

# Reconstruction methods for networks: the case of economic and financial systems

Tiziano Squartini<sup>a</sup>, Guido Caldarelli<sup>a,b,c,\*</sup>, Giulio Cimini<sup>a,b</sup>, Andrea Gabrielli<sup>b,a</sup>, Diego Garlaschelli<sup>a,d</sup>

<sup>a</sup>*IMT School for Advanced Studies Lucca, P.zza San Francesco 19, 55100 Lucca (Italy)*

<sup>b</sup>*Istituto dei Sistemi Complessi (ISC) - CNR, UoS Sapienza, Dipartimento di Fisica, Università "Sapienza", P.le Aldo Moro 5, 00185 Rome (Italy)*

<sup>c</sup>*European Centre for Living Technology (ECLT) San Marco 2940, 30124 Venezia, ITALY*

<sup>d</sup>*Lorentz Institute for Theoretical Physics, Leiden Institute of Physics, University of Leiden, Niels Bohrweg 2, 2333 CA Leiden (The Netherlands)*

---

## Abstract

The study of social, economic and biological systems is often (when not always) limited by the partial information about the structure of the underlying networks. An example of paramount importance is provided by financial systems: information on the interconnections between financial institutions is privacy-protected, dramatically reducing the possibility of correctly estimating crucial systemic properties such as the resilience to the propagation of shocks. The need to compensate for the scarcity of data, while optimally employing the available information, has led to the birth of a research field known as *network reconstruction*. Since the latter has benefited from the contribution of researchers working in disciplines as different as mathematics, physics and economics, the results achieved so far are still scattered across heterogeneous publications. Most importantly, a systematic comparison of the network reconstruction methods proposed up to now is currently missing. This review aims at providing a unifying framework to present all these studies, mainly focusing on their application to economic and financial networks.

*Keywords:* entropy maximization, network reconstruction, statistical inference, economic and financial networks, systemic risk evaluation

*PACS:* 02.50.Tt, 89.65.Gh, 89.70.Cf, 89.75.Hc

---

---

\*Corresponding author: [guido.caldarelli@imtlucca.it](mailto:guido.caldarelli@imtlucca.it)

*Email addresses:* [tiziano.squartini@imtlucca.it](mailto:tiziano.squartini@imtlucca.it) (Tiziano Squartini), [giulio.cimini@imtlucca.it](mailto:giulio.cimini@imtlucca.it) (Giulio Cimini), [andrea.gabrielli@roma1.infn.it](mailto:andrea.gabrielli@roma1.infn.it) (Andrea Gabrielli), [garlaschelli@lorentz.leidenuniv.nl](mailto:garlaschelli@lorentz.leidenuniv.nl) (Diego Garlaschelli)

The study of truth requires a considerable effort  
which is why few are willing to undertake it out of love of knowledge

— Thomas Aquinas, *Summa Contra Gentiles*

## Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Information theory as a basis for network reconstruction</b>	<b>9</b>
2.1	Setting the problem: constraining Shannon entropy . . . . .	11
2.2	Exponential Random Graphs . . . . .	13
<b>3</b>	<b>Reconstruction methods</b>	<b>14</b>
3.1	Dense reconstruction methods . . . . .	16
3.1.1	The MaxEnt algorithm . . . . .	16
3.1.2	Overcoming the problem of missing connections: the IPF algorithm	16
3.1.3	The Directed Weighted Configuration Model . . . . .	17
3.1.4	Combining MaxEnt and ERG frameworks . . . . .	18
3.2	Density-tunable reconstruction methods . . . . .	19
3.2.1	The IPF algorithm: generic formulation . . . . .	19
3.2.2	The Drehmann & Tarashev approach . . . . .	19
3.2.3	The Mastromatteo, Zarinelli & Marsili approach . . . . .	21
3.2.4	The Moussa & Cont approach . . . . .	22
3.3	Exact density methods . . . . .	23
3.3.1	The density-corrected DWCM . . . . .	24
3.3.2	The Directed Enhanced Configuration Model . . . . .	24
3.3.3	Simplifying the DECM: a two-step model . . . . .	26
3.3.4	Fitness-induced Exponential Random Graphs . . . . .	26
3.3.5	Combining fitness-induced DBCM and IPF . . . . .	28
3.3.6	Combining fitness-induced DBCM and DECM . . . . .	28
3.3.7	The degree-corrected gravity model . . . . .	28
3.3.8	Reconciling ERG and gravity models . . . . .	31
3.3.9	A remark on the ensemble methods . . . . .	32
3.4	Shannon-like approaches to reconstruction . . . . .	32
3.4.1	Spectral entropy . . . . .	32
3.4.2	The Cressie-Read family of power divergences . . . . .	33
3.4.3	Other entropic families . . . . .	35
3.5	Beyond Shannon entropy: alternative approaches to reconstruction . . . .	36
3.5.1	The “copula” approach . . . . .	36
3.5.2	A Bayesian approach to network reconstruction . . . . .	37
3.5.3	A comment on the Bayesian approaches to reconstruction . . . . .	39
3.5.4	The Montagna & Lux approach . . . . .	40
3.5.5	The probability map of Hałaj & Kok . . . . .	42
3.5.6	The Minimum Density algorithm . . . . .	42
<b>4</b>	<b>Testing the network reconstruction</b>	<b>44</b>
4.1	Statistical indicators . . . . .	44
4.2	Topological indicators . . . . .	49
4.2.1	Testing weights reconstruction . . . . .	49
4.2.2	Testing higher-order patterns reconstruction . . . . .	49
4.2.3	What may and what may not be reconstructed . . . . .	51

