The statistical physics of real-world networks

Giulio Cimini [ORCID] [1,2], Tiziano Squartini [1], Fabio Saracco [1], Diego Garlaschelli [1,3], Andrea Gabrielli [1,2] and Guido Caldarelli [1,2,4,5] *

Abstract | In the past 15 years, statistical physics has been successful as a framework for modelling complex networks. On the theoretical side, this approach has unveiled a variety of physical phenomena, such as the emergence of mixed distributions and ensemble non-equivalence, that are observed in heterogeneous networks but not in homogeneous systems. At the same time, thanks to the deep connection between the principle of maximum entropy and information theory, statistical physics has led to the definition of null models for networks that reproduce features of real-world systems but that are otherwise as random as possible. We review here the statistical physics approach and the null models for complex networks, focusing in particular on analytical frameworks that reproduce local network features. We show how these models have been used to detect statistically significant structural patterns in real-world networks and to reconstruct the network structure in cases of incomplete information. We further survey the statistical physics models that reproduce more complex, semilocal network features using Markov chain Monte Carlo sampling, as well as models of generalized network structures, such as multiplex networks, interacting networks and simplicial complexes.

Nodes
Also known as vertices. Basic elements in the network or graph under consideration.

[1]IMT School for Advanced Studies, Lucca, Italy.

[2]Istituto dei Sistemi Complessi (CNR) UoS Sapienza, Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy.

[3]Lorentz Institute for Theoretical Physics, Leiden University, Leiden, Netherlands.

[4]European Centre for Living Technology, Università di Venezia Ca' Foscari, Venice, Italy.

[5]London Institute for Mathematical Sciences (LIMS), London, UK.

*e-mail: Guido.Caldarelli@imtlucca.it

The science of networks has exploded in the information age, thanks to the unprecedented production and storage of data on almost all human activities. This is because networks are a simple yet effective way to model a large class of technological, social, economic and biological systems that can be described as a set of entities (nodes) with interactions between them (links). These interactions 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, because 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 (the node's degree) follows a power-law distribution, so that most nodes have few links, but a few of them (the hubs) are highly connected. The distribution of the total weight of connections incident to a node (the node's strength) likewise follows a power law in many cases[3,4]. In addition, most real-world networks are organized into modules or display a community structure[5,6], and they possess a high level of clustering, because nodes tend to create tightly linked groups. However, real-world networks are also small worlds[7–9], that is, the mean distance (in terms of the number of connections) between two nodes 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 of identifying one or more microscopic mechanisms driving the formation of the network and using them to define a dynamic model that 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, although they provide good physical insights, they need refinement to give quantitatively accurate predictions.

The other approach consists of identifying a set of characteristic static properties of real systems and then building networks that have the same properties but are otherwise maximally random. This approach is akin to

Key points

- Statistical physics is a powerful framework to explain properties of complex networks, modelled as systems of heterogeneous entities whose degrees of freedom are their interactions rather than their states.

- The statistical physics of complex networks has brought theoretical insights into physical phenomena that are different in heterogeneous networks than in homogeneous systems.

- From an applied perspective, statistical physics defines null models for real-world networks that reproduce local features but are otherwise as random as possible.

- These models have been used, on the one hand, to detect statistically significant patterns in real-world networks and, on the other, to infer the network structure when information is incomplete.

- These applications are particularly useful in the current information age to make consistent inference from huge streams of continuously produced, high-dimensional, noisy data.

- The statistical mechanics approach has also been extended using numerical techniques to reproduce semilocal network features and, more recently, to encompass structures such as multilayer networks and simplicial complexes.

Links
Also known as edges. Connections or interactions between two nodes or vertices of a network or graph, representing the fundamental degrees of freedom of the system.

Undirected
A type of network for which every link is bidirectional, such as a network of colleagues (Alice works with Bob implies that Bob works with Alice).

Directed
A type of network for which links have a direction, such as an ecological network in which links represent predation (lions eat antelopes, but antelopes do not eat lions).

Binary
A type of network for which links are unweighted, that is, they can be described by either a 1 (the link exists) or a 0 (it does not).

Weighted
A type of network for which links have weights, which represent, for example, carrying capacities or interaction strengths.

Clustering
The tendency of node triples to be connected together, that is, to form triangles.

Graphs
The mathematical abstraction of a network comprising a set of N vertices and a set of E edges, each associated with two nodes.

statistical mechanics and therefore is based on rigorous probabilistic arguments that can lead to accurate and reliable predictions. The mathematical framework is that of the exponential random graph (ERG) model, which was first introduced in the social sciences and statistics[17–25] as a convenient formulation that relies on numerical techniques, such as Markov chain Monte Carlo (MCMC) algorithms. The interpretation of the ERG model in physical terms is due to Park and Newman[26], who showed how to derive the ERG model 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 that best represents the current state of knowledge of a system is the one that maximizes the Shannon entropy, subject, in principle, to any prior information on the system itself. This means making self-consistent inference while assuming maximal ignorance about the unknown degrees of freedom of the system[28]. The maximum entropy principle is conceptually powerful and finds numerous applications in physics and in science in general[29]. In the context of network theory, the maximum entropy approach is used to obtain ensembles of random graphs with given aggregated macroscopic or mesoscopic properties. These ensembles have two related applications. On the one hand, when the microscopic configuration of a real network is not accessible, the random graph ensemble describes the most probable network configuration: as is the case in traditional statistical mechanics, the maximum entropy principle makes it possible to gain maximally unbiased information in the absence of complete knowledge. On the other hand, when the microscopic configuration of the network is known, the ensemble defines a null model that enables assessment of the significance of empirical patterns found in the network against the null hypothesis that the network structure is determined solely by its aggregated structural properties.

This Review presents 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, that is, local, constraints, highlighting the novel physical concepts that can be learned from such models. After that, we present the two main fields of application for these models: detection of statistically significant patterns in empirical networks and reconstruction of network structures from partial information. We then discuss the models obtained by imposing semilocal network features and, finally, the most recent developments on generalized network structures and simplices.

Statistical mechanics of networks

Exponential random graphs. The statistical physics approach defined by the ERG consists of modelling a network system $G^\star$ using an ensemble Ω of graphs with the same number N of nodes and type of links as $G^\star$ (FIG. 1). The model is specified by $P(G)$, the occurrence probability of a graph $G \in \Omega$. According to statistical mechanics and information theory[27,30], the probability distribution that gives the most unbiased expectation of the microscopic configuration of the system under study is the one maximizing the Shannon entropy.

$$S = -\sum_{G \in \Omega} P(G) \ln P(G) \qquad (1)$$

This maximization is subject to the normalization condition $\sum_{G \in \Omega} P(G) = 1$ and a collection of constraints $c^\star$ that represent the macroscopic properties enforced on the system. These constraints define the sufficient statistics of the problem, that is, the model parameters depend only on the values of the constraints.

Imposing hard constraints, that is, assigning a uniform $P(G)$ to each of the graphs G that satisfy $c(G) = c^\star$ and zero probability to graphs that do not, leads to the microcanonical ensemble. Typically, this ensemble is not amenable to analytical treatment beyond steepest descent approximations[31] and 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, $\sum_{G \in \Omega} c(G) P(G) = c^\star$. Introducing the set of Lagrange multipliers θ, the constrained entropy maximization returns

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

where $H(G, \theta) = \theta \cdot c(G)$ is the Hamiltonian, $Z(\theta) = \sum_{G \in \Omega} e^{-H(G,\theta)}$ is the partition function and $\cdot$ represents the scalar product. Thus, the canonical $P(G|\theta)$ depends on G only through $c(G)$, which implies that graphs with constraints of the same value 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, for models of networks with an extensive number of constraints, the microcanonical and canonical ensembles turn out to be inequivalent in the thermodynamic limit $N \to \infty$[33–35]. This is in contrast to the case of traditional statistical physics (except possibly at phase
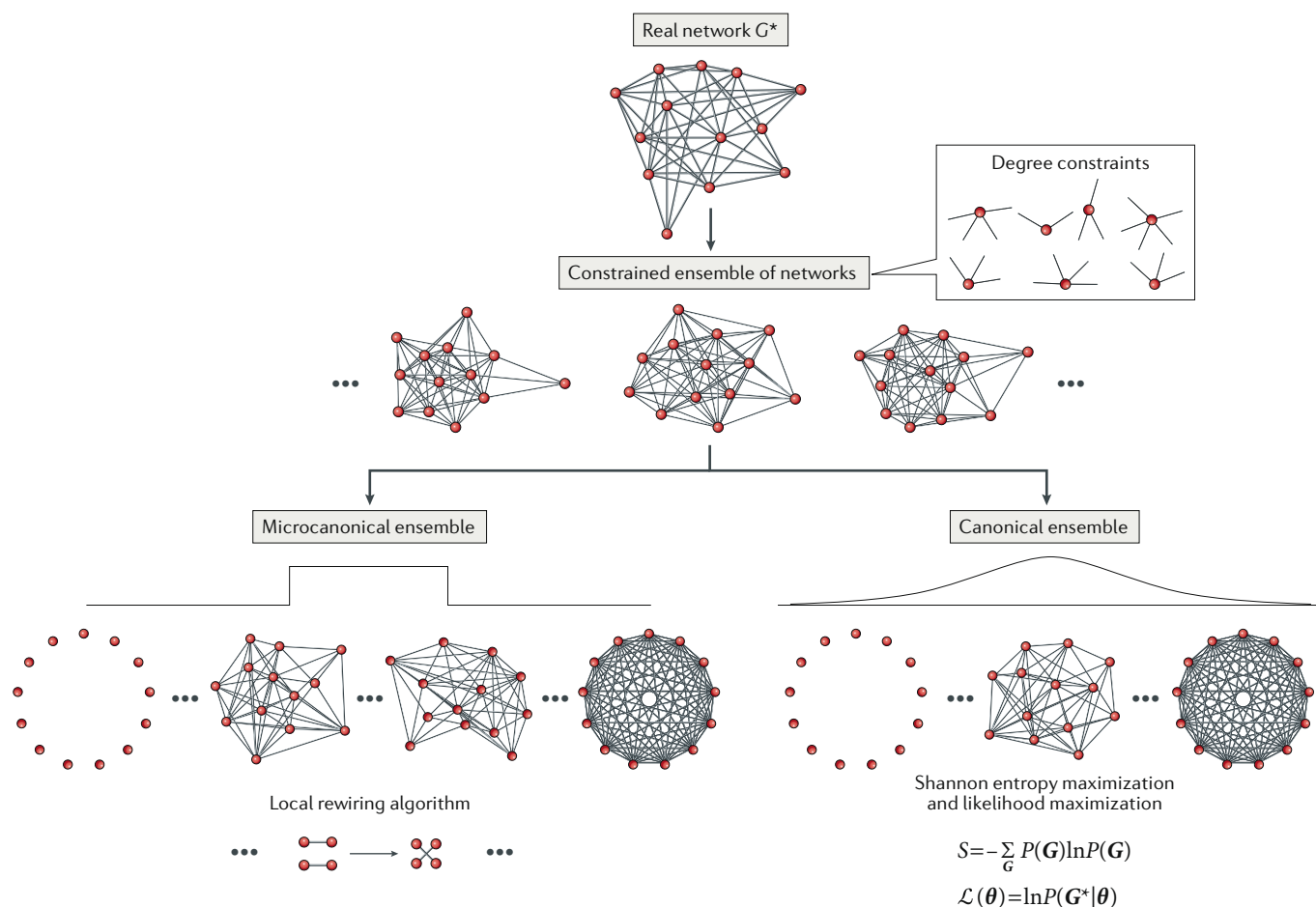
Fig. 1 | **Construction of the microcanonical and canonical ensemble of networks from local constraints.** The real network G^* is the source of constraints on the ensemble. In this case, the constraints are the degrees k^* of G^*. The microcanonical approach relies on the link rewiring method to numerically generate several network configurations, each with exactly the same degree sequence of G^*. The probability $P(G)$ of a network G in the ensemble is non-zero only for the subset of graphs that realize the enforced constraints exactly, as indicated by the schematic probability distribution. Provided the sampling is unbiased, $P(G)$ is uniform for these graphs. The canonical approach obtains $P(G)$ by maximizing the Shannon entropy S while constraining the expected degree values within the ensemble and then maximizing the likelihood $\mathcal{L}$ of $P(G^*)$ to find the ensemble parameters θ^*, such that the expectation values of the degrees match the observations in G^*. Thus, $P(G)$ is non-zero for any graph, ranging from the empty to the complete one, as indicated by the schematic probability distribution.

