



Universiteit
Leiden
The Netherlands

The statistical physics of real-world networks

Cimini, G.; Squartini, T.; Saracco, F.; Garlaschelli, D.; Gabrielli, A.; Caldarelli, G.

Citation

Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., & Caldarelli, G. (2019). The statistical physics of real-world networks. *Nature Reviews Physics*, 1, 58-71.
doi:10.1038/s42254-018-0002-6

Version: Not Applicable (or Unknown)

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

Downloaded from: <https://hdl.handle.net/1887/81728>

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

The Statistical Physics of Real-World Networks

Giulio Cimini,^{1,2} Tiziano Squartini,¹ Fabio Saracco,¹ Diego Garlaschelli,^{1,3} Andrea Gabrielli,^{2,1} and Guido Caldarelli^{1,2,4}

¹*IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca (Italy)*

²*Istituto dei Sistemi Complessi (CNR) UoS Sapienza, Dipartimento di Fisica, "Sapienza" Università di Roma, P.le A. Moro 2, 00185 Rome (Italy)*

³*Lorentz Institute for Theoretical Physics, Leiden University, Niels Bohrweg 2, 2333 CA Leiden (The Netherlands)*

⁴*European Centre for Living Technology, Università di Venezia "Ca' Foscari", S. Marco 2940, 30124 Venice (Italy)*

In the last 15 years, statistical physics has been a very successful framework to model complex networks. On the theoretical side, this approach has brought novel insights into a variety of physical phenomena, such as self-organisation, scale invariance, emergence of mixed distributions and ensemble non-equivalence, that display unconventional features on heterogeneous networks. At the same time, thanks to their deep connection with information theory, statistical physics and the principle of maximum entropy have led to the definition of null models for networks reproducing some features of real-world systems, but otherwise as random as possible. We review here the statistical physics approach and the various null models for complex networks, focusing in particular on the analytic frameworks reproducing the local network features. We then show how these models have been used to detect statistically significant and predictive structural patterns in real-world networks, as well as to reconstruct the network structure in case of incomplete information. We further survey the statistical physics models that reproduce more complex, semi-local network features using Markov chain Monte Carlo sampling, as well as the models of generalised network structures such as multiplex networks, interacting networks and simplicial complexes.

The science of networks has exploded in the Information Age thanks to the unprecedented production and storage of data on basically any human activity. Indeed, a network represents the simplest yet extremely effective way to model a large class of technological, social, economic and biological systems: a set of entities (nodes) and of interactions (links) among them. These interactions do represent the fundamental degrees of freedom of the network, and can be of different types—undirected or directed, binary or valued (weighted)—depending on the nature of the system and the resolution used to describe it. Notably, most of the networks observed in the real world fall within the domain of complex systems, as they exhibit strong and complicated interaction patterns, and feature collective emergent phenomena that do not follow trivially from the behaviours of the individual entities [1]. For instance, many networks are *scale-free* [2], meaning that the number of links incident to a node (known as the node's *degree*) follows a power-law distribution: most of the nodes have a few links, but a few of them (the hubs) are highly connected. The same happens for the distribution of the total weight of connections incident to a node (the node's *strength*) [3, 4]. In addition, most real-world networks are organised into modules or display a community structure [5, 6], and they possess high clustering—as nodes tend to create tightly linked groups, but are also *small-world* [7–9] as the mean distance (in terms of number of connections) amongst node pairs scales logarithmically with the system size. The observation of these universal features in complex networks has stimulated the development of a unifying mathematical language to model their structure and understand the dynamical processes taking place on them—such as the flow of traffic on the Internet or the spreading of either diseases or information in a population [10–12].

Two different approaches to network modelling can be pursued. The first one consists in identifying one or more microscopic mechanisms driving the formation of the network, and use them to define a dynamic model which can reproduce some of the emergent properties of real systems. The small-world model [7], the preferential attachment model [2], the fitness model [13–15], the relevance model [16] and many others follow this approach which is akin to kinetic theory. These models can handle only simple microscopic dynamics, and thus while providing good physical insights they need several refinements to give quantitatively accurate predictions.

The other possible approach consists in identifying a set of characteristic static properties of real systems, and then building networks having the same properties but otherwise maximally random. This approach is thus akin to statistical mechanics and therefore is based on rigorous probabilistic arguments that can lead to accurate and reliable predictions. The mathematical framework is that of *exponential random graphs* (ERG), which has been first introduced in the social sciences and statistics [17–25] as a convenient formulation relying on numerical techniques such as Markov chain Monte Carlo algorithms. The interpretation of ERG in physical terms is due to Park and Newman [26], who showed how to derive them from the principle of maximum entropy and the statistical mechanics of Boltzmann and Gibbs.

As formulated by Jaynes [27], the variational principle of maximum entropy states that the probability distribution

best representing the current state of (knowledge on) a system is the one which maximises the Shannon entropy, subject in principle to any prior information on the system itself. This means making self-consistent inference assuming maximal ignorance about the unknown degrees of freedom of the system [28]. The maximum entropy principle is conceptually very powerful and finds almost countless applications in physics and in science in general [29]. In the context of network theory, the ensemble of random graphs with given aggregated (macroscopic or mesoscopic) structural properties derived from the maximum entropy approach has a two-sided important application. On one hand, when the actual microscopic configuration of a real network is not accessible, this ensemble describes the most probable network configuration: as in traditional statistical mechanics, the maximum entropy principle allows to gain maximally unbiased information in the absence of complete knowledge. On the other hand, when the actual microscopic configuration of the network is known, this ensemble defines a null model which allows to assess the significance of empirical patterns found in the network—against the hypothesis that the network structure is determined solely by its aggregated structural properties.

The purpose of this review is to present the theoretical developments and empirical applications for the statistical physics of real-world complex networks. We start by introducing the general mathematical formalism, and then we focus on the analytic models obtained by imposing mesoscopic (*i.e.*, local) constraints, highlighting the novel physical concepts that can be learned from such models. After that we present the two main fields of applications for these models: the detection of statistically significant patterns in empirical networks, and the reconstruction of network structures from partial information. At the end we discuss the models obtained by imposing semi-local network features, as well as the most recent developments on generalised network structures and simplices.

Statistical mechanics of networks

The statistical physics approach defined by ERG consists in modelling a network system $\mathbf{G}^*$ through an ensemble Ω of graphs with the same number N of nodes and type of links of $\mathbf{G}^*$ (Fig. 1). The model is specified by $P(\mathbf{G})$, the occurrence probability of a graph $\mathbf{G} \in \Omega$. According to statistical mechanics and information theory [27, 30], in order to achieve the most unbiased expectation about the microscopic configuration of the system under study, such probability distribution is the one maximising the Shannon entropy

$$\mathcal{S} = - \sum_{\mathbf{G} \in \Omega} P(\mathbf{G}) \ln P(\mathbf{G}), \quad (1)$$

subject to the normalisation condition $\sum_{\mathbf{G} \in \Omega} P(\mathbf{G}) = 1$, as well as to a collection of constraints $\mathbf{c}^*$ representing the macroscopic properties enforced on the system (and thus defining the sufficient statistics of the problem).

Imposing hard constraints, that is, assigning uniform $P(\mathbf{G})$ over the set of graphs that satisfy $\mathbf{c}(\mathbf{G}) = \mathbf{c}^*$ and zero probability to graphs that do not leads to the *microcanonical ensemble*. Typically, this ensemble is resilient to analytic treatment beyond steepest descent approximations [31], and it is thus sampled numerically (see Box 1).

The *canonical ensemble* is instead obtained by imposing soft constraints, that is, by fixing the expected values of the constraints over the ensemble, $\sum_{\mathbf{G} \in \Omega} \mathbf{c}(\mathbf{G}) P(\mathbf{G}) = \mathbf{c}^*$. Introducing the set of related Lagrange multipliers $\boldsymbol{\theta}$, the constrained entropy maximisation returns

$$P(\mathbf{G}|\boldsymbol{\theta}) = e^{-H(\mathbf{G},\boldsymbol{\theta})} / Z(\boldsymbol{\theta}) \quad (2)$$

where $H(\mathbf{G}, \boldsymbol{\theta}) = \boldsymbol{\theta} \cdot \mathbf{c}(\mathbf{G})$ is the Hamiltonian and $Z(\boldsymbol{\theta}) = \sum_{\mathbf{G} \in \Omega} e^{-H(\mathbf{G},\boldsymbol{\theta})}$ is the partition function. Thus the canonical $P(\mathbf{G}|\boldsymbol{\theta})$ depends on $\mathbf{G}$ only through $\mathbf{c}(\mathbf{G})$, which automatically implies that graphs with the same value of the constraints have equal probability. This means that the canonical ensemble is maximally non-committal with respect to the properties that are not enforced on the system [32].

Remarkably, the microcanonical and canonical ensembles turn out to be non-equivalent in the thermodynamic limit $N \rightarrow \infty$ for models of networks with an extensive number of constraints [33–35]. This is in contrast to what happens in traditional statistical physics (except possibly at phase transitions), where the number of constraints—*e.g.*, total energy and total number of particles—is typically finite. In this sense, complex networks not only provide an important application domain for statistical physics, but can even expand our fundamental understanding of statistical physics itself, leading to novel theoretical insight. Additionally, from a practical point of view, the breaking of ensemble equivalence in networks implies that the choice between microcanonical and canonical ensembles cannot be based solely on mathematical convenience as usually done, and rather should follow from a principled criterion. In particular, since the canonical ensemble describes systems subject to statistical fluctuations, its use is more appropriate

