CrossMark

From Ecology to Finance (and Back?): A Review on Entropy-Based Null Models for the Analysis of Bipartite Networks

Mika J. Straka[1] · Guido Caldarelli[1,2,3,4] ·
Tiziano Squartini[1] · Fabio Saracco[1]

Abstract Bipartite networks provide an insightful representation of many systems, ranging from mutualistic networks of species interactions to investment networks in finance. The analyses of their topological structures have revealed the ubiquitous presence of properties which seem to characterize many—apparently different—systems. Nestedness, for example, has been observed in biological plant-pollinator as well as in country-product exportation networks. Due to the interdisciplinary character of complex networks, tools developed in one field, for example ecology, can greatly enrich other areas of research, such as economy and finance, and vice versa. With this in mind, we briefly review several entropy-based bipartite null models that have been recently proposed and discuss their application to real-world systems. The focus on these models is motivated by the fact that they show three very desirable features: analytical character, general applicability, and versatility. In this respect, entropy-based methods have been proven to perform satisfactorily both in providing benchmarks for testing evidence-based null hypotheses and in reconstructing unknown network configurations from partial information. Furthermore, entropy-based models have been successfully employed to analyze ecological as well as economic systems. As an example, the application of entropy-based null models has detected early-warning signals, both in economic and financial systems, of the 2007–2008 world crisis. Moreover, they have revealed a statistically-significant export specialization phenomenon of country export baskets in international trade, a result that seems to reconcile Ricardo's hypothesis in classical economics with recent findings on the (empirical) diversification industrial production at the national level. Finally, these null models have shown that the information contained in the nestedness is already accounted for by the degree sequence of the corresponding graphs.

✉ Mika J. Straka
   mika.straka@imtlucca.it

[1] IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca, Italy

[2] Istituto dei Sistemi Complessi, CNR, Dip. Fisica Università "Sapienza", P.le A. Moro 2, 00185 Rome, Italy

[3] London Institute of Mathematical Sciences, 35a South St, Mayfair, London W1K 2XF, UK

[4] European Centre for Living Technology, University of Venice, San Marco 2940, 30124 Venice, Italy

 Springer

1 Introduction

"Data is the New Oil" has become the unofficial slogan for the enthusiasts of technological progress in the recent decade [95]. New data sources have created new economic and political possibilities. Catalyzed by the need for new analytical and numerical tools, the theory of *complex networks* has gained much attention, since the interplay between different data agents can often be expressed in the shape of a network, and new methods for the analyses of such structures have been designed.

A prominent network type found in many real-world systems is the so-called *bipartite network*, which is characterized by the presence of two distinct types of nodes. Examples are user-movie databases, plant-pollinator ecosystems, author-article collaborations, or financial bank-asset networks. Although purely data-based analyses provide valuable insights into interaction mechanisms, recent results have shown that such network structures contain more information than is apparent at first sight. In particular, several techniques have been designed based on statistical physics and information theory, which provide the possibility to filter statistically relevant signals from the network that otherwise remain hidden when the data is taken at face value [50,81,82,87,90].

Network theory is by nature interdisciplinary and has thus created a vast vocabulary and a plethora of tools. Due to the interaction patterns of many biological systems, the analysis of bipartite networks has been very popular in ecology and its methodologies have spread to other ares of research. Here, we present a brief review of insights that have been gained in the areas of ecological, economic, and financial networks. Our focus, however, lies on bipartite network modeling, with a particular attention to *entropy-based null models* and their applications. We shall show that seemingly genuine network characteristics, such as nestedness, can be traced back to basic properties like the degree sequence of the nodes. For this purpose, appropriately defined benchmarks models are used that are as unbiased and general as possible. Furthermore, we review how such null models can be used to reconstruct networks when only partial or noisy data is available, and how this method can be applied to asses systemic risk in financial networks [31,50,87]. In addition, we illustrate how Ricardo's specialization hypothesis in international trade [77] can be reconciled with the apparently contradicting export diversification signal that has been observed [21], and how early warning signs are revealed preceding the financial crisis of 2008 [81].

Network theory has found wide-spread applications in different fields of scientific research. In this article, we focus on *bipartite undirected networks*, in which we can distinguish between two distinct types of nodes. They are ordered in two separate layers such that links only exists between, but not within, layers. The structure of the network can be expressed in a *biadjacency matrix*, with nodes of one type along the rows and nodes of the other type along the columns. Due to the ubiquity of network structures in science, different fields have created different vocabularies. For instance, in ecology the biadjacency matrix is commonly known as the *interaction matrix* when the interaction between species is studied. Analogously, research on the occurrence of organisms in different environments uses the *presence-absence matrix*. Furthermore, in economic and financial networks one may use the expression *ownership matrix*. We will use the general term *biadjacency matrix* in the

following. Regarding the number of connections attached to each node, we refer to them as *degrees*, which is more common than *marginal totals* in ecology.

In the following sections, we review major insights that have emerged from the study of bipartite networks. We shall focus initially on several open problems in ecological, economic, and financial networks. Subsequently, we present entropy-based null models for bipartite structures that have appeared in recent literature. Finally, we shall discuss how their application has helped in finding solutions for the aforementioned issues.

2 Open Problems in Ecological Networks

The analysis of networks has a long tradition in the field of biology and ecology. Research on food webs, for instance, can be dated back to the pioneering works of Elton in 1927 [38]. Food webs capture the predator-prey relationships between different species: squirrels eat plants but are hunted by snakes, which fall prey to foxes. Directed links in these networks express the flow of biomass, and species can be order in hierarchical layers (known as trophic levels) according to their position in the food chain.

Ecology studies the interactions among species, or between species and their natural environments. Some typical examples are plants and pollinators, or organisms and their habitats. In these cases, one can distinguish between two different types of nodes that populate two distinct layer of a bipartite network. If the interactions between the species or environments are mutually beneficial and cooperative, for example in the case of pollinators and plants, such bipartite networks are often referred to as *mutualistic networks*.

2.1 Bipartite Motifs

Motifs are defined as n-node subgraphs that are overrepresented in empirical networks and have been labeled as "the building blocks of complex networks" [64]. In directed networks, such as food webs, the smallest nontrivial motifs can be built out of three nodes, leading to 13 distinct patterns [64]. Different motifs are assumed to serve different functions in the network. In genetic transcription networks, for example, is has been observed that certain motifs regulate the expression of genes [4] (for an overview of the motifs and their function, see, e.g., [85]).
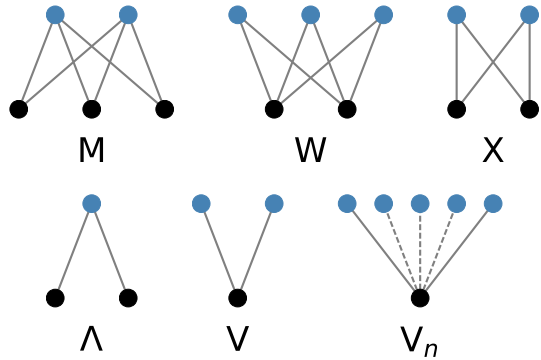
Analyzing networks from ecology, engineering, biochemistry and neurobiology, in [64] it has been observed that different network types show distinct motif abundances. Hence, the question arises whether one can predict global network characteristics from the presence and temporal changes of such structures. Finding motifs in monopartite networks can generally be computationally intensive and different algorithms have been proposed (see [98] for a survey).

Here, we will concentrate on bi-cliques, i.e. motifs in undirected bipartite networks. We shall use the vocabulary presented in [80], since, in our opinion, this nomenclature makes it easier to grasp the shape of the motifs. As an example, the M-, W-, X-, as well as the V-, Λ-, and the V_n-motifs, are shown in Fig. 1.

The simplest motifs are the bi-cliques $K_{1,2}$ and $K_{2,1}$, also known as Λ- and V-motifs, that are composed of two nodes in the same and one node in the opposite layer. They draw exactly a "Λ" and a "V" between the layers, as illustrated in Fig. 1.

Let us call the two layers L and Γ and indicate the respective nodes with $i \in L$ and $\alpha \in \Gamma$. A binary bipartite network can be represented by a *biadjacency matrix*, i.e. a rectangular matrix **M** of dimension $N_L \times N_\Gamma$, whose elements $m_{i\alpha}$ are 1 if nodes i and α are connected,

Fig. 1 Illustration of several undirected bipartite motifs. The nomenclature is based on the visual shape of the structures. Top: closed motif in which all nodes of one layer are connected to those of the other. Bottom: open motif that capture node similarities in terms of common nearest neighbors in the opposite bipartite layer

0 otherwise. We can easily express the number of V-motifs between the nodes i and j of the upper layer L as

$$V^{ij} = \sum_{\alpha \in \Gamma} m_{i\alpha} m_{j\alpha}, \quad i, j \in L. \tag{1}$$

The $\Lambda_{\alpha\beta}$-motifs are defined analogously for the nodes of the lower layer, $\alpha, \beta \in \Gamma$. V^{ij} captures the number of neighbors that the node couple (i, j) has in common. The motifs can be easily generalized to more than two nodes by including n legs that are all attached to the same node in the opposite layer, as shown in Fig. 1. We will call them V_n and Λ_n (with $V = V_2$ and $\Lambda = \Lambda_2$), or in standard graph theory $K_{2,n}$ and $K_{n,2}$. V- and Λ-motifs thus represent the number of connections shared between 2 or more nodes belonging to the same layer.

A more complex class of motifs is represented by the so called *closed motifs*. The M-, W- and X-motifs are illustrated on the top in Fig. 1 and are referred to as $K_{2,2}$, $K_{3,2}$ and $K_{2,3}$ in graph theory. We can express them in terms of the biadjacency matrix and write, for instance, for the total number of X-motifs

$$X = \sum_{i<j} \sum_{\alpha<\beta} m_{i\alpha} m_{j\alpha} m_{i\beta} m_{j\beta}. \tag{2}$$

The other mentioned closed motifs can be described similarly.

Bipartite motifs can even account for non-existing links, which is the case, for example, of the popular *checkerboards*, introduced by Diamond [32] for the study of the avifauna of Vanuatu's islands. A checkerboard considers the case of mutual exclusions of two species. The total number of checkerboards is in the biadjacency matrix is thus

$$C = \sum_{i<j} \sum_{\alpha<\beta} m_{i\alpha}(1 - m_{j\alpha})(1 - m_{i\beta}) m_{j\beta}. \tag{3}$$

Togetherness, T, is defined in a similar way and counts how many times two species interact together with the same species, avoiding, at the same time, the interaction with other ones. In formulas,

$$T = \sum_{i<j} \sum_{\alpha<\beta} m_{i\alpha} m_{j\alpha}(1 - m_{i\beta})(1 - m_{j\beta}). \tag{4}$$

In [89], the authors show that C and T differ by a constant term.

As a final comment to the present section, note that although all the motifs so far involve several links, they are all multi-linear in the corresponding biadjacency matrix. This fact is particular convenient for analytical calculations, as we will see in the following.

2.1.1 Motifs Analysis in Mutualistic Networks

In ecology, finding patterns that explain the distribution of species in different habitats, e.g. islands, forest, or even parasitic hosts, or why certain organisms interact with others, is a major concern. In this context, the frequency of motifs permits to highlight hidden structures in the architecture of the biological system. A clear example is the case of the study of the avifauna of Vanuatu's islands, [32]. Considering birds and their island habitats as the two layers of a bipartite network, the abundance of checkerboards is analyzed in order to understand co-existing behaviors. This question triggered a long debate about the correct null model to test statistically significance of the measurements [27,33,48,78]. An agreement was achieved with the work of [78], who proposed a rewiring randomization in which the degree sequence is fixed for both layers, as well as the average bird population of the islands bird species occupy. The authors observe a statistically significant abundance of checkerboards, suggesting a peculiar colonization pattern that increases the mutual exclusions of some species.

Checkerboards and togetherness in a bipartite network can be measured using, for example, the package released by Dormann et al. [35].

2.2 Nestedness

From the study of ecological systems, the insight has emerged that species in sites of lower biodiversity also populate environments with larger biodiversity. This concept is called *nestedness* and in mutualistic networks translates into the fact that specialists' interactions, i.e. organisms that interact only with a small number of other species, are a subset of those of generalist organisms. This phenomenon is also present in other biological bipartite systems [8]: for example, it has been observed that species occupying different island of an archipelago describe a nested structure. Larger islands are those able to sustain a greater number of species and smaller islands host just a subset of them. It has been hypothesized that this observation is due to an original ecosystem comprising all islands, which has subsequently been divided into an archipelago by disruptive geological events. From a general heterogeneity of species in the original environment, the limited resources of the new ecosystems force some species to (local) extinction, if their population exceed the sustainable limit on the respective island. Thus, smaller islands host a subset of the species hosted by bigger islands, depending on the sustainable population of different species. Studying the effect of a fragmented habitat on the population of animals permits to design the best strategies for the conservation of rare species [23,34,41,63].

The concept of nestedness is reflected in the structure of the biadjacency matrix: rows and columns can be sorted in such a way that the matrix is approximately triangular, as shown in Fig. 2. The role of such a structure is debated, as we will see in the following, but nevertheless it is constantly present in different mutualistic or antagonistic system.

In the last decades, several metrics to capture the nestedness phenomenon have been proposed in literature, with the first attempt dating back to the *nestedness temperature* [8]. After ordering rows and columns in the biadjacency matrix into a state of "maximum packing", a line is drawn on the matrix representing the boundary of the expected fully nested matrix. Then, a quantity called "temperature" is defined by considering the absence in the packed part and the presence in the empty side of interactions, weighted by their distance from the boundary. In [3], the authors show that the nestedness temperature is not maximal for disordered system, since random matrices have a intermediate value of nestedness, and proposed the NODF ("Nestedness metric based on Overlap and Decreasing Fill") to solve the problem.