4.3	Dynamical indicators . . . . .	54
4.3.1	Balance sheets and financial networks . . . . .	54
4.3.2	Counterparty risk and credit shocks . . . . .	55
4.3.3	Rollover risk and liquidity shocks . . . . .	56
4.3.4	Overlapping portfolios and fire-sales spillovers . . . . .	57
<b>5</b>	<b>Model selection criteria</b>	<b>58</b>
5.1	The Likelihood Ratio Test . . . . .	59
5.2	The Akaike Information Criterion . . . . .	60
5.3	The Bayesian Information Criterion . . . . .	60
5.4	A quick look at multimodel averaging . . . . .	61
<b>6</b>	<b>Conclusions and perspectives</b>	<b>62</b>
6.1	Comparing different reconstruction algorithms on real-world networks . . .	62
6.2	Is link density really needed? . . . . .	65
6.3	Policy-making implications: the case of the OTC market . . . . .	66
	<b>Appendix A. A combinatorial derivation of Shannon entropy</b>	<b>67</b>
	<b>Appendix B. Sketching a principled derivation of Shannon entropy</b>	<b>68</b>
	<b>Appendix C. Two relevant properties of Shannon entropy</b>	<b>69</b>
	<b>Appendix D. A notable, continuous case</b>	<b>70</b>

### List of the symbols employed in the review

Throughout this work we shall employ top-hatted letters (*e.g.*,  $\hat{s}_i^{\text{out}}$ ,  $\hat{s}_i^{\text{in}}$ ,  $\hat{w}_{ij}$ , etc.) for the measured values of the empirical networks to be reconstructed, while we shall use the same letters (*e.g.*,  $s_i^{\text{out}}$ ,  $s_i^{\text{in}}$ ,  $w_{ij}$ , etc.) without any addition to indicate the same quantities when considered as (deterministic or stochastic) variables. In particular:

- $\hat{\mathbf{G}}$ : the (observed) network to reconstruct;
- $\mathcal{G}$ : ensemble of network configurations used to reconstruct  $\hat{\mathbf{G}}$ ;
- $N$ : (fixed) number of nodes of  $\hat{\mathbf{G}}$  (and of all networks in the ensemble  $\mathcal{G}$ );
- $\mathbf{G}$ : a generic network configuration belonging to  $\mathcal{G}$ ;
- $P(\mathbf{G})$ : probability measure on the ensemble of configurations  $\mathcal{G}$ ;
- $I(\mathbf{G})$ : information content of the configuration  $\mathbf{G}$ ;
- $H(\mathbf{G})$ : network Hamiltonian of the configuration  $\mathbf{G}$ , *i.e.* linear combination of the constraints determining its probability  $P(\mathbf{G})$ ;
- $S$ : Shannon entropy;
- $\mathcal{L}$ : Lagrangian functional, *i.e.*, the constrained Shannon entropy;
- $M$ : number of constraints (excluding normalization) in  $\mathcal{L}$ ;
- $C_m$ :  $m$ -th constraint;
- $\lambda_m$ :  $m$ -th Lagrange multiplier, controlling for the  $m$ -th constraint;
- $\hat{\lambda}_m$ : estimation of the  $m$ -th Lagrange multiplier for  $\hat{\mathbf{G}}$ ;
- $\mathcal{L}(\mathbf{G}|\vec{\lambda})$ : log-likelihood of  $\mathbf{G}$  given the Lagrange multipliers defining  $P(\mathbf{G})$ ;
- $\langle X \rangle$ : expected value over the ensemble  $\mathcal{G}$  of a generic quantity of interest  $X$ ;
- $\sigma_X$ : ensemble standard deviation of a generic quantity of interest  $X$ ;
- $X^{\text{name}}$ : estimate of quantity  $X$  by the algorithm “name”;
- $L$ : number of links in a network;
- $\rho = L/[N(N-1)]$ : density of links in a (directed) network, defined as the fraction of possible connections in the network that are actually realized;
- $\mathbf{W}$ : weighted adjacency matrix representing a network, with generic element  $w_{ij}$  denoting the weight of the link from node  $i$  to node  $j$ ;
- $\mathbf{A}$ : adjacency matrix representing a binary version of  $\mathbf{W}$ , with generic element  $a_{ij} = 1$  if  $w_{ij} > 0$  (and 0 otherwise);

- $s_i^{\text{out}} = \sum_{j=1}^N w_{ij}$  and  $s_i^{\text{in}} = \sum_{j=1}^N w_{ji}$ : out-strength and in-strength of node  $i$ , or equivalently the *marginal* row and column sums of  $\mathbf{W}$ ;
- $k_i^{\text{out}} = \sum_{j=1}^N a_{ij}$  and  $k_i^{\text{in}} = \sum_{j=1}^N a_{ji}$ : out-degree and in-degree of node  $i$ , or equivalently the *marginal* row and column sums of  $\mathbf{A}$ ;

## 1. Introduction

*Networks: the why and how of a theory.* There is nowadays an overwhelming evidence that a large deal of complex systems around us can be successfully described by means of complex networks [1, 2, 3]. Graph theory, from which complex networks theory originates, was firstly developed in the XVIII century as an application of discrete mathematics to the well-known “Königsberg bridge problem” [4] and for many years it remained relegated to merely solving puzzling topological problems [5].

A new burst of activity was registered in 1920, to provide mathematical support to the analysis of social networks [6, 7]. This laid the foundation for “sociometry”, a discipline characterized by a mathematical description of social sciences<sup>1</sup>. Later on, the seminal papers of Erdős and Rényi [9, 10], lately extended to combinatorics by Bollobás [11], opened the field of “random graph theory” [12].

Only after the digital revolution, complex networks became part of everybody’s life. Indeed current technology has made the presence of personal computers pervasive and also constantly reduced the cost of backup memories. These two features have produced immense databases, collecting information about a wide range of relationships. Just to name a few we have: the Internet wiring [13, 14, 15], the set of WWW connections [16, 17], e-mail exchanges [18, 19] [20], mobile communication networks [21, 22, 23], online social networks [24, 25, 26, 27], protein-protein interactions [28, 29, 30], food-webs and ecological networks [31, 32, 33], production activities [34, 35, 36, 37, 38, 39], financial exchanges and stock investments [40, 41, 42].