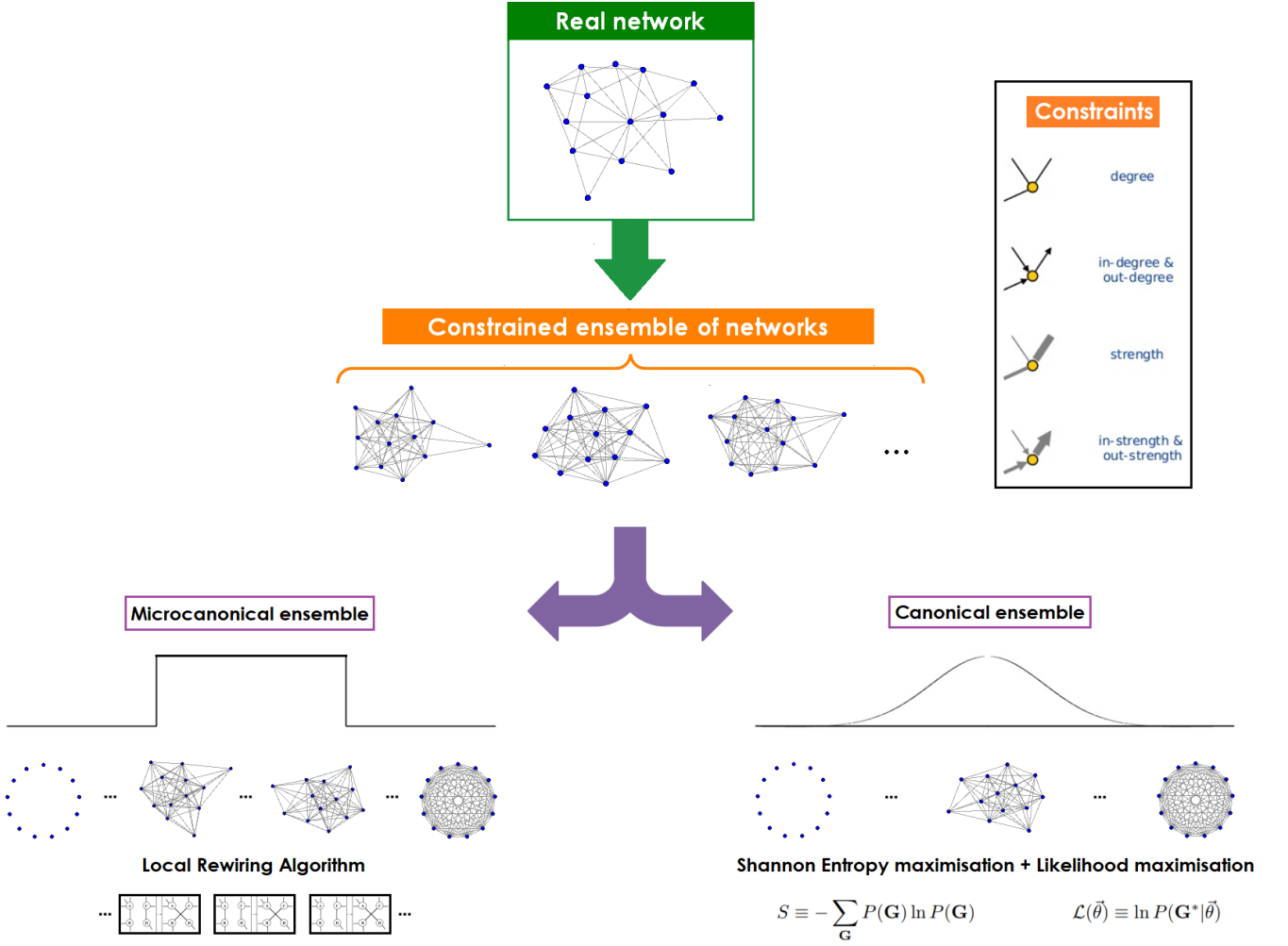


FIG. 1. Construction of the microcanonical and canonical ensemble of networks from local constraints—in this case, the degrees $\mathbf{k}^*$ of a real network $\mathbf{G}^*$. On the left, the microcanonical approach relies on the link rewiring method to numerically generate several network configurations, each with exactly the same degree sequence of $\mathbf{G}^*$: $P(\mathbf{G})$ is non-zero (and uniform, provided the sampling is unbiased) only for the subset of graphs that realise the enforced constraints exactly. On the right, the canonical approach obtains $P(\mathbf{G})$ by maximising the Shannon entropy constraining the expected degree values within the ensemble, and then maximising the Likelihood of $P(\mathbf{G}^*)$ to find the ensemble parameters $\vec{\theta}^*$ such that the expected degree values match the observations in $\mathbf{G}^*$. Thus $P(\mathbf{G})$ is non-zero for any graph (ranging from the empty to the complete one). Figure adapted from [38].

when the observed values of the constraints can be affected by measurement errors, missing and spurious data, or simply stochastic noise. Luckily, as we shall see, this criterion leads to the more analytically tractable ensemble.

The definition of the canonical ensemble of eq. (2) specifies the functional form of $P(\mathbf{G}|\vec{\theta})$, but leaves the Lagrange multipliers as parameters to be determined by the constraints equations $\sum_{\mathbf{G} \in \Omega} \mathbf{c}(\mathbf{G})P(\mathbf{G}|\vec{\theta}) = \mathbf{c}^*$. Since in practical applications the average values of the constraints are seldom available, a possible strategy is to draw the Lagrange multipliers from chosen probability densities inducing archetypal classes of networks (*e.g.*, regular graphs, scale-free networks, and so on) [26, 31, 36]. When instead the task is to fit the model to the observations $\mathbf{c}^* \equiv \mathbf{c}(\mathbf{G}^*)$ for a *given* empirical network $\mathbf{G}^*$, the optimal choice to find the values $\vec{\theta}^*$ is to maximise the likelihood functional [37, 38]

$$\mathcal{L}(\vec{\theta}) = \ln P(\mathbf{G}^*|\vec{\theta}). \quad (3)$$

This results in the matching $\sum_{\mathbf{G} \in \Omega} \mathbf{c}(\mathbf{G})P(\mathbf{G}|\vec{\theta}^*) \equiv \mathbf{c}(\mathbf{G}^*)$ between the ensemble average and the observed value of each constraint.

Box 1: alternative ensemble constructions

Various methods to define ensembles of graphs with local constraints, alternative to maximum entropy, have been proposed in the literature. Here we briefly present them for the case of binary undirected graphs, which is by far the simplest and most frequently explored situation.

Computational methods explicitly generate several random networks with the desired degree sequence. The ‘bottom-up’ approach initially assigns to each node a number of link stubs equal its target degree; then, pairs of stubs are randomly matched avoiding the formation of self-loops and multi-links [157–159]. Unfortunately, very often this procedure gets stuck in configurations where nodes requiring additional connections have no more eligible partners, leading to unacceptably many sample rejections [45]. The ‘top-down’ approach instead starts from a realised network and generates a set of randomised variants by iteratively applying a *link rewiring algorithm* that preserves the degree distribution [42, 58, 160, 161]. The drawbacks here is that the numbers of rewirings needed to generate a single configuration is very large and not rigorously specified [162]. Additionally, the algorithm may fail to sample the ensemble uniformly, unless employing a rewiring acceptance probability depending on the current network configuration [126–128, 163]. Other methods rely on theorems setting necessary and sufficient conditions for a degree sequence to be *graphic*, *i.e.*, realised by at least one graph, and exploit such conditions to define biased sampling algorithms and sampling probabilities [164–166]. These approaches are however rather costly, especially for highly heterogeneous networks.

Analytic methods instead define (usually approximated) explicit expressions for expected values of network properties as a function of the imposed constraints. A standard approach relies on the *generating function* $g(z) = \sum_k z^k P(k)$ for the degree distribution [157, 167], yet is rigorously defined only for infinite and locally tree-like networks—even if it often works surprisingly well for real networks [168]. An alternative popular approach is based on the explicit expression $p_{ij} = \frac{1}{2} k_i^* k_j^* / E^*$ for the connection probability between any two nodes i and j in the randomised ensemble (E^* is the total number of links) [67]. This model actually defines systems with self-loops as well as multilinks [37, 40, 43]. In particular, since p_{ij} may exceed 1 for pairs of high-degree nodes, the model requires that i and j should be connected by more than one link to actually realise the imposed constraints. The occurrence of these events is not negligible in scale-free networks with $P(k) \sim k^{-\gamma}$ for which the natural cutoff $\sim N^{1/(\gamma-1)}$ is larger than the structural cut-off ($\sim N^{1/2}$ for uncorrelated networks) [169, 170].

Imposing local constraints — Unlike most alternative approaches (briefly outlined in Box 1), the maximum entropy method is general and works for networks that are either binary or weighted, undirected or directed, sparse or dense, tree-like or clustered, small or large. However, the specification of the occurrence probability of a graph in the ensemble can be a challenging task. The point is that, like in conventional equilibrium statistical mechanics, whether the partition function can be analytically computed depends on the particular constraints imposed. In a handful of lucky cases such computation is indeed feasible, so that expectation values and higher moments of any quantity in the ensemble can be analytically derived. Besides very simple models like the well-known Erdős-Rényi random graph [39], this happens for the important constraints describing the *local* network structure from the viewpoint of each individual node—namely, degrees $\mathbf{k}^*$ and strengths $\mathbf{s}^*$ [26]. The scale-free behaviour observed for these quantities in real-world systems is indeed the most elemental signature distinguishing networks from systems typically studied in physics such as gases, liquids, lattices, and cannot be obtained from simple models with global constraints. For instance, the Erdős-Rényi model is obtained in the ERG formalism by constraining the expected total number of links. This leads to an ensemble in which each pair of nodes is connected with fixed probability p , so that the degree distribution follows a binomial law. Thus in order to construct ERG ensembles that are both practically useful and theoretically sound (by accurately replicating the observed heterogeneity of real-world networks), imposing local constraints separately for each node is a minimum requirement. And local constraints make the method analytic because of the *independence of dyads*: $P(\mathbf{G})$ factorises into link-specific terms, whose contribution to the partition function sum can be evaluated independently from the rest of the network. Note that local constraints lay at the *mesoscopic* level between the microscopic degrees of freedom of the network (the individual links) and the macroscopic aggregation of all degrees of freedom into global quantities like the total number of links—corresponding for instance to the total energy of the system in traditional statistical physics.

The ERG model obtained by constraining the degrees $\mathbf{c}^* \equiv \mathbf{k}^*$ is known as the (canonical) *binary configuration model* (BCM). In the simplest undirected case, the entropy maximisation procedure returns an ensemble connection

probability between any two nodes i and j given by

$$p_{ij} = \frac{x_i x_j}{1 + x_i x_j}, \quad (4)$$

where $\mathbf{x}$ are the (exponentiated) Lagrange multipliers [26]. The *weighted configuration model* (WCM) [40] is instead obtained by constraining the strengths $\mathbf{c}^* \equiv \mathbf{s}^*$. Again in the simpler undirected case, and considering integer weights, the connection probability between any two nodes i and j is given by $p_{ij} = y_i y_j$, where $\mathbf{y}$ are the (exponentiated) Lagrange multipliers. The probability distribution and the ensemble average for the weight of that link (or, equivalently, for how many links are established between the two nodes) are $q_{ij}(w) = (y_i y_j)^w (1 - y_i y_j)$ and

$$\langle w_{ij} \rangle = \frac{y_i y_j}{1 - y_i y_j}. \quad (5)$$

These models naturally recall the traditional statistical mechanics for systems of non-interacting particles, once connections are interpreted as particles in a quantum gas and pairs of nodes as single-particle states. Indeed in binary networks each single-particle state can be occupied by at most one particle, so the resulting statistics of eq. (4) is fermionic, whereas, in weighted networks single-particle states can be occupied by an arbitrary number of particles, so that eq. (5) better describes a system of bosons—where Bose-Einstein condensation can occur between very strong nodes for which $y_i y_j \rightarrow 1$ [26]. Notably, a mixed Bose-Fermi statistics is obtained when degrees and strengths are imposed simultaneously [36], as in the *enhanced configuration model* (ECM) [41]. Again in the simplest undirected case and using (exponentiated) Lagrange multipliers $\mathbf{x}$ and $\mathbf{y}$ respectively for degrees and strengths, in this case one gets $p_{ij} = (x_i x_j y_i y_j) / (x_i x_j y_i y_j - y_i y_j + 1)$ and $q_{ij}(w > 0) = p_{ij} (y_i y_j)^{w-1} (1 - y_i y_j)$. Hence the ECM differs from the WCM in the way the first link established between any two nodes is treated: the processes of creating a connection from scratch and that of reinforcing an existing one obey intrinsically different rules, the former meant to satisfy the degree constraints and the second to fix the values of the strengths. Like the ensemble non-equivalence, this mechanism and the resulting mixed statistics constitutes a novel physical phenomenon that the statistical physics approach to networks can unveil.

Patterns validation

Validating models, that is comparing their statistical properties with measurements of real-world systems, is an essential step in the activity of theoretical physicists. Specifically, in the context of networks and complex systems, besides looking for what a model is able to explain, much research has been devoted to identify the empirical properties which deviate from a benchmark model [42–50]. This is because possible deviations likely bear important information about the unknown formation process or a particular function of the empirical network.