Fig. 2 Illustration of three different matrices of the same dimensions and number of links (filled squares). The left-most matrix can be packed more densely into a triangular shape than the other two and has the highest nestedness. Notice how "shorter" rows (columns) are completely contained in "longer" rows (columns). The nestedness clearly decreases to the right. Figure courtesy of [67] published under CC license

Using the definition Eq. (1), let us define the number of V-motifs for two nodes (i, j) with different degrees as $\tilde{V}_{ij} = (1 - \delta_{k_i, k_j}) V_{ij}$, where δ is the Kroenecker delta and k_i the degree of node i. The NODF can be written as

$$\text{NODF} = 2 \frac{\sum_{i<j} \frac{\tilde{V}_{ij}}{\min\{k_i, k_j\}} + \sum_{\alpha<\beta} \frac{\tilde{\Lambda}_{\alpha\beta}}{\min\{k_\alpha, k_\beta\}}}{N_\Gamma(N_\Gamma - 1) + N_L(N_L - 1)}, \tag{5}$$

where $\tilde{\Lambda}_{\alpha\beta}$ is analogous to $\tilde{V}_{ij}$ and defined on the layer Γ. Although Eq. (5) may seem slightly mysterious at first sight, we can observe some interesting properties. First, it is independent of the order of elements in the matrix and receives a contribution from each node couple of the same layer, if the respective nodes have different degrees. Second, normalizing each contribution by the minimum degree restricts it value to the range 0 (no overlap between neighbors) to 1 (if all the neighbors of the node with the smaller degree are also shared by the higher degree node). Finally, the overall denominator normalizes the NODF by accounting for the number of couples considered in the calculation. It can be shown that such a normalization induces a bias in the NODF towards the longer layer [80]: if, without loss of generality, $N_L \gg N_\Gamma$, then the main contribution to the NODF will come from the layer L. Because of this peculiarity, sometimes the nestedness is captured by considering separately the contribution from each of the two layers. Nevertheless, some scholars argued about the possibility of disregarding the contribution of couple of nodes with the same degree and, for instance, Bastolla et al. [13] provided a different measure considering such terms.

For nested species abundances in different habitats, a certain agreement about the role of nestedness has been reached: the geological division of a rich original ecosystem in an archipelago forces species to survive with the limited resources of smaller islands. Thus, species on smaller islands are more exposed to the risk of extinction, and the surviving species are a subset of those hosted on bigger islands [8]. Instead, the role of nestedness for the properties of ecological mutualistic networks underwent a deeper debate. On the one hand, it has been argued that nestedness generally increases biodiversity by reducing competition [13] and favors the stability of the network [94]. On the other hand, the authors of [88] claim that nested interactions are inherently less stable compared to random interactions. Although predator-prey interactions seem to stabilize the networks, mutualistic and competitive interactions do not [2].

An important contribution to the discussion was put forward in [91]: the authors show that the attempts of a species to increase its abundance in a mutualistic network drives the system to a more nested configuration. In this scenario, species abundances start from general initial

conditions and growth is shown to be higher if the number of mutualistic interactions is lower. Moreover, the abundance of the rarest species is connected to the resilience of the network, i.e. the speed at which the system, after small perturbations, returns to an equilibrium.

Despite these efforts, no consensus about the importance of nestedness has yet been reached. Moreover, also the origin of nestedness is highly debated: James et al. [57] show that the correlation between persistence and nestedness is present when nestedness correlates even with the connectance of the network. Hence, it is not clear which variable , i.e. whether connectance or nestedness, captures at best the properties of the system.

2.3 Monopartite Projections and Communities

When studying mutualistic networks, the question naturally arises whether one can find groups of highly cooperative species, or groups of organisms that compete for the same resources. In plant-pollinator networks, an example for the former would be a community of plants and pollinators that live in symbiosis and benefit from cooperation. Contrary to that, an example for the latter would be a collection of insects that compete for the same pollen. In ecology, these substructures are referred to as *compartments* [93]. In the following, we shall adopt the network vocabulary and call them *modules* or *communities*. They describe collections of nodes that are more closely related to each other than to individuals in other communities.

The problem of finding communities between nodes of the same layer can be found throughout different fields of complex networks analysis, from ecological, to financial, to economic networks. For this problem, several tools have been presented in the literature (for an overview, see, e.g. [42]). A popular approach is to perform a *monopartite projection*, i.e. to project the bipartite network on one of its layers. In the resulting graph, nodes are connected if they share at least one neighbor in the original bipartite network. Note that the procedure discards parts of the original information—in general, it is not possible to reconstruct the original bipartite network from the projection. Moreover, there is no clear guideline on how to set the link weights in the projection. It has been shown that the communities found in binary projections can be incorrect and misleading and that weighted projections should generally be preferred [51]. Nonetheless, simply setting the weights equal to the number of neighbors in the original network is quantitatively biased [102]. Inspired by the importance of collaborations in the author-article network of scientific coauthors, Newman proposed that links in the author projection should be corrected by a factor $1/(d-1)$, where d is the degree of the collaboration paper [68,69]. Despite these efforts, a systematic exploration of how weight should be set remains open. At the same time, the question of which links carry statistically relevant information is neglected.

3 Open Problems in Economic Networks

Seminal works in classical economics date back to Adam Smith's fundamental "The Wealth of Nations" in 1776 [86]. In the wake of Smith's publication, David Ricardo devoted parts of his intellectual endeavors to economics, which culminated in his famous "Principles of Political Economy and Taxation" [77]. His most important legacy is probably the concept of *comparative advantage*, which expresses the fact that some nations can produce certain products more efficiently than others. As a result, Ricardo advocated the idea that nations should concentrate their resources only on their most advantageous industries. According to

him, combining industrial specialization with free trade would be favorable for all countries and foster national economic growth.

Nowadays, international exportations and importations are recorded on a yearly basis and made available by the UN Comtrade Database.[1] This allows us to scrutinize trade relations and test hypotheses of classical economics with the help of state-of-the-art tools in data analysis and network theory. In fact, the global structure of trade interactions can be expressed as the so-called International Trade Network (ITN), also known as World Trade Web, in which nodes correspond to countries and link weights to trade volumes in US$. Countries can share directed links with different weights, corresponding to products of different categories.

International trade is one of the main global stages on which countries interact, and the ITN has been extensively studied due to its importance for economic growth and to address questions like globalization and the spreading of economic shocks [36]. For example, regarding the number of trade partners, it has been shown that the network is generally disassortative, i.e. that countries with few trade partners tend to interact with nations with many partners [46,83]. When trade volumes are taken into account, however, it has been observed that high-degree countries trade most intensively with other high-degree countries [40]. Although product-specific trade volumes are very heterogeneous [12], the aggregate link weights distribution is almost log-normal [6,12]. Country-specific trade volumes depend strongly on national GDP and their distributions reach from truncated log-normality to Pareto-log-normality [6].

Trade can also be studied at an even finer level, when links are drawn among regional industries instead of countries. Using the World Input-Output Database, it has been shown that global production systems are still regionally organized and industries are asymmetrically connected, leading to possible shock amplification from regional fluctuations to the global scale [24].

3.1 Diversification in Trade

Recent developments on the ITN have been triggered by the suggestion that trade networks should be considered as bipartite with countries in one layer and products in the other layer [45]. The setup is illustrated in Fig. 3. The proposal is motivated by the observation that importers and exporters have intrinsically different motivations for connecting to trade partners [45]. In particular, the rationale is that a strong exporter should also be a good producer, such that the productive capabilities of a country can be inferred from its exported products. This statement, however, has its caveats: for instance, a strong exporter may indeed be good producer, but the inverse may not necessarily be true. In fact, isolated nations have more problems in exporting wares to other countries, even if the internal production prospers. Despite these particular cases, in general studying the exports of different countries can give us information about their industrial production.

For their analyses, the authors of [45] have made use of methodologies developed for mutualistic networks and analyzed the properties of the country-product network using the *revealed comparative advantage* (RCA), also knows as Balassa index [11]. The RCA compares the relative monetary importance of a particular product among all exports of a country (its *export basket*) to the global average and assigns a value to each link accordingly, as explained in detail in Appendix A. By pruning links sequentially for different RCA threshold values, in [45] the authors separate the core and periphery of the network and showed that degree distributions are truncated power laws. The networks emerging from the pruning

---

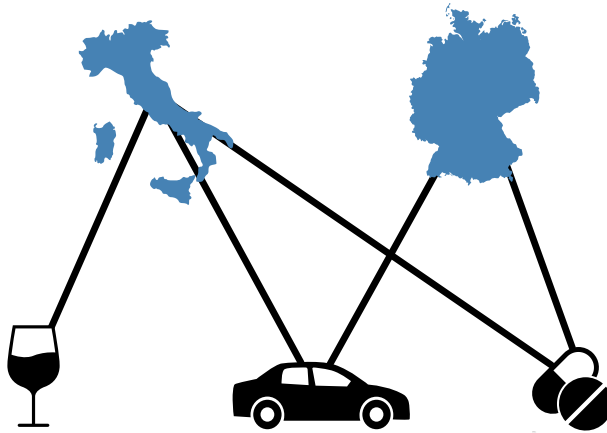[1] The Comtrade Database can be found at https://comtrade.un.org/.

Fig. 3 Illustration of a part of the country-product exportation network. Both Italy and Germany are strong exporters of cars and pharmaceutical products, whereas only Italy has a comparative advantage in wine production ("Wine" by Thengakola, "Car" and "pills" by alrigel from the Noun Project. All icons are under the CC license)

procedure are considered as binary, since each existing link expresses the fact that a certain country is a relevant exporter of a particular product at some threshold value.

A fundamental observation that emerges from the binarized ITN, when only relevant exportations with RCA ≥ 1[2] are kept, is the approximately triangular structure of its biadjacency matrix, as illustrated in Fig. 4: some countries have large export basket and others have small ones, just like some product have only few exporters and others many. The crucial fact is that the smaller export baskets are contained in the bigger ones. The ITN therefore exhibits the nestedness property [22,29,53–55,92,101], which we have already observed for mutualistic networks in the previous section. In the context of the bipartite trade network, this observation is striking, since it contradicts theories of classical economics. As mentioned above, according to Ricardo one would expect a specialization of exportation, which should be observable through a block-diagonal structure of the biadjacency matrix. Instead, the matrix is approximately triangular, as shown in Fig. 4, which corresponds to an increasing diversification of exportations, as has also been mentioned in [21]. The most developed countries export all products, from the most sophisticated to the most basic ones, whereas less developed countries are able to export just few low technology items.

3.2 Product and Country Space

A considerable amount of work on the bipartite trade network has been devoted to the analyses of relations among products and among countries. An intuitive approach would be to project the bipartite network on its two layers, respectively. However, this approach is generally problematic—in fact, in the case of the ITN the projected networks are almost completely connected with link densities of over 93% [82], leading to trivial properties.

---

[2] As we will appreciate better in the following, setting the threshold for the RCA values to 1 is quite natural. In fact, the RCA does nothing more than to compare the observed export values to their expectations provided by an approximated weighted configuration model, which is known as CAPM (capital asset pricing model) in the financial context, or "lift" in data science. Thus, the threshold RCA ≥ 1 means that an observed value is greater than its expectation.
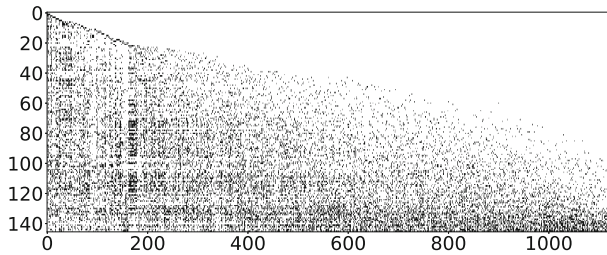
Fig. 4 Biadjacency matrix of the International Trade Network for the year 2000 with countries sorted from top to bottom and products from left to right in increasing fitness and complexity, respectively. Links in the network are shown as black dots. The overall triangularity of the matrix is correlated with the nestedness of the system

To address this question, in [22] the authors apply Minimal Spanning Forests to the country and the product projections. Unexpectedly, they find that neighboring countries compete over the same market rather than diversifying their export baskets [22].

A different approach has been chosen by Hidalgo et al. [55], who construct the "product space" by connecting products that are similar according to a specific metric. The distance between two products is essentially measured as the conditional probability that a country exports both of them as measured on the data [55]. They observe that more sophisticated goods, such as vehicles and machinery, occupy the core of the network, whereas less sophisticated ones, e.g. vegetables or crude oil, populate the periphery. Given the topology of the product space, they argue that less developed countries get trapped in the periphery because of a lack of connections to the more valuable products in the core [55].

Another proposal for inferring the possible evolution of the industrialization of countries is proposed in [101]: from the binary bipartite network of trade, the authors are able to obtain a forest of products, discounting the degree sequence of both layers.

All methods revised here do not rely on the application of unbiased null model, but use different ingredients in order to highlight possible dynamics that may describe the industrialization of countries. None of them discusses the statistical significance of their findings, but rather use some of the features of the bipartite network to propose an explanation for their observations. In order to correctly project the information contained in the bipartite network, more involved methodologies are needed.

3.3 Economic Complexity

The bipartite structure of the ITN encodes information on non-tradable capabilities of countries [5], such as their infrastructure, education system, patent rights, and industry-specific knowledge. As mentioned above, the fact that a country is capable of exporting a certain product signals that its industry is advanced enough to compete in global markets [5]. Consequently, the country has the necessary latent capabilities to manufacture the product, which should be reflected in the composition of its export basket.

In order to capture how the industrial capabilities shape national economies, Hidalgo and Hausmann proposed the so-called *method of reflections* [53,54]. This method is linear in the biadjacency matrix and consists of iteratively assigning a quantity to each node that depends on those of its neighbors and their degrees. As the authors point out, the resulting "complexities" of countries correlate with their GDPs. Unfortunately, the convergence of the algorithm is not guaranteed [76].

This problem was remediated in [22,29,92] and a non-linear recursive algorithm was proposed, which gave rise to the so-called *economic complexity* framework. The capabilities of countries were labeled as their *fitness* and the level of sophistication of the products as their *complexity*. Although some convergence issues are still present, it has been shown that fitness and complexity rankings of countries and products are stable even in absence of convergence [76].