The network representation clearly highlights qualitative universal behaviors, irrespectively from the specific case-studies [43]. In what follows we name some of the features shared by many real-world networks:

- a *long-tail* (or even *scale-free*) *degree distribution* [13, 44];
- a *small-world effect* [45]: distances are distributed around a characteristic average value, usually “very small”, (i.e. 3 – 10) and scaling as the logarithm of the system size;
- a *large clustering* [46]: we spot the presence of densely-connected subgraphs in many complex networks. The small-world and the large clustering effects co-exist thanks to presence of the so-called “weak-ties” in the social networks literature, allowing for “long-range” interconnections without affecting the locally large density of links;
- a distinct *centrality* structure [47] implying that some nodes appear to have a higher importance than others;
- a well-defined *assortativity structure* [48]: the neighbors of each node have a degree that is either positively or negatively correlated to the degree of the node itself (more intuitively, “my” neighbors have a degree that is either very similar or very different from mine). In the case of bipartite networks, a well-defined *nestedness structure* has been observed, mainly in ecological and economic contexts [49, 50, 51].

---

<sup>1</sup>A similar approach coupled to modern technology constitutes the core of a newborn field, “computational social science” [8], where mathematics is employed to analyze the huge amount of data generated by online social networks.

*The problem of missing information.* After an initial activity aimed at determining the structure of real-world networks by measuring standard topological quantities, a more theoretical activity was started, aiming at both defining new quantities and devising proper models to explain observations [45, 44, 52, 53, 54]. Given the complexity that can arise even from a simple mathematical model based upon graphs, researchers have recently focused on the development of a *topological* theory: loosely speaking, topological quantities are employed to define statistical models, rather than reproduced from microscopic dynamical rules [55, 56, 57, 58, 59].

Unfortunately, when moving to the validation of such models a common problem arises: very often, the data available on the real network are either incomplete or imprecise (or both). This problem is particularly evident in the case of economic and financial networks: in this case, data collection suffers from the problem of partial accounting and the presence of disclosure requirements. In order to illustrate the importance of such an issue, let us think of a bipartite, financial network whose node sets represent investors and the investments they do. Although the knowledge of the whole network structure could help regulators to take immediate countermeasures to stop the propagation of financial distress, this information is seldom available (the knowledge of the whole network of investments would pose immense problems of privacy), thus hindering the possibility of providing a realistic estimate of the extent of the contagion. As confirmed by the analysis of the various papers reported in this review, the incompleteness of network instances seems to be unavoidable [60, 61]: since addressing the problem of estimating the resilience of financial networks cannot be addressed without knowing the structural details of national and cross-countries interbank networks, information theory seems indeed to provide the right framework to tackle this kind of problems.

Finance is not the only domain affected by limitedness of information about nodes interdependencies: biological and ecological systems also exist (*e.g.*, cell metabolic networks and ecological webs) whose interaction network is often only partially accessible due either to experimental limitations or observational constraints<sup>2</sup>.

*Approaching network reconstruction.* In order to deal with the problem of missing information, many different approaches have been attempted so far. Some reconstruction procedures are based on a measured (or expected) statistical self-similarity of the network topology [64]. In this case, the observed behavior of a given topological property (*e.g.*, the degree distribution) is supposed to be induced by some non-topological property assigned to nodes (*e.g.*, a node “fitness” [52]) obeying the same behavior.

More often, the fundamental assumption grounding network reconstruction is *statistical homogeneity*. This means that the structure of the network observed are representative of statistical properties not depending on that specific portion. In the jargon of statistics, supposing that from similar observations we can infer similar regularities is equivalent to requiring that the information available is representative of the whole network structure. This is particularly relevant when only limited information is available, and constitutes the physical reason to be confident that the efforts to define a statistically-grounded network reconstruction algorithm can indeed be successful. By using this “symmetry” it is then possible to provide some likely estimates of the missing

---

<sup>2</sup>In these cases, one is often more interested in the reconstruction of individual links from partial local information, a problem known as “link prediction” [62, 63].

quantities of the network under analysis. Clearly, while the homogeneity assumption minimizes the bias introduced by adopting arbitrary assumptions not supported by the available information (and, in principle, untestable), it also limits the accuracy of the reconstruction of real networks showing, instead, strong structural heterogeneity.

Given these premises, entropy maximization provides the unifying concept underlying all the reviewed methods. Entropy maximization is, in fact, an ubiquitous prescription for obtaining the least biased probability distribution consistent with some imposed constraints (*i.e.*, a probability distribution not encoding other information than that represented by the constraints themselves [65, 66]). This principle has not only found hundreds of applications in statistical mechanics, information theory and statistics [67], but it has been also argued to represent an evolutionary drive of out-of-equilibrium systems [68] (*e.g.*, a relationship has been suggested between the dynamics of intelligent systems and entropy maximization [69]).

The outline of this report is the following. In section 2, *Information theory as a basis for network reconstruction*, we present the tools that can be derived from classical approaches in statistical physics and information theory, mainly Gibbs' ensembles theory, Shannon's works on entropy and Jaynes' interpretation of statistical mechanics. Moving from these theoretical premises, in section 3, *Reconstruction methods*, we present an overview of the different reconstruction methods, dividing them into **a)** dense reconstruction methods, **b)** density-tunable reconstruction methods, **c)** exact-density methods, **d)** alternative approaches. In section 4, *Testing the network reconstruction*, we present in detail a number of indicators and metrics that can be used to test the accuracy of the achieved reconstruction; more specifically, we have distinguished three classes of indicators, of *statistical*, *topological* and *dynamical* nature, respectively aiming at testing the accuracy in reconstructing the microscopic details, the macroscopic topological features and the dynamical properties of a given network. Concerning dynamical indicators, we put particular emphasis on the possibility of estimating the resilience of the system to processes of shocks propagation. In the financial context this is known as systemic risk, *i.e.*, the likelihood that a consistent part of a given financial network may collapse (go bankrupt) as a consequence of a local failure. In section 5, *Model selection criteria*, we describe some of the existing criteria to compare different models and the corresponding recipes for how to choose the most appropriate one. The report ends with section 6, *Conclusions and perspectives*, where we describe future possible applications of the reviewed algorithms.

