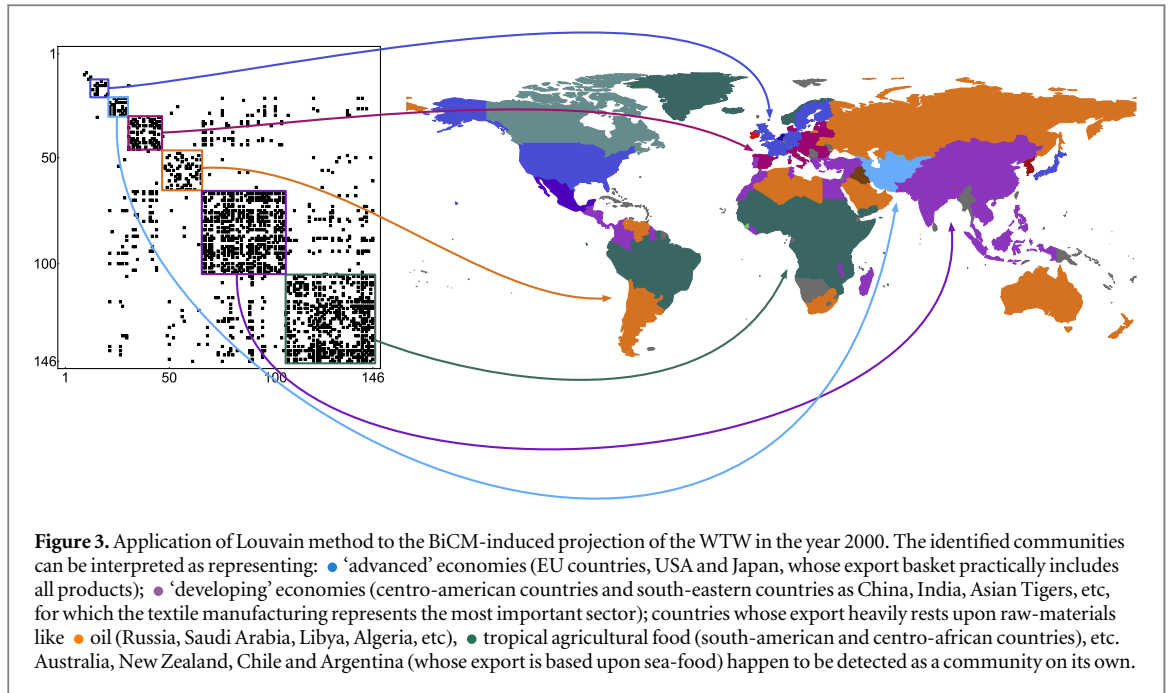
Figure 3. Application of Louvain method to the BiCM-induced projection of the WTW in the year 2000. The identified communities can be interpreted as representing: ● 'advanced' economies (EU countries, USA and Japan, whose export basket practically includes all products); ● 'developing' economies (centro-american countries and south-eastern countries as China, India, Asian Tigers, etc, for which the textile manufacturing represents the most important sector); countries whose export heavily rests upon raw-materials like ● oil (Russia, Saudi Arabia, Libya, Algeria, etc), ● tropical agricultural food (south-american and centro-african countries), etc. Australia, New Zealand, Chile and Argentina (whose export is based upon sea-food) happen to be detected as a community on its own.
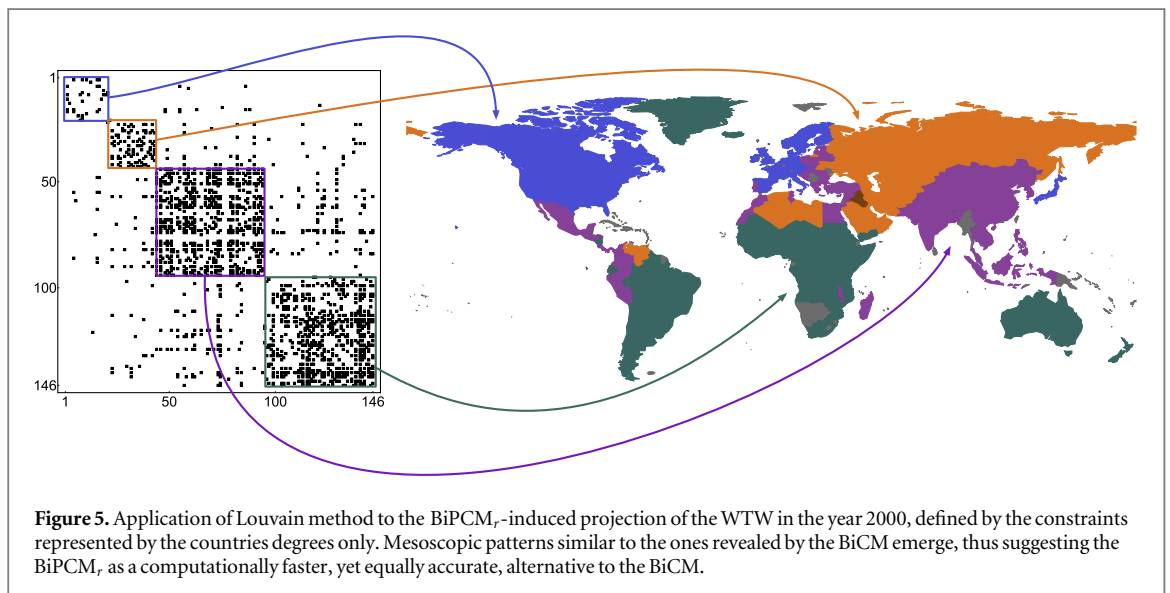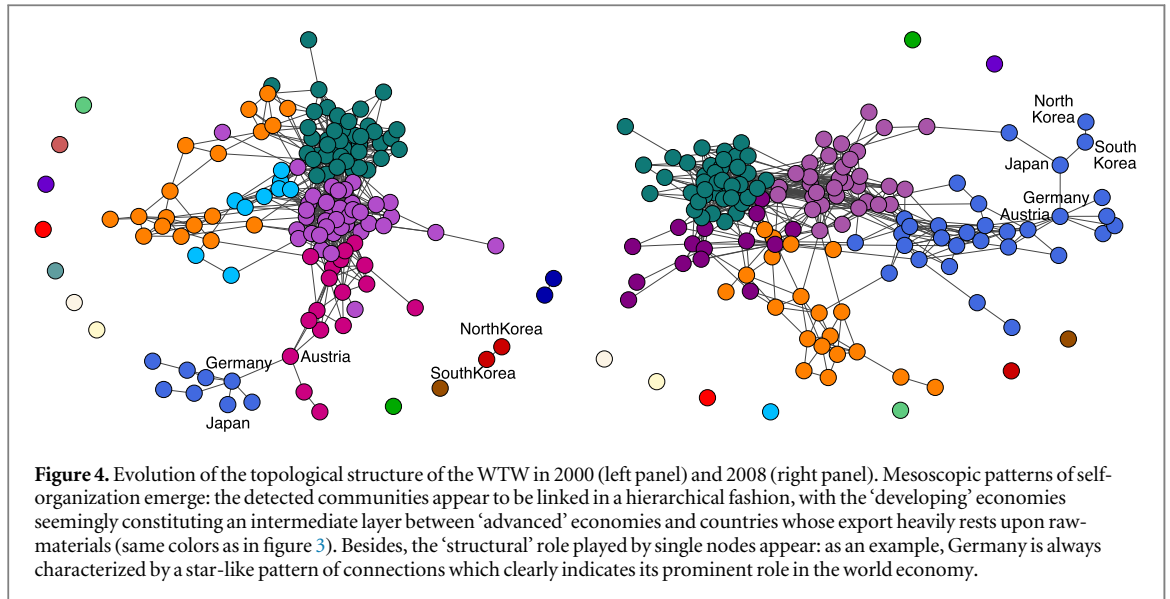
$$f_{\text{Bin}}(V_{rr'} = n) = \binom{N_C}{n}(p_{\text{BiRG}}^2)^n(1 - p_{\text{BiRG}}^2)^{N_C - n}. \tag{3.1}$$

The projection provided by the BiRG individuates a unique connected component of countries (notice that the two blocks at bottom-right and top-left of the panel are linked through off-diagonal connections) beside many disconnected vertices (the big white block in the center of the matrix). Interestingly, the latter represent countries whose economy heavily rests upon the presence of raw-materials (see also figure 3), in turn causing each export basket to be focused around the available country-specific natural resources. As a consequence, the similarity between these countries is not significant enough to allow the corresponding links to pass the validation procedure. In other words, the BiRG-induced projection is able to distinguish between two extreme levels of economic development, thus providing a meaningful, yet too rough, filter.