transitions), in which there is typically a finite number of constraints, such as total energy and total number of particles. In this sense, complex networks not only provide an important domain of application for statistical physics but can also expand our fundamental understanding of statistical physics itself and lead to new theoretical insights. 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 is usually done, but rather should follow from a principled criterion. In particular, because the canonical ensemble describes systems subject to statistical fluctuations, it is the more appropriate ensemble to use when the observed values of the constraints can be affected by measurement errors, missing and spurious data or simply stochastic noise. Fortunately, as we shall see, the canonical ensemble is more analytically tractable than the microcanonical ensemble.

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

$$\mathcal{L}(\theta) = \ln P(G^*|\theta) \qquad (3)$$

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

Box 1 | Alternative ensemble constructions

Various methods have been proposed to define ensembles of graphs with local constraints, as alternatives to maximum entropy. 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 to its target degree, and pairs of stubs are then randomly matched without allowing the formation of self-loops and multilinks[157–159]. Unfortunately, this procedure very often gets stuck in configurations in which nodes requiring additional connections have no more eligible partners, unacceptably leading to many sample rejections[45]. The top-down approach instead starts from a realized network and generates a set of randomized variants by iteratively applying a link rewiring algorithm that preserves the degree distribution[42,58,160,161]. The drawback of the top-down approach is that the number 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 it employs a rewiring acceptance probability that depends on the current network configuration[126–128,163]. Other computational methods rely on theorems that set necessary and sufficient conditions for a degree sequence to be graphic, that is, realized 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 (see REF.[38] and references therein).

Analytic methods instead define 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 $P(k)$[157,167]. The generating function is rigorously defined only for infinite and locally tree-like networks, although it often works surprisingly well for real networks[168]. A popular alternative approach is based on the explicit expression for the connection probability between any two nodes i and j in the randomized ensemble, $p_{ij} = \frac{1}{2}k_i^\star k_j^\star/E^\star$, where $E^\star$ is the total number of links[67]. This model defines systems with self-loops as well as multilinks[37,40,43]. Indeed, because 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 realize the imposed constraints. The model predicts multilinks with non-negligible frequency in scale-free networks with $P(k) \approx k^{-\gamma}$ for which the natural cut-off $\sim N^{1/(\gamma-1)}$ is larger than the structural cut-off ($\sim N^{1/2}$ for uncorrelated networks)[169,170], where N is the number of nodes in the network and γ is the exponent of the power-law.

Density
The fraction of possible connections that are actually realized in a network. Real-world networks are typically sparse, as their density is much smaller than 1.

Erdös–Rényi model
The random graph model in which a link between any two nodes exists with constant probability p, independent of all other links.

Imposing local constraints. Unlike most alternative approaches (briefly outlined in BOX 1), the maximum entropy method is general and works for networks regardless of size, directedness, weighting, density, clustering and other properties. However, determining the occurrence probability of a graph in the ensemble can be a challenging task. In a handful of cases, an analytical calculation of the partition function is feasible, so expectation values and higher moments of any quantity in the ensemble can be analytically derived. As in conventional equilibrium statistical mechanics, whether such an analytical computation is possible depends on the particular constraints imposed. Analytical computation of the partition function is possible in very simple models, such as the well-known Erdös–Rényi random graph[39], as well as when the constraints are the degrees $k^\star$ and strengths s* that describe the local network structure from the viewpoint of each node[26]. In fact, imposing local constraints separately for each node is a minimum requirement to construct ERG ensembles that are both theoretically sound and practically useful, in the sense that they accurately replicate the observed heterogeneity of real-world networks. This is because degree and strength distributions in many real-world systems are scale-free — a property that distinguishes many networks from systems

typically studied in physics, such as gases, liquids and lattices — and scale-free degree and strength distributions cannot be obtained from simple models with global constraints. For instance, the Erdös–Rényi model is obtained in the ERG formalism by constraining the expected total number of links, and this global constraint leads to an ensemble in which each pair of nodes is connected with fixed probability p, implying that the degree distribution follows a binomial law, rather than a power law. Local constraints in turn make the ERG model analytical because of the independence of dyads: $P(\mathbf{G})$ factorizes into link-specific terms, whose contribution to the partition function sum can be evaluated independently of the rest of the network. Note that local constraints lie 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, such as the total number of links, which corresponds, for instance, to the total energy of the system in traditional statistical physics.

The ERG model obtained by constraining the degrees $\mathbf{c}^\star \equiv \mathbf{k}^\star$ is known as the (canonical) binary configuration model (BCM). In the simplest undirected case, the entropy maximization procedure returns an ensemble connection probability between any two nodes i and j:

$$p_{ij} = \frac{x_i x_j}{1 + x_i x_j} \tag{4}$$

where $\mathbf{x}$ are the (exponentiated) Lagrange multipliers[26].

The weighted configuration model (WCM)[40] is instead obtained by constraining the strengths $\mathbf{c}^\star \equiv \mathbf{s}^\star$. 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 the link between i and j (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

$$\left\langle w_{ij} \right\rangle = \frac{y_i y_j}{1 - y_i y_j} \tag{5}$$

These two models recall traditional statistical mechanics for systems of non-interacting particles, if connections are interpreted as particles in a quantum gas and pairs of nodes as single-particle states. Indeed, in binary networks, each pair of nodes can be connected by at most one link, or equivalently, each single-particle state can be occupied by at most one particle. Therefore, for binary networks, equation 4 results in fermionic statistics. However, weighted networks correspond to particle systems for which single-particle states can be occupied by an arbitrary number of particles, so equation 5 describes a system of bosons. For these systems, Bose–Einstein condensation can occur between very strong nodes for which $y_i y_j \to 1$ (REF.[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].

In the simplest undirected case, using (exponentiated) Lagrange multipliers x and y, respectively, for degrees and strengths, 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 of reinforcing an existing one obey intrinsically different rules. The connection creation process serves to satisfy the degree constraints, and the reinforcement process serves to fix the values of the strengths. Like the ensemble nonequivalence mentioned above, this mechanism and the resulting mixed statistics constitute new physical phenomena that the statistical physics approach to networks can unveil.

Pattern validation

Null models for networks. Validating models, that is, comparing their statistical properties with measurements of real-world systems, is an essential activity of theoretical physics. In the context of networks and complex systems, apart from looking for what a model is able to explain, much research has been devoted to identifying properties in real-world networks that deviate from a benchmark model[42–50]. This is because possible deviations are likely to contain important information about the unknown formation process of a network or one of its functions.

Maximum entropy models are perfectly suited for this task. Starting from a real network $G^\star$, they are used to derive the null hypothesis (that is, the benchmark model) using the set of properties $c(G^\star)$ as constraints. Otherwise, no other information about the system is assumed. In other words, the null hypothesis is that the chosen constraints are the only explanatory variables for the network at hand. The other properties of $G^\star$ can then be statistically tested and validated against this null hypothesis. For instance, a null model that is derived by imposing the total number of links as a (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, clustering, assortativity, motifs and so on) are statistically significant beyond what can be expected from the heterogeneity of the degrees or strengths themselves. Because an ensemble given by local constraints can be analytically characterized, the expectation values and standard deviations of most quantities of interest can be explicitly derived, therefore hypothesis testing based on standard scores can be easily performed[38]. When the ensemble distribution of the considered quantity is not normal, sampling of the configuration space using explicit formulas for $P(G)$ makes it straightforward to perform statistical tests based on P values.

In the context of pattern validation, one of the most studied systems, which we also discuss here as an illustrative example, is the World Trade Web (WTW), which is the network of trade relationships between countries in the world[51,52]. The network exhibits disassortativity, in that countries with many trade partners are connected

on average to countries with few partners[53,54]. This pattern is statistically explained to good approximation by the degree sequence[38]. An analogous situation also exists for the clustering coefficient[38]. These two observations exclude the presence of meaningful indirect economic interactions on top of the direct economic interactions in the WTW.