As a general remark, we would like to stress that almost every paper reviewed here has its own nomenclature for the (often similar) quantities of interest. Since we wanted to present the different contributions within a unified framework, our presentation might not reflect the original derivation of the results.

## 2. Information theory as a basis for network reconstruction

Information theory provides the theoretical basis of our formalism [70]. The concept of *information* plays a fundamental role in network reconstruction, since *reconstructing a network* ultimately means making optimal use of the available, partial, information. Otherwise stated, our task is that of inferring as much as possible about the system under analysis from the available data, while limiting the number of unsupported assumptions.

As stated before, real data are very often partial: thus, any data-driven inference procedure is bound to consider an enlarged set of plausible configurations, *i.e.*, *all configurations that are compatible with the available information*. In the language of statistical mechanics, this set is called *ensemble*. Enlarging the set of allowable configurations means, in turn, increasing the degree of uncertainty about the actual one; the description, thus, becomes necessarily *probabilistic*: a probability value must be assigned to each configuration compatible with the known information, that is, to all configurations belonging to the ensemble.

The degree of plausibility of a given configuration can be unambiguously quantified by recalling the concept of *surprise*. Since a “surprising” event (deemed as highly improbable) is assumed to convey a large amount of information [65]. An operative definition of surprise should encode a (negative) correlation with the probability of realization of the event under consideration. The content of information of a given outcome  $\mathbf{G}$  out of the set of possible outcomes  $\mathcal{G}$  can be thus quantified as

$$I(\mathbf{G}) = -\ln P(\mathbf{G}), \quad (1)$$

a definition pointing out that the occurrence of an event that is certain (*i.e.*, characterized by  $P(\mathbf{G}) = 1$ ) brings no information and comes with no surprise, whereas, the occurrence of an (almost) impossible event (*i.e.*, characterized by  $P(\mathbf{G}) \simeq 0$ ) conveys an (almost) infinite amount of information and causes an (almost) “infinite” amount of surprise [65]. The average degree of surprise which accompanies the events belonging to the set  $\mathcal{G}$  can be then quantified by averaging over the ensemble itself. Such an operation leads to the basic concept of the *Shannon entropy*:

$$S = \langle I \rangle = \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) I(\mathbf{G}) = \sum_{\mathbf{G} \in \mathcal{G}} -P(\mathbf{G}) \ln P(\mathbf{G}). \quad (2)$$

Another interpretation of eq. (2) comes from information theory [71, 72]. Given an alphabet of symbols (as a language) to be transmitted across a channel, shorter codes should be assigned to symbols met with larger frequency, while longer codes should be employed for symbols met with smaller frequency (see Appendix A). Looking for the average code length needed to transmit a given message leads to  $S$  as well.

From an axiomatic point of view, Shannon entropy is the *only* functional that satisfies a number of properties known as the Shannon-Khinchin axioms [66]:

1. Shannon entropy is a continuous functional of all its arguments: this ensures that small deformations  $\delta P(\mathbf{G})$  of the probability distribution  $P(\mathbf{G})$  induce small changes in  $S$ ;
2. Shannon entropy attains its maximum in correspondence of the uniform distribution over the set of possible configurations;
3. Shannon entropy is invariant under the addition of events with zero probability;
4. For a system composed by two independent subsystems  $A$  and  $B$ , whose ensembles of possible configurations  $\mathcal{G}_A$  and  $\mathcal{G}_B$  have probability measures  $W_A$  and  $W_B$ , the entropy is additive:  $S(W_{A+B}) = S(W_A W_B) = S(W_A) + S(W_B)$ , *i.e.* the entropy of the whole system is the sum of the entropies of the two subsystems<sup>3</sup>. On the

---

<sup>3</sup>In [73] this axiom is replaced by the so-called “composition law”.

other hand, if the two subsystems are not independent,  $S(W_{A+B}) = S(W_A) + S(W_{B|A})$ , with  $W_{B|A}$  indicating the probability measure for the configurations of the subsystem  $B$ , conditioned to the realization of subsystem  $A$ .

In other words, Shannon entropy is a functional of the probability distribution of an arbitrary set of random variables (in our case, the configurations  $\mathbf{G}$  within the aforementioned ensemble  $\mathcal{G}$ ) and quantifies the (un)evenness of the distribution itself [74]. As an example, if no information on the system is available, uncertainty about it is maximal and Shannon entropy prescribes to assign a uniform distribution over  $\mathcal{G}$ . By converse, any statistical information gained on the system reduces the uniform character of the probability distribution, which becomes progressively more peaked in correspondence of the configurations conveying the given information.

### 2.1. Setting the problem: constraining Shannon entropy

E. T. Jaynes first pointed out the possibility of using Shannon entropy to define a novel inference procedure [73], by extending the recipe proposed by Gibbs in the context of statistical mechanics. Jaynes proposed to carry out a constrained maximization of Shannon entropy, *i.e.*, to maximize the functional

$$\mathcal{L}[P] = S - \lambda_0 \left[ \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) - 1 \right] - \sum_{m=1}^M \lambda_m \left[ \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) C_m(\mathbf{G}) - \langle C_m \rangle \right], \quad (3)$$

with the  $M$  quantities  $\{C_m(\mathbf{G})\}_{m=1}^M$  that sum up the available knowledge on the system acting as constraints to be satisfied by the probability distribution  $P(\mathbf{G})$  itself. Maximizing Shannon entropy ensures that *our ignorance about the system to be reconstructed is maximized, except for what is known* or, equivalently, that the number of unjustified assumptions about the system itself is minimized. Indeed, it can be proven that the probability distribution that maximizes Shannon entropy is maximally non-committal with respect to the unknown information [73] (see Appendix B).