Maximum entropy models are perfectly suited for this task. Starting from a real network $\mathbf{G}^*$, they derive the null hypothesis (that is, the benchmark model) from the set of properties $\mathbf{c}(\mathbf{G}^*)$ imposed as constraints, and otherwise assuming no other information on the system. This means formulating the null hypothesis that these constraints are the only explanatory variables for the network at hand. The other properties of $\mathbf{G}^*$ can then be statistically tested and possibly validated against this null hypothesis. For instance, a null model derived from imposing the total number of links as (macroscopic) constraint is typically used to reject a homogeneity hypothesis for the degree distribution. Instead, when imposing local constraints the aim is to check whether higher-order patterns of a real network—such as reciprocity (the tendency of nodes in a directed network to be mutually linked), clustering (the tendency of node triples to be connected together forming triangles), (dis-)assortativity (the tendency of nodes to be linked to other nodes with (dis-)similar degrees)—are statistically significant beyond what can be expected from the heterogeneity of degrees or strengths themselves. Note that the possibility to analytically characterise the ensemble given by the use of local constraints means that expectation values and standard deviations of most quantities of interests can be explicitly derived, so that hypothesis testing based on standard scores can be easily performed [38]. And when the ensemble distribution of the considered quantity is not normal, sampling of the configuration space using the explicit formulas for $P(\mathbf{G})$ easily allows to perform statistical tests based on p -values.

In the context of patterns validation, one of the most studied systems—that we also discuss here as an illustrative example—has been the World Trade Web (WTW), namely the network of trade relationships between countries in the world [51, 52]. The observed disassortative pattern, for which countries with many trade partners are connected on average to countries with few partners [53, 54], turns out to be statistically explained with good approximation by the degree sequence [38]. The same happens with the clustering coefficient [38]. These two observations exclude the presence of meaningful indirect (economic) interactions on top of a concatenation of direct (economic) interactions.

Things change however when the weighted version of the network is analysed—although no unique extension of binary quantities exists in this case [44, 55, 56]. Indeed, the disassortative pattern is still observed, but is not compatible with that of the WCM null model. Concerning the weighted clustering coefficient [57], the agreement between empirical network and model is only partial. These findings point to the fact that, unlike in the binary case, the knowledge of the strengths conveys only limited information about the higher-order weighted structure of the network. This, together with the fact that also basic topological properties like the link density (*i.e.*, the fraction of possible connections that are actually realised in the network) are not reproduced by the WCM, suggests that even in weighted analyses the binary structure plays an important role, irreducible to what local weighted properties can explain.

Network motifs and communities — Motifs [58, 59] are patterns of interconnections involving a small subset of nodes in the network, thus generalising the clustering coefficient. The typical null model used to study motifs in directed networks is obtained by constraining, in addition to degrees, also the number of reciprocal links per node [60–62], which is meaningful in many contexts. For instance, in food webs the presence of bi-directed predator-prey relations between two species strongly characterises an ecosystem [63]. In interbank networks, the presence of mutual loans between two banks is a signature of trust between them, and in fact the appearance of the motif corresponding to three banks involved in a circular lending loop with no reciprocation provides significant early warnings of financial turmoil [64].

Community structure instead refers to the presence of large groups of nodes that are densely connected internally but sparsely connected externally [6]. Most of the methods to find communities in networks are based on the optimisation of a functional, the *modularity* [5] being the most prominent example, that compares the number of links falling within and between groups with the expectation of such numbers under a given null network model. This comparison with the null model is a fundamental step of the procedure, since even random graphs possess an intrinsic yet trivial community structure [65, 66]. Indeed, in its original formulation the modularity is defined on top of the configuration model by Chung and Lu [67] (see Box 1). This model is fast and analytic but generates self-loops and multilinks, hence it gives an accurate benchmark only when these events are rare (*i.e.*, in very large and sparse networks). More generally the maximum entropy approach (*e.g.*, the configuration model of eq. (4)), while more demanding from a practical viewpoint, can provide the proper null model discounting the degree heterogeneity as well as other properties of the network [38, 68]. Notably, the ERG framework can be directly used to generate networks with a community structure, by specifying the average numbers of links within and between each community. In this way the ensemble becomes equivalent to the classical *block-model*, in which each node is assigned to one of B blocks (communities), and links are independently drawn between pairs of nodes with probabilities that are a function only of the block membership of the nodes [69]. This means that eq. (4) becomes $p_{ij} = q_{b_i b_j}$, where b_i denotes the block membership of node i , or $p_{ij} = (x_i x_j q_{b_i b_j}) / (x_i x_j q_{b_i b_j} + 1 - q_{b_i b_j})$ for the degree-corrected block-model [70–72] in which also nodes' total degrees are constrained.

Bipartite networks and one-mode projections — Bipartite networks are a particular class of networks whose nodes can be divided into two disjoint sets, such that links exist only between nodes belonging to different sets [73]. Typical examples of these systems include affiliation networks, where individuals are connected with the groups they are member of, and ownership networks where individuals are connected with the items they collected. The *Bipartite Configuration Model* (BiCM) [74] extends the BCM to this class of networks. This method has been used, for instance, to study the network of countries and products they export (*i.e.*, the bipartite representation of the WTW) [75, 76] and detect temporal variations related to the occurrence of global financial crises [77]. More recently, the BiCM has been applied to show that the degree sequence of interacting species in mutualistic ecological networks already induces a certain amount of *nestedness* of the interactions [78].

In order to directly show the relation structure among one of the two sets of nodes, bipartite networks can be compressed into one-mode projection, namely a network containing only the nodes of the considered set—connected according to how many common neighbours they have in the other set [79]. The problem of building a statistically validated projection of a bipartite network is similar in spirit to that of extracting the backbone from standard weighted networks [80–83], and has been typically addressed by determining which links are significant using a threshold—either unconditional or depending on the degree of nodes in the projected set [84–88]. However, differently from weighted networks, one-mode projections should be assessed against null models constraining the local information of both sets of the original bipartite network. Unfortunately, these models are much more difficult to derive. *Degree sequence models* used in the social sciences are based on computational link rewiring methods [89, 90] (which in this case are even more impractical and biased [91] than for standard networks) or require multiple observations of the empirical network [92]. The null model for bipartite network projections derived from the maximum entropy principle is instead obtained by one-mode projecting the BiCM, that is, by computing the expected distribution for the number of common neighbours between nodes on the same layer [93, 94] (Fig. 2). Null models of these sort have been used, for instance, to analyse the one-mode projection of the bipartite WTW, allowing to detect modules of countries with

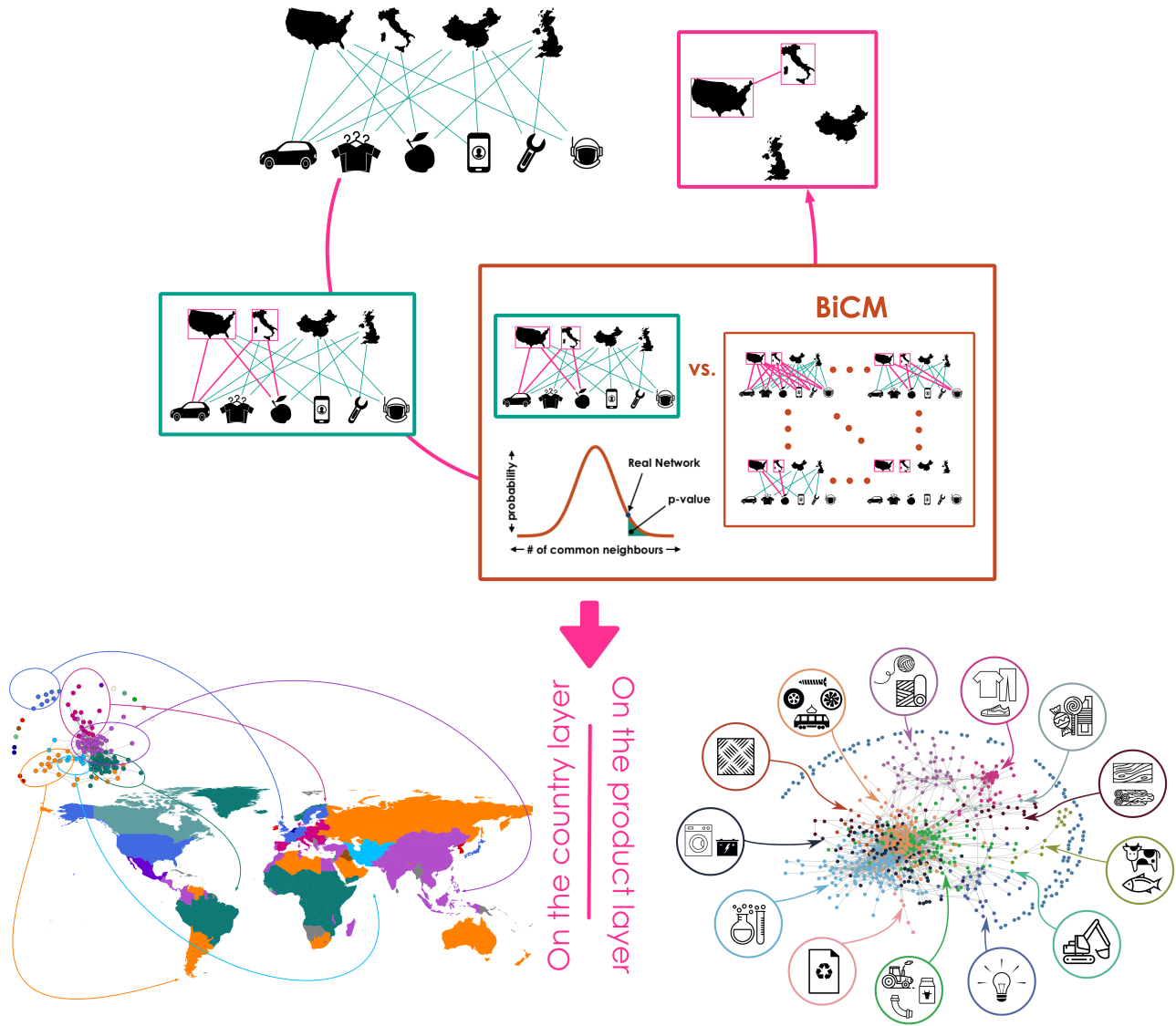


FIG. 2. One-mode projection of the network of countries and products they export (the bipartite representation of the WTW) and its statistical validation against a null hypothesis derived from the bipartite configuration model (BiCM). In the upper part of the figure, we illustrate the procedure applied on the countries set. First of all, the BiCM ensemble is derived constraining the degrees of both countries and products. The result is a connection probability between any country-product pair analogous to that of eq. (4). Then, taken two countries, the actual number of products they both export (cyan box) is compared with the expectation and probability distribution obtained from the BiCM (orange box). If this observed value is statistically significant (*i.e.*, the null hypothesis that such value is explained by the degrees of both countries and products is rejected), a link connecting the two countries is established in the one-mode validated projection (magenta box). On the bottom left, we report results of the projection performed on the country set of the real WTW. The validation procedure helps to identify communities of countries with a similar industrial system. On the bottom right, we instead show that when the projection is performed on the product set, the statistical test highlights products requiring similar technological capabilities. Figure readapted from [94]. All icons are from the Noun Project and under the CC licence available at <http://thenounproject.com>.^a

^a “China” by Alexander Skowalsky; “Italy” by Mehmet I K Berker; “United Kingdom” by Luke Peek; “United States” by DeeAnn Gray; “Apple” by Kimberly Chandler; “Car” by Olivier Guin; “Clothes” by Daniel Hanly; “Smartphone” by Gregor Cresnar; “Screwdriver” by FR; “Astronaut Helmet” by Yoshe; “Cow” by Nook Fulloption; “Fish” by Iconic; “Excavator” by Kokota; “Light bulb” by Hopkins; “Milk” by Artem Kovyazin; “Curved Pipe” by Oliviu Stoian; “Tractor” by Iconic; “Recycle” by Agus Purwanto; “Experiment” by Made; “Accumulator” by Aleksandr Vector; “Washing Machine” by Tomas Knopp; “Metal” by Leif Michelsen; “Screw” by Creaticca Creative Agency; “Tram” by Gleb Khorunzhiy; “Turbine” by Luigi Di Capua; “Tire” by Rediffusion; “Ball Of Yarn” by Denis Sazhin; “Fabric” by Oliviu Stoian; “Shoe” by Giuditta Valentina Gentile; “Clothing” by Marvdrock; “Candies” by Creative Mania; “Wood Plank” by Cono Studio Milano; “Wood Logs” by Alice Noir.

similar industrial systems and a hierarchical structure of products [94], as well as traces of specialisations emerging from the baseline diversification strategy of countries [95]. Or to study the patterns of assets ownership by financial institutions, identify significant portfolio overlaps bearing the highest riskiness for fire sales liquidation, and forecast market crashes and bubbles [93]. More recently, a null model obtained by pairwise projecting multiple bipartite networks has been successfully applied to identify significant innovation patterns involving the interplay of scientific, technological and economic activities [96].