On the other hand, the BiCM-induced projection (shown in the third panel of figure 2), allows for a definite structure of clusters to emerge. The economic meaning of the detected diagonal blocks can be made explicit by running the Louvain algorithm on the projected network. As figure 3 shows, our algorithm reveals a partition into communities enclosing countries characterized by similar economic development [31]. In particular, we recognize the 'advanced' economies (EU countries, USA and Japan—whose export basket is practically constituted by all products [8, 30, 32–36]), the 'developing' economies (as centro-american countries and south-eastern countries as China, India, Asian Tigers, etc, for which the textile manufacturing represents the most important sector) and countries whose export heavily rests upon raw-materials like oil (Russia, Saudi Arabia, Libya, Algeria, etc), tropical agricultural food (south-american and centro-african countries), etc. An additional group of countries whose export is based upon sea-food is constituted by Australia, New Zealand, Chile and Argentina, which happen to be detected as a community on its own in partitions with comparable values of modularity.

Our algorithm is also able to highlight the structural changes that have affected the WTW topology across the temporal period considered for the present analysis. Figure 4 shows two snapshots of the WTW, referring to the years 2000 and 2008. While in 2000 EU countries were split into two different modules, with the north-european countries (as Germany, UK, France) grouped together with USA and Japan and the south-eastern european countries constituting a separate cluster, this is no longer true in 2008. Furthermore, the structural role played by single nodes is also pointed out. As an example, Austria and Japan emerge as two of the countries with highest betweenness, indicating their role of bridges respectively between western and eastern european countries and western and eastern world countries. A second example is provided by Germany, whose star-like pattern of connections clearly indicates its prominent role in the global trade.

The block diagonal structure of the BiCM-induced adjacency matrix reflects another interesting pattern of the world economy self-organization: the detected communities appear to be linked in a hierarchical fashion, with the 'developing' economies seemingly constituting an intermediate layer between the 'advanced' economies and those countries whose export heavily rests upon raw-materials. Interestingly, such a mesoscopic organization persists across all years of our data set, shedding new light on the WTW evolution.
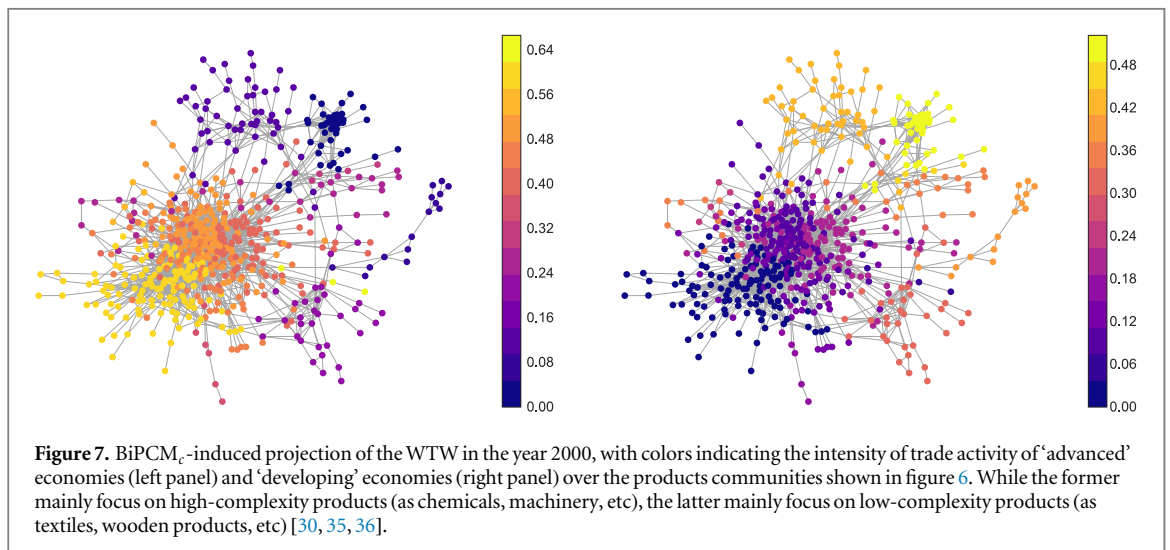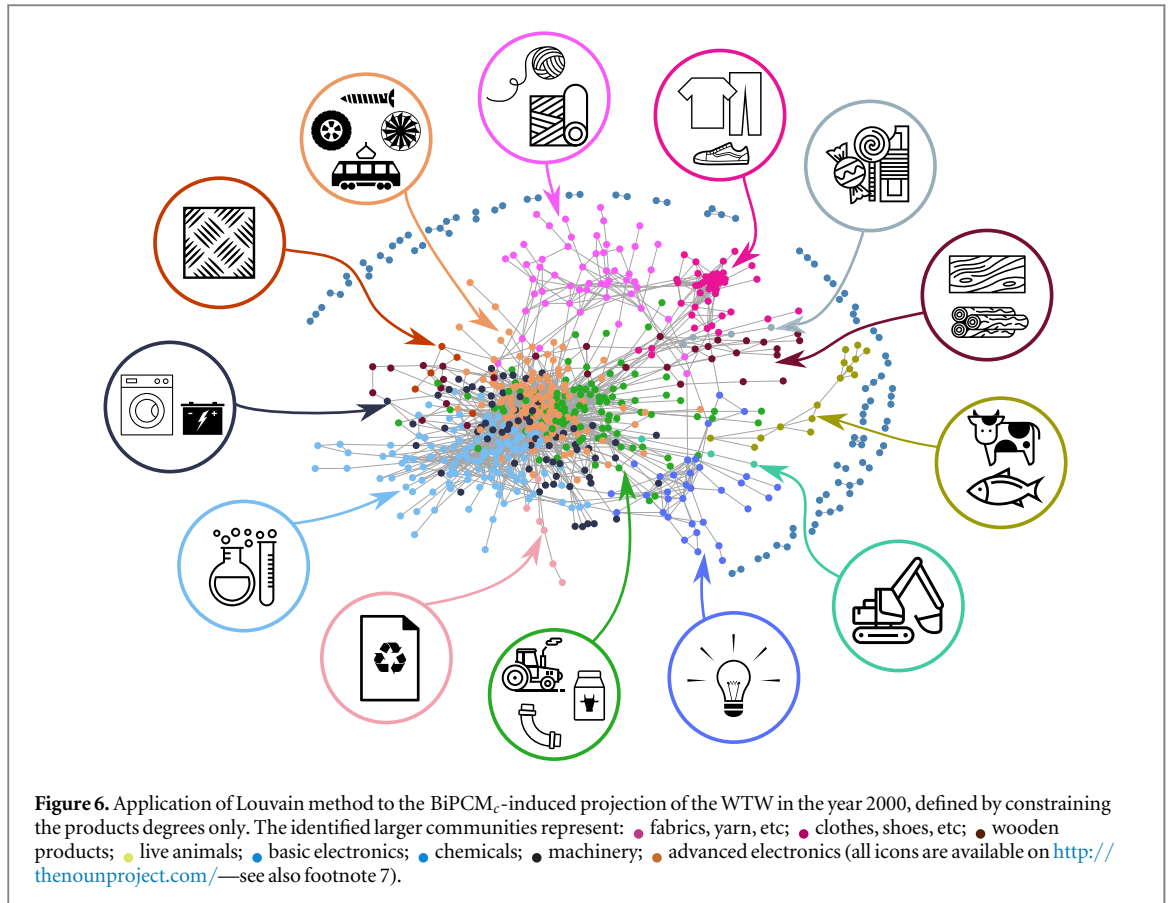
Figure 4. Evolution of the topological structure of the WTW in 2000 (left panel) and 2008 (right panel). Mesoscopic patterns of self-organization emerge: the detected communities appear to be linked in a hierarchical fashion, with the 'developing' economies seemingly constituting an intermediate layer between 'advanced' economies and countries whose export heavily rests upon raw-materials (same colors as in figure 3). Besides, the 'structural' role played by single nodes appear: as an example, Germany is always characterized by a star-like pattern of connections which clearly indicates its prominent role in the world economy.