The situation changes, however, when the weighted version of the WTW network is analysed[44,55,56]. The network is still disassortative, in that countries with large export volumes are connected on average to countries with small export volumes. However, this pattern is not compatible with that of the WCM null model. Concerning the weighted clustering coefficient[57], the agreement between the empirical network and the model is only partial. These findings point to the fact that, unlike in the binary case, knowledge of the strengths conveys only limited information about the higher-order weighted structure of the network. Together with the fact that basic topological properties, such as the link density, are not reproduced by the WCM (as discussed below), this 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 generalizing the concept of clustering. Typically, the null model used to study motifs in directed networks is obtained by constraining, in addition to the degrees, the number of reciprocal links per node[60–62]. Indeed, reciprocity is meaningful in many contexts. For instance, in food webs, the presence of bidirected predator–prey relations between two species strongly characterizes 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 important early warnings of financial turmoil[64].

Community structure instead refers to the presence of large groups of nodes that are densely connected within the group but sparsely connected to nodes outside the group[6]. Most methods to find communities in networks are based on optimization of a functional. The most prominent example of such a functional is the modularity[5], which compares the number of links falling within and between groups with the expectation of such numbers under a null network model. The comparison with a null model is a fundamental step in the procedure, because 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, and thus it gives an accurate benchmark only when these events are rare, such as in very large and sparse networks. More generally, the maximum entropy approach (for example, the configuration model of equation 4), although more demanding from a practical viewpoint, can provide an appropriate null

model that discounts the degree heterogeneity as well as other properties of the network[38,68].

The ERG framework can be directly used to generate networks with a community structure by specifying the average number of links within and between each community. In this way, the ensemble becomes equivalent to the stochastic 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 of only the block membership of the nodes[69]. This means that equation 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 the node degrees are also constrained.

Bipartite networks and one-mode projections. In bipartite networks, 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, in which individuals are connected with the groups of which they are a member, and ownership networks, in which individuals are connected with the items they collected. The bipartite configuration model (BiCM)[74] extends the BCM to this class of networks. The BiCM method has been used, for instance, to study the network of countries and products they export, that is, the bipartite representation of the WTW[75,76], and to 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 is sufficient to induce a certain amount of nestedness of the interactions[78].

To directly show the structure of relationships among one of the two sets of nodes, a bipartite network can be compressed into its one-mode projection. This is a network containing only the nodes of the considered set, connected with weights that depend on how many common neighbours the nodes have in the other set[79]. The problem of building a statistically validated one-mode projection of a bipartite network is similar in spirit to that of extracting the backbone from standard weighted networks[80–83].

The typical approach involves determining which links are significant using a threshold, which is either unconditional or dependent on the degree of nodes in the projected set[84–88]. However, unlike weighted networks, one-mode projections should be assessed against null models that constrain the local information of both sets of the original bipartite network. Unfortunately, these models are difficult to derive. For instance, the use of computational link rewiring methods for one-mode projections[89,90] is even more impractical and biased[91] than for standard networks. Other null models 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). This null model has been used, for instance, to

analyse the one-mode projection of the bipartite WTW. This analysis allows detection of modules of countries that have similar industrial systems, construction of a hierarchical structure of products[94] and tracing of specializations that emerge from the baseline diversification strategy of countries[95]. The same null model has also been used to study the patterns of asset ownership by financial institutions. In this case, the analysis allows identification of significant portfolio overlaps that bear the highest risk of fire-sale liquidation and forecast of market crashes and bubbles[93]. More recently, a null model obtained by pairwise projection of multiple bipartite networks has been successfully applied to identify significant innovation patterns involving the interplay of scientific, technological and economic activities[96].

Network reconstruction

The problem of partial information. Many dynamical processes of critical importance, such as disease spreading or information diffusion, are sensitive to the topology of the network of interactions on which they occur[97]. However, in many situations, the structure of the network is at least partially unknown. A classic example is that of financial networks. Financial institutions publicly disclose their aggregate exposures in their balance sheets, but individual exposures (who is lending how much to whom) remain confidential[98–100]. Another example is that of social networks, which are too large in scale to allow exhaustive crawling and for which only aggregate information is typically released owing to privacy issues[101,102]. For natural and biological networks, measuring all possible interactions is difficult because of technological limitations or high experimental costs[103,104]. Thus, reconstruction of the network structure when only limited information is available is a problem that 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 partially 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]. As with the other problems discussed in this Review, the key to success is to make optimal use of what is known about the system and to make the most unbiased estimate of what is not known. This is naturally achieved using techniques based on the maximum entropy principle: the probability distribution that best represents the current state of knowledge on the network is the one with the largest uncertainty that also satisfies the constraints corresponding to the available information. Note that the ERG approach has the additional advantage of defining not a single reconstructed network instance but instead 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.

However, different kinds of local constraints lead to substantially different outcomes for the reconstruction

---

Nestedness
The pattern in which the interactions of nodes with low degree are a subset of the interactions of nodes with high degree.

Backbone
The core component of the network that is extracted by filtering redundant information.
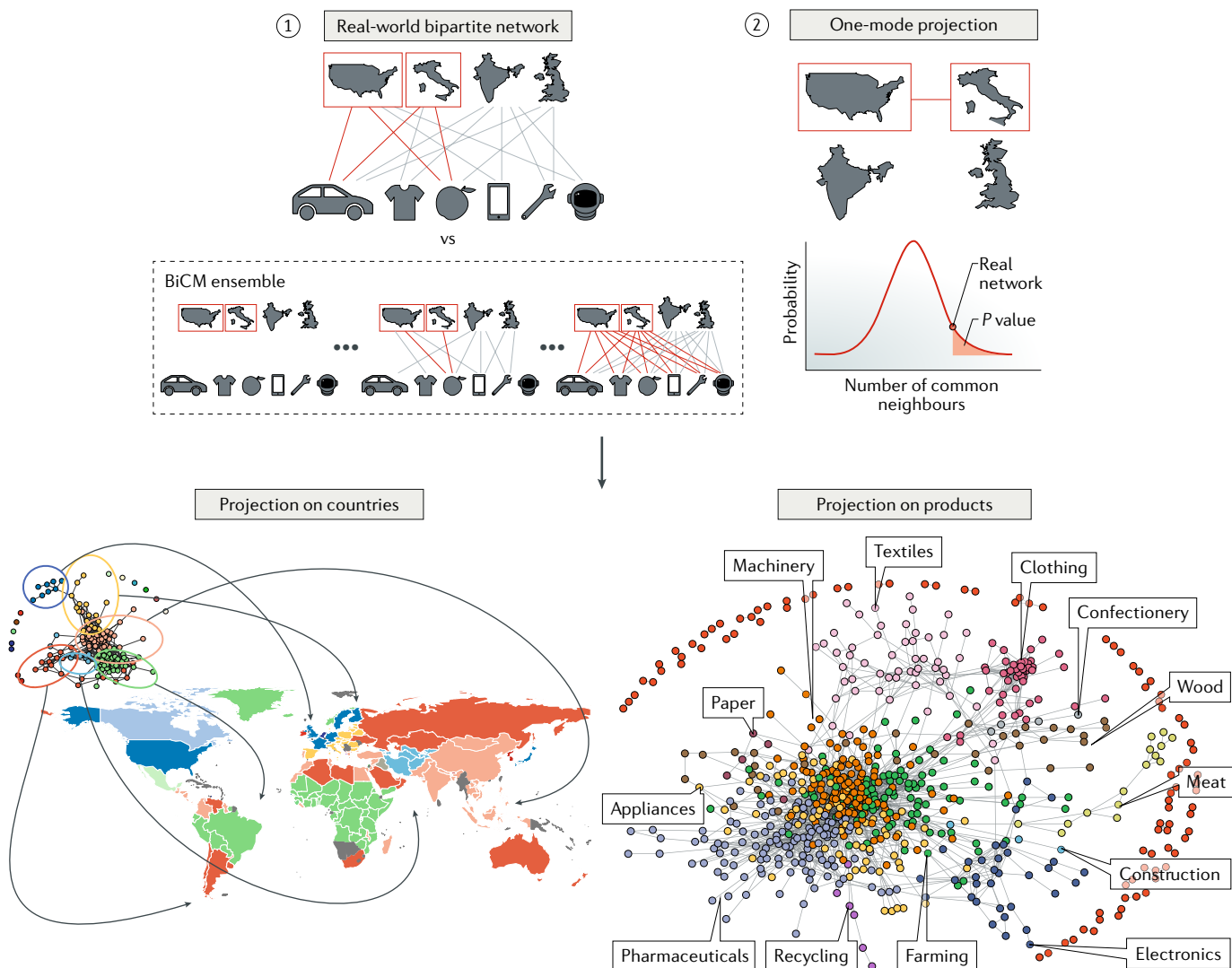
Fig. 2 | **One-mode projection of the network of countries and products they export and its statistical validation against a null hypothesis derived from the BiCM.** In this example, the existence of a link between the US and Italy (red boxes in the networks) in the one-mode projection is determined. Step 1: the real-world bipartite network is compared with the bipartite configuration model (BiCM) ensemble (dashed box), which is derived by constraining the degrees of both countries and products. The ensemble gives a connection probability for any country–product pair analogous to that of equation 4. Step 2: for the pair of countries under consideration, the actual number of products they both export (red links on the real-world network) is assessed using the probability distribution obtained from the BiCM. In this example, the observed value is statistically significant, that is, the null hypothesis that the value is explained by the degrees of both countries and products is rejected, and thus a link connecting the two countries is placed in the one-mode validated projection. This procedure can be repeated for each pair of countries to obtain the projection on the country set of the real-world trade web. This projection identifies communities of countries that have similar industrial systems. Instead, the projection performed on the product set highlights classes of products that require similar technological capabilities. Adapted from REF.[94], CC-BY-4.0.

process. For a variety of networks of different nature (such as economic, financial, social or ecological networks), constraining the degrees, as in the BCM, typically returns a satisfactory reconstruction of the binary network features. However, constraining the strengths, as in the WCM, almost always leads to a poor weighted reconstruction[38,41]. This is because the entropy maximization procedure is unbiased by not assuming any relationship 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 the probability of assigning zero weight to a link is extremely small; the reconstructed network becomes almost fully connected, regardless of the link density of the original network. This phenomenon shows that degrees and strengths carry different kinds of information and constrain the network in fundamentally different ways. To reconstruct sparse weighted networks, both are required[47,49], as in the ECM[41]. This is quantitatively shown using information-theoretic criteria (see REF.[41] and BOX 2).