Network reconstruction

Many dynamical processes of critical importance, from the spread of infectious diseases to the diffusion of information and the propagation of financial losses, are highly sensitive to the topology of the underlying network of interactions [97]. However in many situations the structure of the network is at least partially unknown. A classical example is that of financial networks: financial institutions publicly disclose their aggregate exposures in their balance sheets, whereas, individual exposures (*who* is lending to whom and *how much*) remain confidential [98–100]. Another example is that of social networks, for which only aggregate information is released due to privacy issues, while their large scales deny the possibility of exhaustive crawling [101, 102]. For natural and biological networks, collecting all kinds of interactions is highly demanding because of technological limitations or high experimental costs [103, 104]. Thus reconstructing the network structure when only limited information is available is relevant across several domains, and represents one of the major challenges for complexity science.

When the task is to predict individual missing connections in partly known networks one talks about *link prediction* [105]. Here we instead discuss the fundamentally different task of reconstructing a whole network from partial information on the system, aggregated at the mesoscopic and macroscopic level [106]. The key to success is of course to make optimal use of what is known on the system, but also to make the most unbiased guess about what is not known. This is naturally achieved through the maximum entropy principle: the probability distribution which best represents the current state of knowledge on the network is the one with the largest uncertainty but satisfying the constraints corresponding to the available information. Note that the ERG approach has the additional advantage of not defining a single reconstructed network instance but an ensemble of plausible configurations with related probabilities. In this way, it can handle spurious or fluctuating data, and obtain robust confidence intervals for the outcomes of a given dynamical process on the (unknown) network.

Different kinds of local constraints lead to substantially diverse outcomes for the reconstruction process, though. Indeed, for a variety of networks of different nature (*e.g.*, economic, financial, social, ecological networks), constraining the degrees as in the BCM typically returns a satisfactory reconstruction of the binary network features, whereas, constraining the strengths as in the WCM leads almost always to a very bad weighted reconstruction [38, 41]. The latter result is due to the entropy maximisation procedure being unbiased by not assuming any dependency between the strength of a node and the number of connections that node can establish. Hence, out of the many possible ways to redistribute the strength of each node over all possible links, the method chooses the most even one, so that the probability of assigning zero weight to a link is extremely small: the reconstructed network becomes almost fully connected—whatever the link density of the original network. This shows that degrees and strengths do carry different kinds of information, and constrain the network in a fundamentally different way. In order to reconstruct sparse weighted networks, both of them are required [47, 49], as in the ECM [41]. This is quantitatively revealed using information-theoretic criteria (see [41] and Box 2).

Box 2: comparing models from different constraints

Various likelihood-based statistical criteria exist to compare competing models resulting from different choices of the constraints, which are thus useful to assess the informativeness of different network properties. Consider two models $\mathcal{M}_1$ and $\mathcal{M}_2$. When these models are *nested*—meaning that $\mathcal{M}_2$ contains extra parameters with respect to $\mathcal{M}_1$, or equivalently that $\mathcal{M}_1$ is a special case of $\mathcal{M}_2$, comparison can be made through the *Likelihood Ratio Test* [171]: if $\text{LRT}_{\mathcal{M}_1/\mathcal{M}_2} = -2[\mathcal{L}_{\mathcal{M}_1}(\theta_1^*) - \mathcal{L}_{\mathcal{M}_2}(\theta_2^*)]$ is smaller than a given significance level, $\mathcal{M}_2$ should be rejected even though its likelihood is by definition higher than that of $\mathcal{M}_1$, because $\mathcal{M}_2$ over-fits the data via redundant parameters that have inter-correlations and thus provide spurious information on the system [172]. When $\mathcal{M}_1$ and $\mathcal{M}_2$ are not nested, the *Akaike Information Criterion* [173] ranks them in increasing order

of $\text{AIC}_{\mathcal{M}_r} = 2[K_{\mathcal{M}_r} - \mathcal{L}_{\mathcal{M}_r}(\boldsymbol{\theta}_r^*)]$, where $K_{\mathcal{M}_r}$ is the number of parameters of model $\mathcal{M}_r$. For R competing models, *Akaike weights* $\omega_{\mathcal{M}_r} = e^{-(\text{AIC}_{\mathcal{M}_r}/2)} / \sum_{r=1}^R e^{-(\text{AIC}_{\mathcal{M}_r}/2)}$ quantify the probability that each model is the most appropriate to describe the data [174]. The *Bayesian Information Criterion* [175] is similar to AIC, but accounts for the number n of empirical observations as $\text{BIC}_{\mathcal{M}_r} = K_{\mathcal{M}_r} \ln n - 2\mathcal{L}_{\mathcal{M}_r}(\boldsymbol{\theta}_r^*)$. *Bayesian weights* can be also defined, analogously to Akaike weights. BIC is believed to be more restrictive than AIC [172], yet which criteria performs best and under which conditions is still debated. Other model-selection methods have also been proposed, such as *multimodel inference* where a sort of average over different models is performed [172].

The fitness ansatz — Unfortunately in many situations, typically in financial networks, also the node degrees are unknown. A possible solution here comes from the observation that, in many real networks, the connection probabilities between nodes can be expressed in terms of *fitnesses*, which are node “masses” typical of the system under analysis, so that node degrees can be mapped to their fitness values [14, 52, 107–109]. The strengths themselves work well as fitnesses in many cases [44]. The *fitness ansatz* then assumes that the strength of a node is a monotonic function of the Lagrange multiplier of the BCM controlling the degree of that node. Assuming the simplest linear dependency $\mathbf{s}^* \propto \mathbf{x}$ (but other functional forms are in principle allowed), eq. (4) becomes [110–112]

$$p_{ij} = \frac{zs_i^* s_j^*}{1 + zs_i^* s_j^*}. \quad (6)$$

The proportionality constant z can be easily found using maximum likelihood arguments, together with a bootstrapping approach to assess the network density [113]. Node degrees estimated from such a fitness ansatz are then used to feed the ECM and obtain the ensemble of reconstructed networks [111]. Heuristic techniques can be alternatively employed to reduce the complexity of the method, by replacing the construction of the ECM ensemble with a *density-corrected gravity model* [112]. The methodology is also ready extendable to bipartite networks [114].

Note that despite having only the strengths as input, the described reconstruction method (Fig. 3) is different from the WCM by using these strengths not to directly reconstruct the network, but to estimate the degrees first, and only then to build the maximum entropy ensemble. In such a way it can generate sparse and non-trivial topological structures, and can faithfully be used to reconstruct complex networked systems [100, 106].

Beyond local constraints

ERG models are analytically tractable or not depending on whether a closed-form expression of the partition function Z can be derived. As we have seen, this is indeed the case of local linear constraints, for which Z factorises into link-specific terms. In some other cases, it may be possible to get approximate analytic solutions using a variety of techniques (mean-field theory, saddle-point approximation, diagrammatic perturbation theory, path integral representations). These situations have been vastly explored in the literature, and include the degree-correlated network model [115], the reciprocity model and the two-star model [116, 117], the Strauss model of clustering [118], models of social collaboration [119], models of community structure [31], hierarchical topologies [120], models with spatial embedding [121] and rich-club features [122], and finally model constraining both the degree distribution and degree-degree correlations—which are known under the name of *Tailored Random Graphs* [123–125]. At last, when any analytic approach for computing Z becomes intractable, the ensemble can be still populated using Monte Carlo simulations, either to explicitly sample the configuration space—taking care of avoiding sampling biases through the use of ergodic Markov chains fulfilling detailed balance [126–128], or to derive approximate maximum likelihood estimators—taking care of avoiding degenerate regions of the phase space leading to frequent trapping in local minima [129–135]. Such a variety of possible techniques is what makes ERG an extremely ductile and powerful framework for complex network modelling.

Markov chain Monte Carlo — The usual Markov chain Monte Carlo (MCMC) method for ERG works as follows. Starting from a network $\mathbf{G} \in \Omega$, a new network $\mathbf{G}' \in \Omega$ is proposed by choosing two links at random and shuffling them as to preserve a given network property (such as the degree sequence of the network). The proposed network is accepted with the Metropolis-Hastings probability $Q_{\mathbf{G} \rightarrow \mathbf{G}'} = \min\{1, e^{H(\mathbf{G}) - H(\mathbf{G}')}\}$ [136], where H is the Hamiltonian containing the Lagrange multipliers—the latter taking the role of inverse temperature(s) used for simulated annealing. The process is repeated from $\mathbf{G}'$ ($\mathbf{G}$) if the proposal is accepted (rejected). Since the moves fulfil ergodicity and detailed balance, for sufficiently long times the values of the constraints in the sampled networks are distributed according to what the canonical ensemble prescribes. However, despite this asymptotic guarantee, in practice this method often fails because the time to approximate the probability distribution grows exponentially