Figure 5. Application of Louvain method to the $BiPCM_r$-induced projection of the WTW in the year 2000, defined by the constraints represented by the countries degrees only. Mesoscopic patterns similar to the ones revealed by the BiCM emerge, thus suggesting the $BiPCM_r$ as a computationally faster, yet equally accurate, alternative to the BiCM.

As shown in figure 5, the results obtained by running the $BiPCM_r$ (defined by constraining only the degrees of countries) are, although less detailed, compatible with the ones obtained by running the BiCM. In this case, the $BiPCM_r$ constitutes an approximation to the BiCM, providing a computationally faster, yet equally accurate, alternative to it. On the other hand, the $BiPCM_c$ induces a projection which is close to the BiRG one, thus adding little information with respect to the latter.
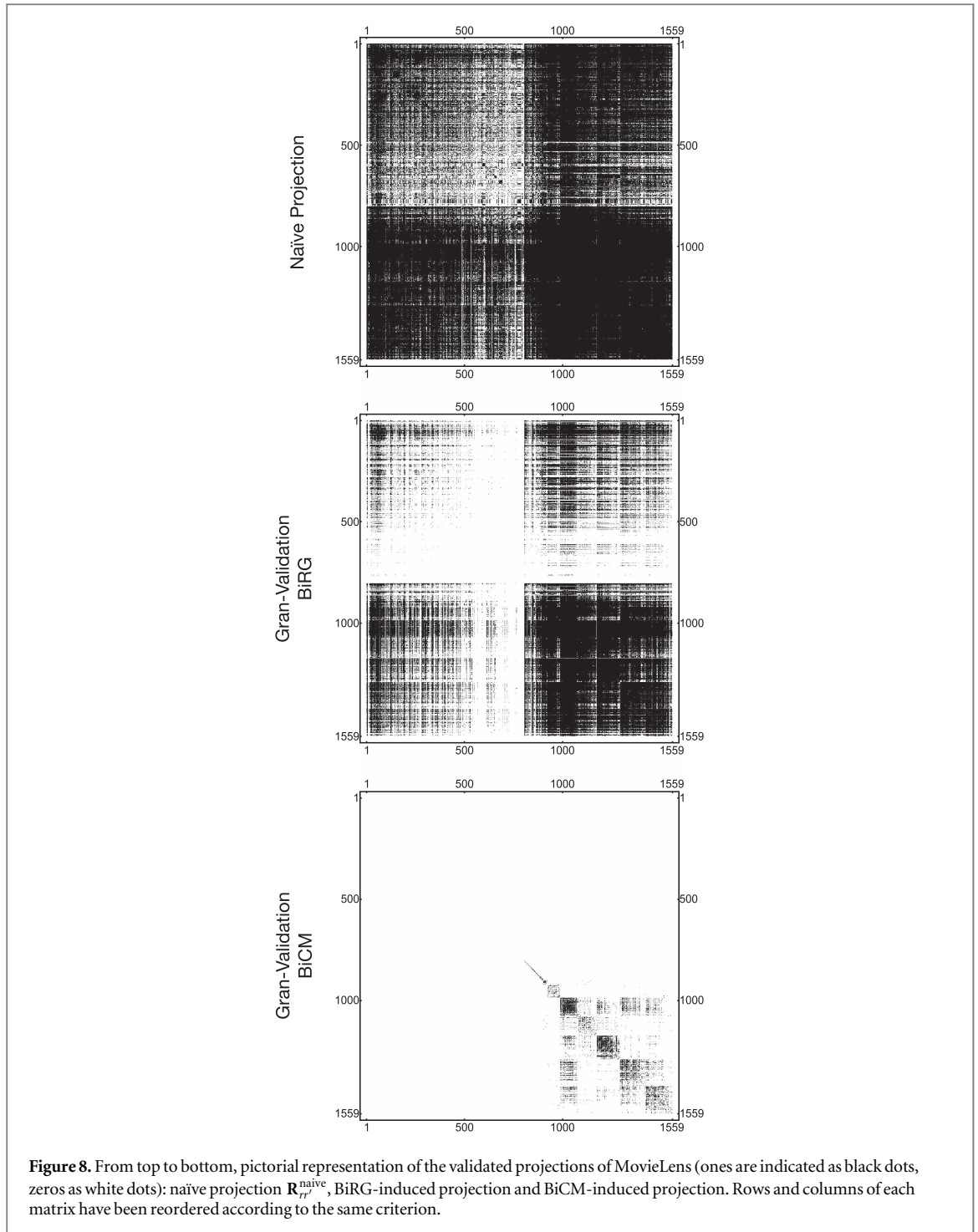
Products layer. While the BiCM provides an informative benchmark to infer the presence of significant connections between countries, this is not the case when focusing on products. For this reason, we consider the $BiPCM_c$, i.e. the null model defined by constraining only the products degrees: figure 6 shows the $BiPCM_c$-induced projection of the WTW on the layer of products[7]. Several communities appear, the larger ones being machinery, transportation, chemicals, electronics, textiles and live animals (a partition that seems to be stable across time).

The detected communities seem to be organized into two macro-groups: 'high-complexity' products (on the left of the figure), including machinery, chemicals, advanced electronics, etc and 'low-complexity' products (on the right of the figure), including live animals, wooden products, textiles, basic electronics, etc. This macroscopic

---

[7] 'Cow' by Nook Fulloption; 'Fish' by Iconic; 'Excavator' by Kokota; 'Light bulb' by Hopkins; 'Milk' by Artem Kovyazin; 'Curved Pipe' by Oliviu Stoian; 'Tractor' by Iconic; 'Recycle' by Agus Purwanto; 'Experiment' by Made by Made; 'Accumulator' by Aleksandr Vector; 'Washing Machine' by Tomas Knopp; 'Metal' by Leif Michelsen; 'Screw' by Creaticca Creative Agency; 'Tram' by Gleb Khorunzhiy; 'Turbine' by Luigi Di Capua; 'Tire' by Rediffusion; 'Ball Of Yarn' by Denis Sazhin; 'Fabric' by Oliviu Stoian; 'Shoe' by Giuditta Valentina Gentile; 'Clothing' by Marvdrock; 'Candies' by Creative Mania; 'Wood Plank' by Cono Studio Milano; 'Wood Logs' by Alice Noir from the Noun Project. All icons are under the CC licence.

Figure 6. Application of Louvain method to the BiPCM$_c$-induced projection of the WTW in the year 2000, defined by constraining the products degrees only. The identified larger communities represent: ● fabrics, yarn, etc; ● clothes, shoes, etc; ● wooden products; ● live animals; ● basic electronics; ● chemicals; ● machinery; ● advanced electronics (all icons are available on http://thenounproject.com/—see also footnote 7).

Figure 7. BiPCM$_c$-induced projection of the WTW in the year 2000, with colors indicating the intensity of trade activity of 'advanced' economies (left panel) and 'developing' economies (right panel) over the products communities shown in figure 6. While the former mainly focus on high-complexity products (as chemicals, machinery, etc), the latter mainly focus on low-complexity products (as textiles, wooden products, etc) [30, 35, 36].

separation reflects the level of economic development of the countries trading these products. As figure 7 clarifies, the 'advanced' economies focus their trading activity on products characterized by high complexity, while 'developing' economies are preferentially active on low-complexity products [30, 35, 36]. A simple topological index captures this tendency: $I_{\mathcal{R}\mathcal{C}} = \frac{\sum_{r\in\mathcal{R}}\sum_{c\in\mathcal{C}} m_{rc}}{|\mathcal{R}||\mathcal{C}|}$, i.e. the link density between groups of nodes $\mathcal{R}$ and $\mathcal{C}$, indicating one of the aforementioned communities of countries and one of the aforementioned communities of products, respectively. For example, as evident upon inspecting figure 7, 'advanced' economies (left panel) and 'developing' economies (right panel) are active on different clusters of products: while the trading activity of the former is mainly constituted by, e.g. chemicals and machinery, the latter mainly trade textiles, wooden products, etc. A more in-depth analysis of the grand canonical projection of the World Trade Web can be found in [37].