In [30] the authors study the evolution of countries in terms of their fitness (intangible assets assessing competitiveness) and GDP per capita (GDPpc, a monetary measure). They observe a strong heterogeneity in the country dynamics and identify several regimes, such as a "poverty trap" in the low fitness regime, and a laminar region for high fitness countries. In conclusion, they argue that the overall heterogeneous evolution dynamics cannot be assessed with classical regression tools and that methods from dynamical systems theory would be more appropriate [30].

In a recent study, the evolution of products has been analyzed in an analogous way [5]. Similar to countries, the dynamic of products is observed in the complexity-logPRODY space, with logPRODY being a monetary measure defined as the average weight of a product exporter's GDPpc [5]. As the authors observe, products tend to move towards an asymptotic zone with product-specific asymptotic markets. Interestingly, the asymptotic markets seem to be determined by the product complexities and are characterized by high competition [5].

Even though the study of the International Trade Network has enjoyed much attention in the last decade, it is striking that the topic of early warning signals detection in economic networks has received relatively little attention. This comes as a surprise, since financial and trade relation are strongly connected: in the aftermath of the crisis, world merchandise exports fell by 22% [100]. As we shall see in the next section, the problem of detecting early structural changes in economic systems can be tackled by adopting the same methodology employed for financial systems, i.e. by comparing the observed networks with properly-defined benchmarks.

4 Open Problems in Financial Networks

Financial institutions form a global system of investments and money lending. In the aftermath of the 2008 financial crisis, correctly assessing systemic risk and shock propagation has become a top priority for policy makers and regulators. Contrary to previous beliefs, the financial network has revealed itself to be more unstable than expected due to the complex structure of its connections [7,14,19,25,59].

Financial stress can be transmitted through two main channels: direct exposure due to bilateral agreements, such as credit swap contracts [49], and indirect exposure due to portfolio overlaps [1,37,44]. Whereas the first gives rise to a monopartite inter-bank network, the second presents itself naturally as a bipartite financial institutions-assets network.

Interest in the inter-bank network has surged in the fields of public administration and academic research ever since the bankruptcy of Lehman Brothers and the subsequent turmoil. An important contribution of network theory has been to shift the paradigm from the dogma "too big to fail" to "too central to fail" [15]. To quantify the financial risk associated to different institutions including network effects, the so-called "DebtRank" was introduced in [15].

Indirect exposures, on the other hand, can be created through bank portfolio overlaps. In a bank-asset network, financial institutions are ordered along one layer and assets (or asset

classes) along the other. Financial contagion can be created through the effects of *fire sales* spillovers: a sudden drop in the value of an asset can trigger a cascade of sell-orders, which leads to asset illiquidity [20,28,49,50,84,87]. This phenomenon can put banks in distress, who may react by selling other assets, thereby causing further devaluation dynamics.

In an recent article, a dynamical model for the analysis of shocks in the bank-asset network has been presented and applied to the Venezuelan banking system [60]. The authors show that their model is able to capture temporal changes in the structure of the network and that some assets with small capitalization can cause significant global shocks [60]. Fire sale spillovers have also been analyzed by [49], who have introduced a metric to asses the systemic risk of the bank-asset network.

Despite these significant advancements, the analysis of financial network is often hindered by a lack of detailed data. The model in [60], for instance, uses balance sheets for the model construction – but often, such information is available only in aggregate and detailed asset holdings are undisclosed. Many tools of financial analysis therefore rely on aggregate data, resulting in unrealistically dense networks and a biased underestimation of systemic risk [87]. As a consequence, improved methods are necessary that reconstructed the network in a more realistic way while avoiding systematic bias [87].

4.1 Systemic Risk

Be $\mathbf{W}$ the weighted biadjacency matrix of a bipartite network in which connections carry weights. We can obtain the node strengths by summing over the rows and columns, respectively. Let us index the banks with $i \in L$ and the assets with $\alpha \in \Gamma$. In the financial context, the vertex strengths are often described as the *total asset size* of a bank (or *market value* of their portfolio), $V_i = \sum_\alpha w_{i\alpha}$, and the *market capitalization* of an asset, $C_\alpha = \sum_i w_{i\alpha}$ [31,87]. It is possible to remap the matrix entries in such a way that banks choose their portfolio weights proportional to their market value and the asset's capitalization:

$$w_{i\alpha}^{CAPM} = \frac{V_i C_\alpha}{w}, \tag{6}$$

where we have used $w = \sum_{i',\alpha'} w_{i'\alpha'}$ This model is called *capital asset pricing model* (CAPM, [61,66]). As has been shown in [31] and is illustrated in Fig. 5, these matrix weights give a good approximation of the systemic risk of the system measured in terms of the metric introduced by [49], despite the fact that networks of return price correlations show little agreement with real cases [16,17]. However, without the use of a null model, little can be said about the precision of the risk predictions [31].

5 Bipartite Exponential Random Graph

Statistical null models can be used as comparison benchmarks in order to verify whether real systems show unusual properties. For this purpose, they should be unbiased and formulated as general as possible. This notwithstanding, null models may maintain certain characteristics of the empirical network that should be discounted. Let us define the set of all possible networks, called *ensemble*, with the same node composition as the empirical network, but different number of links. To each member of the ensemble we will assign a probability that depends on some property of the original network that should be maintained. By comparing the ensemble characteristics with the actual network, we can observe how constraining certain information on the ensemble captures some features of the real system. If the quantities measured on
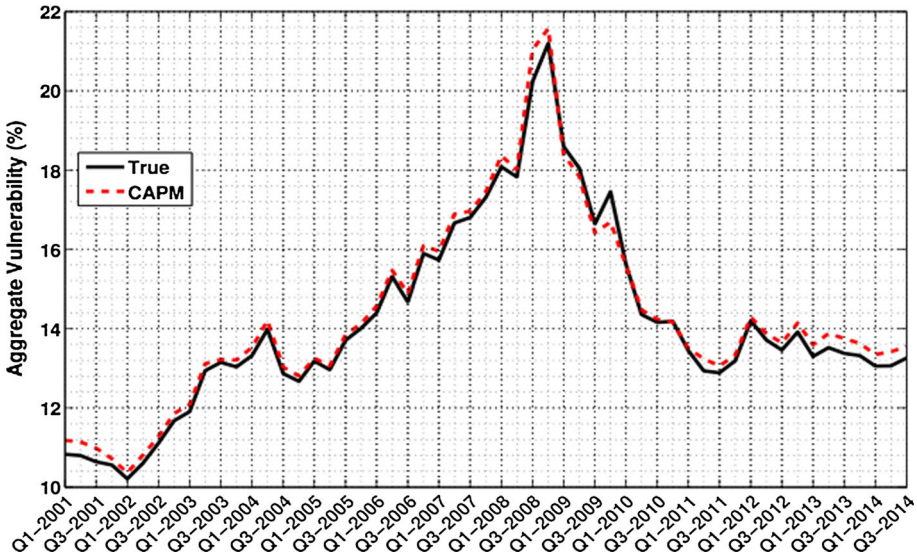
Fig. 5 Aggregate vulnerability of the bank-asset network calculated with the metric calculated by Greenwood et al. [49] using the whole data (solid black line) and the CAPM matrix weights, which require only the node strengths (dotted red line). Figure courtesy of [31] (Color figure online)

the real network are correctly reproduced, then the constraints are enough to explain such a behavior. If, however, the real network shows a statistically significant deviation, this information cannot be explained by the constraints and represent a non-trivial information about the structure of the empirical network. In these sections, we review the extension of the *exponential random graph model* (ERGM) to bipartite networks, motivated by the fact that its monopartite version has enjoyed considerable success in the past [47,58,71]. More detailed derivations of the null models can be found in the Appendix B and in [80,82,90].

5.1 Bipartite Erdős-Rényi Random Graph

Let us consider an empirical binary bipartite network, expressed by its biadjacency matrix $\mathbf{M}^*$ with layer dimensions N_L and N_Γ. Quantities measured on the real system will be marked with an asterisk. We start by constructing the set of all possible networks with the same layer dimensions: this set, the *ensemble* $\mathcal{G}_B$, runs from the empty network (without any links) to the fully connected network (in which all possible $N_L \times N_\Gamma$ links are realized).

Be $G_B \in \mathcal{G}_B$ an element of the ensemble $\mathcal{G}_B$ of bipartite networks with fixed layer dimensions. The most general and unbiased probability distribution over the ensemble can be obtained by maximizing the Shannon entropy [71], defined as

$$S = - \sum_{G_B \in \mathcal{G}_B} P(G_B) \ln \left(P(G_B) \right). \tag{7}$$

Assume now that we have measured some quantities $\mathbf{C}$ of the real network, for example the number of edges. We want the corresponding ensemble expectation value $\langle \mathbf{C} \rangle$ to reflect the same value, i.e. we constrain the expectation value of the observable in such a way that $\langle \mathbf{C} \rangle \equiv \mathbf{C}^*$. It can be shown that the probability of observing the generic graph $G_B \in \mathcal{G}_B$ is,

as in [58],

$$P(G_{\mathrm{B}}|\boldsymbol{\theta}) = \frac{e^{-\boldsymbol{\theta} \cdot \mathbf{C}(G_{\mathrm{B}})}}{\mathcal{Z}(\boldsymbol{\theta})}, \tag{8}$$

where $\boldsymbol{\theta}$ is the vector of Lagrange multipliers associated to the constraints $\mathbf{C}$, and $\mathbf{C}(G_{\mathrm{B}})$ is the value of the constraints on the graph G_{B}. $\mathcal{Z}(\boldsymbol{\theta})$ is the partition function known from statistical physics,

$$\mathcal{Z}(\boldsymbol{\theta}) = \sum_{G_{\mathrm{B}}' \in \mathcal{G}_{\mathrm{B}}} e^{-\boldsymbol{\theta} \cdot \mathbf{C}(G_{\mathrm{B}}')}, \tag{9}$$

and its exponent the graph Hamiltonian, $\mathcal{H} = \boldsymbol{\theta} \cdot \mathbf{C}$ [71]. Assuming that the network quantities $\mathbf{C}$ can be expressed analytically in terms of the biadjacency matrix, $\mathbf{C} \equiv \mathbf{C}(\mathbf{M})$, we can see that the probability $P(G_B|\boldsymbol{\theta})$ for a given graph only depends on the Lagrange multipliers. The trick is to derive their values by maximizing the likelihood of observing the real network in the ensemble, $\mathcal{L} \equiv \ln P(G_B^*)$ [47]. This is equivalent to explicitly imposing $\langle \mathbf{C} \rangle = \mathbf{C}^*$ on the ensemble. Summarizing, the ERGM formulation essentially relies on two steps: first, the maximization of the entropy under some constraints, and second the likelihood maximization. The result of the former is a functional relation that describes how the probabilities per link depend on the Lagrangian multipliers. In the latter, on the other hand, the Lagrangian multipliers are explicitly calculated, essentially deriving them from the real network. In principle, different approaches can be used in order to obtain the values of the Lagrange multipliers $\boldsymbol{\theta}$; the likelihood maximization permits to find the most similar configuration to the actual network.

The formalism above extends the exponential random graph [71] to networks with bipartite structure. It is well known that constraining the number of links E^* in the ERGM framework returns the Erdős-Rényi random graph [39]. In analogy, imposing the same constraints on the ensemble $\mathcal{G}_B$ gives us the *bipartite random graph* (BiRG, as it is called in [82]), in which all links have the same probability $p = E/N_L N_\Gamma$. The derivation of the BiRG is shown in the Appendix B.

5.2 Bipartite Configuration Model

In the previous paragraphs, we have derived the bipartite exponential random graph by focusing on the simplest constraint possible, the total number of edges E. In this section, we shall proceed by considering the degree sequence as a local, node specific property.

In the realm of monopartite networks, the *configuration model* (CM, [26,65,68,71]) has enjoyed a variety of application. It is constructed using the degree sequence of the real network and can be extended to the bipartite case, giving rise to the *bipartite configuration model* (BiCM, [80]). Constraining the degree sequence of both layers, L and Γ, corresponds to imposing two series of Lagrange multipliers, $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$, respectively. With some algebra, it can be shown that the probability distribution becomes [80]

$$P(G_B|\boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{i,\alpha} (p_{\mathrm{BiCM}})_{i\alpha}^{m_{i\alpha}} \left(1 - (p_{\mathrm{BiCM}})_{i\alpha}\right)^{1-m_{i\alpha}}, \tag{10}$$

where the probability per link reads

$$(p_{\mathrm{BiCM}})_{i\alpha} = \frac{e^{-(\theta_i + \rho_\alpha)}}{1 + e^{-(\theta_i + \rho_\alpha)}}, \quad i \in L, \alpha \in \Gamma \tag{11}$$

The Lagrange multiplicators can be recovered through the maximization of the likelihood $\mathcal{L}$ of observing the empirical network, as shown in Appendix B.

Note that Eq. (10) factorizes into the single link probabilities. This is very convenient for analytical calculations, for example for the multi-linear bipartite motifs, such as the V-motif in Eq. (1) and the X-motif in Eq. (2).

Other less strict null models can be defined through the relaxation of the constraints. For instance, imposing only the degree sequence of one layer leads to the *bipartite partial configuration model* (BiPCM, [82]). The choice of the null model generally depends on the information that one wishes to discount.

5.2.1 Validated Projections

The BiCM can be used to safely project the information contained in the bipartite network on one network layer, discounting the information from the degree sequence [50,82,90]. The idea is to compare the observed co-occurrence of links between nodes on the same layer (or, otherwise stated, the number of V-motifs between nodes on the same layer) with the expectations from the null model. In the BiCM, given a node couple, the probabilities for the V-motifs insisting on them are independent and, in general, different [82]. Thus, the distribution of such V-motifs is Poisson-Binomial, the generalization of a binomial distribution to independent events with different success probability [56]. The comparison between the observations with the expectation of the null model can be captured by a p-value, such that we obtain one p-value for every node couple: p-values are analyzed through a multiple hypothesis test and only statistically significant V-motif abundances are validated, leading to a monopartite network with only statistically relevant links. This methods is known in literature as the *grand canonical projection algorithm* and is described in detail in [50,82,90].