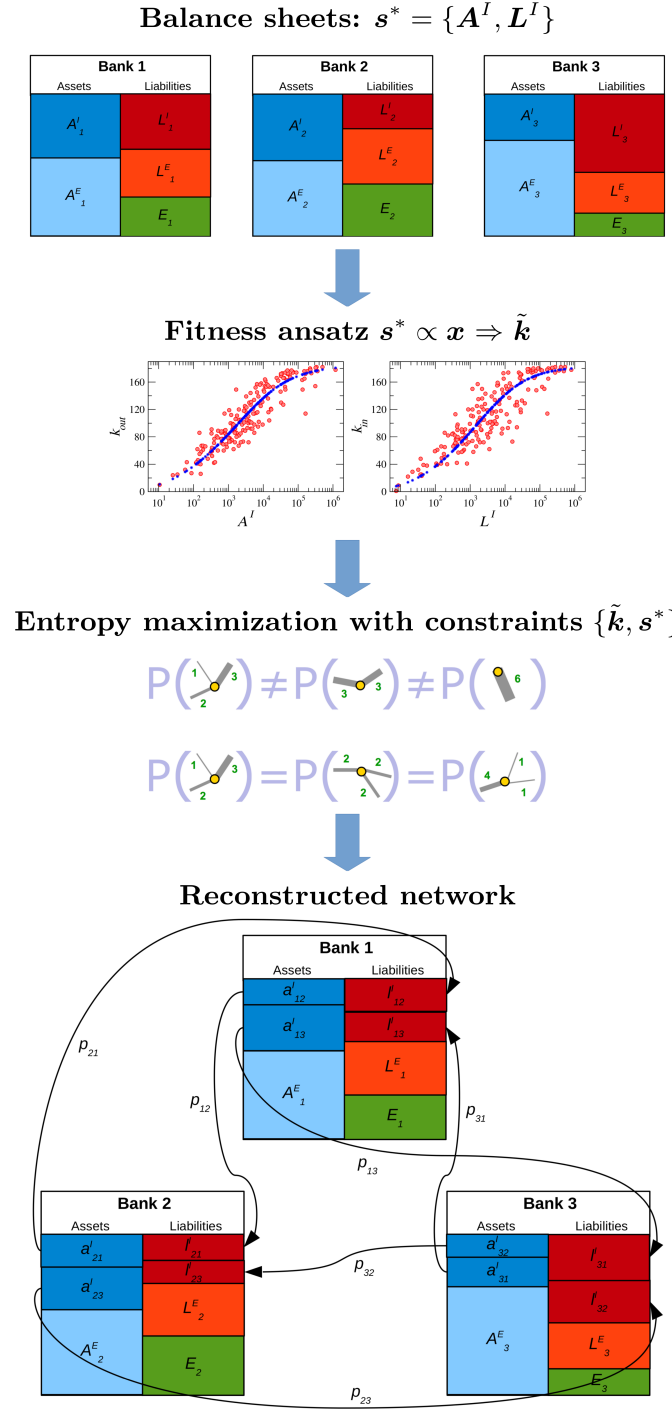


FIG. 3. Statistical reconstruction of an interbank network given by bilateral exchanges among banks. The method takes as input the exposure of each bank (*i.e.*, its total interbank assets A^I and liabilities L^I , obtained from publicly available balance sheets, which corresponds to the strength s^* of the node), and then performs two entropy maximisation steps. The first step estimates the (unknown) node degrees $\tilde{k}$ and reconstructs the binary topology of the network using the fitness assumption that the number of connections and the total exposure of each bank scale together—resulting in a connection probability between nodes of the form of eq. (6). This step requires as additional input an estimate of the density of the network, which can be obtained using a bootstrapping approach [113] relying on the hypothesis of statistical homogeneity (subsets of the network are representative of the whole system, regardless of the specific portion that is observed). The second step reconstructs the full weighted topology of the network, using either an ECM entropy maximisation constraining both degrees $\tilde{k}$ estimated from the first step and empirical strengths s^* [111], or a heuristic density-corrected gravity approach [112] to place weights on realised connections. The outcome of the reconstruction process is a probability distribution over the space of networks compatible with the constraints. Scatter plots adapted from [112].

with the system size. This happens whenever $P(\mathbf{G})$ possesses more than one local maximum, for instance when aiming at generating ensembles of networks with desired degree distribution, degree-degree correlations and clustering coefficient (within the so-called *dk-series* approach) [137, 138]. Indeed rewiring methods biased by aiming at a given clustering coefficient display strong *hysteresis* phenomena: cluster cores of highly interconnected nodes do emerge during the process, but once formed they are very difficult to remove in realistic sampling time scales—leading to a break of ergodicity [139]. Multi-canonical sampling has been recently proposed to overcome this issue of phase transitions [140]. The idea is to explore the original canonical ensembles without being restricted to the most probable regions, which means sampling networks uniformly on a predefined range of constraints values. This is achieved using Metropolis-Hastings steps based on a microcanonical density of states estimated using the Wang-Landau algorithm [141].

Generalised network structures

Networks of networks — Many complex systems are not just isolated networks, but are better represented by a “network of networks” [142–144] (Fig. 4). The simplest and most studied situation is the so-called *multiplex*, where the same set of nodes interacts on several layers of networks. This is the case, for instance, of social networks where each individual has different kinds of social ties, or urban systems where locations can be connected by different means of transportation, or financial markets where institutions can exchange different kinds of financial instruments.

Mathematically, a multiplex $\tilde{\mathbf{G}}$ is a system of N nodes and M layers of interactions, each layer $\alpha = 1, \dots, M$ consisting a network $\mathbf{G}^\alpha$. When modelling such a system, the zero-th order hypothesis is that the various layers are uncorrelated. In ERG terms, this means that the probability of the multiplex factorises into the probabilities of each network layer: $P(\tilde{\mathbf{G}}) = \prod_{\alpha=1}^M P_\alpha(\mathbf{G}^\alpha)$ [145, 146]. This happens whenever the constraints imposed on the multiplex are linear combination of constraints on individual network layers. Imposing local constraints separately for each network layer falls into this category. For instance, imposing the degrees in each layer leads to M independent BCMs (one for each layer). The connection probability between any two nodes i and j in layer α reads $p_{ij}^\alpha = (x_i^\alpha x_j^\alpha) / (1 + x_i^\alpha x_j^\alpha)$ where $\mathbf{x}^\alpha$ are layer-specific Lagrange multipliers, meaning that the existence of a link is independent on the presence of *any* other link in the multiplex. In this situation, the overlap of links between pairs of sparse network layers vanishes in the large N limit.

More realistic models of correlated multiplexes—in which the existence of a link in one layer is correlated with the existence of a link in another layer—can be generated by constraining the *multilink* structure of the system [145]. A *multilink* $\mathbf{m}$ is an M -dimensional binary vector $\mathbf{m}$ indicating a given pattern of connections between a generic pair of nodes in the various layers. The *multidegree* $k(\mathbf{m})$ of a node in a given graph configuration is then the total number of other nodes with which the multilink $\mathbf{m}$ is realised. Constraining the multidegree sequence of the network (*i.e.*, imposing 2^M constraints per node) leads to a probability for a multilink $\mathbf{m}$ between node i and node j given by $p_{ij}^{\mathbf{m}} = (x_i^{\mathbf{m}} x_j^{\mathbf{m}}) / (\sum_{\mathbf{m}} x_i^{\mathbf{m}} x_j^{\mathbf{m}})$, which can be used to build systems made up of sparse layers with non-vanishing overlap

Concerning weighted multiplexes, imposing the strength sequence (or degrees and strengths) on each layers is simple and again leads to uncorrelated layers [147]. Models of correlated weighted multiplexes are instead obtained by constraining nodes multistrength, the weighted analogous of the multilink [148]. An interesting and related case of study is provided by systems of aggregated multiplexes, *i.e.*, simple networks given by the sum of the various layers of the multiplex. In this case, constraining the aggregated local structure leads to the traditional canonical ensemble. For instance, constraining the sum of the weights of connections incident to each node in each layer leads to the standard WCM. However, if we have information about the layered structure of the system—meaning that we know the number of layers of the original multiplex, the model can be built using a layer-degeneracy term for each link counting the number of ways its weight can be split across the layers. By doing so, one obtains a $\text{WCM}(M)$ model where the weight distribution is a negative binomial—the geometric distribution of the standard WCM being the special case $M = 1$. Additionally, if we know that links in each layer possess different nature, a good modelling framework is equivalent to that of *multiedge* networks (ME) [149, 150], where links belonging to different layers can be distinguished. To deal with this situation, it is necessary to introduce another degeneracy term counting all the network configurations giving rise to the same $P(\mathbf{G})$. The solution of the model is then obtained using a mixed ensemble with hard constraint for the total network weight and soft constraints for node strengths. This leads to a Poisson statistics (independent on M) for link weights and $\langle w_{ij} \rangle = M y_i y_j \propto s_i^* s_j^*$, a rather different situation than the outcome of either WCM or $\text{WCM}(M)$ [151]. Now, it can be tricky to decide which statistics to use in a specific case. For instance, mobility or origin-destination networks, consisting of number of travels between locations (nodes) aggregated over observation periods (layers), are better modelled by ME networks—assuming that each trip is distinguishable. For the aggregated

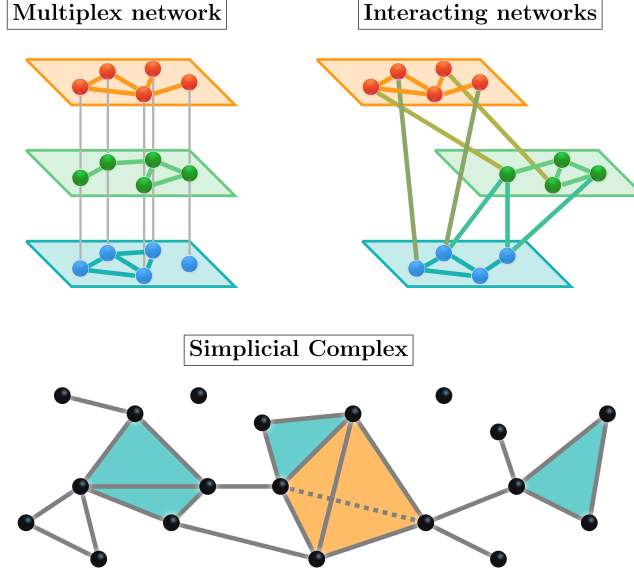


FIG. 4. Generalised network structures. A *multiplex* (or link-coloured) network consists of various layers with the same set of nodes and specific interaction patterns in each layer. An *interacting* (or multi-layer) network consists of various layers each with its own nodes, and of interactions existing within and between layers. A *simplicial complex* represents the different kind of interactions (simplices) between groups of d nodes: isolated nodes ($d = 0$), pairwise interactions ($d = 1$, in grey), triangles ($d = 2$, in cyan), tetrahedra ($d = 3$, in yellow) and so on.

WTW the situation is less clear: while commodities are in principle distinguishable, trade transactions are much less so, and in fact neither the WCM, WCM(M) and ME model can well reproduce it. A possible solution here is again to constrain both strengths and degrees simultaneously [152].

Simplicial complexes — Simplicial complexes are generalised network structures able to encode interactions occurring between more than two nodes, and allow to describe a large variety of complex interacting systems: collaboration networks where artefacts results from two or more actors working together, protein-interaction networks where complexes often consist of more than two proteins, economic systems of financial transactions often involving several parties, and social systems where groups of people are united by common motives or interest. Simplicial complexes can involve any number of nodes. For instance, simplices of dimension $d = 0, 1, 2, 3$ are, respectively, nodes, links, triangles, and tetrahedra, and in general a d -dimensional simplex is formed by a set of $d + 1$ interacting nodes and includes all the simplices formed by subsets of $\delta + 1$ nodes (with $\delta < d$), which are called the δ -dimensional faces of the simplex. A simplicial complex is then a collection of simplices of different dimensions properly “glued” together (Fig. 4).

Exponential random simplicial complexes (ERSC) have been recently introduced as a higher dimensional generalisations of ERG, and allow to generate random simplicial complexes where each simplex has its own independent probability of appearance, conditioned to the presence of simplex boundaries (which thus become additional constraints) [153]. In particular, the explicit calculation of the partition function is possible when considering random simplicial complexes formed exclusively by d -dimensional simplices [154]. Indeed, by constraining the generalised degree (the number of d -dimensional simplices incident to a given δ -dimensional face), the graph probability becomes a product of marginal probabilities for individual d -dimensional simplices. Alternatively, an appropriate Markov chain Monte Carlo sampling can be used to populate a microcanonical ensemble of simplicial complexes formed by simplices of any dimension [155].

Perspectives and Conclusion

Complex networks have a twofold key difference with respect to the systems traditionally studied in equilibrium statistical physics. Firstly, their microscopic degrees of freedom are the interactions between the nodes making up the system, and not the states of the nodes themselves. Secondly, nodes are so heterogeneous, both in terms of intrinsic characteristics and connectivity features, that networks can hardly be assigned a typical scale. Maximum entropy

models of networks based on local constraints are grounded on these two facts, by defining probability distributions on the network connections, and by not distinguishing nodes beyond their (heterogeneous) local features.

As we reviewed here, these models have found an extremely wide range of practical applications. This is due to their analytic characterisation, and their versatility to encompass also higher-order network characteristics (*e.g.*, assortativity, clustering, community structure) using stochastic sampling, as well as even more complex structures such as networks of networks and simplicial complexes. There are of course limitations of this approach, though. First of all, the constraints that can be imposed have to be static topological properties of the network. Only recently, cases of dynamical constraints have been considered using either the principle of Maximum Caliber—which is to dynamical pathways what the principle of maximum entropy is to equilibrium states [29, 156], or definitions of the entropy functional alternative to Shannon’s (see Box 3). Second, the possibility to consider in these models semi-local network properties heavily relies on numerical sampling, which becomes unfeasible or much biased for non-trivial patterns involving more than two or three nodes. These patterns can however be important in situations in which the network structure is determined by complex optimisation principles: sub-units of an electrical circuit, biochemical reactions in a cell, neuron firing patterns in the brain. Statistical physics of networks is bound to face the challenge of developing more complex network models for these kind of structures. Nevertheless, maximum entropy models based on local constraints do represent effective benchmarks to detect and validate them.