Figure 8. From top to bottom, pictorial representation of the validated projections of MovieLens (ones are indicated as black dots, zeros as white dots): naïve projection $\mathbf{R}^{\text{naive}}_{rr'}$, BiRG-induced projection and BiCM-induced projection. Rows and columns of each matrix have been reordered according to the same criterion.
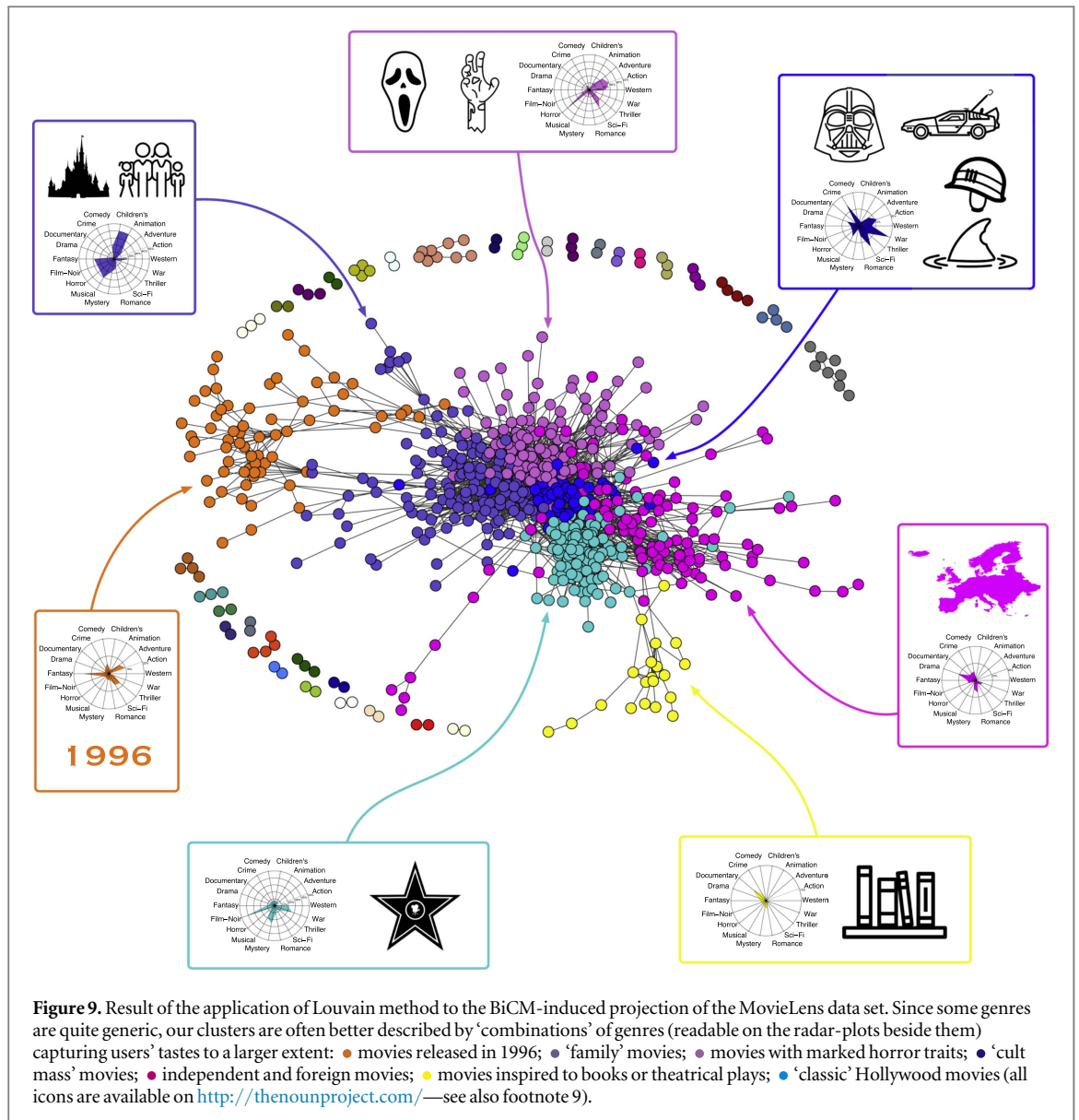
3.2. MovieLens

Let us now consider the second data set: MovieLens 100 k. MovieLens is a project by GroupLens [38], a research lab at the University of Minnesota. Data (collected from 19th September 1997 through 22nd April 1998) consist of 10^5 ratings—from 1 to 5—given by $N_C = 943$ users to $N_R = 1559$ different movies[8]; information about the movies (date of release and genre) and about the users (age, gender, occupation and US zip code) is also provided. We binarize the dataset by setting $m_{rc} = 1$ if user c rated movie r at least 3, providing a favorable recension.

In what follows we will be interested into projecting this network on the layer of movies. Figure 8 shows the three projections already discussed for the WTW. As for the latter, $\mathbf{R}^{\text{naive}}_{rr'} = \Theta[V_{rr'}]$ is still a very dense network, whose connectance amounts to 0.58. Similarly, the projection induced by the BiRG provides a rather rough

Figure 9. Result of the application of Louvain method to the BiCM-induced projection of the MovieLens data set. Since some genres are quite generic, our clusters are often better described by 'combinations' of genres (readable on the radar-plots beside them) capturing users' tastes to a larger extent: ● movies released in 1996; ● 'family' movies; ● movies with marked horror traits; ● 'cult mass' movies; ● independent and foreign movies; ● movies inspired to books or theatrical plays; ● 'classic' Hollywood movies (all icons are available on http://thenounproject.com/—see also footnote 9).

filter, producing a unique large connected component, to which only the most popular movies (i.e. the ones with a large degree in the original bipartite network) belong.

While both the naïve and the BiRG-induced projections only allow for a trivially-partitioned structure to be observed, this is not the case for the BiCM. By running the Louvain algorithm, we found a very composite community structure (characterized by a modularity of $Q \simeq 0.58$), pictorially represented by the diagonal blocks visible in the third panel of figure 8. The BiCM further refines the results found by the BiRG, allowing for the internal structure of the blocks to emerge: in our dicussion, we will focus on the bottom-right block, which shows the richest internal organization.

Figure 9 shows the detected communities within the aforementioned block, beside the genres (provided together with the data)[9]: action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Horror, Musical, Mystery, Noir, Romance, Sci-Fi, Thriller, War, western[10]. Since some genres are quite generic and, thus, appropriate for several movies (e.g. Adventure, Comedy and Drama), our clusters are often better described by 'combinations' of genres, capturing the users' tastes to a larger extent: the detected

---

[9] 'DeLorean' by Aaron Humphreys; 'Darth Vader' by Jake Dunham; 'Castle' by Olly Banham; 'Movie Star' by Nikita Kozin; 'Books on a Shelf' by Lucas Glenn; 'Shark' by Randomhero; 'Mask' by Gorka Cestao; 'Zombie Hand' by Valery; 'Army Helmet' by Henry Ryder; 'Family' by abeldb, from the Noun Project. All icons are under the CC licence.

[10] Every movie is assigned an array of 17 entries, representing the aforementioned genres. Each entry can be either zero or one, depending if that movie is considered as belonging to that genre or not (the number of ones in the vector can vary from 1 to a maximum of 6, if the selected film falls under several genres).

communities, in fact, partition the set of movies quite sharply, once appropriate combinations of genres are considered.