Since *constrained* Shannon entropy maximization represents a sort of “guessing” process about the unknown information, characterized by the least amount of statistical bias, it can be viewed as an updated version of the Laplace principle of insufficient reason [73]. The latter states that in absence of any information about the system under analysis, there is no reason to prefer any particular configuration which, thus, is equiprobable to any other. This principle is nothing else than a particular case of eq. (3) with no constraints but the normalization condition—which actually leads to the uniform distribution over the ensemble  $P(\mathbf{G}) = \frac{1}{|\mathcal{G}|}$ ,  $\forall \mathbf{G} \in \mathcal{G}$ . As already mentioned, Shannon entropy attains its maximum in this situation.

The framework sketched above is general enough to allow physical systems as well as networks to be analyzed. While in the first case the  $M$  constraints represent physical quantities (*e.g.*, the mean energy of the system), in the second case they represent purely topological quantities as nodes degrees, the network’s reciprocity, and so on. We would also like to stress that the Gibbs-Jaynes approach has been originally defined within the realm of equilibrium statistical mechanics, and this is also the spirit that has guided its application to networks. Generally speaking, however, networks do not satisfy the equilibrium conditions valid for thermodynamic systems. Consequently, in the networks realm the adoption of an entropy-maximization approach for the reconstruction of

higher-order statistical properties from lower-order constraints is rather justified by the minimization of arbitrary statistical assumptions on the network structure not supported by the available information.

Let us now explicitly consider  $\mathcal{G}$  as the set of all possible network configurations with  $N$  vertices. In most cases we consider networks that are both weighted and directed, which implies that the interlinkages characterizing them can be represented by an asymmetric  $N \times N$  matrix with real entries:

$$\mathbf{W} = \begin{pmatrix} w_{11} & \dots & w_{1i} & \dots & w_{1N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{i1} & \dots & w_{ii} & \dots & w_{iN} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{N1} & \dots & w_{Ni} & \dots & w_{NN} \end{pmatrix} \quad (4)$$

where  $w_{ij} \geq 0$  represents the weight of the link from node  $i$  to node  $j$ . Matrix  $\mathbf{W}$  induces a second matrix  $\mathbf{A}$ , known as the *adjacency matrix* of the network. Formally, the generic element  $a_{ij}$  of  $\mathbf{A}$  is equal to 1 if  $w_{ij} > 0$ , and 0 otherwise. In other words, the matrix  $\mathbf{A}$  simply indicates the presence/absence of connections between node pairs.

The problem of network reconstruction arises whenever the weights  $\hat{w}_{ij}$  of an empirical network  $\hat{\mathbf{W}}$  are not directly observable, and instead only aggregate (*marginal*) information on the network is accessible. More precisely, only the sum of the rows and/or the columns are typically known:

$$\begin{cases} \hat{s}_i^{\text{out}} = \sum_{j=1}^N \hat{w}_{ij} & \text{(out-strength)} \\ \hat{s}_i^{\text{in}} = \sum_{j=1}^N \hat{w}_{ji} & \text{(in-strength)} \end{cases} \quad \forall i. \quad (5)$$

Note that rows and column sums of  $\hat{\mathbf{A}}$ , namely the out-degree  $\hat{k}_i^{\text{out}} = \sum_{j=1}^N \hat{a}_{ij}$  and in-degree  $\hat{k}_i^{\text{in}} = \sum_{j=1}^N \hat{a}_{ji}$  of each node  $i$  are typically unknown.

In the financial context,  $\hat{\mathbf{W}}$  typically represents a matrix of interbank exposures, also named *liability matrix*. The entries  $\hat{w}_{ij}$  of this matrix are the loans and borrowings between banks, protected by privacy issues, while marginals are publicly released in balance sheets.  $\hat{s}_i^{\text{out}}$  then quantifies the total interbank *assets* of node  $i$ , and  $\hat{s}_i^{\text{in}}$  its total interbank *liabilities* (see section 4.3.1). Another classical example is the World Trade Network (WTN), where these quantities represent the total export and import of countries.

Generally speaking, any algorithm aimed at reconstructing a weighted directed network outputs two matrices,  $\mathbf{P} = \{p_{ij}\}_{i,j=1}^N$  and  $\mathbf{W} = \{w_{ij}\}_{i,j=1}^N$ : while the generic entry  $p_{ij}$  of the first matrix describes the probability that any two nodes  $i$  and  $j$  are connected, the generic entry of the second matrix provides an estimate of the weight  $w_{ij}$  of the corresponding link. We can say that, in certain conditions, the probabilities and weights estimates of the methods considered in the present review are functions of the accessible information and can, in general, be written as  $p_{ij}(\hat{s}_i^{\text{out}}, \hat{s}_j^{\text{in}})$  and  $w_{ij}(\hat{s}_i^{\text{out}}, \hat{s}_j^{\text{in}})$ . As a consequence, the entries  $a_{ij}$  can be interpreted as Bernoulli variables that are 1 with a certain probability  $p_{ij}$  and 0 otherwise.

Since the number of available data in the cases we consider here is  $O(N)$  ( $2N$  if only out- and in-strengths are known), the problem of reconstructing an adjacency matrix of  $N^2$  real numbers is under-determined. In what follows, we shall review methods that adopt a probabilistic approach to tackle these kinds of problems, making use of tools and concepts developed within the information-theoretic framework introduced in the previous subsection.