Box 3: Boltzmann, von Neumann, and Kolmogorov entropies

For a microcanonical ensemble, the Boltzmann entropy is the logarithm of the number of network configurations belonging to the ensemble: $\Sigma = \ln \sum_{\mathbf{G} \in \Omega} \delta[\mathbf{c}(\mathbf{G}), \mathbf{c}^*]$. Similarly to the Shannon entropy for a canonical ensemble, the Boltzmann entropy can be used to quantify the complexity of the ensemble, *i.e.*, to assess the role of different constraints in terms of information they carry on the network structure [31, 120]. Indeed, the more informative the constraints in shaping the network structure, the smaller the effective number of graphs with the imposed features, and so the lower the Boltzmann entropy of the corresponding ensemble. Using these arguments, it is possible to show that homogeneous degree distributions are the most likely outcome when the Boltzmann entropy of the microcanonical BCM ensemble is large, whereas, scale-free degree distributions naturally emerge for network ensembles with minimal Boltzmann entropy [121].

The von Neumann entropy provides the amount of information encrypted in a quantum system composed by a mixture of pure quantum states [33, 176–178]. It is given by $\Xi = -\text{Tr}(\boldsymbol{\rho} \ln \boldsymbol{\rho})$ where $\boldsymbol{\rho}$ is the density matrix of the system, and is thus equal to the Shannon entropy of the eigenvalue distribution of $\boldsymbol{\rho}$. For undirected binary networks, a formulation of this entropy which satisfies the sub-additivity property can be given in terms of the combinatorial graph Laplacian $\mathbf{L} = \text{diag}(\mathbf{k}) - \mathbf{G}$, by defining $\boldsymbol{\rho} = e^{-\tau \mathbf{L}} / \text{Tr}(e^{-\tau \mathbf{L}})$ [179]. The resulting von Neuman entropy thus depends on the spectrum of the Laplacian, and can be seen as the result of constraining the properties of diffusion dynamics on the network [180, 181].

Finally, the Kolmogorov entropy generalises the Shannon entropy by describing the rate at which a stochastic process generates information. Considering for simplicity a Markovian ergodic stochastic process, described by the matrix $\boldsymbol{\Phi} = \{\phi_{ij}\}$ of transition rates $\{i \rightarrow j\}$ and by the stationary asymptotic probability distribution $\{\pi_i\}$, the dynamical entropy is defined as $\kappa(\boldsymbol{\Phi}) = -\sum_{ij} \pi_i \phi_{ij} \log \phi_{ij}$, that is the average of the Shannon entropies of rows of $\boldsymbol{\Phi}$ —each weighted by the stationary distribution of the corresponding node [182]. The Kolmogorov entropy turns out to be related on one hand to the capacity of the network to withstand random structural changes [182], and on the other hand to the Ricci curvature of the network [183]. This connection is particularly intriguing, as the Ricci curvature has been used to differentiate stages of cancer from gene co-expression networks [184], as well as to give hallmarks of financial crashes from stock correlation networks [185].

We remark that, from a practical point of view, the possibility of quantifying the relevance of a set of observed features and extracting meaningful information from huge streams of continuously-produced, high-dimensional noisy data is particularly relevant in the present era of Big Data. On one hand, the details and facets of information we are able to extract nowadays has reached levels never seen before, which means that we need more and more complex data structures and models in order to represent and comprehend them. On the other hand, the amount of information available calls for effective and scalable ways to let the signal emerge from the noise originated by the large variety of sources. In this effort, the theoretical framework of statistical physics stands as an essential instrument to make consistent inference from data, whatever the level of complexity we have to face.

Acknowledgements — GCi, TS, FS and GCa acknowledge support from the EU projects CoeGSS (grant no. 676547), Openmaker (grant no. 687941), SoBigData (grant no. 654024) and DOLFINS (grant no. 640772). DG acknowledges support from the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands). AG acknowledges support from the CNR PNR Project CRISISLAB funded by Italian Government. GCa also acknowledges the Israeli-Italian project MAC2MIC financed by Italian MAECI.

-
- [1] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, “Critical phenomena in complex networks,” *Reviews of Modern Physics* **80**, 1275–1335 (2008).
 - [2] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science* **286**, 509–512 (1999).
 - [3] S. H. Yook, H. Jeong, A.-L. Barabási, and Y. Tu, “Weighted evolving networks,” *Physical Review Letters* **86**, 5835–5838 (2001).
 - [4] A. Barrat, M. Barthélemy, and A. Vespignani, “Weighted evolving networks: Coupling topology and weight dynamics,” *Physical Review Letters* **92**, 228701 (2004).
 - [5] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” *Physical Review E* **69**, 026113 (2004).
 - [6] S. Fortunato, “Community detection in graphs,” *Physics Reports* **486**, 75–174 (2010).
 - [7] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-world networks,” *Nature* **393**, 440–442 (1998).
 - [8] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley, “Classes of small-world networks,” *Proceedings of the National Academy of Sciences* **97**, 11149–11152 (2000).
 - [9] F. Chung and L. Lu, “The average distances in random graphs with given expected degrees,” *Proceedings of the National Academy of Sciences* **99**, 15879–15882 (2002).
 - [10] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of Modern Physics* **74**, 47–97 (2002).
 - [11] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Review* **45**, 167–256 (2003).
 - [12] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics Reports* **424**, 175–308 (2006).
 - [13] G. Bianconi and A.L. Barabási, “Bose-einstein condensation in complex network,” *Physical Review Letters* **86**, 5632–5635 (2001).
 - [14] G. Caldarelli, A. Capocci, P. De Los Rios, and M. a Muñoz, “Scale-free networks from varying vertex intrinsic fitness,” *Physical Review Letters* **89**, 258702 (2002).
 - [15] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, “Structure of growing networks with preferential linking,” *Physical Review Letters* **85**, 4633–4636 (2000).
 - [16] M. Medo, G. Cimini, and S. Gualdi, “Temporal effects in the growth of networks,” *Physical Review Letters* **107**, 238701 (2011).
 - [17] P. W. Holland and S. Leinhardt, “An exponential family of probability distributions for directed graphs,” *Journal of the American Statistical Association* **76**, 33–50 (1981).
 - [18] O. Frank and D. Strauss, “Markov graphs,” *Journal of the American Statistical Association* **81**, 832–842 (1986).
 - [19] D. Strauss, “On a general class of models for interaction,” *SIAM Review* **28**, 513–527 (1986).
 - [20] S. Wasserman and P. Pattison, “Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p ,” *Psychometrika* **61**, 401–425 (1996).
 - [21] C. J. Anderson, S. Wasserman, and B. Crouch, “A p^* primer: Logit models for social networks,” *Social Networks* **21**, 37–66 (1999).
 - [22] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, “New specifications for exponential random graph models,” *Sociological Methodology* **36**, 99–153 (2006).
 - [23] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, “An introduction to exponential random graph (p^*) models for social networks,” *Social Networks* **29**, 173–191 (2007).
 - [24] S. J. Cranmer and B. A. Desmarais, “Inferential network analysis with exponential random graph models,” *Political Analysis* **19**, 6686 (2011).
 - [25] T. A. B. Snijders, “Statistical models for social networks,” *Annual Review of Sociology* **37**, 131–153 (2011).
 - [26] J. Park and M. E. J. Newman, “Statistical mechanics of networks,” *Physical Review E* **70**, 066117 (2004).
 - [27] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical Review* **106**, 620–630 (1957).
 - [28] J. Shore and R. Johnson, “Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy,” *IEEE Transactions on Information Theory* **26**, 26–37 (1980).
 - [29] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, “Principles of maximum entropy and maximum caliber in statistical physics,” *Reviews of Modern Physics* **85**, 1115–1141 (2013).
 - [30] E. T. Jaynes, “On the rationale of maximum-entropy methods,” *Proceedings of the IEEE* **70**, 939–952 (1982).
 - [31] G. Bianconi, “The entropy of randomized network ensembles,” *Europhysics Letters* **81**, 28005 (2008).
 - [32] T. Squartini, R. Mastrandrea, and D. Garlaschelli, “Unbiased sampling of network ensembles,” *New Journal of Physics* **17**, 023052 (2015).
 - [33] K. Anand and G. Bianconi, “Entropy measures for networks: Toward an information theory of complex topologies,” *Physical Review E* **80**, 045102 (2009).

- [34] T. Squartini, J. de Mol, F. den Hollander, and D. Garlaschelli, “Breaking of ensemble equivalence in networks,” *Physical Review Letters* **115**, 268701 (2015).
- [35] T. Squartini and D. Garlaschelli, “Reconnecting statistical physics and combinatorics beyond ensemble equivalence,” <https://arxiv.org/abs/1710.11422> (2018).
- [36] D. Garlaschelli and M. I. Loffredo, “Generalized bose-fermi statistics and structural correlations in weighted networks,” *Physical Review Letters* **102**, 038701 (2009).
- [37] D. Garlaschelli and M. I. Loffredo, “Maximum likelihood: Extracting unbiased information from complex networks,” *Physical Review E* **78**, 015101(R) (2008).
- [38] T. Squartini and D. Garlaschelli, “Analytical maximum-likelihood method to detect patterns in real networks,” *New Journal of Physics* **13**, 083001 (2011).
- [39] P. Erdős and A. Rényi, “On random graphs,” *Publicationes Mathematicae Debrecen* **6**, 290–297 (1959).
- [40] M. Á. Serrano and M. Boguñá, “Weighted configuration model,” *AIP Conference Proceedings* **776**, 101–107 (2005).
- [41] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, “Enhanced reconstruction of weighted networks from strengths and degrees,” *New Journal of Physics* **16**, 043022 (2014).
- [42] S. Maslov and K. Sneppen, “Specificity and stability in topology of protein networks,” *Science* **296**, 910–913 (2002).
- [43] J. Park and M. E. J. Newman, “Origin of degree correlations in the internet and other networks,” *Physical Review E* **68**, 026112 (2003).
- [44] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks,” *Proceedings of the National Academy of Sciences* **101**, 3747–3752 (2004).
- [45] S. Maslov, K. Sneppen, and A. Zaliznyak, “Detection of topological patterns in complex networks: Correlation profile of the internet,” *Physica A: Statistical Mechanics and its Applications* **333**, 529–540 (2004).
- [46] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, “Detecting rich-club ordering in complex networks,” *Nature Physics* **2**, 110 (2006).
- [47] M. Á. Serrano, M. Boguñá, and R. Pastor-Satorras, “Correlations in weighted networks,” *Physical Review E* **74**, 055101 (2006).
- [48] R. Guimerá, M. Sales-Pardo, and L. A. N. Amaral, “Classes of complex networks defined by role-to-role connectivity profiles,” *Nature Physics* **3**, 63 (2006).
- [49] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna, “The international trade network: weighted network analysis and modelling,” *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P02002 (2008).
- [50] T. Opsahl, V. Colizza, P. Panzarasa, and J. J. Ramasco, “Prominence and control: The weighted rich-club effect,” *Physical Review Letters* **101**, 168702 (2008).
- [51] M. Á. Serrano and M. Boguñá, “Topology of the world trade web,” *Physical Review E* **68**, 015101 (2003).
- [52] D. Garlaschelli and M. I. Loffredo, “Fitness-dependent topological properties of the world trade web,” *Physical Review Letters* **93**, 188701 (2004).
- [53] D. Garlaschelli and M. I. Loffredo, “Structure and evolution of the world trade network,” *Physica A: Statistical Mechanics and its Applications* **355**, 138–144 (2005).
- [54] G. Fagiolo, J. Reyes, and S. Schiavo, “World trade web: Topological properties, dynamics, and evolution,” *Physical Review E* **79**, 036115 (2009).
- [55] M. E. J. Newman, “Analysis of weighted networks,” *Physical Review E* **70**, 056131 (2004).
- [56] S. E. Ahnert, D. Garlaschelli, T. M. A. Fink, and G. Caldarelli, “Ensemble approach to the analysis of weighted networks,” *Physical Review E* **76**, 016101 (2007).
- [57] J. Saramäki, M. Kivelä, J.-P. Onnela, K. Kaski, and J. Kertész, “Generalizations of the clustering coefficient to weighted complex networks,” *Physical Review E* **75**, 027105 (2007).
- [58] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, “Network motifs: Simple building blocks of complex networks,” *Science* **298**, 824–827 (2002).
- [59] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, “Network motifs in the transcriptional regulation network of *escherichia coli*,” *Nature Genetics* **31**, 64 (2002).
- [60] D. Garlaschelli and M. I. Loffredo, “Patterns of link reciprocity in directed networks,” *Physical Review Letters* **93**, 268701 (2004).
- [61] D. Garlaschelli and M. I. Loffredo, “Multispecies grand-canonical models for networks with reciprocity,” *Physical Review E* **73**, 015101(R) (2006).
- [62] T. Squartini and D. Garlaschelli, “Triadic motifs and dyadic self-organization in the world trade network,” in *Self-Organizing Systems*, edited by F. A. Kuipers and P. E. Heegaard (Springer Berlin Heidelberg, 2012) pp. 24–35.
- [63] D. B. Stouffer, J. Camacho, W. Jiang, and L. A. N. Amaral, “Evidence for the existence of a robust pattern of prey selection in food webs,” *Proceedings of the Royal Society of London B: Biological Sciences* **274**, 1931–1940 (2007).
- [64] T. Squartini, I. van Lelyveld, and D. Garlaschelli, “Early-warning signals of topological collapse in interbank networks,” *Scientific Reports* **3**, 3357 (2013).
- [65] R. Guimerà, M. Sales-Pardo, and L. A. N. Amaral, “Modularity from fluctuations in random graphs and complex networks,” *Physical Review E* **70**, 025101 (2004).
- [66] J. Reichardt and S. Bornholdt, “Partitioning and modularity of graphs with arbitrary degree distribution,” *Physical Review E* **76**, 015102 (2007).
- [67] F. Chung and L. Lu, “Connected components in random graphs with given expected degree sequences,” *Annals of Combinatorics* **6**, 125–145 (2002).