5.3 Bipartite Configuration Models for Weighted Networks

The ERGM bipartite framework can be easily extended from binary to weighted networks. In weighted bipartite networks, nodes are characterized by their degrees and strengths, i.e. the sum over the weight of their edges. One may intuitively be inclined to extend the BiCM to its weighted counterpart, the *bipartite weighted configuration model* (BiWCM [31], see Appendix B), by simply exchanging the degree with strength constraints. However, it has been shown for monopartite networks that such a null model performs very badly in capturing the essential features of the actual network [62]. This is due to the fact that it ignores the information on the network topology that is contained in the binary degree sequence. As the authors of [62] point out, non-trivial degree and strength sequences complement each other in the network reconstruction. The constraints should thus be modified accordingly: in Sect. 8 we shall see how a bipartite system guides us to the definition of a proper null model in the financial context.

6 Results in Ecological Networks

6.1 Degree Sequence in Biological Bipartite Networks

Entropy-based approaches for the analysis of biological systems are well present in literature [9,52], but they have rarely been employed for the analysis of bipartite networks. Contrary to this general tendency, Williams [97] has used the aforementioned BiRG to assess the significance of the degree distribution in mutualistic networks. The author has sampled the ensemble of the BiRG and compared the observed degree distribution with the frequencies

expected from the null model by implementing the likelihood ratio statistics. The calculation is repeated for every element of a sample of the BiRG ensemble and the values are compared. The comparison shows that the degree distribution of mutualistic networks, besides being strongly skewed, can be usually explained just by the total number of links. The result is even more striking, considering that its monopartite analogous has not such a good performance [96].

6.2 Nestedness (Reprise)

The study of nestedness has enjoyed considerable attention in ecology. Yet despite those efforts, doubts about its origin and significance persist, as we have seen in Sect. 2. To address this matter, Johnson et al. [67] highlight that a null model should be implemented in order to state if the nestedness of a particular system is a genuine quantity, or whether it is already captured by the degree sequence. By applying an approximation of the configuration model presented in [26,65,68,71], which is valid for sparse networks, they argue that nestedness naturally derives from degree heterogeneities and disassortative degree-degree correlation, i.e. the tendency of low degree nodes to connect to high degree nodes. As they point out, finite null models, such as the approximated configuration model [26,65,68,71], tend to be dissassortative and nested. Studying 60 empirical networks, they conclude that in almost 90% of them the observed nestedness can be described by a degree-conserving null model.

The results of Johnson et al. [67] have been refined and re-interpreted in a recent work by Payrató Borrás et al. [72], which has been published during the writing of the present review. Based on a comparison between the NODF values and the BiCM expectations of 167 different biological systems, they conclude that the observed nestedness can indeed be explained by the null model and are thus accounted for by the degree sequence.

7 Results in Economic Networks

7.1 Network Validation in Trade

In the following paragraphs, we review some recent results that have been obtained through the application of the binary BiCM in the area of international trade.

To compare the empirical network with the null model, we shall make us of z-scores. Be $Q(G_B)$ a quantities that we can measure on the network G_B. Eq. (11) gives us the tools to calculate its expectation value $\langle Q \rangle$ and the standard deviation σ_Q on the BiCM ensemble. The z-score is defined as

$$z(Q) = \frac{Q^* - \langle Q \rangle}{\sigma_Q} \tag{12}$$

and expresses the discrepancy between the observed and the expectation value in terms of standard deviations.

7.1.1 Nestedness and Specialization

The bipartite export network of countries and products is nested, meaning that smaller export baskets are contained in larger ones [22,29,53–55,92,101]. As we have seen in the previous paragraph, by using the BiCM it is possible to show that the nestedness of mutualistic networks is explained by their degree sequence [67,72]. The same idea has been tested for the trade
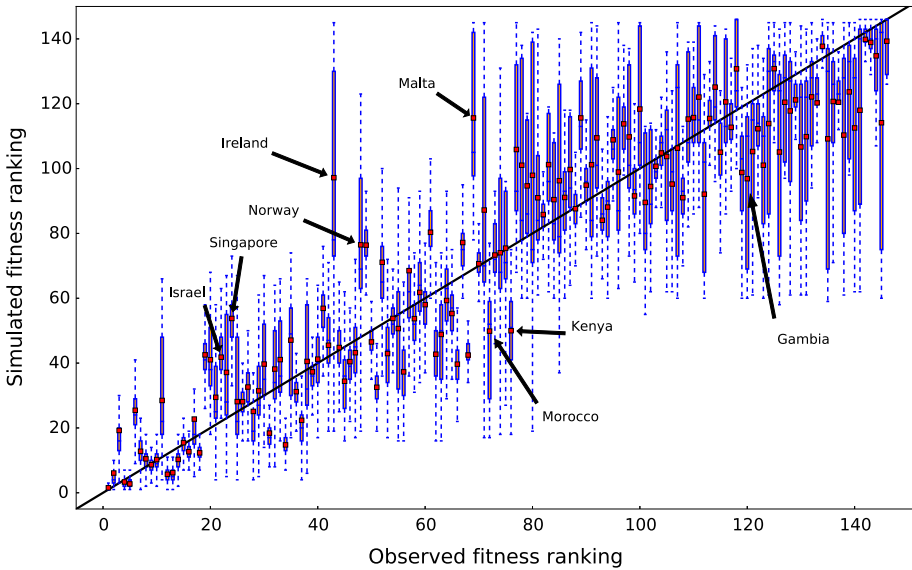
Fig. 6 Boxplots for the expected fitness rankings against the actual values for the year 2000. The country with the highest rank 1 (Germany) has the highest fitness, the one with lowest rank 146 (Iraq) has the lowest fitness. Red dots represent the averages on the ensemble. While the expected ranks are close to the observed ones, the vertical distribution intervals are quite large. We annotate those countries whose ranks diverge the most from the expectations of the null model: the countries whose real fitness values are much higher than expected are above the identity line. Ireland, Norway, Singapore and Israel have thus much higher empirical ranks, whereas Morocco's, Kenya's and Gambia's are much lower (Color figure online)

network [80]: the authors show that the nestedness of the International Trade Network (ITN) could be reproduced by the BiCM, thereby highlighting that the observed connection between degree sequence and nestedness also holds in non-ecological systems. However, the authors also argue that the BiCM cannot fully reproduce the disassortativity of the ITN, leading to the conclusion that the degree sequence alone is not enough to explain why strong exporters preferably connect to weak exporters.

In the nested representation of the bipartite trade matrix (as shown in Fig. 4), countries and products are typically ordered in increasing fitness and complexity along the rows and columns, respectively. Although the algorithm for calculating their values may not converge, it has been shown that the rankings are stable nonetheless [75]. Since the degree-constrained BiCM recovers the nestedness of the network, one may thus question its ability to reproduce the fitness and complexity rankings. This matter was addressed in [80], in which the authors show that the degree distributions for countries and products can be accurately reproduced when they are ranked by fitness and complexity, respectively. However, the model does not reproduce the exact fitness and complexity rankings: even when the observed values are close to the null model expectations, the distributions over the BiCM ensemble are broad, as shown in Fig. 6. Despite the fact that fitness is positively correlated with the degree sequence, highlighted by the tendency of the expectations to follow the real ranks in Fig. 6, it contains information that is not accounted for by the latter.[3] A similar behavior is observed for the product complexity rankings.

---

[3] We thank the anonymous reviewer for suggesting this test.
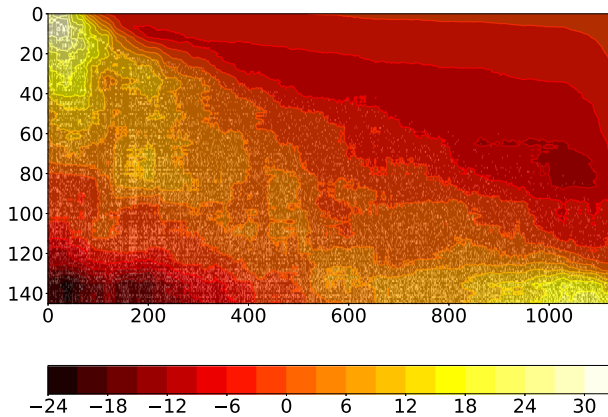
Fig. 7 The images shows the biadjacency matrix in Fig. 4 with links as white dots. The z-scores of the connectivity with respect to the BiCM are shown as a superimposed color map (gray shading). Higher link abundances than expected are shown as lighter colors, and lower abundances as dark colors. The smaller exporters on the top focus on the most basic products, as shown in the upper-left corner by the z-scores ($z \sim 30$). The most developed countries on the bottom, on the other hand, specialize on the most sophisticated products, as measured by the scores in the lower-right ($z \sim 25$). Moreover, they export basic products much less than expected, as can be seen in the the lower-left corner ($z \sim -20$). This suggests that countries diversify their export baskets as much as possible while specializing on the most sophisticated products they are able to export. Figure modified from [90] (Color figure online)
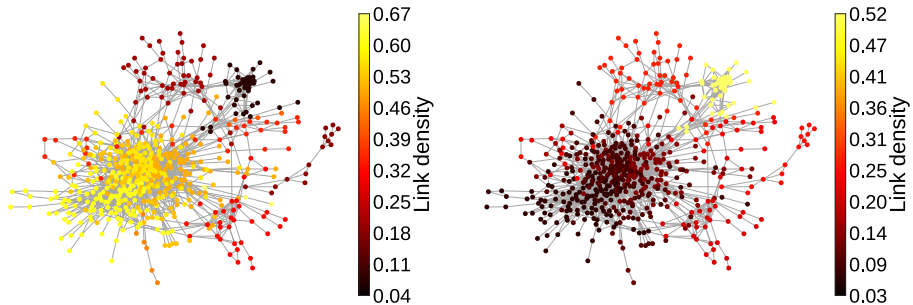

Fig. 8 Illustration of the product network with colors according to the occupation density of country communities, i.e. the density of links existing between the corresponding product and country communities [90]. Left: Advanced economies occupy the core of the network, which contains high technology items. Right: Developing economies occupy much more the periphery with less sophisticated product rather than the center. Figure courtesy of [90] (Color figure online)

The analysis of the triangular structure of the country-product matrix has been extended in [90]. Discounting the degree sequence and comparing empirical with expected link abundances, the authors observed a specialization signal within the overall export diversification tendency. Figure 7 shows the phenomenon quantitatively using the empirical biadjacency matrix and the z-scores of the expected number of links as a heat map. "Hotter" (whitish) colors represent higher z-score values, "colder" (dark) colors lower values. The high z-scores stretch from the upper-left to the lower-right and illustrate that countries concentrate their exports on more complex products more than expected, while still exporting basic goods as well. As a consequence, stating that exports baskets are nested is only partially true, since the

density in the baskets of more developed countries remains biased towards more exclusive products.

These conclusions can also be drawn from validated monopartite projection of the bipartite trade network (see Sect. 5.2.1). Figure 8 shows the link density between different areas of the product network and advanced economies (left) and developing economies (right): in both cases, the network is not populated uniformly, since the former tend to occupy preferentially the highly technological items in the core of the network, whereas the latter focus their export on lower complexity products that belongs to the periphery of the validated product network. In this sense, developed countries, given the size of their export basket, tend to specialize their exports towards the most exclusive products.

The previous result on the biadjacency matrix and the product network therefore reconcile the apparent conflict between Ricardo's argument of export specification and the overall diversification reported in literature [21]. Nevertheless, let us stress that the specialization phenomenon is a higher-order feature of the bipartite network of trade: in fact, it can be uncovered only by discounting the degree sequence information.

7.1.2 Motif Validation in Trade

Evolution of Bipartite Motifs of Countries In the economic literature, acronyms are often used to refer to countries that supposedly share similar features in their economic development and institutional frameworks. Famous examples are the G7 (Canada, the USA, Italy, France, Germany, the UK, Japan), which share a large part of the global GDP, and the five rising BRICS economies (Brazil, Russia, India, China, South Africa). Further groups are, e.g., the MINT countries (Mexico, Indonesia, Nigeria, Turkey) that show interesting economic developments [70] and the south European "PIGS" (Portugal, Italy Greece, Spain) that were struggling during the 2008 financial crisis [43].

Using the bipartite International Trade Network, it is possible to quantify the similarities within these country groups in terms of their V_n-motifs, n being the number of members. In [81], the authors compare the real trade network with the randomized BiCM ensemble to observe if such similarities are genuine or can just be attributed to the dimension of the export baskets, i.e. the country node degrees. They applied the BiCM to the product-country trade network and calculated the number of V_n-motifs for each country group, where n is the number of members.

Figure 9 compares the number of V_n-motifs for different country groups [81]: green dots represent the values observed in the data, whereas the box-plots capture the probability distribution of the ensemble. In panel (c) we can see that the observations for the "PIGS" lie clearly above the box-plot whiskers, which indicates that the similarities are not merely due to their export baskets sizes. This is true even before the 2008 crisis, although the discrepancy gradually increases over time. The BRICS in panel (a), on the contrary, show export basket overlaps that are compatible with the null model. The MINT countries in panel (f) do not have a single export item in common. As a consequence, both MINT and BRICS groupings cannot be justified from the observation of similar industrial capabilities alone (see also [29,92]).