As an example, the orange block on the left side of our matrix is composed by movies released in 1996 (i.e. the year before the survey). Remarkably, our projection algorithm is able to capture the peculiar 'similarity' of these movies, not trivially related to the genres to which they are ascribed to (that are quite heterogeneous: Action, Comedy, Fantasy, Thriller, Sci-Fi) but to the curiosity of users towards the yearly new releases.

Proceeding clockwise, the violet block next to the orange one is composed by movies classified as Animation, Children's, Fantasy and Musical (e.g. 'Mrs. Doubtfire', 'The Addams Family', 'Free Willy', 'Cinderella', 'Snow White'). In other words, we are detecting the so-called 'family movies', a more comprehensive definition accounting for all elements described by the single genres above.

The next purple block is composed by genres Action, Adventure, Horror, Sci-Fi and Thriller: examples are provided by 'Stargate', 'Judge Dredd', 'Dracula', 'The Evil Dead'. This community encloses movies with marked horror traits, including titles far from 'mainstream' movies. This is the main difference with respect to the following blue block: although characterized by similar genres (but with Crime replacing Horror and Thriller) movies belonging to it are more popular: 'cult mass' movies, in fact, can be found here. Examples are provided by 'Braveheart', 'Blade Runner' and sagas as 'Star Wars' and 'Indiana Jones'.

The following two blocks represent niche movies for US users. The module in magenta is, in fact, composed by foreign movies (mostly European—French, German, Italian, English—which usually combine elements from Comedy and elements from Drama), as well as US independent films (as titles by Jim Jarmush); the yellow module, on the other hand, is composed by movies inspired by books or theatrical plays and documentaries.

The last, cyan block is composed by movies which are considered as 'classic' Hollywood movies (because of the presence of either iconic actors or master directors): examples are provided by 'Casablanca', 'Ben Hur', 'Taxi Driver', 'Vertigo' (and all movies directed by Hitchcock), 'Manhattan', 'Annie Hall'.

As in the WTW case, running the BiPCM$_r$ (defined by constraining only the degrees of movies) leads us to obtain a coarse-grained (i.e. still informative, although less detailed) version of the aforementioned results. Only three macro-groups of movies are, in fact, detected: 'authorial' movies (as 'classic' Hollywood movies, Hitchcock's, Kubrick's, Spielberg's movies), recent mainstream 'blockbusters' (as 'Star Trek', 'Star Wars', 'Indiana Jones', 'Batman' sagas) and independent/niche movies (as Spike Lee's and European movies).

As a final remark, we point out that projecting on the users layer with the BiCM indeed allows several communities to be detected. However, interestingly enough, none of them seems to be accurately described by the provided indicators (age, gender, occupation and US zip code), thus suggesting that users tastes are correlated with hidden (sociometric) variables yet to be individuated.

4. Discussion

Projecting a bipartite network on one of its layers poses a number of problems for which several solutions have been proposed so far [3, 8–10, 12–14, 32], differing from each other in the way the information encoded into the bipartite structure is dealt with.

The present paper proposes an algorithm that prescribes to, first, quantify the similarity of any two nodes belonging to the layer of interest and, then, link them if, and only if, this value is found to be statistically significant. The links constituting the monopartite projection are, thus, inferred from the co-occurrences observed in the original bipartite network, by comparing them with a proper statistical benchmark.

Since the null models considered for the present analysis retain a different amount of information, the induced projections are characterized by a different level of detail. In particular, the BiRG represents a very rough filter which employs the same probability distribution to validate the similarity between any two nodes, thereby preferentially connecting nodes with large degree than nodes with small degree. By enforcing stronger constraints (increasing the amount of retained information), stricter benchmark models are obtained.

The two partial configuration models constitute the simplest examples of benchmarks retaining also the information on the nodes degrees. However, it should be noticed that the two BiPCMs perform quite differently. In fact, the BiPCM constraining the degrees of the *opposite* layer we are interested in finding a projection of, provides an homogeneous benchmark as well (i.e. the same Poisson–Binomial distribution for all pairs of nodes —see also the appendix), whence the expected little difference with respect to the BiRG performance; on the other hand, the BiPCM constraining the degrees of nodes belonging to the *same* layer we are interested in finding a projection of, provides a performance which is halfway between the BiRG one and the BiCM one. The reason lies in the fact that a (Binomial) pair-specific distribution is now induced by the constraints, i.e. a benchmark properly taking into account the heterogeneity of the considered nodes. As shown in the results section, this often allows one to obtain an accurate enough approximation to the BiCM, i.e. the null model constraining the whole degree sequence.

As also suggested in [3], the use of a benchmark which ensures that the heterogeneity of all nodes is correctly accounted for is recommended: in other words, any suitable null model for projecting a network on a given layer should (at least) constrain the degree sequence of the same layer. The use of partial null models is allowed in case of constraints redundancy, e.g. when node degrees are well described by their mean (as indicated by the coefficient of variation, for example—see also the appendix): in cases like these, specifying the whole degree sequence is actually unnecessary.

As a final remark, we explicitly notice that implementing the BiCM can be computationally demanding: this is the reason why several approximations to the Poisson–Binomial distribution have been proposed so far. However, the applicability of each approximation is limited and, whenever employed to find the projection of a real, bipartite network, they may even fail to a large extent (see the appendix). With the aim of speeding up the numerical computation of the *p*-values induced by any of the null models discussed in the paper—while retaining the *exact* expression of the corresponding distributions—a Python code has been made publicly available by the authors at [25].

Remarkably, our method can be extended in a variety of directions, e.g. to analyze directed and weighted bipartite networks, and generalized to account for co-occurrences between more than two nodes, a study that constitutes the subject of future work.

Acknowledgments

Appendix A. The Poisson–Binomial distribution

The Poisson–Binomial distribution is the generalization of the usual Binomial distribution when the single Bernoulli trials are characterized by different probabilities.

More formally, let us consider N Bernoulli trials, each one described by a random variable x_i, $i = 1 \ldots N$, characterized by a probability of success equal to $f_{\text{Ber}}(x_i = 1) = p_i$: the random variable described by the Poisson–Binomial distribution is the sum $X = \sum_i x_i$. Notice that if all p_i are equal the Poisson–Binomial distribution reduces to the usual Binomial distribution.

Since every event is supposed to be independent, the expectation value of X is simply

$$\langle X \rangle = \sum_{i=1}^{N} p_i = \mu \tag{A.1}$$

and higher-order moments read

$$\sigma^2 = \sum_{i=1}^{N} p_i(1 - p_i), \quad \gamma = \sigma^{-3} \sum_{i=1}^{N} p_i(1 - p_i)(1 - 2p_i), \tag{A.2}$$

where σ^2 is the variance and γ is the skewness.

In the problem at hand, we are interested in calculating the probability of observing a number of V-motifs larger than the measured one, i.e. the *p*-value corresponding to the observed occurrence of V-motifs. This translates into requiring the knowledge of the survival distribution function (SDF) for the Poisson–Binomial distribution, i.e. $S_{\text{PB}}(X^*) = \sum_{X \geqslant X^*} f_{\text{PB}}(X)$. Reference [39] proposes a fast and precise algorithm to compute the Poisson–Binomial distribution, which is based on the characteristic function of the Poisson–Binomial distribution. Let us will briefly review the main steps of the algorithm in [39]. If we have observed exactly X^* successes, then

$$p\text{-value}(X^*) = S_{\text{PB}}(X^*) = \sum_{X \geqslant X^*} f_{\text{PB}}(X) = \sum_{X = X^*} \sum_{C_X} \left[\prod_{c_i \in C_X} p_{c_i} \prod_{c_j \notin C_X} (1 - p_{c_j}) \right], \tag{A.3}$$

where summing over C_X means summing over each set of X-tuples of integers.