Box 2 | Comparing models from different constraints

Competing models that result from different choices of constraints can be compared using various likelihood-based statistical criteria. This, in turn, means assessing 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 $\boldsymbol{\theta}^{\star}$ with respect to $\mathcal{M}_1$ or equivalently that $\mathcal{M}_1$ is a special case of $\mathcal{M}_2$, a comparison can be made through the likelihood ratio test (LRT)[171]; if $\mathrm{LRT}_{\mathcal{M}_1/\mathcal{M}_2} = -2[\mathcal{L}_{\mathcal{M}_1}(\boldsymbol{\theta}_1^{\star}) - \mathcal{L}_{\mathcal{M}_2}(\boldsymbol{\theta}_2^{\star})]$ is smaller than a given significance level, $\mathcal{M}_2$ should be rejected even though its likelihood is, by definition, higher than that of $\mathcal{M}_1$. This is because $\mathcal{M}_2$ overfits the data by using redundant parameters that have intercorrelations and that thus provide spurious information on the system[172].

When $\mathcal{M}_1$ and $\mathcal{M}_2$ are not nested, the Akaike information criterion (AIC)[173] ranks them in increasing order of $\mathrm{AIC}_{\mathcal{M}_r} = 2[K_{\mathcal{M}_r} - \mathcal{L}_{\mathcal{M}_r}(\boldsymbol{\theta}_r^{\star})]$, where $K_{\mathcal{M}_r}$ is the number of parameters of model $\mathcal{M}_r$. For R competing models, Akaike weights $\omega_{\mathcal{M}_r} = e^{-\left(\mathrm{AIC}_{\mathcal{M}_r}/2\right)} / \sum_{r=1}^{R} e^{-\left(\mathrm{AIC}_{\mathcal{M}_r}/2\right)}$ quantify the probability that each model is the most appropriate to describe the data[174].

The Bayesian information criterion (BIC)[175] is similar to the AIC but accounts for the number n of empirical observations as $\mathrm{BIC}_{\mathcal{M}_r} = K_{\mathcal{M}_r} \ln n - 2\mathcal{L}_{\mathcal{M}_r}(\boldsymbol{\theta}_r^{\star})$. Bayesian weights can be also defined, analogously to Akaike weights. The BIC is believed to be more restrictive than the AIC, but which criterion performs best and under which conditions is still debated (see (REF.[172]) and references therein).

We note that other model selection methods have also been proposed, such as multimodel inference, in which a sort of average over different models is performed[172].

The fitness ansatz. Unfortunately, in many situations — and typically for financial networks — the node degrees are also unknown. A possible solution comes from the observation that in many real networks, the nodes can be allocated fitnesses, from which connection probabilities between nodes and node degrees can be obtained[14,52,107–109]. The strengths themselves work well as fitnesses in many cases[44]. The fitness ansatz assumes that the strength of a node is a monotonic function of the Lagrange multiplier x controlling the degree of that node. Assuming the simplest linear dependence $s^{\star} \propto x$ (other functional forms are, in principle, allowed), equation 4 becomes[110–112]

$$p_{ij} = \frac{z s_i^{\star} s_j^{\star}}{1 + z s_i^{\star} s_j^{\star}} \qquad (6)$$

The proportionality constant z can be easily found using maximum likelihood arguments together with a bootstrapping approach to assess the network density[113], which relies on the hypothesis that subsets of the network are representative of the whole system, regardless of the specific portion that is observed. Node degrees estimated from the fitness ansatz are then used as inputs to the ECM to obtain the ensemble of reconstructed networks[111] (FIG. 3). Alternatively, heuristic techniques can be employed to reduce the complexity of the method by replacing the construction of the ECM ensemble with a density-corrected gravity model that obtains weights as $w_{ij} \propto z^{-1} + s_i^{\star} s_j^{\star}$ with probability p_{ij} (REF.[112]). The method is also readily extended to bipartite networks[114].

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

Beyond local constraints

Approximate and numerical methods. Whether an ERG model is analytically tractable depends on whether a closed-form expression of its partition function Z can be derived. As we have seen, this is indeed possible in the case of local linear constraints, for which Z factorizes into link-specific terms. In some other cases, it may be possible to obtain approximate analytic solutions using a variety of techniques, such as mean-field theory, saddle-point approximation, diagrammatic perturbation theory and path integral representations. These situations have been 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 models constraining both the degree distribution and degree–degree correlations, which are known under the name of tailored random graphs[123–125]. If analytic approaches for computing Z are intractable, the ensemble can still be populated using Monte Carlo simulations. These simulations can be used either to explicitly sample the configuration space — taking care to avoid sampling biases by using ergodic Markov chains fulfilling detailed balance[126–128] — or to derive approximate maximum likelihood estimators — taking care to avoid degenerate regions of the phase space, which often lead to trapping in local minima[129–135]. Such a variety of techniques is what makes the ERG model an extremely flexible and powerful framework for complex network modelling.

Markov chain Monte Carlo. In the MCMC method for ERG models, a new network $\boldsymbol{G}' \in \Omega$ is proposed by taking a network $\boldsymbol{G} \in \Omega$ and shuffling two links chosen at random. The shuffling is performed so 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 $Q_{\boldsymbol{G} \to \boldsymbol{G}'} = \min\{1, e^{H(\boldsymbol{G}) - H(\boldsymbol{G}')}\}$[136], where H is the Hamiltonian containing the Lagrange multipliers that take the role of inverse temperatures used for simulated annealing. The process is repeated from $\boldsymbol{G}'$ if the proposal is accepted and from $\boldsymbol{G}$ if the proposal is rejected. The moves fulfil ergodicity and detailed balance, and thus, for sufficiently long times, the values of the constraints in the sampled networks are distributed according to the prescription of the canonical ensemble.

However, the time to reach the correct distribution grows exponentially with system size, leading to failure of the MCMC method in practice. This happens whenever $P(\boldsymbol{G})$ possesses more than one local maximum, which happens for instance when using ERG models to generate 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 that are biased by aiming at a given level of clustering 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 timescales, leading to a breaking of ergodicity[139]. Multicanonical
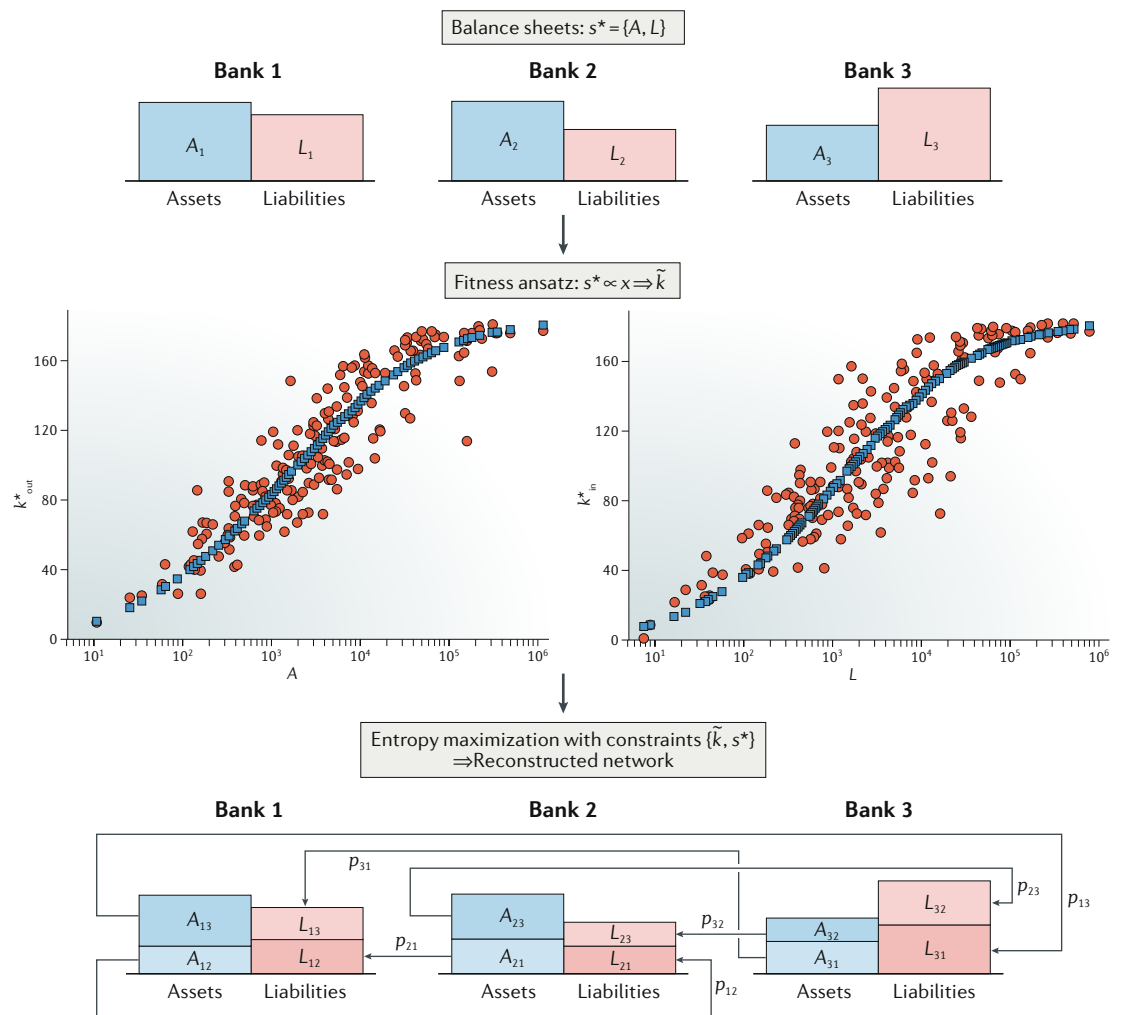
Fig. 3 | **Statistical reconstruction of an interbank network given by bilateral exchanges among banks.** The method takes as input the total exposure of each bank, that is, its total interbank assets A and liabilities L, obtained from publicly available balance sheets (upper panel), which correspond to the out-strength and in-strength s^* of the node. Two entropy maximization steps are performed. The first step is to estimate the node degrees $\tilde{k}$, that is, the number of loans for each bank. These are unknown because the balance sheets list only the total amount of assets and liabilities. The estimate is performed using the fitness assumption that the Lagrange multiplier x controlling the out-degree and in-degree k^* of a node scale with the total interbank assets and liabilities, respectively (middle panel), which results in a connection probability p_{ij} between nodes i and j of the form of equation 6. Node degrees can then be estimated using these probabilities. The second step reconstructs the full weighted topology of the network. One method for doing this is to use an enhanced configuration model with constraints on both the degrees $\tilde{k}$ estimated from the first step and empirical strengths s^* (REF.[111]). Another method is to use a heuristic density-corrected gravity approach[112] to place weights on realized connections. The outcome of the reconstruction process is a probability distribution over the space of networks compatible with the constraints (lower panel). Scatter plots adapted from REF.[112], CC-BY-4.0.

sampling has been 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 requires sampling networks uniformly on a predefined range of constraint values. This is achieved using Metropolis–Hastings steps based on a microcanonical density of states estimated using the Wang–Landau algorithm[141].