Contrary to that, strong similarities can be observed in the *Tiger Cubs* (Thailand, Indonesia, Malaysia, Philippines, Vietnam), which have experienced a recent industrialization process similar to the original *Four Asian Tigers* (Hong Kong, Singapore, South Korea, Taiwan). Panel (d) in Fig. 9 shows a statistically significant signal of V_n-motifs, which gradually diminishes in intensity. This indicates that their recent industrial developments have begun to diverge, progressively turning into a differentiation in their exports. Similarly, the impact of a common communist industrialization program can be observed in the exports of *ex-*
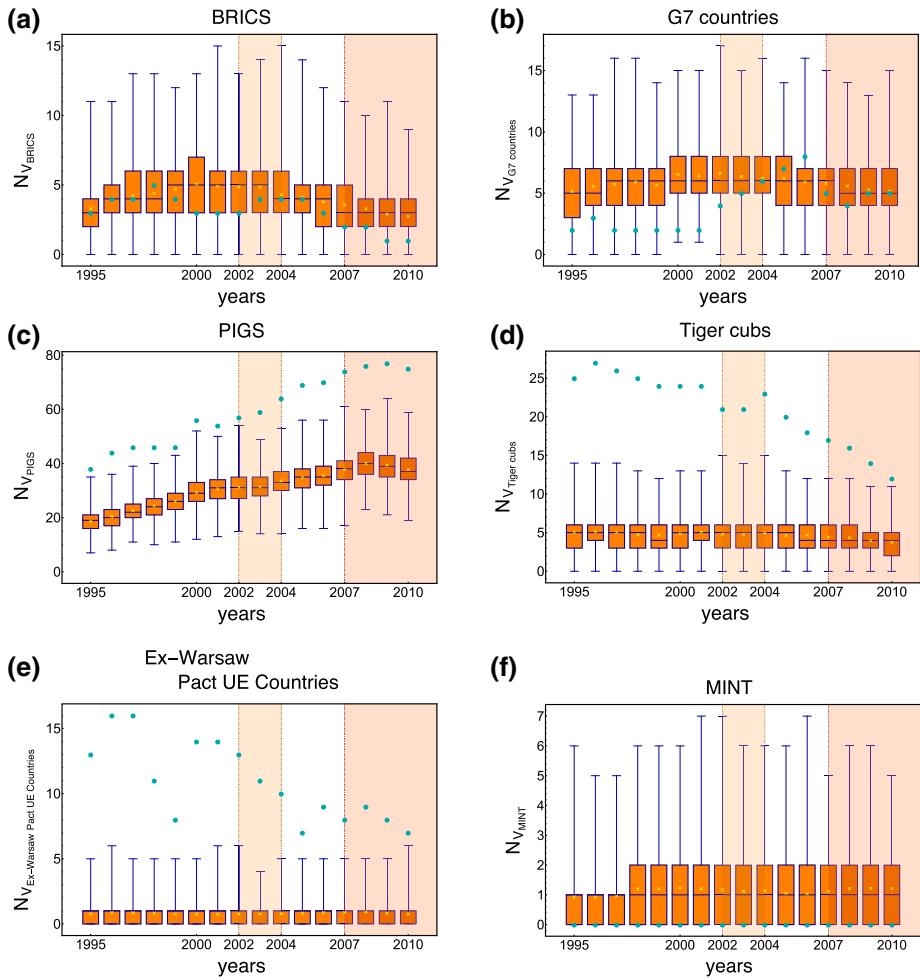
Fig. 9 Comparison of the observed numbers of V_n-motifs (see Sect. 2.1). The green dots represent the empirical quantity measured on the real network, the box plot the expectation value distribution according to the BiCM. The whiskers capture the 0.15th and the 99.85th percentile. Whereas the BRICS and G7 are compatible with the null model predictions, PIGS, Tiger Cubs and the ex-Warsaw Pact countries show abundances that are not explainable in terms of the degree sequences alone. Figure courtesy of [81] (Color figure online)

Warsaw Pact countries that are now part of the European Union (such as Poland, Romania, and Hungary) well into the years 2000, see panel (e) in Fig. 9. After joining the EU, the signal has progressively declined. The composition of the G7 group, on the other hand, can be simply attributed to their degrees, i.e. to the dimensions of their export baskets: panel (b) shows that no statistically significant signal can be detected.

Closed Motifs evolution Closed motifs are more complex combinations of links and capture mesoscopic structural properties of the network. For instance, a X-motif measures how many times two countries compete on the world market by sharing more than two products in their export basket. A high number of X-motifs indicates that their export baskets are very similar on a global level. Several scholars [99] suggest that an excessive degree of similarity in
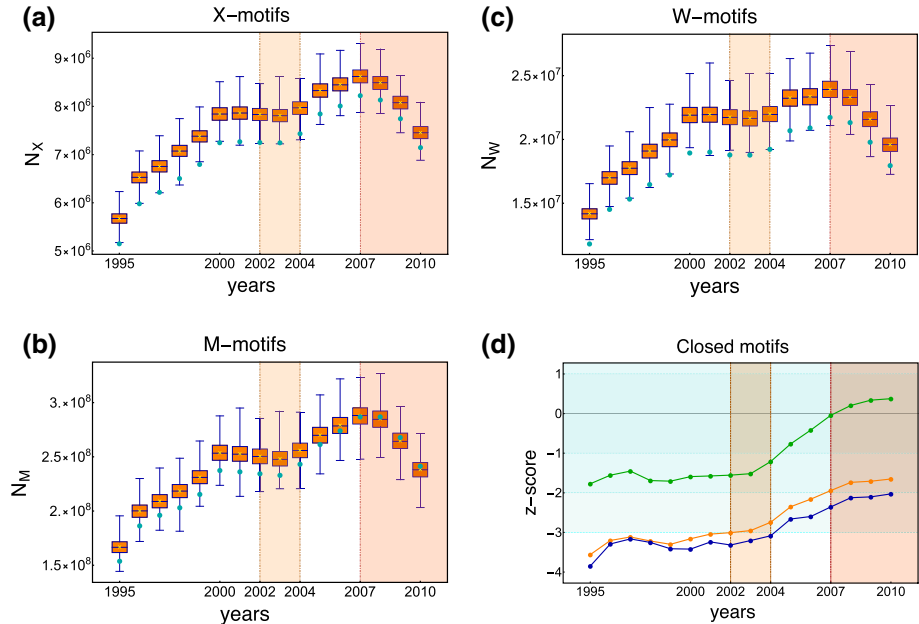
Fig. 10 **a**–**c** show the abundance of closed motifs (see Sect. 2.1) in the ITN network from 1995 until 2010. It is apparent that they increased until the financial crisis struck and dropped after 2007. The values are compared with the BiCM expectations in (**d**) using z-scores. Before the crisis, abundances were strongly underrepresented and increasing significantly already from 2003 onwards, indicating a global change of the network. Figure courtesy of [81] (Color figure online)

industrialization and exportation weakens the International Trade Network and makes is more prone to stress. A diversification of industrial capabilities, on the other hand, would make the system more resilient.

Following the trends in Fig. 10a–c, we can observe an increase in the number of closed motifs, i.e. of the similarity of export baskets, before the financial crisis [81]. This development comes to an abrupt halt in 2008 and abundances drop from that point on. Notice that this evolution only illustrates what happened in the wake of the crisis and does not provide any early-warning signal. However, comparing the observations with the null model, a very different picture arises, as illustrated in panel (d): all three motifs occur less in the network than would be expected from the BiCM. Z-scores are predominantly negative, with values as small as -3 to -4 for the X- and M-motifs. They are relatively stable until 2003, from which point on we observe a clear trend towards greater z-scores, i.e. towards a better agreement with the null model. Thus when the crisis struck in 2008, the ITN had already undergone significant structural changes in the preceding five years, which eventually fade out around 2010.

Note that the latter results do not imply any causal relation, as discussed in [81]. However, the dramatic evolution of the closed motifs anticipated the crisis. In that sense it can be regarded as an early-warning signal, since it informs us that the network is changing globally such that the resilience of the whole system is modified, making it more vulnerable to external shocks.

Moreover, it is worth underlining that these observations cannot be made without the benchmark of the null model. Considering only the data limits us to the simple observation

of abrupt trend changes in motif abundances. Only the comparison with the BiCM reveals a clear signal of anticipating structural change. Analyses as the one presented here provide effective tools to monitor systems such as the ITN, and offer deep insights to policy makers.

In summary, recent results on the International Trade Network obtained through the comparison with the BiCM have shown that many apparently genuine properties can be traced back to the degree sequence, i.e. the number of exporters of a product and the size of the export basket of a country.

8 Results in Financial Networks

8.1 Systemic Risk (Reprise)

Financial stress can propagate among banks through asset portfolio overlaps. As indicated in Sect. 4.1, the capital asset pricing model (CAPM) performs well at assessing the systemic risk due to fire sales spillover effects in the bipartite bank-asset network of financial US institutions [31]. Their data is derived from quarterly reports which disclose the single positions in the bank portfolios. Hence, the authors could compare the risk estimations due to aggregate exposures, considering only the node strengths, with measures that take also the degrees into account. Risk is measured using the metric defined by Greenwood et al. [49].

Although the CAPM framework replicates the systemic risk well (see Fig. 5), without a proper model little can be said about the precision of the risk prediction. Hence, the *maximum entropy capital asset pricing model* has been proposed (MECAPM, [31], see Appendix B), for which the portfolio weights are constrained as $\langle w_{i\alpha} \rangle = w_{i\alpha}^{CAPM}$. With N_L and N_Γ being the numbers of banks and portfolios, the MECAPM thus uses $N_L \times N_\Gamma$ conditions. As shown in Appendix B, the single link probabilities read [31,87]

$$(p_{CAPM})_{i\alpha} = \frac{w_{i\alpha}^{CAPM}}{1 + w_{i\alpha}^{CAPM}} \tag{13}$$

and the probability for a particular graph, $P(G_B)$, is geometrically distributed [31].

Since strength and degree information can complement each other, in analogy to the monopartite case in [62] the authors of [31] have also included the so-called *bipartite enhanced configuration m odel* (BiECM, [31]), on which degrees and strength constraints are imposed.

The results of the analysis are summarized in Fig. 11 for the banks with the highest systemic exposures in the data. The BiWCM refers to the weighted counterpart of the BiCM, in which node strengths instead of node degrees are constrained (see Appendix B and [31]). Although all three models systematically underestimate risk, the MECAPM clearly outperforms the other two models [31]. Notice that the BiWCM performs very badly, underestimating the risk as much as -80%. Errors are relatively large, as we can see from the shaded areas.

A possible reason for the large error intervals in Fig. 11 has been suggested in [87], pointing to the fact that the MECAPM predicts very dense network configurations. In fact, from Eq. (13) we can see that the link probabilities quickly approach 1 for $w_{i\alpha}^{CAPM} \gg 1$. This issue has been taken up in [87], which proposes the so-called *enhanced capital asset pricing model* (ECAPM, [87]). Using only the strength sequences C_i and V_α of the asset capitalization and portfolio market value, respectively, this approach aims at reconstructing the network topology while imposing the CAPM link weights.
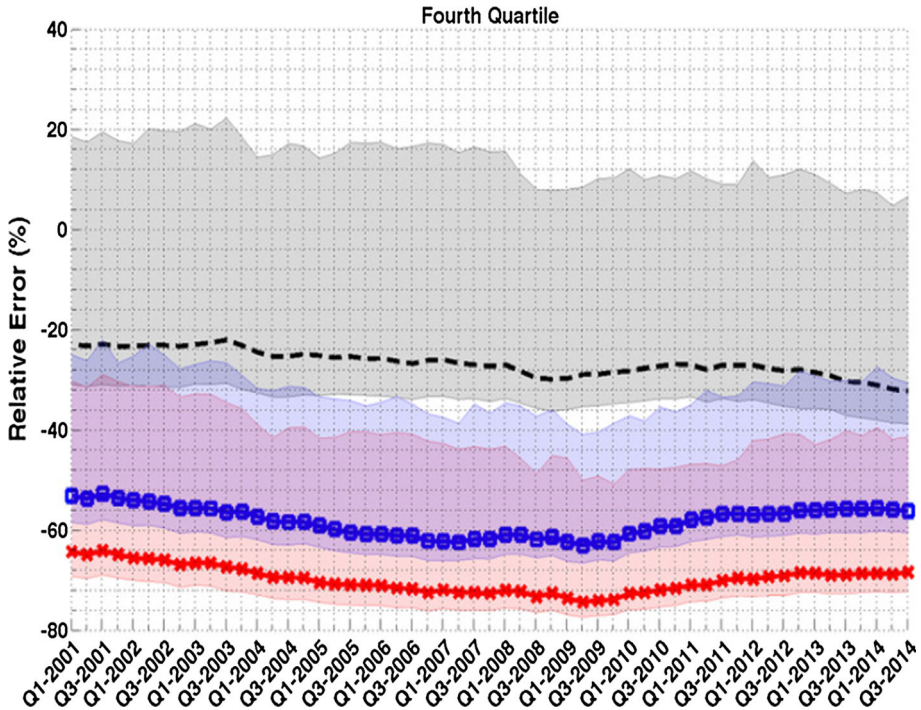
Fig. 11 Quartile of banks with the highest indirect risk as measured by [31] during the interval 2001–2014. Top: MECAPM (dashed line, grey shading); Middle: BiECM (blue squares, blue shading); Bottom: BiWCM (red crosses, red shading). The shaded areas and their partial overlaps illustrate the corresponding distributions. Although all three models systematically underestimate the systemic risk, MECAPM outperforms the best. Figure courtesy of [31] (Color figure online)

Firstly, the topology of the network is established by using the BiCM under the assumption that the Lagrange multipliers are proportional to node-specific fitness values, represented by their strengths [87]. In analogy to the BiCM (see Appendix B), this yields

$$(p_{ECAPM})_{i\alpha} = \frac{z V_i C_\alpha}{1 + z V_i C_\alpha}, \quad \forall i \in L, \alpha \in \Gamma, \tag{14}$$

where z absorbs the proportionality constants.