## 2.2. Exponential Random Graphs

Exponential Random Graphs (ERG) [55, 59] occupy a central role in most of network reconstruction algorithms. Indeed, ERG are defined as the ensemble of graphs whose probability  $P(\mathbf{G})$  is obtained by maximizing of the constrained entropy functional of eq. (3). More specifically, solving the functional differential equation  $\frac{\delta \mathcal{L}[P]}{\delta P(\mathbf{G})} = 0$  with respect to  $P(\mathbf{G})$  leads to the following formula:

$$P(\mathbf{G}|\vec{\lambda}) = e^{-1-\lambda_0-\sum_{m=1}^M \lambda_m C_m(\mathbf{G})} \quad (6)$$

that describes an exponential distribution (whence the name of the formalism) over the set  $\mathcal{G}$  of all possible network configurations. Notice that the coefficients  $\{\lambda_m\}_{m=0}^M$  are nothing else than the Lagrange multipliers in eq. (3), whose values are fixed by the set of equations

$$\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}|\vec{\lambda}) C_m(\mathbf{G}) = \langle C_m \rangle \quad (7)$$

$\forall m = 0 \dots M$ , where  $C_0(\mathbf{G}) = 1$  sets the normalization condition

$$e^{1+\lambda_0} \equiv Z(\vec{\lambda}) = \sum_{\mathbf{G} \in \mathcal{G}} e^{-\sum_{m=1}^M \lambda_m C_m(\mathbf{G})}. \quad (8)$$

Using the above relation, we can eliminate  $\lambda_0$  and obtain the standard expression of the ERG probability distribution that is analogous to that of the canonical ensemble in statistical physics:

$$P(\mathbf{G}|\vec{\lambda}) = \frac{e^{-\sum_{m=1}^M \lambda_m C_m(\mathbf{G})}}{Z(\vec{\lambda})}. \quad (9)$$

The quantity  $H(\mathbf{G}|\vec{\lambda}) = \sum_{m=1}^M \lambda_m C_m(\mathbf{G})$  at the exponent is called *graph Hamiltonian* and  $Z(\vec{\lambda}) = \sum_{\mathbf{G} \in \mathcal{G}} e^{-\sum_{m=1}^M \lambda_m C_m(\mathbf{G})}$  is the *partition function*, which properly normalizes the probability distribution. It can be easily shown that eq. (6) not only makes the first functional derivative of  $\mathcal{L}[P]$  vanish, but also makes the second derivative  $\frac{\delta^2 \mathcal{L}[P]}{\delta P(\mathbf{G}) \delta P(\mathbf{G}')}$  negative definite, so that  $P(\mathbf{G})$  is indeed a maximum of  $\mathcal{L}[P]$  (see Appendix C).

The ERG formalism can be fruitfully used to analyze real-world networks by supposing that a given observed configuration  $\hat{\mathbf{G}}$  has been drawn from  $\mathcal{G}$  and, as such, can be consistently assigned the probability coefficient  $P(\hat{\mathbf{G}}|\vec{\lambda})$ . However, we still need a recipe to set the mean values  $\{\langle C_m \rangle\}_{m=1}^M$  or, equivalently, the Lagrange multipliers  $\{\lambda_m\}_{m=1}^M$ , in

an optimal way. To this aim, we can invoke the *maximum-likelihood principle*, prescribing to maximize  $P(\hat{\mathbf{G}}|\vec{\lambda})$  as a function of  $\vec{\lambda}$ , or equivalently the *log-likelihood function*

$$\mathcal{L}(\hat{\mathbf{G}}|\vec{\lambda}) = \ln P(\hat{\mathbf{G}}|\vec{\lambda}) = - \sum_{m=1}^M \lambda_m C_m(\hat{\mathbf{G}}) - \ln Z(\vec{\lambda}). \quad (10)$$

Using eqs. (7) and (9), it is possible to show that the set of values  $\{\hat{\lambda}_m\}_{m=1}^M$  satisfying the set of equations  $\frac{\partial \mathcal{L}(\hat{\mathbf{G}}|\vec{\lambda})}{\partial \lambda_m} = 0$ ,  $m = 1 \dots M$  are those ensuring that

$$\langle C_m \rangle = \hat{C}_m \quad (11)$$

$\forall m = 1 \dots M$ , namely that the ensemble averages of constraints match their values observed in  $\hat{\mathbf{G}}$ . [58]. Remarkably, likelihood maximization consistently prescribes that the only information usable to make inference is the one we have access to. Moreover, it is simple to show that the same choice makes the Hessian of  $\mathcal{L}(\hat{\mathbf{G}}|\vec{\lambda})$  negative definite, implying that the position in eq. (11) corresponds to a maximum of the likelihood.

The whole ERG recipe can be thus summarized as joining two optimization principles:

- entropy maximization, which guarantees that the derived probability distribution encodes information only from the chosen constraints;
- likelihood maximization, which guarantees that the value of the imposed constraints matches the observed one, without any statistical bias.

In what follows, we shall dwell into the details of various network reconstruction methods. Note however that, generally speaking, the ERG formalism can be also fruitfully used for a second purpose, *i.e.*, analyzing known real-world networks in order to detect the level of randomness affecting their topological structure. Indeed, suppose to have an empirical network  $\hat{\mathbf{G}}$  whose structure is completely known. We can then ask whether this network is a “typical” configuration of a particular ERG derived from imposing an arbitrary set of network observables  $\{C_m(\hat{\mathbf{G}})\}_{m=1}^M$ . In other words, we can check if the network  $\hat{\mathbf{G}}$  is “maximally random”—the level of randomness being determined by the chosen constraints, so that the information brought by any other network observable can be reduced to the information encoded in the constraints. We shall briefly discuss this approach in subsection 4.2.3. <sup>4</sup>

### 3. Reconstruction methods

The reconstruction methods reviewed in the present work can be classified according to the link density of the output configurations. Indeed, the reconstructed networks can be fully connected (or, at least, very dense), with a tunable density, or exactly reproducing the observed number of links.