Generalized network structures
Networks of networks. Many complex systems are not simply isolated networks but are better represented by networks of networks[142–144] (see FIG. 4). The simplest and

most studied situation is the so-called multiplex (FIG. 4a), in which the same set of nodes interacts on several layers of networks. For example, in social networks, each individual has different kinds of social ties; in urban systems, locations can be connected by different means of transportation; and in financial markets, institutions can exchange different kinds of financial instruments.

Mathematically, a multiplex $\vec{G}$ is a system of N nodes and M layers of interactions. Each layer, labelled $\alpha = 1, \ldots, M$, consists of a network G^α. When modelling such a system, the simplest hypothesis is that the various layers are uncorrelated. In ERG terms, this means that the probability of the multiplex factorizes into the probabilities
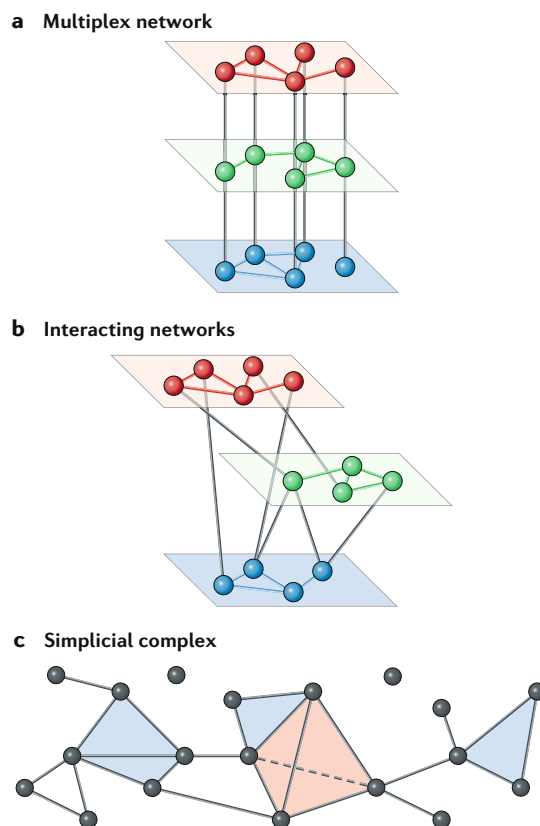
a Multiplex network

b Interacting networks

c Simplicial complex

Fig. 4 | **Generalized network structures. a** | A multiplex (or 'link-coloured') network consists of various layers with the same set of nodes and specific interaction patterns in each layer. **b** | An interacting (or multilayer) network consists of various layers, each with its own nodes, with interactions existing within and between layers. **c** | A simplicial complex represents the different kinds of interactions (simplices) between groups of d nodes: isolated nodes ($d = 0$), pairwise interactions ($d = 1$, grey lines), triangles ($d = 2$, blue) and tetrahedra ($d = 3$, red).

of each network layer: $P(\vec{G}) = \prod_{\alpha=1}^{M} P_{\alpha}(G^{\alpha})$[145,146]. This happens whenever the constraints imposed on the multiplex are linear combinations of constraints on individual network layers, which includes the situation in which local constraints are imposed separately for each network layer. 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 $p_{ij}^{\alpha} = \left(x_i^{\alpha} x_j^{\alpha}\right) / \left(1 + x_i^{\alpha} x_j^{\alpha}\right)$, where x^{α} are layer-specific Lagrange multipliers. As a result, the existence of a link is independent of 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.

In more realistic models of correlated multiplexes, the existence of a link in one layer is correlated with the existence of a link in another layer. Such models can be generated by constraining the multilink structure of the system[145]. A multilink m is an M-dimensional binary vector m indicating a given pattern of connections between a generic pair of nodes in the various layers. The multidegree $k(m)$ of a node in a given graph configuration is then the total number of other nodes

with which the multilink m is realized. Constraining the multidegree sequence of the network, which requires imposing 2^M constraints per node, leads to a probability for a multilink m between node i and node j given by $p_{ij}^{m} = (x_i^m x_j^m)/(\sum_m x_i^m x_j^m)$. This method can be used to build systems made up of sparse layers with non-vanishing overlap.

For weighted multiplexes, it is simple to impose either the strengths alone or both degrees and strengths on each layer. Doing this leads to uncorrelated layers[147]. However, the situation is far more complicated if layers are correlated: no closed-form solution for $P(\vec{G})$ has been obtained to date, and thus a sampling procedure has been devised[148]. An interesting related case is provided by systems of aggregated multiplexes, which are simple networks given by the sum of the various layers of the multiplex. For aggregated multiplexes, 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 the number of layers of the original multiplex is known, the model can be built using a layer-degeneracy term for each link, which counts the number of ways its weight can be split across the layers. Doing this leads to a WCM(M) model, for which the weight distribution is a negative binomial, with the geometric distribution of the standard WCM being the special case $M = 1$.

Furthermore, if it is known that links in each layer have a different nature, a good modelling framework is equivalent to that of multi-edge networks[149,150], for which links belonging to different layers can be distinguished (FIG. 4b). To deal with this distinguishability, it is necessary to introduce another degeneracy term that counts all the network configurations giving rise to the same $P(G)$. The solution of the model is then obtained using a mixed ensemble with a hard constraint for the total network weight and soft constraints for node strengths. This leads to Poisson statistics (independent of M) for link weights and $\langle w \rangle_{ij} = M y_i y_j \propto s_i^{\star} s_j^{\star}$, a rather different situation than the outcome of either WCM or WCM(M)[151].

We note that 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 trips between locations (nodes) aggregated over observation periods (layers), are better modelled by multi-edge networks, as long as each trip is distinguishable. For the aggregated WTW, the situation is less clear: whereas commodities are, in principle, distinguishable, trade transactions are much less so, and in fact, neither the WCM, WCM(M) or the multi-edge model can reproduce it well. A possible solution here is again to constrain both strengths and degrees simultaneously[152].

Simplicial complexes. Simplicial complexes are generalized network structures that describe interactions between more than two nodes. They can be used to describe a wide variety of complex interacting systems, such as collaboration networks in which works result from two or more actors working together, protein

---

Box 3 | Boltzmann, von Neumann and Kolmogorov entropies

Alternatives to Shannon entropy are used in statistical physics of networks for various purposes.

The Boltzmann entropy is the logarithm of the number of network configurations belonging to a microcanonical ensemble. Like the Shannon entropy for a canonical ensemble, the Boltzmann entropy can be used to quantify the complexity of the ensemble, that is, to assess the role of different constraints in terms of information they carry about the network structure[31,120]. Indeed, the more informative the constraints in shaping the network structure are, the smaller the effective number of graphs with the imposed features, and thus 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 configuration model ensemble is large, whereas scale-free degree distributions naturally emerge for network ensembles with minimal Boltzmann entropy[121].

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

Finally, the Kolmogorov entropy generalizes the Shannon entropy by describing the rate at which a stochastic process generates information. Consider for simplicity a Markovian ergodic stochastic process described by the matrix $\Phi = \{\phi_{ij}\}$ of transition rates $\{i \to j\}$ and by the stationary asymptotic probability distribution $\{\pi_i\}$. In this case, the dynamical entropy is defined as $-\sum_{ij}\pi_i\phi_{ij}\log\phi_{ij}$, that is, the average of the Shannon entropies of rows of Φ, each weighted by the stationary distribution of the corresponding node[182]. The Kolmogorov entropy is related, on the 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 because 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].

---

interaction networks in which complexes often consist of more than two proteins, economic systems of financial transactions often involving several parties and social systems in which groups of people are united by common motives or interests. Simplicial complexes can involve any number of nodes. For instance, simplices of dimension $d = 0, 1, 2$ and 3 are nodes, links, triangles and tetrahedra, respectively, and d-dimensional simplices are their d-dimensional generalizations. The δ-dimensional faces of a d-dimensional simplex ($\delta < d$) are all the simplices formed by subsets of $\delta + 1$ nodes. A simplicial complex represents the different kinds of interactions within groups of d nodes. It is a collection of simplices of different dimensions that are connected such that every face of a simplex in the complex is in the complex and the intersection of every pair of simplices in the complex is a face of both simplices (FIG. 4c).

Exponential random simplicial complexes have been recently introduced as a higher-dimensional generalization of ERGs. They enable generation of random simplicial complexes in which each simplex has its own independent probability of appearance, conditioned to the presence of simplex boundaries, which become additional constraints on the model[153]. Explicit calculation of the partition function is possible when considering random simplicial complexes formed exclusively by d-dimensional simplices[154]. Indeed, by constraining

the generalized 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 MCMC sampling scheme can be used to populate a microcanonical ensemble of simplicial complexes formed by simplices of any dimension[155].

Perspectives and conclusion

Complex networks are different from the systems traditionally studied in equilibrium statistical physics in two key ways. One is that the microscopic degrees of freedom of networks are the interactions between the nodes making up the system and not the states of the nodes themselves. The other is that in real-world networks, nodes are typically heterogeneous, both in terms of intrinsic characteristics and connectivity features, such that networks cannot be assigned a typical scale. Maximum entropy models of networks based on local constraints are founded on these two facts, because they define probability distributions on the network connections and do not distinguish nodes beyond their local features. As we reviewed here, these models have found a wide range of practical applications. This is because they can often be analytically characterized; they are able to include higher-order network features, such as assortativity, clustering and community structure, using stochastic sampling; and they can model even more complex structures such as networks of networks and simplicial complexes.