The problem lies in calculating C_X. In order to avoid to explicitly consider all the possible ways of extracting a number of X integers from a given set, let us consider the Iinverse discrete Fourier transform of $f_{\text{PB}}(X)$, i.e.

$$\chi_l = \sum_{X=0}^{N} f_{\text{PB}}(X) e^{i\omega Xl}, \tag{A.4}$$

with $\omega = \frac{2\pi}{N+1}$. By comparing χ_l with the inverse discrete Fourier transform of the characteristic function of f_{PB}, it is possible to prove (see [39] for more details) that the real and the imaginary part of χ_l can be easily computed in terms of the coefficients $\{p_i\}_{i=1}^N$, which are the data of our problem: more specifically, if $z_i(l) = 1 - p_i + p_i \cos(\omega l) + \mathbf{i}(p_i \sin(\omega l))$, it is possible to prove that

$$\mathrm{Re}(\chi_l) = \mathrm{e}^{\sum\limits_{j=1}^{N} \log|z_j(l)|} \cos\left(\sum_{i=1}^{N} \arg[z_i(l)]\right), \tag{A.5}$$

$$\mathrm{Im}(\chi_l) = \mathrm{e}^{\sum\limits_{j=1}^{N} \log|z_j(l)|} \sin\left(\sum_{i=1}^{N} \arg[z_i(l)]\right), \tag{A.6}$$

where $\arg[z_i(l)]$ is the principal value of the argument of $z_i(l)$ and $|z_i(l)|$ represents its modulus. Once all terms of the discrete Fourier transform of χ_l (i.e. the coefficients $f_{PB}(X)$) have been derived, $S_{PB}(X)$ can be easily calculated. To the best of our knowledge, the approach proposed by [39] does not suffer from the numerical instabilities which, instead, affect [40].

Appendix B. Approximations of the Poisson–Binomial distribution

Binomial approximation. Whenever the probability coefficients of the N Bernoulli trials coincide (i.e. $p_i = p$ as in the case of the BiRG—see later), each pair-specific Poisson–Binomial distribution reduces to the usual Binomial distribution. Notice that, in this case, all distributions coincide since the parameter is the same.

However, the Binomial approximation may also be employed whenever the distribution of the probabilities of the single Bernoulli trials is not too broad (i.e. $\sigma/\mu < 0.5$): in this case, all events can be assigned the same probability coefficient $\overline{p}$, coinciding with their average $\overline{p} = \frac{\mu}{N}$. In this case,

$$S_{PB}(X) = S_{Bin}(X; \overline{p}, N), \tag{B.1}$$

where $S_{Bin}(X; \overline{p}, N)$ is the SDF for the random variable X following a Binomial distribution with parameter $\overline{p}$.

Whenever the aforementioned set of probability coefficients can be partitioned into homogeneous subsets (i.e. subsets of coefficients assuming the same value), the Poisson–Binomial distribution can be computed as the distribution of a sum of Binomial random variables [13]. Such an algorithm is particularly useful when the number of subsets is not too large, a condition which translates into requiring that the heterogeneity of the degree sequences is not too high. However, when considering real networks this is often not the case and different approximations may be more appropriate.

Poissonian approximation. According to the error provided by Le Cam's theorem (stating that $\sum_{X=0}^{N} |f_{PB}(X) - f_{Poiss}(X)| < 2\sum_{i=1}^{N} p_i^2$), Poisson approximation is known to work satisfactorily whenever the expected number of successes is small. In this case

$$S_{PB}(X) \simeq S_{Poiss}(X), \tag{B.2}$$

where the considered Poisson distribution is defined by the parameter μ [39].

Gaussian approximation. The Gaussian approximation consists in considering

$$S_{PB}(X) \simeq S_{Gauss}\left(\frac{X + 0.5 - \mu}{\sigma}\right), \tag{B.3}$$

where μ and σ have been defined in (A.1) and (A.2). The value 0.5 represents the continuity correction [39]. Since the Gaussian approximation is based upon the central limit theorem, it works in a complementary regime with respect to the Poissonian approximation: more precisely, when the expected number of successes is large.

Skewness-corrected Gaussian approximation. Based on the results of [41, 42], the Gaussian approximation of the Poisson–Binomial distribution can be further refined by introducing a correction based on the value of the skewness. Upon defining

$$G(x) \equiv S_{Gauss}(x) - \gamma\left(\frac{1 - x^2}{6}\right) f_{Gauss}(x), \tag{B.4}$$

where $f_{Gauss}(x)$ is the probability density function of the standard normal distribution and γ is defined by (A.2), then

$$S_{PB}(X) \simeq G\left(\frac{X + 0.5 - \mu}{\sigma}\right). \tag{B.5}$$

The refinement described by formula (B.4) provides better results than the Gaussian approximation when the number of events is small.

However, upon comparing the WTW projection (at the level $t = 0.01$, for the year 2000) obtained by running the skewness-corrected Gaussian approximation with the projection based on the full Poisson–Binomial

distribution, we found that $\simeq$20% of the statistically-significant links are lost in the Gaussian-based validated projection. The limitations of the Gaussian approximations are discussed in further detail in [42, 43].

Appendix C. Null models

C.1. BiRG model

The BiRG model is the random graph model solved for bipartite networks. This model is defined by a probability coefficient for any two nodes, belonging to different layers, to connect which is equal for all pairs of nodes. More specifically, $p_{\text{BiRG}} = \frac{L}{N_R \cdot N_C}$, where $L = \sum_{r=1}^{N_R} \sum_{c=1}^{N_C} m_{rc}$ is the observed number of links and N_R and N_C indicate, respectively, the number of rows and columns of our network. Since all probability coefficients are equal, the probability of a single V-motif (defined by the pair of nodes r and r' belonging to the same layer and node c belonging to the second one) reads

$$p(V_{rr'}^c) = p_{\text{BiRG}}^2. \tag{C.1}$$

Thus, the probability distribution of the number of V-motifs shared by nodes r and r' is simply a Binomial distribution defined by a probability coefficient equal to p_{BiRG}^2:

$$f_{\text{Bin}}(V_{rr'} = n) = \binom{N_C}{n}(p_{\text{BiRG}}^2)^n (1 - p_{\text{BiRG}}^2)^{N_C - n}. \tag{C.2}$$

C.2. Bipartite configuration model

The BiCM [16], represents the bipartite version of the configuration model [18–20]. The BiCM is defined by two degree sequences. Thus, our Hamiltonian is

$$H(\vec{\theta}, \vec{C}(\mathbf{M})) = \sum_{r=1}^{N_R} \alpha_r k_r + \sum_{c=1}^{N_C} \beta_c h_c, \tag{C.3}$$

where $k_r = \sum_{c=1}^{N_C} m_{rc}$ and $h_c = \sum_{r=1}^{N_R} m_{rc}$ are the degrees of nodes on the top and bottom layer, respectively; α_r and β_c, instead, are the Lagrangian multipliers associated with the constraints.

The probability of the generic matrix $\mathbf{M}$ thus reads

$$P(\mathbf{M}) = \frac{e^{-H(\vec{\theta}, \vec{C}(\mathbf{M}))}}{Z(\vec{\theta})}, \tag{C.4}$$

where $Z(\vec{\theta})$ is the grand canonical partition function. It is possible to show that

$$P(\mathbf{M}) = \prod_{r=1}^{N_R} \prod_{c=1}^{N_C} p_{rc}^{m_{rc}} (1 - p_{rc})^{(1 - m_{rc})}, \tag{C.5}$$

where

$$p_{rc} = p_{rc}(\alpha_r, \beta_c) = \frac{e^{-(\alpha_r + \beta_c)}}{1 + e^{-(\alpha_r + \beta_c)}} \equiv \frac{x_r y_c}{1 + x_r y_c} \tag{C.6}$$

is the probability for a link between nodes r and c to exist.

In order to estimate the values for x_r and y_c, let us maximize the probability of observing the given matrix $\mathbf{M}^*$, i.e. the likelihood function $\mathcal{L} = \ln P(\mathbf{M}^*)$ [19]. It is thus possible to derive the Lagrangian multipliers $\{x_r\}_{r=1}^{N_R}$ and $\{y_c\}_{c=1}^{N_C}$ by solving

$$k_r^* = \langle k_r \rangle = \sum_{c=1}^{N_C} \frac{x_r y_c}{1 + x_r y_c}, \quad h_c^* = \langle h_c \rangle = \sum_{r=1}^{N_R} \frac{x_r y_c}{1 + x_r y_c}, \tag{C.7}$$

where $\{k_r^*\}_{r=1}^{N_R}$, $\{h_c^*\}_{c=1}^{N_C}$ are the observed degree sequences.