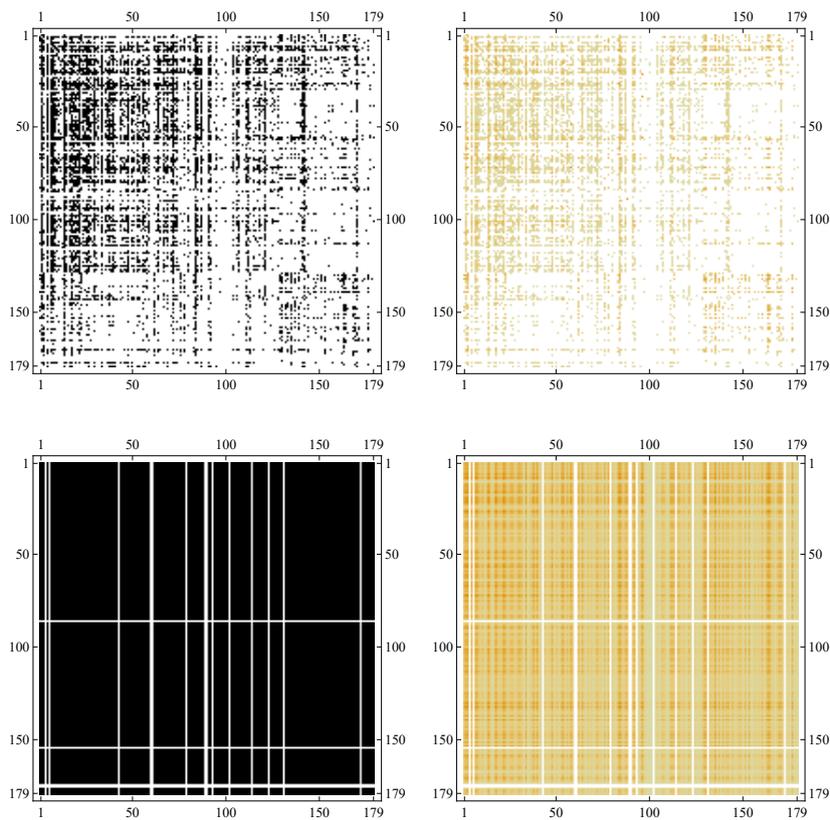


Figure 1: Comparison between the observed adjacency matrix of the eMID network in 2003 (top panels) and its reconstructed version according to the MaxEnt method described in section 3.1.1 (bottom panels). Left panels represent binary adjacency matrices with black/white denoting the presence/absence of connections, whereas, right panels represent weighted adjacency matrices with color intensity denoting the weight of connections

### 3.1. Dense reconstruction methods

#### 3.1.1. The MaxEnt algorithm

The MaxEnt algorithm [76, 77] represents the simplest and, probably, the best known method for reconstructing networks. It prescribes to maximize the functional

$$S = - \sum_{i=1}^N \sum_{j=1}^N w_{ij} \ln w_{ij} \quad (12)$$

under the constraints represented by eqs. (5). Equation (12) defines a particular kind of entropy in which the random variables are the matrix entries themselves. However, the absence of a proper normalization condition prevents eq. (12) from returning a genuine probability distribution. The solution to the aforementioned constrained maximization problem is, in fact,

$$w_{ij}^{\text{ME}} = \frac{\hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}{\hat{W}} \quad \forall i, j, \quad (13)$$

where  $\hat{W} = \sum_{i=1}^N \hat{s}_i^{\text{out}} \equiv \sum_{j=1}^N \hat{s}_j^{\text{in}}$  is the total weight of the observed network  $\hat{\mathbf{G}}$ . It is easy to verify that the constraints are satisfied, since  $\hat{s}_i^{\text{out}} = \sum_{j=1}^N w_{ij}^{\text{ME}}$  and  $\hat{s}_i^{\text{in}} = \sum_{j=1}^N w_{ji}^{\text{ME}}$   $\forall i$ . The summation index, however, has to run over all values  $j = 1 \dots N$ , including the ones corresponding to the diagonal entries. Note that eq. (13) implies that, unless either  $\hat{s}_i^{\text{out}} = 0$  or  $\hat{s}_i^{\text{in}} = 0$  for some nodes, no entries can be zero and the resulting matrix is fully connected (see fig. 1 where this algorithm has been applied to a snapshot of the Italian electronic market for interbank deposits eMID [40]). This feature represents the main limitation of the method, for a twofold reason. The first one is that real-world networks are often very sparse, and thus MaxEnt cannot possibly reproduce their topology. The second one is that systemic risk is underestimated in dense networks [78, 79]. Yet, the MaxEnt prescription provides quite accurate estimates whenever only the magnitude of weights are considered [80, 81]. The latter is the reason why MaxEnt is widely used in economics—the simple *gravity model* (without distance) has the same functional form of the MaxEnt estimate [82], and in finance—where it takes the same form of the *capital asset pricing model* (CAPM) [83, 84]. As a final remark, we stress that the MaxEnt algorithm generates a unique reconstructed configuration, thus being classifiable as a deterministic algorithm.

#### 3.1.2. Overcoming the problem of missing connections: the IPF algorithm

A first step in the description of networks which are not fully connected (*i.e.*, characterized by some null entries of the corresponding adjacency matrix) is given by the *iterative proportional fitting* (IPF) procedure. It is a simple recipe to obtain a matrix  $\mathbf{W}$  that

- lies at the “minimum distance” from the MaxEnt matrix  $\mathbf{W}^{\text{ME}}$  defined by eq. (13);

---

<sup>4</sup>Entropy-maximization can be also used to assign probabilities to (dynamical) pathways. In such a context, the principle is known as maximum-caliber [67]. An application of this principle concerns the so-called *origin-destination* networks, *i.e.*, graphs defined by an origin node, a destination node and a set of connections linking them, representing the pathways along which information flows [75].

- satisfies eqs. (5);
- admits the presence of a set of zero entries—in the typical case, the diagonal ones.

Let us call  $\mathbf{W}^{\text{IPF}}$  the matrix satisfying these conditions. In the case of null diagonal entries, the method formally defines  $\mathbf{W}^{\text{IPF}}$  as:

$$\min_{\mathbf{W}} \left\{ \sum_{i=1}^N \sum_{j(\neq i)=1}^N w_{ij} \ln \left( \frac{w_{ij}}{w_{ij}^{\text{ME}}} \right) \right\} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N w_{ij}^{\text{IPF}} \ln \left( \frac{w_{ij}^{\text{IPF}}}{w_{ij}^{\text{ME}}} \right) \quad (14)$$

*i.e.*, the matrix that minimizes the Kullback-Leibler (KL) divergence [85] between a generic non-negative  $\mathbf{W}$  with null diagonal entries and  $\mathbf{W}^{\text{ME}}$ . The KL divergence is an asymmetric measure of “distance” between any two probability distributions and quantifies the amount of information lost when  $\mathbf{W}$  is approximated by  $\mathbf{W}^{\text{ME}}$ .