Secondly, the link weights are reconstructed using the CAPM model while taking the network topology into consideration. Instead of setting $w_{i\alpha} = V_i C_\alpha / w$, a correction factor is applied [87]

$$
\begin{aligned}
w_{i\alpha} &= m_{i\alpha} \frac{V_i C_\alpha}{w \, (p_{ECAPM})_{i\alpha}} \\
&= (V_i C_\alpha + z^{-1}) \frac{m_{i\alpha}}{w},
\end{aligned}
\tag{15}
$$

where $m_{i\alpha}$ is 0 or 1 depending on the link presence. As pointed out in [87], the weight expectations of the ECAPM correspond with those of the MECAPM and thus the CAPM [87]. However, the former reconstructs the network topology separately, which compensates the high network densities for the latter.
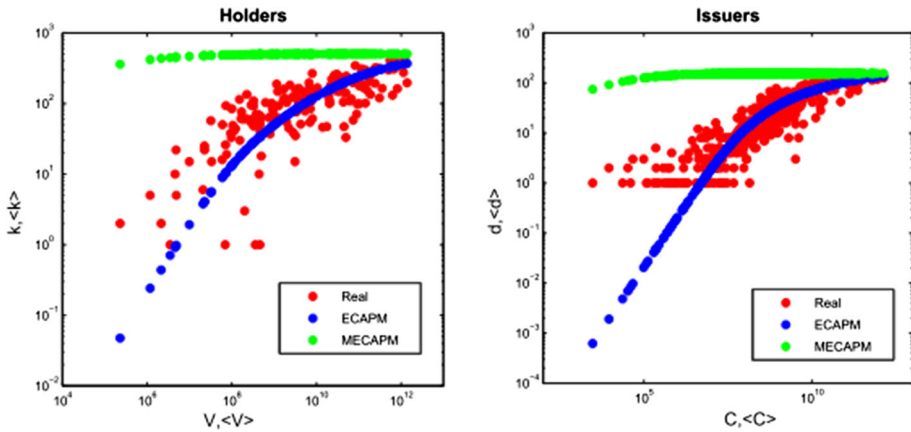
Fig. 12 Relation between the strengths and degrees of the nodes in [87]. The security holder portfolios have certain market values V_i (the "banks") and issued security assets have a market capitalization C_α (the "assets"). The ECAPM follows the real empirical data much closer than the MECAPM, which systematically overestimates the degrees. Figure courtesy of [87] (Color figure online)

The difference between the ECAPM and the MECAPM has been tested in [87] on a data set of security holdings of European institutions (Security Holding Statistics, SHS) collected by the European Central Bank from 2009 to 2015. The empirical difference is visible in Fig. 12, which compares the node degrees with node strengths. The MECAPM shows continuously high degrees and does not capture the real distribution, which illustrates the observation made for Eq. (13). Even though the ECAPM underestimates degrees for small strength values, it better reproduces the data.

Since both MECAPM and ECAPM reproduce the same weights, they estimate the same systemic risks as measured with the metric introduced in [49]. However, reconstructing the topology as in [87] significantly decreases the uncertainty of the risk metric. In particular, comparing the errors of the MECAPM and ECAPM yields

$$\sigma_{S_i}^{MECAPM} \propto V_i^{1/2} \sigma_{S_i}^{ECAPM}, \tag{16}$$

where V_i is the value of institution i and S_i its systematicness [87]. The fluctuations in the ECAPM framework are thus systematically smaller than in MECAPM and motivate the application of a degree as well as strength reconstruction.

8.2 Portfolio Overlap Projection

The risk of fire sales has also been treated in [50] using a different methodology. Contrary to [31] and [87], they do not consider weighted networks but instead the binary bipartite network between financial institutions and their assets holding in the years 1991–2013. Instead of calculating the systemic risk measure on the null model, they focus on the overlap matrix of portfolios, which expresses the number of assets that financial institutions share, i.e. the number of their V-motifs (see Eq. (1)). By applying the grand canonical projection algorithm [82] with a Bonferroni correction for the p-value testing, and thus comparing the observed values V* with their BiCM expectations and validating only statistically significant links, they obtain a network of financial institutions containing only relevant links that are not accounted for by the degree sequences.
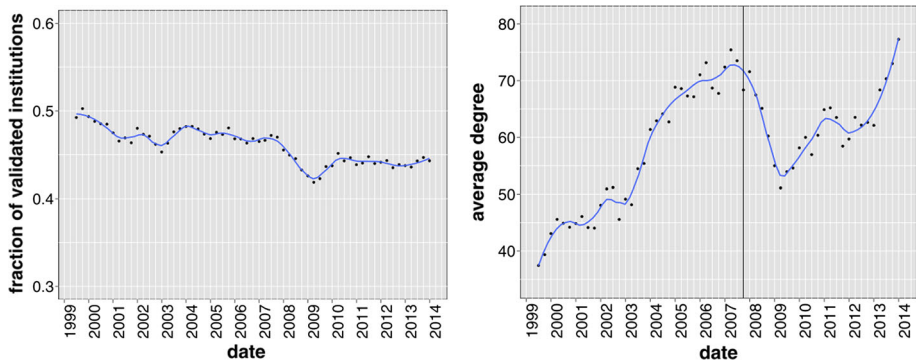
Fig. 13 Left: Fraction of financial institutions in the validated monopartite projection. The values decrease only slightly, with temporary reduction around 2009. Right: Average degree in the monopartite projection of financial institutions. Notice the salient dip in 2009 after the beginning of the crisis. Figure courtesy of [50] (Color figure online)

As Fig. 13 shows, the fraction of institutions that have at least one significant edge in the validated network remains relatively constant. In spite of this seemingly innocuous development, the similarity of these nodes increase very quickly, as illustrated in Fig. 13. In particular, notice how the similarity increases before the 2007–2008 financial crisis. After a drop in 2009, it took up pace and has reached levels even higher than before the crisis [50].

In conclusion, the authors point out that the validated projection method can recover those financial institutions that would be at risk of suffering the greatest losses in cases of financial distress [50].

9 Conclusions and Outlook

In this review, we have presented several recent methods designed for the statistical analysis of bipartite network. Interestingly, despite the fact that bipartite networks have been considered deeply in ecology, and many analytical quantities of interest originate from the study of, e.g., species interaction networks, recent techniques for statistical significance testing have been formulated in a -seemingly- completely different context, namely economic and financial systems. However, while the latter has benefited greatly from the progress made in ecology, many of the analytical tools commonly employed in socio-economic networks seem not to have been recognized yet as potentially useful for ecological systems, apart from few exceptions [67,72,97]. However, the recent application of the bipartite configuration model, originally formulated for the study of international trade, has already shed some light on a better understanding of the origin of nestedness. In a similar way, validating similarities among insect species could provide insights on their competition in isolated environments. Furthermore, the resilience studies of economic and financial bipartite networks can catalyze the study of extinction spillovers in the standard ecological framework. This work aims at bridging the gap between these two—apparently—distant areas which can benefit greatly from sharing advancements across scientific fields.

A straightforward example is provided by the detection of mesoscale structures in bipartite networks, such as communities and motifs. Although motifs have been popularized in biological science, recently proposed techniques to overcome the limitations of modular-

ity maximization have not yet crossed the borders of graph theory, and their application in ecological networks is still in an embryonic stage [10]. The same holds true for the multiplex formalism [73]. The cross-disciplinary importance of these topics is clearly shown by the idea of considering social networks as "mutualistic" information ecosystems. In this context, has been found that mesoscale structures may be correlated to the emergence of collective attention [18], in particular when a transition from a modular to a nested structure is observed. These results also underline the general need to develop dynamical models to study the evolution of bipartite networks.

An application that has benefited from both, the concepts developed in ecology and the methods designed for social networks, are so-called recommender systems [102]. Briefly speaking, these algorithms suggest "users" their next "choices", be they movies to watch or items to buy, among others. Although many recommendation algorithms exist, an interesting example is provided by those which are built upon the idea of a *resource-allocation* dynamic taking place on the network. Other concepts such as *specialization* and *interaction* have inspired models able to reproduce observed patterns in both ecological and social systems [79].

Some of the statistical methods presented in this review have been recently employed and extended for the study of *tripartite* structures to asses the relationship between technology and economic development [74]. In this network, the three layers consist of technologies, countries, and products, and the analysis aims at quantifying the probability of jumping from a given technology in one layer to a given product in another one, while accounting for all possible paths through the intermediate countries layer. Although the null model employed in this approach is, in fact, a combination of two distinct bipartite configuration models, the paper certainly represents an interesting direction for future research.

A Appendix: Revealed Comparative Advantage

The *revealed comparative advantage* (RCA, also knows as Balassa index [11]), rescales the product export volumes in order to determine whether countries are relevant exporters of products. Be $e(c, p)$ the export value of product p in country c's export basket. The RCA is calculated by comparing the monetary importance of p in c's export basket to the global average,

$$\text{RCA}_{c,p} = \frac{e(c, p)}{\sum_{p'} e(c, p')} \Bigg/ \frac{\sum_{c'} e(c', p)}{\sum_{c',p'} e(c', p')} . \qquad (17)$$

A country is a relevant exporter if $\text{RCA} \geq 1$. Using the RCA, the weighted country-product biadjacency matrix can be binarized by keeping only those matrix entries that identify relevant exports and setting them to 1.

B Appendix: Bipartite Exponential Random Graph Model

We report some of the null models that have been obtained through entropy maximization and which have been applied to binary and weighted bipartite networks. In the following, all

quantities marked with an asterisk refer to the real networks, expressed by their binary ($\mathbf{M}^*$) or weighted ($\mathbf{W}^*$) biadjacency matrix. The layer dimensions are N_L and N_Γ.

B.1 Bipartite Random Graph

Constraining the expected number of links in the graph ensemble yields an extension of the Erdős-Rényi random graph to bipartite networks, the *bipartite random graph* (BiRG). The constraint $C \equiv E = \sum_{i,\alpha} m_{i\alpha}$, and thus the Lagrange multiplier θ as well, is scalar. The partition function can be calculated easily:

$$\mathcal{Z}_{\text{BiRG}}(\theta) = \sum_{G_B \in \mathcal{G}_B} e^{-\theta E(G_B)}$$
$$= (1 + e^{-\theta})^{N_L N_\Gamma}. \tag{18}$$

The probability per graph reads

$$P(G_G|\theta) = \frac{e^{-\theta E}}{(1 + e^{-\theta})^{N_L N_\Gamma}}$$
$$= (p_{\text{BiRG}})^E (1 - p_{\text{BiRG}})^{N_L N_\Gamma - E}, \tag{19}$$

where $p_{\text{BiRG}} \equiv \frac{e^{-\theta}}{1 + e^{-\theta}}$ is the probability of observing a bipartite link between any node couple $i \in L$, $\alpha \in \Gamma$. Notice that p_{BiRG} is uniform and independent of the links. Since Eq. (19) is a Binomial distribution, we see that the probability of observing a generic graph G_B in the ensemble reduces to the problem of observing $E(G_B)$ successful trials with the same probability p_{BiRG}. We can obtain an analytical expression for the Lagrange multiplier θ and thus for the link probability by maximizing the likelihood, which reads

$$\mathcal{L} = \ln P(G^*|\theta) = -\theta \, E^* - N_L N_\Gamma \ln(1 + e^{-\theta}), \tag{20}$$

and returns

$$p_{\text{BiRG}} = \frac{E^*}{N_L N_\Gamma}. \tag{21}$$

B.2 Bipartite Partial Configuration Model

Without loss of generality, we constrain the degree sequence on the layer L such that $\langle k_i \rangle = k_i^*$, $\forall i \in L$. For each node degree k_i, we introduce one associated Lagrange multiplier, θ_i. This gives us the *bipartite partial configuration model* (BiPCM, [82]). Following the same procedure as in Eq. (18), we can obtain

$$\mathcal{Z}_{\text{BiPCM}}(\boldsymbol{\theta}) = \prod_{i,\alpha} 1 + e^{-\theta_i}. \tag{22}$$

The probability per graph reads

$$P(G_B|\boldsymbol{\theta}) = \prod_{i,\alpha} (p_{\text{BiPCM}})_i^{m_{i\alpha}} \left(1 - (p_{\text{BiPCM}})_i\right)^{1 - m_{i\alpha}}$$
$$= \prod_i (p_{\text{BiPCM}})_i^{k_i} \left(1 - (p_{\text{BiPCM}})_i\right)^{N_\Gamma - k_i}, \tag{23}$$

where $(p_{\text{BiPCM}})_i = \frac{e^{-\theta_i}}{1 + e^{-\theta_i}}$ is the probability of connecting the node i with any of the node of the opposite layer Γ. The link probabilities are not uniform, but depend on the Lagrange

multipliers of the nodes $i \in L$. The factors in the product in Eq. (23) express the probabilities of observing exactly the constrained node degrees: the probability of the degree k_i of the node $i \in L$ is given by the probability of observing k_i successes trials of a binomial distribution with probability $(p_{\text{BiPCM}})_i$. Maximizing the likelihood $\mathcal{L}$ returns the explicit expressions for the link probabilities:

$$(p_{\text{BiPCM}})_i = \frac{k_i^*}{N_\Gamma}. \tag{24}$$

B.3 Bipartite Configuration Model

In the monopartite configuration model, the degrees of all the nodes are constrained. Analogously, in the *bipartite configuration model* (BiCM, [80]) the degrees of the two layer degree sequences are constrained, such that $\langle k_i \rangle = k_i^*$, $\forall i \in L$, and $\langle k_\alpha \rangle = k_\alpha^*$, $\forall \alpha \in \Gamma$. If $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ are the corresponding Lagrange multipliers, the partition function reads [80]

$$\mathcal{Z}_{\text{BiCM}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{i,\alpha} 1 + e^{-(\theta_i + \rho_\alpha)}, \tag{25}$$

following essentially the same strategy used in Eq. (18). Again, the probability per graph factorizes in a product of probabilities per link:

$$\begin{aligned} P(G_B | \boldsymbol{\theta}, \boldsymbol{\rho}) &= \prod_{i,\alpha} \frac{e^{-(\theta_i + \rho_\alpha) m_{i\alpha}}}{1 + e^{-(\theta_i + \rho_\alpha)}} \\ &= \prod_{i,\alpha} (p_{\text{BiCM}})_{i\alpha}^{m_{i\alpha}} \left(1 - (p_{\text{BiCM}})_{i\alpha}\right)^{1 - m_{i\alpha}}, \end{aligned} \tag{26}$$

where the probability per link reads

$$(p_{\text{BiCM}})_{i\alpha} = \frac{e^{-(\theta_i + \rho_\alpha)}}{1 + e^{-(\theta_i + \rho_\alpha)}}, \quad i \in L, \alpha \in \Gamma \tag{27}$$

Compared to the probability distributions of the BiRG and BiPCM, we can see that the BiCM distribution is more general and corresponds to the product of different Bernoulli events with link-specific success probabilities. Note that the distribution factorizes and link probabilities are independent. Maximizing the likelihood returns the equation system [80]

$$\begin{cases} \sum_\alpha \dfrac{e^{-(\theta_i + \rho_\alpha)}}{1 + e^{-(\theta_i + \rho_\alpha)}} = k_i^*, & \forall i \in L, \\[4mm] \sum_i \dfrac{e^{-(\theta_i + \rho_\alpha)}}{1 + e^{-(\theta_i + \rho_\alpha)}} = k_\alpha^*, & \forall \alpha \in \Gamma. \end{cases} \tag{28}$$

Solving this system allows us to evaluate the Lagrange multipliers and ultimately obtain the graph probabilities.