However, there are limitations to the maximum entropy model approach. In the models discussed here, the constraints that can be imposed must be static topological properties of the network. Dynamical constraints have been considered only recently, 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 by using definitions of the entropy functional alternative to Shannon's definition (see BOX 3). Another limitation of maximum entropy network models is that the possibility of considering semilocal network properties relies heavily on numerical sampling, which becomes unfeasible or strongly biased for non-trivial patterns involving more than two or three nodes. However, these patterns can be important in situations in which the network structure is determined by complex optimization principles, such as subunits of an electrical circuit, biochemical reactions in a cell or neuron firing patterns in the brain. The field of statistical physics of networks will need to face the challenge of developing more sophisticated network models for these kinds of structures. Nevertheless, maximum entropy models based on local constraints provide effective benchmarks to detect and validate such structures.

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 the one hand, the details and facets of

information that can now be extracted have reached levels never seen before, which means that increasingly complex data structures and models are needed in order to represent and comprehend them. On the other hand, the quantity of information available requires effective and scalable ways to let the signal emerge from the noise originating from the large variety of sources. The theoretical framework of statistical physics stands as an essential instrument to make consistent inference from data, whatever the level of complexity faced.

Published online 8 January 2019

1. Dorogovtsev, S. N., Goltsev, A. V. & Mendes, J. F. F. Critical phenomena in complex networks. *Rev. Mod. Phys.* **80**, 1275–1335 (2008).
2. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
3. Yook, S. H., Jeong, H., Barabási, A.-L. & Tu, Y. Weighted evolving networks. *Phys. Rev. Lett.* **86**, 5835–5838 (2001).
4. Barrat, A., Barthelemy, M. & Vespignani, A. Weighted evolving networks: coupling topology and weight dynamics. *Phys. Rev. Lett.* **92**, 228701 (2004).
5. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).
6. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174 (2010).
7. Watts, D. J. & Strogatz, S. H. Collective dynamics of small-world networks. *Nature* **393**, 440–442 (1998).
8. Amaral, L. A. N., Scala, A., Barthélémy, M. & Stanley, H. E. Classes of small-world networks. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149–11152 (2000).
9. Chung, F. & Lu, L. The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15879–15882 (2002).
10. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys0.* **74**, 47–97 (2002).
11. Newman, M. E. J. The structure and function of complex networks. *SIAM Rev. Soc. Ind. Appl. Math.***45**, 167–256 (2003).
12. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: structure and dynamics. *Phys. Rep.***424**, 175–308 (2006).
13. Bianconi, G. & Barabási, A. L. Bose-einstein condensation in complex network. *Phys. Rev. Lett.* **86**, 5632–5635 (2001).
14. Caldarelli, G., Capocci, A., De Los Rios, P. & Muñoz, M. A. Scale-free networks from varying vertex intrinsic fitness. *Phys. Rev. Lett.* **89**, 258702 (2002).
15. Dorogovtsev, S. N., Mendes, J. F. F. & Samukhin, A. N. Structure of growing networks with preferential linking. *Phys. Rev. Lett.* **85**, 4633–4636 (2000).
16. Medo, M., Cimini, G. & Gualdi, S. Temporal effects in the growth of networks. *Phys. Rev. Lett.* **107**, 238701 (2011).
17. Holland, P. W. & Leinhardt, S. An exponential family of probability distributions for directed graphs. *J. Am. Stat. Assoc.* **76**, 33–50 (1981).
   This paper introduces ERGs as a formalism to define probability distributions for the structures of social networks.
18. Frank, O. & Strauss, D. Markov graphs. *J. Am. Stat. Assoc.* **81**, 832–842 (1986).
19. Strauss, D. On a general class of models for interaction. *SIAM Rev. Soc. Ind. Appl. Math.* **28**, 513–527 (1986).
20. Wasserman, S. & Pattison, P. Logit models and logistic regressions for social networks: I. An introduction to markov graphs and *p. Psychometrika* **61**, 401–425 (1996).
21. Anderson, C. J., Wasserman, S. & Crouch, B. A p* primer: logit models for social networks. *Soc. Networks* **21**, 37–66 (1999).
22. Snijders, T. A. B., Pattison, P. E., Robins, G. L. & Handcock, M. S. New specifications for exponential random graph models. *Sociol. Methodol.* **36**, 99–153 (2006).
23. Robins, G., Pattison, P., Kalish, Y. & Lusher, D. An introduction to exponential random graph (p*) models for social networks. *Soc. Networks* **29**, 173–191 (2007).
24. Cranmer, S. J. & Desmarais, B. A. Inferential network analysis with exponential random graph models. *Polit. Anal.* **19**, 6686 (2011).
25. Snijders, T. A. B. Statistical models for social networks. *Annu. Rev. Sociol.* **37**, 131–153 (2011).
26. Park, J. & Newman, M. E. J. Statistical mechanics of networks. *Phys. Rev. E* **70**, 066117 (2004).
   In this paper, ERGs are interpreted for the first time as the statistical physics framework for complex networks.

27. Jaynes, E. T. Information theory and statistical mechanics. *Phys. Rev.* **106**, 620–630 (1957).
   In this milestone paper, Jaynes shows that equilibrium statistical mechanics provides an unbiased prescription to make inferences from partial information.
28. Shore, J. & Johnson, R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **26**, 26–37 (1980).
29. Pressé, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141 (2013).
30. Jaynes, E. T. On the rationale of maximum-entropy methods. *Proc. IEEE* **70**, 939–952 (1982).
31. Bianconi, G. The entropy of randomized network ensembles. *Europhys. Lett.* **81**, 28005 (2008).
   This paper derives the Boltzmann entropy of a variety of network ensembles to assess the role of structural network properties.
32. Squartini, T., Mastrandrea, R. & Garlaschelli, D. Unbiased sampling of network ensembles. *New J. Phys.* **17**, 023052 (2015).
33. Anand, K. & Bianconi, G. Entropy measures for networks: toward an information theory of complex topologies. *Phys. Rev. E* **80**, 045102 (2009).
34. Squartini, T., de Mol, J., den Hollander, F. & Garlaschelli, D. Breaking of ensemble equivalence in networks. *Phys. Rev. Lett.* **115**, 268701 (2015).
35. Squartini, T. & Garlaschelli, D. Reconnecting statistical physics and combinatorics beyond ensemble equivalence. Preprint at https://arxiv.org/abs/1710.11422 (2018).
36. Garlaschelli, D. & Loffredo, M. I. Generalized bose-fermi statistics and structural correlations in weighted networks. *Phys. Rev. Lett.* **102**, 038701 (2009).
   This paper develops the ERG approach for a general class of weighted networks.
37. Garlaschelli, D. & Loffredo, M. I. Maximum likelihood: extracting unbiased information from complex networks. *Phys. Rev. E* **78**, 015101(R) (2008).
38. Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. *New J. Phys.* **13**, 083001 (2011).
   This paper turns ERGs into null models for empirically observed networks using the maximum likelihood principle.
39. Erdos, P. & Rényi, A. On random graphs. *Publ. Math. Debr.* **6**, 290–297 (1959).
   This paper introduces the first statistical ensemble of random graphs.
40. Serrano, M. Á. & Boguñá, M. Weighted configuration model. *AIP Conf. Proc.* **776**, 101–107 (2005).
41. Mastrandrea, R., Squartini, T., Fagiolo, G. & Garlaschelli, D. Enhanced reconstruction of weighted networks from strengths and degrees. *New J. Phys.* **16**, 043022 (2014).
42. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science***296**, 910–913 (2002).
   This paper introduces the local link rewiring method to build a null network model.
43. Park, J. & Newman, M. E. J. Origin of degree correlations in the internet and other networks. *Phys. Rev. E* **68**, 026112 (2003).
44. Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3747–3752 (2004).
45. Maslov, S., Sneppen, K. & Zaliznyak, A. Detection of topological patterns in complex networks: correlation profile of the internet. *Phys. A Stat. Mech. Appl.* **333**, 529–540 (2004).
46. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, A. Detecting rich-club ordering in complex networks. *Nat. Phys.***2**, 110 (2006).
47. Serrano, M. Á., Boguñá, M. & Pastor-Satorras, R. Correlations in weighted networks. *Phys. Rev. E* **74**, 055101 (2006).

48. Guimera, R., Sales-Pardo, M. & Amaral, L. A. N. Classes of complex networks defined by role-to-role connectivity profiles. *Nat. Phys.* **3**, 63 (2006).
49. Bhattacharya, K., Mukherjee, G., Saramäki, J., Kaski, K. & Manna, S. S. The international trade network: weighted network analysis and modelling. *J. Stat. Mech. Theory Exp.* **2008**, P02002 (2008).
50. Opsahl, T., Colizza, V., Panzarasa, P. & Ramasco, J. J. Prominence and control: the weighted rich-club effect. *Phys. Rev. Lett.* **101**, 168702 (2008).
51. Serrano, M. Á. & Boguñá, M. Topology of the world trade web. *Phys. Rev. E* **68**, 015101 (2003).
52. Garlaschelli, D. & Loffredo, M. I. Fitness-dependent topological properties of the world trade web. *Phys. Rev. Lett.***93**, 188701 (2004).
53. Garlaschelli, D. & Loffredo, M. I. Structure and evolution of the world trade network. *Phys. A Stat. Mech. Appl.* **355**, 138–144 (2005).
54. Fagiolo, G., Reyes, J. & Schiavo, S. World trade web: topological properties, dynamics, and evolution. *Phys. Rev. E* **79**, 036115 (2009).
55. Newman, M. E. J. Analysis of weighted networks. *Phys. Rev. E* **70**, 056131 (2004).
56. Ahnert, S. E., Garlaschelli, D., Fink, T. M. A. & Caldarelli, G. Ensemble approach to the analysis of weighted networks. *Phys. Rev. E* **76**, 016101 (2007).
57. Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K. & Kertész, J. Generalizations of the clustering coefficient to weighted complex networks. *Phys. Rev. E* **75**, 027105 (2007).
58. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
59. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of escherichia coli. *Nat. Genet.* **31**, 64 (2002).
60. Garlaschelli, D. & Loffredo, M. I. Patterns of link reciprocity in directed networks. *Phys. Rev. Lett.* **93**, 268701 (2004).
61. Garlaschelli, D. & Loffredo, M. I. Multispecies grand-canonical models for networks with reciprocity. *Phys. Rev. E* **73**, 015101 (2006).
62. Squartini, T. & Garlaschelli, D. in *Self-Organizing Systems* (eds Kuipers, F. A. & Heegaard, P. E.) 24–35 (Springer Berlin, Heidelberg, 2012).
63. Stouer, D. B., Camacho, J., Jiang, W. & Amaral, L. A. N. Evidence for the existence of a robust pattern of prey selection in food webs. *Proc. R. Soc. Lond. B Biol. Sci.* **274**, 1931–1940 (2007).
64. Squartini, T., van Lelyveld, I. & Garlaschelli, D. Early-warning signals of topological collapse in interbank networks. *Sci. Rep.***3**, 3357 (2013).
65. Guimerà, R., Sales-Pardo, M. & Amaral, L. A. N. Modularity from uctuations in random graphs and complex networks. *Phys. Rev. E* **70**, 025101 (2004).
66. Reichardt, J. & Bornholdt, S. Partitioning and modularity of graphs with arbitrary degree distribution. *Phys. Rev. E* **76**, 015102 (2007).
67. Chung, F. & Lu, L. Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **6**, 125–145 (2002).
   This paper defines a very popular analytic model of networks with given degree sequence, admitting self-loops and multilinks.
68. Bargigli, L. & Gallegati, M. Random digraphs with given expected degree sequences: a model for economic networks. *J. Econ. Behav. Organ.* **78**, 396–411 (2011).
69. Fronczak, P., Fronczak, A. & Bujok, M. Exponential random graph models for networks with community structure. *Phys. Rev. E* **88**, 32810 (2013).
70. Lancichinetti, A., Fortunato, S. & Radicchi, F. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* **78**, 046110 (2008).
71. Karrer, B. & Newman, M. E. J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
72. Peixoto, T. P. Entropy of stochastic blockmodel ensembles. *Phys. Rev. E* **85**, 056122 (2012).