- [68] L. Bargigli and M. Gallegati, “Random digraphs with given expected degree sequences: A model for economic networks,” *Journal of Economic Behavior & Organization* **78**, 396–411 (2011).
- [69] P. Fronczak, A. Fronczak, and M. Bujok, “Exponential random graph models for networks with community structure,” *Physical Review E* **88**, 32810 (2013).
- [70] A. Lancichinetti, S. Fortunato, and F. Radicchi, “Benchmark graphs for testing community detection algorithms,” *Physical Review E* **78**, 046110 (2008).
- [71] B. Karrer and M. E. J. Newman, “Stochastic blockmodels and community structure in networks,” *Physical Review E* **83**, 016107 (2011).
- [72] T. P. Peixoto, “Entropy of stochastic blockmodel ensembles,” *Physical Review E* **85**, 056122 (2012).
- [73] P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim, “Network bipartivity,” *Physical Review E* **68**, 056107 (2003).
- [74] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini, “Randomizing bipartite networks: The case of the world trade web,” *Scientific Reports* **5**, 10595 (2015).
- [75] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, and L. Pietronero, “A new metrics for countries’ fitness and products’ complexity,” *Scientific Reports* **2**, 723 (2012).
- [76] G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, and A. Tacchella, “A network analysis of countries’ export flows: Firm grounds for the building blocks of the economy,” *PLoS ONE* **7**, e47278 (2012).
- [77] F. Saracco, R. Di Clemente, A. Gabrielli, and T. Squartini, “Detecting early signs of the 2007-2008 crisis in the world trade,” *Scientific Reports* **6**, 30286 (2016).
- [78] C. Payrató Borrás, L. Hernández, and Y. Moreno, “Breaking the spell of nestedness,” <https://arxiv.org/abs/1711.03134> (2017).
- [79] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang, “Bipartite network projection and personal recommendation,” *Physical Review E* **76**, 046115 (2007).
- [80] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, “A tool for filtering information in complex systems,” *Proceedings of the National Academy of Sciences* **102**, 10421–10426 (2005).
- [81] M. Á. Serrano, M. Boguñá, and A. Vespignani, “Extracting the multiscale backbone of complex weighted networks,” *PNAS* **106**, 6483–6488 (2009).
- [82] P. B. Slater, “A two-stage algorithm for extracting the multiscale backbone of complex weighted networks,” *Proceedings of the National Academy of Sciences* **106**, E66–E66 (2009).
- [83] F. Radicchi, J. J. Ramasco, and S. Fortunato, “Information filtering in complex weighted networks,” *Physical Review E* **83**, 046101 (2011).
- [84] D. S. Goldberg and F. P. Roth, “Assessing experimentally derived interactions in a small world,” *Proceedings of the National Academy of Sciences* **100**, 4372–4376 (2003).
- [85] M. Latapy, C. Magnien, and N. D. Vecchio, “Basic notions for the analysis of large two-mode networks,” *Social Networks* **30**, 31–48 (2008).
- [86] M. Tumminello, S. Miccichè, F. Lillo, J. Piilo, and R. N. Mantegna, “Statistically validated networks in bipartite complex systems,” *PLoS ONE* **6**, e17994 (2011).
- [87] M. Tumminello, F. Lillo, J. Piilo, and R. N. Mantegna, “Identification of clusters of investors from their real trading activity in a financial market,” *New Journal of Physics* **14**, 013041 (2012).
- [88] Z. Neal, “Identifying statistically significant edges in one-mode projections,” *Social Network Analysis and Mining* **3**, 915–924 (2013).
- [89] K. A. Zweig and M. Kaufmann, “A systematic approach to the one-mode projection of bipartite graphs,” *Social Network Analysis and Mining* **1**, 187–218 (2011).
- [90] E.-Á. Horvát and K. A. Zweig, “A fixed degree sequence model for the one-mode projection of multiplex bipartite graphs,” *Social Network Analysis and Mining* **3**, 1209–1224 (2013).
- [91] A. Gionis, H. Mannila, T. Mielikäinen, and P. Tsaparas, “Assessing data mining results via swap randomization,” *ACM Trans. Knowl. Discov. Data* **1** (2007), 10.1145/1297332.1297338.
- [92] Z. Neal, “The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors,” *Social Networks* **39**, 84–97 (2014).
- [93] S. Gualdi, G. Cimini, K. Primicerio, R. Di Clemente, and D. Challet, “Statistically validated network of portfolio overlaps and systemic risk,” *Scientific Reports* **6**, 39467 (2016).
- [94] F. Saracco, M. J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini, “Inferring monopartite projections of bipartite networks: An entropy-based approach,” *New Journal of Physics* **19**, 053022 (2017).
- [95] M. J. Straka, G. Caldarelli, and F. Saracco, “Grand canonical validation of the bipartite international trade network,” *Physical Review E* **96**, 022306 (2017).
- [96] E. Pugliese, G. Cimini, A. Patelli, A. Zaccaria, L. Pietronero, and A. Gabrielli, “Unfolding the innovation system for the development of countries: co-evolution of science, technology and production,” <https://arxiv.org/abs/1707.05146> (2017).
- [97] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, “Epidemic processes in complex networks,” *Reviews of Modern Physics* **87**, 925–979 (2015).
- [98] S. J. Wells, *Financial interlinkages in the United Kingdom’s interbank market and the risk of contagion*, Working Paper 230 (Bank of England, 2004).
- [99] C. Upper, “Simulation methods to assess the danger of contagion in interbank markets,” *Journal of Financial Stability* **7**, 111–125 (2011).

- [100] K. Anand, I. van Lelyveld, Á. Banai, S. Friedrich, R. Garratt, G. Halaj, J. Figue, I. Hansen, S. Martínez-Jaramillo, H. Lee, J. L. Molina-Borboa, S. Nobili, S. Rajan, D. Salakhova, T. C. Silva, L. Silvestri, and S. R. Stancato de Souza, “The missing links: A global study on uncovering financial network structures from partial data,” *Journal of Financial Stability*, (in press) (2017).
- [101] G. Kossinets, “Effects of missing data in social networks,” *Social Networks* **28**, 247–268 (2006).
- [102] C. Lynch, “How do your data grow?” *Nature* **455**, 28 (2008).
- [103] L. A. N. Amaral, “A truer measure of our ignorance,” *Proceedings of the National Academy of Sciences* **105**, 6795–6796 (2008).
- [104] R. Guimerá and M. Sales-Pardo, “Missing and spurious interactions and the reconstruction of complex networks,” *Proceedings of the National Academy of Sciences* **106**, 22073–22078 (2009).
- [105] L. Lu and T. Zhou, “Link prediction in complex networks: A survey,” *Physica A: Statistical Mechanics and its Applications* **390**, 1150–1170 (2011).
- [106] T. Squartini, G. Caldarelli, G. Cimini, A. Gabrielli, and D. Garlaschelli, “Reconstruction methods for networks: The case of economic and financial systems,” *Physics Reports* **757**, 1–47 (2018).
- [107] M. Boguñá and R. Pastor-Satorras, “Class of correlated random networks with hidden variables,” *Physical Review E* **68**, 036112 (2003).
- [108] D. Garlaschelli, S. Battiston, M. Castri, V. D.P. Servedio, and G. Caldarelli, “The scale-free topology of market investments,” *Physica A: Statistical Mechanics and its Applications* **350**, 491–499 (2005).
- [109] G. De Masi, G. Iori, and G. Caldarelli, “Fitness model for the italian interbank money market,” *Physical Review E* **74**, 066112 (2006).
- [110] N. Musmeci, S. Battiston, G. Caldarelli, M. Puliga, and A. Gabrielli, “Bootstrapping topological properties and systemic risk of complex networks using the fitness model,” *Journal of Statistical Physics* **151**, 1–15 (2013).
- [111] G. Cimini, T. Squartini, A. Gabrielli, and D. Garlaschelli, “Estimating topological properties of weighted networks from limited information,” *Physical Review E* **92**, 040802 (2015).
- [112] G. Cimini, T. Squartini, D. Garlaschelli, and A. Gabrielli, “Systemic risk analysis on reconstructed economic and financial networks,” *Scientific Reports* **5**, 15758 (2015).
- [113] T. Squartini, G. Cimini, A. Gabrielli, and D. Garlaschelli, “Network reconstruction via density sampling,” *Applied Network Science* **2**, 3 (2017).
- [114] T. Squartini, A. Almog, G. Caldarelli, I. van Lelyveld, D. Garlaschelli, and G. Cimini, “Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks,” *Physical Review E* **96**, 032315 (2017).
- [115] J. Berg and M. Lässig, “Correlated random networks,” *Physical Review Letters* **89**, 228701 (2002).
- [116] M. E. J. Park, J. and Newman, “Solution of the two-star model of a network,” *Physical Review E* **70**, 066146 (2004).
- [117] Mei Yin and Lingjiong Zhu, “Reciprocity in directed networks,” *Physica A: Statistical Mechanics and its Applications* **447**, 71–84 (2016).
- [118] J. Park and M. E. J. Newman, “Solution for the properties of a clustered network,” *Physical Review E* **72**, 026136 (2005).
- [119] P. Fronczak, A. Fronczak, and J. A. Holyst, “Phase transitions in social networks,” *The European Physical Journal B* **59**, 133–139 (2007).
- [120] G. Bianconi, A. C. C. Coolen, and C. J. Perez Vicente, “Entropies of complex networks with hierarchically constrained topologies,” *Physical Review E* **78**, 016114 (2008).
- [121] G. Bianconi, “Entropy of network ensembles,” *Physical Review E* **79**, 036114 (2009).
- [122] R. J. Mondragón, “Network null-model based on maximal entropy and the rich-club,” *Journal of Complex Networks* **2**, 288–298 (2014).
- [123] A. Annibale, A. C. C. Coolen, L. P. Fernandes, F. Fraternali, and J. Kleinjung, “Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure,” *Journal of Physics A: Mathematical and Theoretical* **42**, 485001 (2009).
- [124] E. S. Roberts, T. Schlitt, and A. C. C. Coolen, “Tailored graph ensembles as proxies or null models for real networks II: results on directed graphs,” *Journal of Physics A: Mathematical and Theoretical* **44**, 275002 (2011).
- [125] E. S. Roberts and A. C. C. Coolen, “Entropies of tailored random graph ensembles: bipartite graphs, generalized degrees, and node neighbourhoods,” *Journal of Physics A: Mathematical and Theoretical* **47**, 435101 (2014).
- [126] Y. Artzy-Randrup and L. Stone, “Generating uniformly distributed random networks,” *Physical Review E* **72**, 056708 (2005).
- [127] A. C. C. Coolen, A. De Martino, and A. Annibale, “Constrained markovian dynamics of random graphs,” *Journal of Statistical Physics* **136**, 1035–1067 (2009).
- [128] E. S. Roberts and A. C. C. Coolen, “Unbiased degree-preserving randomization of directed binary networks,” *Physical Review E* **85**, 046103 (2012).
- [129] D. Strauss and M. Ikeda, “Pseudolikelihood estimation for social networks,” *Journal of the American Statistical Association* **85**, 204–212 (1990).
- [130] M. A. J. van Duijn, K. J. Gile, and M. S. Handcock, “A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models,” *Social Networks* **31**, 52–62 (2009).
- [131] T. A. B. Snijders, J. Koskinen, and M. Schweinberger, “Maximum likelihood estimation for social network dynamics,” *The Annals of Applied Statistics* **4**, 567–588 (2010).
- [132] M. Schweinberger, “Instability, sensitivity, and degeneracy of discrete exponential families,” *Journal of the American Statistical Association* **106**, 1361–1370 (2011).