B.4 Bipartite Weighted Configuration Model

Constraining the node strengths as $\langle s_i \rangle = s_i^*$, $\forall i \in L$, and $\langle s_\alpha \rangle = s_\alpha^*$, $\forall \alpha \in \Gamma$, yields the *bipartite weighted configuration model* (BiWCM, [31]). Be $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ the corresponding Lagrange multipliers. As shown in [31], the partition function is

$$\mathcal{Z}_{\text{BiCM}}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{i,\alpha} \frac{1}{1 - e^{-(\theta_i + \rho_\alpha)}}. \tag{29}$$

The graph probability yields

$$P(G_B|\boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{i,\alpha} \left(e^{-(\theta_i + \rho_\alpha)} \right)^{w_{i\alpha}} (1 - e^{-(\theta_i + \rho_\alpha)}). \tag{30}$$

Similar to the BiCM, the Lagrange multipliers can be obtained by solving an equation system, which reads [31]

$$\begin{cases} \sum_\alpha \dfrac{e^{-(\theta_i + \rho_\alpha)}}{1 - e^{-(\theta_i + \rho_\alpha)}} = s_i^*, & \forall i \in L, \\[4mm] \sum_i \dfrac{e^{-(\theta_i + \rho_\alpha)}}{1 - e^{-(\theta_i + \rho_\alpha)}} = s_\alpha^*, & \forall \alpha \in \Gamma. \end{cases} \tag{31}$$

B.5 Bipartite Enhanced Configuration Model

The *bipartite enhanced configuration model* (BiECM, [31]) is a bipartite extension of the monopartite *enhanced configuration model* introduced in [62]. Both, degrees as well as strengths, are constrained.

Be θ_i and θ_α the constraints associated to the degrees, and ρ_i and ρ_α those associated to the strengths for the nodes $i \in L$ and $\alpha \in \Gamma$, respectively. Using the short-hand notation $\phi_i = e^{-\rho_i}$, $\xi_\alpha = e^{-\rho_\alpha}$, $\psi_i = e^{-\theta_i}$ and $\gamma_\alpha = e^{-\theta_\alpha}$, the partition function reads [31]

$$\mathcal{Z}_{BiECM}(\boldsymbol{\theta}, \boldsymbol{\rho}) = \prod_{i,\alpha} \frac{1 - \phi_i \xi_\alpha (1 - \psi_i \gamma_\alpha)}{1 - \phi_i \xi_\alpha}. \tag{32}$$

Consequently, the network probability is given by

$$P(G_B) = \prod_{i,\alpha} \frac{(1 - \phi_i \xi_\alpha)(\phi_i \xi_\alpha)^{w_{i\alpha}} (\psi_i \gamma_\alpha)^{\Theta(w_{i\alpha})}}{1 - \phi_i \xi_\alpha (1 - \psi_i \gamma_\alpha)} \tag{33}$$

and factorizes in single link probabilities. The values of the Lagrange multipliers can be obtained through a nonlinear system of equations, as shown in the Appendix of [31].

B.6 Maximum Entropy Capital Asset Pricing Model

The elements of the weighted biadjacency matrix can be rescaled to yield the quantities of the *capital asset pricing model* (CAPM, [61,66]). In the financial context, the vertex strengths are often described as the *total asset size* of a bank (or *market value* of their portfolio), $V_i = \sum_\alpha w_{i\alpha}$, and the *market capitalization* of an asset, $C_\alpha = \sum_i w_{i\alpha}$ [31,87]. In the CAPM, banks choose their portfolio weights proportional to their market value and the asset's capitalization:

$$w_{i\alpha}^{CAPM} = \frac{V_i C_\alpha}{w}, \tag{34}$$

where we have used $w = \sum_{i',\alpha'} w_{i'\alpha'}$ The probability distribution for the MECAPM yields [31]

$$P(G_B) = \prod_{i,\alpha} \left[1 - (p_{CAPM})_{i\alpha} \right]^{w_{i\alpha}} (p_{CAPM})_{i\alpha}, \tag{35}$$

where the probability per link reads

$$(p_{CAPM})_{i\alpha} = \frac{w_{i\alpha}^{CAPM}}{1 + w_{i\alpha}^{CAPM}}. \tag{36}$$

Note that $P(G_B)$ is geometrically distributed for $w_{i\alpha} \in \mathbb{N}$ [31]. The link probabilities can be easily calculated using the identity in Eq. (34).

B.7 Enhanced Capital Asset Pricing Model

The so-called *enhanced capital asset pricing model* (ECAPM, [87]) reconstructs the link topology and subsequently the link weights. Their method makes only use of the strength sequence and is composed of two steps.

Firstly, the topology of the network is reconstructed by using the BiCM under the assumption that the exponential Lagrange multipliers $x_i \equiv e^{-\theta_i}$ and $y_\alpha \equiv e^{-\theta_\alpha}$ are proportional to node-specific fitness values, represented by their strengths:

$$
\begin{aligned}
x_i &\equiv \sqrt{z_\Gamma} s_i, \quad \forall i \in L \\
y_\alpha &\equiv \sqrt{z_L} s_\alpha, \quad \forall \alpha \in \Gamma.
\end{aligned}
\tag{37}
$$

Constraining the network density with the total number of links $\langle E \rangle \equiv E^*$, the parameter $z = \sqrt{z_\Gamma z_L}$ can be estimated using [87]

$$
\langle E \rangle = \sum_{i,\alpha} \frac{z V_i C_\alpha}{1 + z V_i C_\alpha}, \quad \forall i \in L, \alpha \in \Gamma,
\tag{38}
$$

Subsequently, the single link probabilities are simply given by the BiCM expression in Eq. (11), substituting the Lagrange multipliers with the expressions (37):

$$
(p_{ECAPM})_{i\alpha} = \frac{z V_i C_\alpha}{1 + z V_i C_\alpha}, \quad \forall i \in L, \alpha \in \Gamma,
\tag{39}
$$

where z absorbs the proportionality constants.

Secondly, the link weights are reconstructed using the CAPM model while taking the network topology into consideration. Instead of setting $w_{i\alpha} = V_i C_\alpha / w$, a correction factor is applied [87]

$$
\begin{aligned}
w_{i\alpha} &= m_{i\alpha} \frac{V_i C_\alpha}{w (p_{ECAPM})_{i\alpha}} \\
&= (V_i C_\alpha + z^{-1}) \frac{m_{i\alpha}}{w},
\end{aligned}
\tag{40}
$$

where $m_{i\alpha}$ is 0 or 1, depending the link is present in the graph or not.

References

1. Allen, F., Gale, D.: Financial contagion. J. Polit. Econ. **108**(1), 1–33 (2000). https://doi.org/10.1086/262109
2. Allesina, S., Tang, S.: Stability criteria for complex ecosystems. Nature. **483**(7388), 205–208 (2012). https://doi.org/10.1038/nature10832
3. Almeida-Neto, M., Guimarães, P., Guimarães, P.R., Loyola, R.D., Ulrich, W.: A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos **117**(8), 1227–1239 (2008). https://doi.org/10.1111/j.0030-1299.2008.16644.x
4. Alon, U.: Network motifs: theory and experimental approaches. Nat. Rev. Genet. **8**(6), 450–461 (2007). https://doi.org/10.1038/nrg2102
5. Angelini, O., Cristelli, M., Zaccaria, A., Pietronero, L.: The complex dynamics of products and its asymptotic properties. PLoS ONE **12**(5), 1–20 (2017). https://doi.org/10.1371/journal.pone.0177360

6. Annunziata, M.A., Petri, A., Pontuale, G., Zaccaria, A.: How log-normal is your country? An analysis of the statistical distribution of the exported volumes of products. Eur. Phys. J. Spec. Top. **1995**(225), 1985–1995 (2016). https://doi.org/10.1140/epjst/e2015-50320-7

7. Arinaminpathy, N., Kapadia, S., May, R.M.: Size and complexity in model financial systems. PNAS **109**(45), 18338–18343 (2012). https://doi.org/10.1073/pnas.1213767109

8. Atmar, W., Patterson, B.D.: The measure of order and disorder in the distribution of species in fragmented habitat. Oecologia **96**(3), 373–382 (1993). https://doi.org/10.1007/BF00317508

9. Azaele, S., Suweis, S., Grilli, J., Volkov, I., Banavar, J.R., Maritan, A.: Statistical mechanics of ecological systems: neutral theory and beyond. Rev. Mod. Phys. (2016). https://doi.org/10.1103/RevModPhys.88.035003

10. Baiser, B., Elhesha, R., Kahveci, T.: Motifs in the assembly of food web networks. Oikos **125**(4), 480–491 (2016). https://doi.org/10.1111/oik.02532

11. Balassa, B.: Trade liberalization and 'revealed' comparative advantage. Manch. Sch. Econ. Soc. Stud. **33**, 99–123 (1965)

12. Barigozzi, M., Fagiolo, G., Garlaschelli, D.: Multinetwork of international trade: a commodity-specific analysis. Phys. Rev. E **81**(4), 046,104 (2010). https://doi.org/10.1103/PhysRevE.81.046104

13. Bastolla, U., Fortuna, Ma., Pascual-García, A., Ferrera, A., Luque, B., Bascompte, J.: The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature **458**(7241), 1018–1020 (2009). https://doi.org/10.1038/nature07950

14. Battiston, S., Farmer, J.D., Flache, A., Garlaschelli, D., Haldane, A., Heesterbeek, H., Hommes, C., Jaeger, C., May, R.M., Scheffer, M.: Complexity theory and financial regulation. Science **351**(6275), 818–819 (2016). https://doi.org/10.1126/science.aad0299

15. Battiston, S., Puliga, M., Kaushik, R., Tasca, P., Caldarelli, G.: DebtRank: too central to Fail? Financial networks, the FED and systemic risk. Sci. Rep. **2**, 1–6 (2012). https://doi.org/10.1038/srep00541

16. Bonanno, G., Caldarelli, G., Lillo, F., Mantegna, R.N.: Topology of correlation based minimal spanning trees in real and model markets. Phys. Rev. E **046130**, 17–20 (2003). https://doi.org/10.1103/PhysRevE.68.046130

17. Bonanno, G., Caldarelli, G., Lillo, F., Micciché, S., Vandewalle, N., Mantegna, R.N.: Networks of equities in financial markets. Eur. Phys. J. B **38**(2), 363–371 (2004). https://doi.org/10.1140/epjb/e2004-00129-6

18. Borge-Holthoefer, J., Baños, R.A., Gracia-lázaro, C., Moreno, Y.: Emergence of consensus as a modular-to-nested transition in communication dynamics. Sci. Rep. **7**, 1–9 (2017). https://doi.org/10.1038/srep41673

19. Brunnermeier, M.K.: Deciphering the liquidity and credit crunch 2007–2008. J. Econ. Perspect. **23**(1), 77–100 (2009). https://doi.org/10.1257/jep.23.1.77

20. Caccioli, F., Shrestha, M., Moore, C., Farmer, J.D.: Stability analysis of financial contagion due to overlapping portfolios. J. Bank. Financ. **46**(Supplement C), 233–245 (2014). https://doi.org/10.1016/j.jbankfin.2014.05.021

21. Cadot, O., Carrere, C., Strauss-Kahn, V.: Export diversification: what's behind the hump? Rev. Econ. Stat. **93**(2), 590–605 (2011). https://doi.org/10.1162/REST_a_00078

22. 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**(10), 1–17 (2012). https://doi.org/10.1371/journal.pone.0047278

23. Cane, J.H., Minckley, R.L., Kervin, L.J., Roulston, T.H., Williams, N.M.: Complex responses within a desert bee guild (Hymenoptera: Apiformes) to urban habitat fragmentation. Ecol. Appl. **16**(2), 632–644 (2006). https://doi.org/10.1890/1051-0761(2006)016[0632:CRWADB]2.0.CO;2

24. Cerina, F., Riccaboni, M.: World input-output network world input-output network. PLoS ONE **10**(7), 1–21 (2015). https://doi.org/10.1371/journal.pone.0134025

25. Chan-Lau, J.A., Espinosa, M., Giesecke, K., Solé, J.A.: Assessing the systemic implications of financial linkages. IMF Glob. Financ. Stab. Rep. **2**, 1–38 (2009). https://ssrn.com/abstract=1417920

26. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. Ann. Comb. **6**, 125–145 (2002). https://link.springer.com/content/pdf/10.1007/PL00012580.pdf

27. Connor, E.F., Simberloff, D.: The assembly of species communities: chance or competition? Ecology **60**(6), 1132 (1979). https://doi.org/10.2307/1936961

28. Cont, R., Wagalath, L.: Fire sales forensics: measuring endogenous risk. Math. Financ. **26**(4), 835–866 (2016). https://doi.org/10.1111/mafi.12071

29. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., Pietronero, L.: Measuring the intangibles: a metrics for the economic complexity of countries and products. PLoS ONE **8**(8), e70726 (2013). https://doi.org/10.1371/journal.pone.0070726

30. Cristelli, M., Tacchella, A., Pietronero, L.: The heterogeneous dynamics of economic complexity. PLoS ONE **10**(2), 1–15 (2015). https://doi.org/10.1371/journal.pone.0117174