C.3. Bipartite partial configuration models

Dealing with bipartite networks allows us to explore two 'partial' versions of the BiCM (hereafter BiPCM), defined by constraining the degree sequences of, say, the top and bottom layer separately. Let us start with the null model BiPCM$_r$, defined by the following Hamiltonian:

$$H(\vec{\theta}, \vec{C}(\mathbf{M})) = \sum_{r=1}^{N_R} \alpha_r k_r, \tag{C.8}$$

where $k_r = \sum_{c=1}^{N_C} m_{rc}$, $\forall r$ are the degrees of nodes on the top layer. Although the probability of the generic matrix $\mathbf{M}$ still reads

$$P(\mathbf{M}) = \prod_{r=1}^{N_R} \prod_{c=1}^{N_C} p_{rc}^{m_{rc}} (1 - p_{rc})^{(1-m_{rc})},\tag{C.9}$$

upon 'switching off' the multipliers $\{\beta_c\}_{c=1}^{N_C}$ the coefficient p_{rc} now assumes the form

$$p_{rc} = p_{rc}(\alpha_r) = \frac{e^{-\alpha_r}}{1 + e^{-\alpha_r}} \equiv \frac{x_r}{1 + x_r}, \forall c.\tag{C.10}$$

Notice that the BiCM probability coefficients in (C.6) exactly reduce to the ones in (C.10) whenever the degrees of all nodes belonging to the bottom layer coincide (i.e. $h_c \equiv h, \forall c$). However, BiPCM$_r$ provides an accurate approximation to the BiCM even when the values $\{h_c\}_{c=1}^{N_C}$ are characterized by a reduced degree of heterogeneity (e.g. as signaled by a coefficient of variation $c_v = s/m < 0.5$, with m and s being, respectively, the mean and the standard deviation of the bottom layer degrees).

In order to estimate the values for x_r, let us maximize the likelihood function $\mathcal{L} = \ln P(\mathbf{M}^*)$ again [19]. It is thus possible to derive the Lagrangian multipliers $\{x_r\}_{r=1}^{N_R}$:

$$k_r^* = \langle k_r \rangle = \sum_{c=1}^{N_C} \frac{x_r}{1 + x_r} \Longrightarrow p_{rc} = \frac{k_r^*}{N_C}.\tag{C.11}$$

Notice that, in this case,

$$p(V_{rr'}^c) = \frac{k_r^* k_{r'}^*}{N_C^2}\tag{C.12}$$

i.e. each V-motif defined by r and r' has the same probability, independently from c. This, in turn, implies that the probability distribution of the number of V-motifs shared by nodes r and r' is again a Binomial distribution defined as

$$f_{\text{Bin}}(V_{rr'} = n) = \binom{N_C}{n} \left(\frac{k_r^* k_{r'}^*}{N_C^2} \right)^n \left(1 - \frac{k_r^* k_{r'}^*}{N_C^2} \right)^{N_C - n}.\tag{C.13}$$

Let us now move to considering the second partial null model, BiPCM$_c$, defined by the Hamiltonian

$$H(\vec{\theta}, \vec{C}(\mathbf{M})) = \sum_{c=1}^{N_C} \beta_c h_c,\tag{C.14}$$

where $h_c = \sum_{r=1}^{N_R} m_{rc}, \forall c$ are the degrees of nodes on the bottom layer. The probability of the generic matrix $\mathbf{M}$ still factorizes, with the coefficient p_{rc} assuming the form

$$p_{rc} = p_{rc}(\beta_c) = \frac{e^{-\beta_c}}{1 + e^{-\beta_c}} \equiv \frac{y_c}{1 + y_c}, \forall r;\tag{C.15}$$

as for the previously-considered BiPCM, the BiCM probability coefficients in (C.6) exactly reduce to the ones in (C.15) whenever the degrees of all nodes belonging to the top layer coincide (i.e. $k_r \equiv k, \forall r$). Again, when the values $\{k_r\}_{r=1}^{N_R}$ are characterized by a reduced degree of heterogeneity, BiPCM$_c$ provides an accurate approximation to the BiCM.

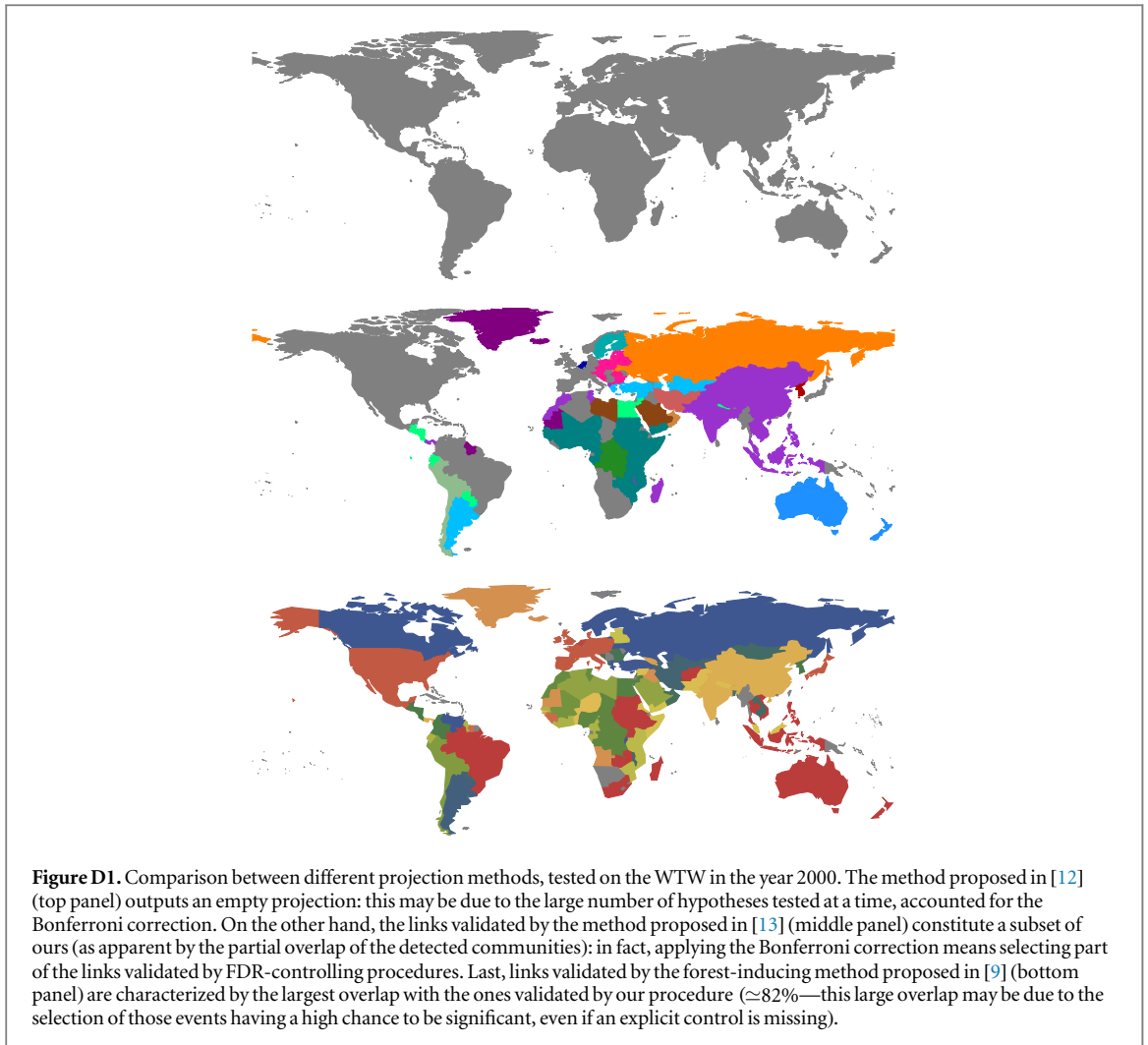The Lagrangian multipliers $\{y_c\}_{c=1}^{N_C}$ are again straightforwardly estimated as

$$h_c^* = \langle h_c \rangle = \sum_{r=1}^{N_R} \frac{y_c}{1 + y_c} \Longrightarrow p_{rc} = \frac{h_c^*}{N_R}.\tag{C.16}$$

In this case, each V-motif defined by r, r' and c has a probability which depends exclusively on c. As a consequence, the probability distribution of the number of V-motifs shared by any two nodes r and r' is the same one, i.e. a Poisson–Binomial whose single Bernoulli trial is defined by a probability reading

$$p(V_{rr'}^c) = \left(\frac{h_c^*}{N_R} \right)^2.\tag{C.17}$$

Appendix D. Comparing different projection algorithms

Available procedures suffer from a number of limitations that our method aims at overcoming. In what follows we compare the performance of some of them in projecting the WTW on the countries layer, for the year 2000, in greater detail, see figure D1 for the results of the comparison.