73. Holme, P., Liljeros, F., Edling, C. R. & Kim, B. J. Network bipartivity. *Phys. Rev. E* **68**, 056107 (2003).

74. Saracco, F., Di Clemente, R., Gabrielli, A. & Squartini, T. Randomizing bipartite networks: the case of the world trade web. *Sci. Rep.* **5**, 10595 (2015).

75. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Sci. Rep.* **2**, 723 (2012).

76. Caldarelli, G. et al. A network analysis of countries' export flows: firm grounds for the building blocks of the economy. *PLoS ONE* **7**, e47278 (2012).

77. Saracco, F., Di Clemente, R., Gabrielli, A. & Squartini, T. Detecting early signs of the 2007–2008 crisis in the world trade. *Sci. Rep.* **6**, 30286 (2016).

78. Payrató Borràs, C., Hernández, L. & Moreno, Y. Breaking the spell of nestedness. Preprint at https://arxiv.org/abs/1711.03134 (2017).

79. Zhou, T., Ren, J., Medo, M. & Zhang, Y.-C. Bipartite network projection and personal recommendation. *Phys. Rev. E* **76**, 046115 (2007).

80. Tumminello, M., Aste, T., Di Matteo, T. & Mantegna, R. N. A tool for filtering information in complex systems. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10421–10426 (2005).

81. Serrano, M. Á., Boguñá, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6483–6488 (2009).

82. Slater, P. B. A two-stage algorithm for extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, E66 (2009).

83. Radicchi, F., Ramasco, J. J. & Fortunato, S. Information filtering in complex weighted networks. *Phys. Rev. E* **83**, 046101 (2011).

84. Goldberg, D. S. & Roth, F. P. Assessing experimentally derived interactions in a small world. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 4372–4376 (2003).

85. Latapy, M., Magnien, C. & Vecchio, N. D. Basic notions for the analysis of large two-mode networks. *Soc. Networks* **30**, 31–48 (2008).

86. Tumminello, M., Micciché, S., Lillo, F., Piilo, J. & Mantegna, R. N. Statistically validated networks in bipartite complex systems. *PLoS ONE* **6**, e17994 (2011).

87. Tumminello, M., Lillo, F., Piilo, J. & Mantegna, R. N. Identification of clusters of investors from their real trading activity in a financial market. *New J. Phys.* **14**, 013041 (2012).

88. Neal, Z. Identifying statistically significant edges in one-mode projections. *Soc. Netw. Anal. Min.* **3**, 915–924 (2013).

89. Zweig, K. A. & Kaufmann, M. A systematic approach to the one-mode projection of bipartite graphs. *Soc. Netw. Anal. Min.* **1**, 187–218 (2011).

90. Horvát, E.-Á. & Zweig, K. A. A fixed degree sequence model for the one-mode projection of multiplex bipartite graphs. *Soc. Netw. Anal. Min.* **3**, 1209–1224 (2013).

91. Gionis, A., & Mannila, H., & Mielikäinen, T. & Tsaparas, P. Assessing data mining results via swap randomization. *ACM Trans. Knowl. Discov. Data* **1**, 14 (2007).

92. Neal, Z. The backbone of bipartite projections: inferring relationships from co-authorship, cosponsorship, co-attendance and other co-behaviors. *Soc. Networks* **39**, 84–97 (2014).

93. Gualdi, S., Cimini, G., Primicerio, K., Di Clemente, R. & Challet, D. Statistically validated network of portfolio overlaps and systemic risk. *Sci. Rep.* **6**, 39467 (2016).

94. Saracco, F. et al. Inferring monopartite projections of bipartite networks: an entropy-based approach. *New J. Phys.* **19**, 053022 (2017).

95. Straka, M. J., Caldarelli, G. & Saracco, F. Grand canonical validation of the bipartite international trade network. *Phys. Rev. E* **96**, 022306 (2017).

96. Pugliese, E. et al. Unfolding the innovation system for the development of countries: co-evolution of science, technology and production. Preprint at https://arxiv.org/abs/1707.05146 (2017).

97. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).

98. Wells, S. J. Financial interlinkages in the United Kingdom's interbank market and the risk of contagion. *Bank of England Working Paper* https://doi.org/10.2139/ssrn.641288 (2004).

99. Upper, C. Simulation methods to assess the danger of contagion in interbank markets. *J. Financ. Stab.* **7**, 111–125 (2011).

100. Anand, K. et al. The missing links: a global study on uncovering financial network structures from partial data. *J. Financ. Stab.* **35**, 107–119 (2018).

101. Kossinets, G. Effects of missing data in social networks. *Soc. Networks* **28**, 247–268 (2006).

102. Lynch, C. How do your data grow? *Nature* **455**, 28 (2008).

103. Amaral, L. A. N. A truer measure of our ignorance. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6795–6796 (2008).

104. Guimerà, R. & Sales-Pardo, M. Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22073–22078 (2009).

105. Lu, L. & Zhou, T. Link prediction in complex networks: a survey. *Phys. A Stat. Mech. Appl.* **390**, 1150–1170 (2011).

106. Squartini, T., Caldarelli, G., Cimini, G., Gabrielli, A. & Garlaschelli, D. Reconstruction methods for networks: the case of economic and financial systems. *Phys. Rep.* **757**, 1–47 (2018).

107. Boguñá, M. & Pastor-Satorras, R. Class of correlated random networks with hidden variables. *Phys. Rev. E* **68**, 036112 (2003).

108. Garlaschelli, D., Battiston, S., Castri, M., Servedio, V. D. P. & Caldarelli, G. The scale-free topology of market investments. *Phys. A Stat. Mech. Appl.* **350**, 491–499 (2005).

109. De Masi, G., Iori, G. & Caldarelli, G. Fitness model for the italian interbank money market. *Phys. Rev. E* **74**, 066112 (2006).

110. Musmeci, N., Battiston, S., Caldarelli, G., Puliga, M. & Gabrielli, A. Bootstrapping topological properties and systemic risk of complex networks using the fitness model. *J. Stat. Phys.* **151**, 1–15 (2013).

111. Cimini, G., Squartini, T., Gabrielli, A. & Garlaschelli, D. Estimating topological properties of weighted networks from limited information. *Phys. Rev. E* **92**, 040802 (2015).

112. Cimini, G., Squartini, T., Garlaschelli, D. & Gabrielli, A. Systemic risk analysis on reconstructed economic and financial networks. *Sci. Rep.* **5**, 15758 (2015).
This paper uses ERGs in combination with the fitness model to reconstruct networks from partial information.

113. Squartini, T., Cimini, G., Gabrielli, A. & Garlaschelli, D. Network reconstruction via density sampling. *Appl. Netw. Sci.* **2**, 3 (2017).

114. Squartini, T. et al. Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks. *Phys. Rev. E* **96**, 032315 (2017).

115. Berg, J. & Lässig, M. Correlated random networks. *Phys. Rev. Lett.* **89**, 228701 (2002).

116. Park, M. E. J. & Newman, J. Solution of the two-star model of a network. *Phys. Rev. E* **70**, 066146 (2004).

117. Yin, M. & Zhu, L. Reciprocity in directed networks. *Phys. A Stat. Mech. Appl.* **447**, 71–84 (2016).

118. Park, J. & Newman, M. E. J. Solution for the properties of a clustered network. *Phys. Rev. E* **72**, 026136 (2005).

119. Fronczak, P., Fronczak, A. & Holyst, J. A. Phase transitions in social networks. *Eur. Phys. J. B* **59**, 133–139 (2007).

120. Bianconi, G., Coolen, A. C. C. & Perez Vicente, C. J. Entropies of complex networks with hierarchically constrained topologies. *Phys. Rev. E* **78**, 016114 (2008).

121. Bianconi, G. Entropy of network ensembles. *Phys. Rev. E* **79**, 036114 (2009).

122. Mondragón, R. J. Network null-model based on maximal entropy and the rich-club. *J. Complex Netw.* **2**, 288–298 (2014).

123. Annibale, A., Coolen, A. C. C., Fernandes, L. P., Fraternali, F. & Kleinjung, J. Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure. *J. Phys. A Math. Theor.* **42**, 485001 (2009).

124. Roberts, E. S., Schlitt, T. & Coolen, A. C. C. Tailored graph ensembles as proxies or null models for real networks II: results on directed graphs. *J. Phys. A Math. Theor.* **44**, 275002 (2011).