31. Di Gangi, D., Lillo, F., Pirino, D.: Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. SSRN Electron. J. (2015). https://doi.org/10.2139/ssrn.2639178. https://ssrn.com/abstract=2639178

32. Diamond, J.M.: Assembly of Species Communities. Belknap Press, Cambridge, MA (1975). https://doi.org/10.2307/1936961

33. Diamond, J.M., Gilpin, M.E.: Examination of the null model of connor and simberloff for species co-occurrences on Islands. Oecologia **52**(1), 64–74 (1982). https://doi.org/10.1007/BF00349013

34. Donnelly, R., Marzluff, J.M.: Importance of reserve size and landscape context to urban bird conservation. Conserv. Biol. **18**(3), 733–745 (2004). https://doi.org/10.1111/j.1523-1739.2004.00032.x

35. Dormann, C.F., Fründ, J., Bluthgen, N., Gruber, B.: Indices, graphs and null models: analysing bipartite ecological networks. Open Ecol. J. **2**, 7–24 (2009). https://doi.org/10.2174/1874213000902010007

36. Dueñas, M., Fagiolo, G.: Modeling the International-Trade Network: a gravity approach. J. Econ. Interact. Coord. **8**(1), 155–178 (2013). https://doi.org/10.1007/s11403-013-0108-y

37. Eisenberg, L., Noe, T.H.: Systemic risk in financial systems. Manag. Sci. **47**(2), 236–249 (2001)

38. Elton, C.S.: Animal Ecology. Sidgwick and Jackson, London (1927)

39. Erdos, P., Rényi, A.: On random graphs I. Publ. Math. Debr. **6**, 290–297 (1959)

40. Fagiolo, G., Reyes, J., Schiavo, S.: World-trade web: topological properties, dynamics, and evolution. Phys. Rev. E (2009). https://doi.org/10.1103/PhysRevE.79.036115

41. Fahrig, L.: Relative effects of habitat loss and fragmentation on population extinction. J. Wildl. Manag. **61**(3), 603–610 (1997). https://doi.org/10.2307/3802168

42. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3–5), 75–174 (2010). https://doi.org/10.1016/j.physrep.2009.11.002

43. Furceri, D., Mourougane, A.: The effect of financial crises on potential output: new empirical evidence from OECD countries. J. Macroecono. **34**(3), 822–832 (2012). https://doi.org/10.1016/j.jmacro.2012.05.010

44. Gai, P., Kapadia, S.: Contagion in financial networks. Proc. R. Soc. **466**(2120), 2401–2423 (2010). https://doi.org/10.1257/jel.20151228

45. Galeano, J., Fernandez, M., Hidalgo, C.: ipartite networks provide new insights on international trade markets. Am. Inst. Math. Sci. **7**(3), 399–413 (2012)

46. Garlaschelli, D., Loffredo, M.I.: Fitness-dependent topological properties of the world trade web. Phys. Rev. Lett. **93**, 188,701, (2004). https://doi.org/10.1103/PhysRevLett.93.188701

47. Garlaschelli, D., Loffredo, M.I.: Maximum likelihood: extracting unbiased information from complex networks. Phys. Rev. E **78**(1), 1–5 (2008). https://doi.org/10.1103/PhysRevE.78.015101

48. Gilpin, M.E., Diamond, J.M.: Factors contributing to non-randomness in species Co-occurrences on Islands. Oecologia **52**(1), 75–84 (1982). https://doi.org/10.1007/BF00349014

49. Greenwood, R., Landier, A., Thesmar, D.: Vulnerable banks. J. Financ. Econ. **115**(3), 471–485 (2015). https://doi.org/10.1016/j.jfineco.2014.11.006

50. Gualdi, S., Cimini, G., Primicerio, K., di Clemente, R., Challet, D.: Statistically validated network of portfolio overlaps and systemic risk. Sci. Rep. **6**, 39,467 (2016). https://doi.org/10.1038/srep39467

51. Guimerà, R., Sales-Pardo, M., Amaral, L.A.N.: Module identification in bipartite and directed networks. Phys. Rev. E **76**, 036,102 (2007). https://doi.org/10.1103/PhysRevE.76.036102

52. Harte, J.: Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics. Oxford University Press, Oxford (2011)

53. Hausmann, R., Hidalgo, C.A.: The network structure of economic output. J. Econ. Growth **16**(October), 309–342 (2011). https://doi.org/10.1007/s10887-011-9071-4

54. Hidalgo, C.A., Hausmann, R.: The building blocks of economic complexity. Proc. Natl. Acad. Sci. USA. **106**(26), 10570–10575 (2009). https://doi.org/10.1073/pnas.0900943106

55. Hidalgo, C.A., Klinger, B., Barabasi, A.L., Hausmann, R.: The product space conditions the development of nations. Science (80-.). **317**(5837), 482–487 (2007). http://www.sciencemag.org/cgi/doi/10.1126/science.1144581

56. Hong, Y.: On computing the distribution function for the poisson binomial distribution. Comput. Stat. Data Anal. **59**(1), 41–51 (2013). https://doi.org/10.1016/j.csda.2012.10.006

57. James, A., Pitchford, J.W., Plank, M.J.: Disentangling nestedness from models of ecological complexity. Nature **487**(7406), 227–230 (2012). https://doi.org/10.1038/nature11214

58. Jaynes, E.: Information theory and statistical mechanics. Phys. Rev. **106**, 620–630 (1957). https://doi.org/10.1103/PhysRev.106.620

59. Krause, A., Giansante, S.: Interbank lending and the spread of bank failures: a network model of systemic risk. J. Econ. Behav. Org. **83**(3), 583–608 (2012). https://doi.org/10.1016/j.jebo.2012.05.015

60. Levy-Carciente, S., Kenett, D.Y., Avakian, A., Stanley, H.E., Havlin, S.: Dynamical macroprudential stress testing using network theory. J. Bank. Financ. **59**, 164–181 (2015). https://doi.org/10.1016/j.jbankfin.2015.05.008
61. Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Rev. Econ. Stat. **47**(1), 13–37 (1965)
62. Mastrandrea, R., Squartini, T., Fagiolo, G., Garlaschelli, D.: Enhanced reconstruction of weighted networks from strengths and degrees. New J. Phys. (2014). https://doi.org/10.1088/1367-2630/16/4/043022
63. McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P.: Species abundance distributions: moving beyond single prediction theories to integrate within an ecological framework. Ecol. Lett. **10**, 995–1015 (2007). https://doi.org/10.1111/j.1461-0248.2007.01094.x
64. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.: Network motifs: simple building blocks of complex networks. Sci. Rep. **298**(October), 11–14 (2002). www.sciencemag.org/cgi/content/full/298/5594/824/DC1
65. Molloy, M., Reed, B.: The critical phase for random graphs with a given degree sequence. Random Struct. Algorithms. **6**, 161–179 (1995). https://doi.org/10.1017/S096354830700867X
66. Mossin, J.: Equilibrium in a capital asset market. Econometrica **34**(4), 768–783 (1966)
67. Munoz, M.A., Jonhson, S., Dominquez-Garcia, V.: Factors determining nestedness in complex networks. PLoS ONE **8**(9), e74025 (2013). https://doi.org/10.1371/journal.pone.0074025
68. Newman, M.E.J.: Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. Phys. Rev. E **64**, 016,132 (2001). https://doi.org/10.1103/PhysRevE.64.016132
69. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E **69**(22), 026,113 (2004). https://doi.org/10.1103/PhysRevE.69.026113
70. O'Neill, J.: Who You Calling a BRIC? Bloomberg, New York (2013). https://www.bloomberg.com/view/articles/2013-11-12/who-you-calling-a-bric-. Accessed 05 Sep, 2017
71. Park, J., Newman, M.E.J.: Statistical mechanics of networks. Phys. Rev. E **70**(6), 66,117 (2004). https://doi.org/10.1103/PhysRevE.70.066117
72. Payrató Borrás, C., Hernández, L., Moreno, Y.: Breaking the spell of nestedness. ArXiv e-prints (2017). arXiv:1711.03134
73. Pilosof, S., Porter, M.A., Pascual, M., Kéfi, S.: The multilayer nature of ecological networks. Nat. Ecol. Evol. **1**(March), 1–9 (2017). https://doi.org/10.1038/s41559-017-0101
74. Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., Gabrielli, A.: Unfolding the innovation system for the development of countries: co-evolution of Science, Technology and Production. ArXiv e-prints (2017). arXiv:1707.05146
75. Pugliese, E., Zaccaria, A., Pietronero, L.: On the convergence of the Fitness-Complexity Algorithm. ArXiv e-prints (2014). arXiv:1410.0249
76. Pugliese, E., Zaccaria, A., Pietronero, L.: On the convergence of the Fitness-Complexity Algorithm. Eur. Phys. J. Spec. Top. **225**(10), 1893–1911 (2016). https://doi.org/10.1140/epjst/e2015-50118-1
77. Ricardo, D.: On the Principles of Political Economy, and Taxation. John Murray, London (1817)
78. Roberts, A., Stone, L.: Island-sharing by archipelago species. Oecologia **83**(4), 560–567 (1990). https://doi.org/10.1007/BF00317210
79. Saavedra, S., Reed-Tsochas, F., Uzzi, B.: Common Organizing Mechanisms in Ecological and Socio-economic Networks. ArXiv e-prints (2011) arXiv:1110.0376
80. Saracco, F., Di Clemente, R., Gabrielli, A., Squartini, T.: Randomizing bipartite networks: the case of the World Trade Web. Sci. Rep. **5**, 10,595 (2015). http://www.nature.com/articles/srep10595
81. Saracco, F., Di Clemente, R., Gabrielli, A., Squartini, T.: Detecting early signs of the 2007–2008 crisis in the world trade. Sci. Rep. **6**, 30,286 (2016). https://doi.org/10.1038/srep30286
82. Saracco, F., Straka, M.J., Di Clemente, R., Gabrielli, A., Caldarelli, G., Squartini, T.: Inferring monopartite projections of bipartite networks: an entropy-based approach. New J. Phys. **19**(5), 053,022 (2016). https://doi.org/10.1088/1367-2630/aa6b38
83. Serrano, M.A., Boguñá, M.: Topology of the world trade web. Phys. Rev. E **68**, 015,101 (2003). https://doi.org/10.1103/PhysRevE.68.015101
84. Shleifer, A., Vishny, R.W.: Fire sales in finance and macroeconomics. Working Paper 16642, National Bureau of Economic Research (2010). https://doi.org/10.3386/w16642
85. Shoval, O., Alon, U.: SnapShot: network motifs. Cell **143**(2), 326–326.e1 (2010). https://doi.org/10.1016/j.cell.2010.09.050
86. Smith, A.: An Inquiry into the Nature and Causes of the Wealth of Nations. W. Strahan and T. Cadell, London (1776)

87. Squartini, T., Almog, A., Caldarelli, G., van Lelyveld, I., Garlaschelli, D., Cimini, G.: Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks. Phys. Rev. E **96**, 032,315 (2017). https://doi.org/10.1103/PhysRevE.96.032315

88. Staniczenko, P.P.A., Kopp, J.C., Allesina, S.: The ghost of nestedness in ecological networks. Nat. Commun. **4**, 1391–1396 (2013). https://doi.org/10.1038/ncomms2422

89. Stone, L., Roberts, A.: The checkerboard score and species distributions. Oecologia **85**(1), 74–79 (1990). https://doi.org/10.1007/BF00317345

90. Straka, M.J., Caldarelli, G., Saracco, F.: Grand canonical validation of the bipartite International Trade Network. Phys. Rev. E **96**(022306), 1–12 (2017). https://doi.org/10.1103/PhysRevE.96.022306

91. Suweis, S., Simini, F., Banavar, J.R., Maritan, A.: Emergence of structural and dynamical properties of ecological mutualistic networks. Nature **500**(7463), 449–52 (2013). https://doi.org/10.1038/nature12438

92. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., Pietronero, L.: A new metrics for countries' fitness and products' complexity. Sci. Rep. **2**, 1–4 (2012). https://doi.org/10.1038/srep00723

93. Thebault, E.: Identifying compartments in presence-absence matrices and bipartite networks: insights into modularity measures. J. Biogeogr. (2012). https://doi.org/10.1111/jbi.12015. http://wileyonlinelibrary.com/journal/jbi

94. Thébault, E., Fontaine, C.: Stability of ecological communities and the architecture of mutualistic and trophic networks. Science **329**, 853–856 (2010). https://doi.org/10.1126/science.1188321

95. Toonders, J.: Data is the new oil of the digital economy. WIRED (2014). https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/. Accessed 10 Sep, 2017

96. Williams, R.J.: Simple MaxEnt models explain food web degree distributions. Theor. Ecol. (2010). https://doi.org/10.1007/s12080-009-0052-6

97. Williams, R.J.: Biology, methodology or chance? The degree distributions of bipartite ecological networks. PLoS ONE (2011). https://doi.org/10.1371/journal.pone.0017645

98. Wong, E., Baur, B., Quader, S., Huang, C.H.: Biological network motif detection: principles and practice. Brief. Bioinform. **13**(2), 202–215 (2012). https://doi.org/10.1093/bib/bbr033

99. World Economic Forum: Building Resilience in Supply Chains. Tech. Rep. January (2013). http://www3.weforum.org/docs/WEF_RRN_MO_BuildingResilienceSupplyChains_Report_2013.pdf

100. World Trade Organization: Trade in goods and services has fluctuated significantly over the last 20 years. Tech. rep. (2015). https://www.wto.org/english/res_e/statis_e/its2015_e/its15_highlights_e.pdf

101. Zaccaria, A., Cristelli, M., Tacchella, A., Pietronero, L.: How the taxonomy of products drives the economic development of countries. PLoS ONE **9**(12), 1–17 (2014). https://doi.org/10.1371/journal.pone.0113770

102. Zhou, T., Ren, J., Medo, M., Zhang, Y.C.: Bipartite network projection and personal recommendation. Phys. Rev. E (2007). https://doi.org/10.1103/PhysRevE.76.046115