- [133] B. A. Desmarais and S. J. Cranmer, “Statistical mechanics of networks: Estimation and uncertainty,” *Physica A: Statistical Mechanics and its Applications* **391**, 1865–1876 (2012).
- [134] S. Chatterjee and P. Diaconis, “Estimating and understanding exponential random graph models,” **41**, 2428–2461 (2013).
- [135] S. Horvát, É. Czabarka, and Z. Toroczkai, “Reducing degeneracy in maximum entropy models of networks,” *Physical Review Letters* **114**, 158701 (2015).
- [136] W. K. Hastings, “Monte carlo sampling methods using markov chains and their applications,” *Biometrika* **57**, 97–109 (1970).
- [137] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, “Systematic topology analysis and generation using degree correlations,” *SIGCOMM Comput. Commun. Rev.* **36**, 135–146 (2006).
- [138] C. Orsini, M. M. Dankulov, P. Colomer-de Simón, A. Jamakovic, P. Mahadevan, A. Vahdat, K. E. Bassler, Z. Toroczkai, M. Boguñá, G. Caldarelli, S. Fortunato, and D. Krioukov, “Quantifying randomness in real networks,” *Nature Communications* **6**, 8627 (2015).
- [139] D. Foster, J. Foster, M. Paczuski, and P. Grassberger, “Communities, clustering phase transitions, and hysteresis: Pitfalls in constructing network ensembles,” *Physical Review E* **81**, 046115 (2010).
- [140] R. Fischer, J. C. Leitão, T. P. Peixoto, and E. G. Altmann, “Sampling motif-constrained ensembles of networks,” *Physical Review Letters* **115**, 188701 (2015).
- [141] Fugao Wang and D. P. Landau, “Efficient, multiple-range random walk algorithm to calculate the density of states,” *Physical Review Letters* **86**, 2050–2053 (2001).
- [142] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, “Multilayer networks,” *Journal of Complex Networks* **2**, 203–271 (2014).
- [143] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin, “The structure and dynamics of multilayer networks,” *Physics Reports* **544**, 1–122 (2014).
- [144] M. De Domenico, C. Granell, M. A. Porter, and A. Arenas, “The physics of spreading processes in multilayer networks,” *Nature Physics* **12**, 901–906 (2016).
- [145] G. Bianconi, “Statistical mechanics of multiplex networks: Entropy and overlap,” *Physical Review E* **87**, 062806 (2013).
- [146] V. Gemmetto and D. Garlaschelli, “Multiplexity versus correlation: the role of local constraints in real multiplexes,” *Scientific Reports* **5**, 9120 (2015).
- [147] G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondragón, and G. Bianconi, “Weighted multiplex networks,” *PLoS ONE* **9**, e97857 (2014).
- [148] G. Menichetti, D. Remondini, and G. Bianconi, “Correlations between weights and overlap in ensembles of weighted multiplex networks,” *Physical Review E* **90**, 062817 (2014).
- [149] O. Sagarra, C. J. Pérez Vicente, and A. Díaz-Guilera, “Statistical mechanics of multiedge networks,” *Physical Review E* **88**, 062806 (2013).
- [150] O. Sagarra, F. Font-Clos, C. J. Pérez-Vicente, and A. Díaz-Guilera, “The configuration multi-edge model: Assessing the effect of fixing node strengths on weighted network magnitudes,” *Europhysics Letters* **107**, 38002 (2014).
- [151] O. Sagarra, C. J. Pérez Vicente, and A. Díaz-Guilera, “Role of adjacency-matrix degeneracy in maximum-entropy-weighted network models,” *Physical Review E* **92**, 052816 (2015).
- [152] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, “Reconstructing the world trade multiplex: The role of intensive and extensive biases,” *Physical Review E* **90**, 062804 (2014).
- [153] K. Zuev, O. Eisenberg, and D. Krioukov, “Exponential random simplicial complexes,” *Journal of Physics A: Mathematical and Theoretical* **48**, 465002 (2015).
- [154] O. T. Courtney and G. Bianconi, “Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes,” *Physical Review E* **93**, 062311 (2016).
- [155] J.-G. Young, G. Petri, F. Vaccarino, and A. Patania, “Construction of and efficient sampling from the simplicial configuration model,” *Physical Review E* **96**, 032312 (2017).
- [156] P. D. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, “Perspective: Maximum caliber is a general variational principle for dynamical systems,” *The Journal of Chemical Physics* **148**, 010901 (2018).
- [157] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, “Random graphs with arbitrary degree distributions and their applications,” *Physical Review E* **64**, 026118 (2001).
- [158] S. Itzkovitz, R. Milo, N. Kashtan, M. E. J. Newman, and U. Alon, “Reply to ‘comment on ‘subgraphs in random networks’ ’,” *Physical Review E* **70**, 058102 (2004).
- [159] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, “Generation of uncorrelated random scale-free networks,” *Physical Review E* **71**, 027103 (2005).
- [160] G. Zamora-López, V. Zlatić, C. Zhou, H. Štefančić, and J. Kurths, “Reciprocity of networks with degree correlations and arbitrary degree sequences,” *Physical Review E* **77**, 016106 (2008).
- [161] V. Zlatić, G. Bianconi, A. Díaz-Guilera, D. Garlaschelli, F. Rao, and G. Caldarelli, “On the rich-club effect in dense and weighted networks,” *The European Physical Journal B* **67**, 271–275 (2009).
- [162] L. Tabourier, C. Roth, and J.-P. Cointet, “Generating constrained random graphs using multiple edge switches,” *Journal of Experimental Algorithmics* **16**, 1.1–1.15 (2011).
- [163] C. J. Carstens and K. J. Horadam, “Switching edges to randomize networks: what goes wrong and how to fix it,” *Journal of Complex Networks* **5**, 337–351 (2017).
- [164] C. I. Del Genio, H. Kim, Z. Toroczkai, and K. E. Bassler, “Efficient and exact sampling of simple graphs with given arbitrary degree sequence,” *PLoS ONE* **5**, e10012 (2010).

- [165] J. Blitzstein and P. Diaconis, “A sequential importance sampling algorithm for generating random graphs with prescribed degrees,” *Internet Mathematics* **6**, 489–522 (2011).
- [166] H. Kim, C. I. Del Genio, K. E. Bassler, and Z. Toroczkai, “Constructing and sampling directed graphs with given degree sequences,” *New Journal of Physics* **14**, 023012 (2012).
- [167] M. E. J. Newman, “Random graphs with clustering,” *Physical Review Letters* **103**, 058701 (2009).
- [168] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, “The unreasonable effectiveness of tree-based theory for networks with clustering,” *Physical Review E* **83**, 036112 (2011).
- [169] Z. Burda and A. Krzywicki, “Uncorrelated random networks,” *Physical Review E* **67**, 046118 (2003).
- [170] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, “Cut-offs and finite size effects in scale-free networks,” *The European Physical Journal B* **38**, 205–209 (2004).
- [171] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests of statistical hypotheses,” *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **231**, 289–337 (1933).
- [172] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag New York, 2002).
- [173] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control* **19**, 716–723 (1974).
- [174] E.-J. Wagenmakers and S. Farrell, “Aic model selection using akaike weights,” *Psychonomic Bulletin & Review* **11**, 192–196 (2004).
- [175] K. P. Burnham and D. R. Anderson, “Multimodel inference: Understanding aic and bic in model selection,” *Sociological Methods & Research* **33**, 261–304 (2004).
- [176] S. L. Braunstein, S. Ghosh, and S. Severini, “The laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states,” *Annals of Combinatorics* **10**, 291–317 (2006).
- [177] K. Anand, G. Bianconi, and S. Severini, “Shannon and von neumann entropy of random networks with heterogeneous expected degree,” *Physical Review E* **83**, 036109 (2011).
- [178] K. Anand, D. Krioukov, and G. Bianconi, “Entropy distribution and condensation in random networks with a given degree distribution,” *Physical Review E* **89**, 062807 (2014).
- [179] M. De Domenico and J. Biamonte, “Spectral entropies as information-theoretic tools for complex network comparison,” *Physical Review X* **6**, 041062 (2016).
- [180] J.-C. Delvenne, R. Lambiotte, and L. E. C. Rocha, “Diffusion on networked systems is a question of time or structure,” *Nature Communications* **6**, 7366 (2015).
- [181] N. Masuda, M. A. Porter, and R. Lambiotte, “Random walks and diffusion on networks,” *Physics Reports* **716-717**, 1–58 (2017).
- [182] L. Demetrius and T. Manke, “Robustness and network evolution—an entropic principle,” *Physica A: Statistical Mechanics and its Applications* **346**, 682–696 (2005).
- [183] J. Lott and C. Villani, “Ricci curvature for metric-measure spaces via optimal transport,” *Annals of Mathematics* **169**, 903–991 (2009).
- [184] R. Sandhu, T. Georgiou, E. Reznik, L. Zhu, I. Kolesov, Y. Senbabaoglu, and A. Tannenbaum, “Graph curvature for differentiating cancer networks,” *Scientific Reports* **5**, 12323 (2015).
- [185] R. Sandhu, T. Georgiou, and A. Tannenbaum, “Ricci curvature: An economic indicator for market fragility and systemic risk,” *Science Advances* **2** (2016), 10.1126/sciadv.1501495.