125. Roberts, E. S. & Coolen, A. C. C. Entropies of tailored random graph ensembles: bipartite graphs, generalized degrees, and node neighbourhoods. *J. Phys. A Math. Theor.* **47**, 435101 (2014).

126. Artzy-Randrup, Y. & Stone, L. Generating uniformly distributed random networks. *Phys. Rev. E* **72**, 056708 (2005).

127. Coolen, A. C. C., De Martino, A. & Annibale, A. Constrained markovian dynamics of random graphs. *J. Stat. Phys.* **136**, 1035–1067 (2009).
This paper introduces Monte Carlo processes for uniform sampling of network ensembles.

128. Roberts, E. S. & Coolen, A. C. C. Unbiased degree-preserving randomization of directed binary networks. *Phys. Rev. E* **85**, 046103 (2012).

129. Strauss, D. & Ikeda, M. Pseudolikelihood estimation for social networks. *J. Am. Stat. Assoc.* **85**, 204–212 (1990).

130. van Duijn, M. A. J., Gile, K. J. & Handcock, M. S. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Soc. Networks* **31**, 52–62 (2009).

131. Snijders, T. A. B., Koskinen, J. & Schweinberger, M. Maximum likelihood estimation for social network dynamics. *Ann. Appl. Stat.* **4**, 567–588 (2010).

132. Schweinberger, M. Instability, sensitivity, and degeneracy of discrete exponential families. *J. Am. Stat. Assoc.* **106**, 1361–1370 (2011).

133. Desmarais, B. A. & Cranmer, S. J. Statistical mechanics of networks: estimation and uncertainty. *Phys. A Stat. Mech. Appl.* **391**, 1865–1876 (2012).

134. Chatterjee, S. & Diaconis, P. Estimating and understanding exponential random graph models. *Ann. Stat.* **41**, 2428–2461 (2013).

135. Horvát, S., Czabarka, É. & Toroczkai, Z. Reducing degeneracy in maximum entropy models of networks. *Phys. Rev. Lett.* **114**, 158701 (2015).

136. Hastings, W. K. Monte carlo sampling methods using markov chains and their applications. *Biometrika* **57**, 97–109 (1970).

137. Mahadevan, P., Krioukov, D., Fall, K. & Vahdat, A. Systematic topology analysis and generation using degree correlations. *SIGCOMM Comput. Commun. Rev.* **36**, 135–146 (2006).

138. Orsini, C. et al. Quantifying randomness in real networks. *Nat. Commun.* **6**, 8627 (2015).
This paper uses the *dk*-series approach to show that degree distributions, degree correlations and clustering often represent sufficient statistics to describe a network.

139. Foster, D., Foster, J., Paczuski, M. & Grassberger, P. Communities, clustering phase transitions, and hysteresis: pitfalls in constructing network ensembles. *Phys. Rev. E* **81**, 046115 (2010).

140. Fischer, R., Leitão, J. C., Peixoto, T. P. & Altmann, E. G. Sampling motif-constrained ensembles of networks. *Phys. Rev. Lett.* **115**, 188701 (2015).

141. Fugao Wang & Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**, 2050–2053 (2001).

142. Kivelä, M. et al. Multilayer networks. *J. Complex Netw.* **2**, 203–271 (2014).

143. Boccaletti, S. et al. The structure and dynamics of multilayer networks. *Phys. Rep.* **544**, 1–122 (2014).

144. De Domenico, M., Granell, C., Porter, M. A. & Arenas, A. The physics of spreading processes in multilayer networks. *Nat. Phys.* **12**, 901–906 (2016).

145. Bianconi, G. Statistical mechanics of multiplex networks: entropy and overlap. *Phys. Rev. E* **87**, 062806 (2013).
This paper develops the ERG framework for multiplex networks.

146. Gemmetto, V. & Garlaschelli, D. Multiplexity versus correlation: the role of local constraints in real multiplexes. *Sci. Rep.* **5**, 9120 (2015).

147. Menichetti, G., Remondini, D., Panzarasa, P., Mondragón, R. J. & Bianconi, G. Weighted multiplex networks. *PLoS ONE* **9**, e97857 (2014).

148. Menichetti, G., Remondini, D. & Bianconi, G. Correlations between weights and overlap in ensembles of weighted multiplex networks. *Phys. Rev. E* **90**, 062817 (2014).

149. Sagarra, O., Pérez Vicente, C. J. & Díaz-Guilera, A. Statistical mechanics of multiedge networks. *Phys. Rev. E* **88**, 062806 (2013).

150. Sagarra, O., Font-Clos, F., Péerez-Vicente, C. J. & Díaz-Guilera, A. The configuration multiedge model: assessing the effect of fixing node strengths on weighted network magnitudes. *Europhys. Lett.* **107**, 38002 (2014).

151. Sagarra, O., Pérez Vicente, C. J. & Díaz-Guilera, A. Role of adjacency-matrix degeneracy in maximum-entropy-weighted network models. *Phys. Rev. E* **92**, 052816 (2015).

152. Mastrandrea, R., Squartini, T., Fagiolo, G. & Garlaschelli, D. Reconstructing the world trade multiplex: the role of intensive and extensive biases. *Phys. Rev. E* **90**, 062804 (2014).

153. Zuev, K., Eisenberg, O. & Krioukov, D. Exponential random simplicial complexes. *J. Phys. A Math. Theor.* **48**, 465002 (2015).

154. Courtney, O. T. & Bianconi, G. Generalized network structures: the configuration model and the canonical ensemble of simplicial complexes. *Phys. Rev. E* **93**, 062311 (2016).

155. Young, J.-G., Petri, G., Vaccarino, F. & Patania, A. Construction of and efficient sampling from the simplicial configuration model. *Phys. Rev. E* **96**, 032312 (2017).
156. Dixit, P. D. et al. Perspective: maximum caliber is a general variational principle for dynamical systems. *J. Chem. Phys.* **148**, 010901 (2018).
157. Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118 (2001).
158. Itzkovitz, S., Milo, R., Kashtan, N., Newman, M. E. J. & Alon, U. Reply to comment on 'subgraphs in random networks'. *Phys. Rev. E* **70**, 058102 (2004).
159. Catanzaro, M., Boguñá, M. & Pastor-Satorras, R. Generation of uncorrelated random scalefree networks. *Phys. Rev. E* **71**, 027103 (2005).
160. Zamora-Lopez, G., Zlatic, V., Zhou, C., Stefancic, H. & Kurths, J. Reciprocity of networks with degree correlations and arbitrary degree sequences. *Phys. Rev. E* **77**, 016106 (2008).
161. Zlatic, V. et al. On the rich-club effect in dense and weighted networks. *Eur. Phys. J. B* **67**, 271–275 (2009).
162. Tabourier, L., Roth, C. & Cointet, J.-P. Generating constrained random graphs using multiple edge switches. *J. Exp. Algorithm.* **16**, 1.1–1.15 (2011).
163. Carstens, C. J. & Horadam, K. J. Switching edges to randomize networks: what goes wrong and how to fix it. *J. Complex Netw.* **5**, 337–351 (2017).
164. Del Genio, C. I., Kim, H., Toroczkai, Z. & Bassler, K. E. Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PLoS ONE* **5**, e10012 (2010).
165. Blitzstein, J. & Diaconis, P. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6**, 489–522 (2011).
166. Kim, H., Del Genio, C. I., Bassler, K. E. & Toroczkai, Z. Constructing and sampling directed graphs with given degree sequences. *New J. Phys.* **14**, 023012 (2012).
167. Newman, M. E. J. Random graphs with clustering. *Phys. Rev. Lett.* **103**, 058701 (2009).
168. Melnik, S., Hackett, A., Porter, M. A., Mucha, P. J. & Gleeson, J. P. The unreasonable effectiveness of tree-based theory for networks with clustering. *Phys. Rev. E* **83**, 036112 (2011).
169. Burda, Z. & Krzywicki, A. Uncorrelated random networks. *Phys. Rev. E* **67**, 046118 (2003).
170. Boguñá, M., Pastor-Satorras, R. & Vespignani, A. Cut-offs and finite size effects in scale-free networks. *Eur. Phys. J. B* **38**, 205–209 (2004).
171. Neyman, J. & Pearson, E. S. On the problem of the most efficient tests of statistical hypotheses. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* **231**, 289–337 (1933).
172. Burnham, K. P. & Anderson, D. R. (eds) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer-Verlag, New York, 2002).
173. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
174. Wagenmakers, E.-J. & Farrell, S. Aic model selection using akaike weights. *Psychon. Bull. Rev.* **11**, 192–196 (2004).
175. Burnham, K. P. & Anderson, D. R. Multimodel inference: understanding aic and bic in model selection. *Sociol. Methods Res.* **33**, 261–304 (2004).
176. Braunstein, S. L., Ghosh, S. & Severini, S. The laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Ann. Comb.* **10**, 291–317 (2006).
177. Anand, K., Bianconi, G. & Severini, S. Shannon and von neumann entropy of random networks with heterogeneous expected degree. *Phys. Rev. E* **83**, 036109 (2011).
178. Anand, K., Krioukov, D. & Bianconi, G. Entropy distribution and condensation in random networks with a given degree distribution. *Phys. Rev. E* **89**, 062807 (2014).
179. De Domenico, M. & Biamonte, J. Spectral entropies as information-theoretic tools for complex network comparison. *Phys. Rev. X* **6**, 041062 (2016).
180. Delvenne, J.-C., Lambiotte, R. & Rocha, L. E. C. Diffusion on networked systems is a question of time or structure. *Nat. Commun.* **6**, 7366 (2015).
181. Masuda, N., Porter, M. A. & Lambiotte, R. Random walks and diusion on networks. *Phys. Rep.* **716-717**, 1–58 (2017).
182. Demetrius, L. & Manke, T. Robustness and network evolution-an entropic principle. *Phys. A Stat. Mech. Appl.* **346**, 682–696 (2005).
183. Lott, J. & Villani, C. Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.* **169**, 903–991 (2009).
184. Sandhu, R. et al. Graph curvature for differentiating cancer networks. *Sci. Rep.* **5**, 12323 (2015).
185. Sandhu, R. S., Georgiou, T. T. & Tannenbaum, A. R. Ricci curvature: an economic indicator for market fragility and systemic risk. *Sci. Adv.* **2**, e1501495 (2016).