Figure D1. Comparison between different projection methods, tested on the WTW in the year 2000. The method proposed in [12] (top panel) outputs an empty projection: this may be due to the large number of hypotheses tested at a time, accounted for the Bonferroni correction. On the other hand, the links validated by the method proposed in [13] (middle panel) constitute a subset of ours (as apparent by the partial overlap of the detected communities): in fact, applying the Bonferroni correction means selecting part of the links validated by FDR-controlling procedures. Last, links validated by the forest-inducing method proposed in [9] (bottom panel) are characterized by the largest overlap with the ones validated by our procedure ($\simeq$82%—this large overlap may be due to the selection of those events having a high chance to be significant, even if an explicit control is missing).

The method proposed in [12] outputs an empty network for all years of our dataset: we suspect the reason to lie in the very large number of hypotheses tested at a time, leading to a too-severe correction. A similar result is obtained when applying the recipe proposed in [7]: only a tenth of links (among the group of advanced economies) are validated.

Although similar-in-spirit to ours, the method proposed in [13] prescribes to implement the Bonferroni correction as well. All links validated by applying this kind of correction are always a subset of the links validated when controlling for the FDR: this is the reason underlying the less informative community structure obtained when this algorithm is run on the WTW.

The third comparison we have explicitly carried out is the one with the forest-inducing method proposed in [9]. Links validated by such a method are characterized by the largest overlap ($\simeq$82%) with the ones validated by our procedure. This may be due to the selection of those events which have the higher chance to be significant (i.e. the largest number of shared co-occurrences): anyway, no statistical control is explicitly provided (e.g. the forest-like topology is not *per se* guaranteed to encode the most significant events).

As a final remark, we explicitly notice that the problem of spurious clustering does not affect our method, by definition. In fact, the presence of a node simultaneously connected to several nodes on the opposite layer does not imply the latter to be connected in the projection: this is the case if, and only if, the similarity between the involved nodes passes the test of statistical significance. An extreme example is provided by a network having a node c (on one layer) which is connected to every other node (on the opposite layer), projected by employing the BiCM: since the fully-connected node is, actually, a 'deterministic' node (its links are described by probability coefficients which are 1), any V-motif having it as a vertex (e.g. $V_{rr'}^c$) is deterministic as well. Thus, $P(V_{rr'} = 0) = 0$ (one V-motif is surely present) and the distribution describing the overlap between r and r' is shifted, as a whole, by one. In other words, the set of events which determine the presence of a link between r and r' does not include the deterministic V-motif (even more so, deterministic nodes can be discarded from the validation process carried out by the BiCM from the very beginning).

References

[1] Caldarelli G 2007 *Scale-free Networks: Complex Webs in Nature and Technology* (Oxford: Oxford University Press)

[2] Newman M 2010 *Networks: An Introduction* (Oxford: Oxford University Press)

[3] Neal Z 2014 *Soc. Networks* **39** 84

[4] Latapy M, Magnien C and del Vecchio N 2008 *Soc. Networks* **30** 31

[5] Watts D J and Strogatz S H 1998 *Nature* **393** 440

[6] Derudder B and Taylor P 2005 *Glob. Networks* **5** 71

[7] Serrano M A, Boguná M and Vespignani A 2009 *Proc. Natl Acad. Sci.* **106** 6483

[8] Caldarelli G, Cristelli M, Gabrielli A, Pietronero L, Scala A and Tacchella A 2012 *PLoS One* **7** e47278

[9] Zaccaria A, Cristelli M, Tacchella A and Pietronero L 2014 *PLoS One* **9** e113770

[10] Linden G, Smith B and York J 2003 *IEEE Internet Comput.* **7** 76

[11] Bonacich P 1972 *Soc. Methodologies* **4** 176

[12] Tumminello M, Micciché S, Lillo F, Piilo J and Mantegna R N 2011 *PLoS One* **6** e17994

[13] Gualdi S, Cimini G, Primicerio K, Di Clemente R and Challet D 2016 *Sci. Rep.* **6** 39461

[14] Dianati N 2016 arXiv:1607.01735

[15] Guillaume J-L and Latapy M 2003 arXiv:0307095

[16] Saracco F, Di Clemente R, Gabrielli A and Squartini T 2015 *Sci. Rep.* **5** 10595

[17] Diestel R 2006 *Graph Theory* (Heidelberg: Springer)

[18] Park J and Newman M E J 2004 *Phys. Rev.* E **70** 066117

[19] Garlaschelli D and Loffredo M I 2008 *Phys. Rev.* E **78** 015101

[20] Squartini T and Garlaschelli D 2011 *New J. Phys.* **13** 083001

[21] Fronczak A 2014 *Encyclopedia of Social Network Analysis and Mining* ed R Alhajj and J Rokne (New York: Springer) p 500

[22] Mastrandrea R, Squartini T, Fagiolo G and Garlaschelli D 2014 *New J. Phys.* **16** 043022

[23] Hoeffding W 1956 *Ann. Math. Stat.* **27** 713

[24] Wang Y H 1993 *Stat. Sin.* **3** 295

[25] Benjamini Y and Hochberg Y 1995 *J. R. Stat. Soc.* B **57** 289

[26] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 *J. Stat. Mech.* P10008

[27] Fortunato S 2010 *Phys. Rep.* **486** 75

[28] Staudt C L and Meyerhenke H 2016 *IEEE Trans. Parallel Distrib. Syst.* **27** 171

[29] Gaulier S G 2013 BACI: International Trade Database at the Product Level www.cepii.fr/anglaisgraph/workpap/pdf/2010/wp2010-23.pdf (last accessed: 19/04/2017)

[30] Tacchella A, Cristelli M, Caldarelli G, Gabrielli A and Pietronero L 2012 *Sci. Rep.* **2** 723

[31] Di Clemente R, Chiarotti G L, Cristelli M, Tacchella A and Pietronero L 2014 *PLoS One* **9** e112525

[32] Hidalgo C A, Klinger B, Barabási A-L and Hausmann R 2007 *Science* **317** 482

[33] Hidalgo C A and Hausmann R 2009 *Proc. Natl Acad. Sci.* **106** 10570

[34] Hausmann R and Hidalgo C 2010 *HKS Working Paper No. RWP* 10-045

[35] Cristelli M, Gabrielli A, Tacchella A, Caldarelli G and Pietronero L 2013 *PLoS One* **8** e70726

[36] Tacchella A, Cristelli M, Caldarelli G, Gabrielli A and Pietronero L 2013 *J. Econ. Dyn. Control* **37** 1683

[37] Straka M J, Saracco F and Caldarelli G 2017 arXiv:1703.04090

[38] Resnick P, Iacovou N, Suchak M, Bergstrom P and Riedl J 1994 *Proc. 1994 ACM Conf. on Computer Supported Cooperative Work, CSCW '94* (New York: ACM) pp 175–86

[39] Hong Y 2013 *Comput. Stat. Data Anal.* **59** 41

[40] Chen S X *et al* 1998 *Ann. Stat.* **26** 1894

[41] Deheuvels P, Puri M L and Ralescu S S 1989 *J. Multivariate Anal.* **28** 282

[42] Volkova A Y 1996 *Theory Probab. Appl.* **40** 791

[43] Mikhailov V G 1990 *Theory Probab. Appl.* **38** 479