A numerical recipe guaranteeing that the three requests above are met is provided by the iterative process whose basic steps at the  $n$ -th and  $(n+1)$ -th iterations are

$$w_{ij}^{(n+1)} = \hat{s}_j^{\text{in}} \left( \frac{w_{ij}^{(n)}}{\sum_{k(\neq j)} w_{kj}^{(n)}} \right), \quad w_{ij}^{(n)} = \hat{s}_i^{\text{out}} \left( \frac{w_{ij}^{(n-1)}}{\sum_{k(\neq i)} w_{ik}^{(n-1)}} \right), \quad (15)$$

so that  $w_{ij}^{\text{IPF}} = \lim_{n \rightarrow \infty} w_{ij}^{(n)}$ , and  $w_{ij}^{(0)}$  represents the matrix used to initialize the algorithm. In a nutshell, the IPF algorithm iteratively distributes the known matrix marginals across the non-zero entries of the initial matrix. As long as this initial matrix is irreducible (meaning that it cannot be permuted into a block upper triangular matrix, or equivalently that the network it represents is strongly connected), eqs. (15) always yield a unique matrix that satisfies the marginals [86]. As a first consistency check, let us consider the case in which the initial matrix is defined by  $w_{ij}^{(0)} = w_{ij}^{\text{ME}} \forall i, j$ . Without restricting the sum to the non-diagonal terms, we would obtain  $w_{ij}^{(1)} = w_{ij}^{(2)} = \dots = w_{ij}^{\text{ME}} \forall i, j$ . As a second check, let us consider the case in which the initial matrix is taken to be  $w_{ij}^{(0)} = 1 \forall i, j$ , a position that is equivalent to immediately maximizing the functional in eq. (13). We obtain  $w_{ij}^{(1)} = \frac{\hat{s}_i^{\text{out}}}{N}$  and  $w_{ij}^{(2)} = \frac{\hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}{\bar{W}} \forall i, j$ , hence the MaxEnt estimation is correctly recovered after just two iterations.

### 3.1.3. The Directed Weighted Configuration Model

Like MaxEnt, the IPF has the major drawback of generating a single deterministic configuration, so that it is difficult to statistically evaluate the accuracy of the provided reconstruction. A more rigorous statistical method to evaluate the probability that nodes  $i$  and  $j$  are connected by a link is the ERG-based approach known as Directed Weighted Configuration Model (DWCM) [58]. The method constrains the out-strength  $s_i^{\text{out}}$  and in-strength  $s_i^{\text{in}}$  (defined as in eqs. (5)) of each node  $i$  of the network, and the Hamiltonian takes the form

$$H(\mathbf{W}|\vec{\gamma}, \vec{\delta}) = \sum_{i=1}^N (\gamma_i s_i^{\text{out}} + \delta_i s_i^{\text{in}}). \quad (16)$$

Substituting the definitions of  $s_i^{\text{out}}$  and  $s_i^{\text{in}}$  in eq. (16) leads to a probability distribution  $P(\mathbf{W}|\vec{\gamma}, \vec{\delta})$  which factorizes into the product of  $N(N-1)$  pair-specific distributions

$$P(\mathbf{W}|\vec{\gamma}, \vec{\delta}) = \prod_{i=1}^N \prod_{j(\neq i)=1}^N q_{ij}^{\text{DWCM}}(w). \quad (17)$$

In the simple case of weights  $w_{ij}$  taking only non-negative integer values, the probability distribution governing the behavior of the random variable  $w_{ij}$  is geometric [58]:

$$q_{ij}^{\text{DWCM}}(w) = (y_i^{\text{out}} y_j^{\text{in}})^w (1 - y_i^{\text{out}} y_j^{\text{in}}) \text{ for } w \in \mathbb{Z}_+ \quad (18)$$

where  $y_i^{\text{out}} = e^{-\gamma_i}$  and  $y_i^{\text{in}} = e^{-\delta_i}$ . From eq. (18), we immediately find that

$$\langle w_{ij} \rangle^{\text{DWCM}} = \frac{y_i^{\text{out}} y_j^{\text{in}}}{1 - y_i^{\text{out}} y_j^{\text{in}}} \quad (19)$$

and, by definition, the probability  $p_{ij}$  that a directed link from node  $i$  to  $j$  is present is  $p_{ij} \equiv \sum_{w=1}^{\infty} q_{ij}(w)$ , which in view of eq. (18) becomes:

$$p_{ij}^{\text{DWCM}} = y_i^{\text{out}} y_j^{\text{in}}. \quad (20)$$

Finally, the Lagrange multipliers are found by solving the corresponding  $2N$  equations deriving from the likelihood-maximization principle:  $\forall i, \hat{s}_i^{\text{out}} = \langle s_i^{\text{out}} \rangle \equiv \sum_{j(\neq i)} \langle w_{ij} \rangle^{\text{DWCM}}$  and  $\hat{s}_i^{\text{in}} = \langle s_i^{\text{in}} \rangle \equiv \sum_{j(\neq i)} \langle w_{ji} \rangle^{\text{DWCM}}$ .

The DWCM falls into the category of dense reconstruction methods because the observed marginals are usually so large that the induced link probability between any two nodes  $i$  and  $j$  becomes very close to 1. However, differently from the MaxEnt and the IPF, the DWCM algorithm produces a whole ensemble of networks, by treating link as independent variables and drawing the corresponding weights from the geometric distributions described by eq. (19).

### 3.1.4. Combining MaxEnt and ERG frameworks

An approach combining the MaxEnt and the ERG frameworks has been recently developed, under the name of Maximum Entropy CAPM (MECAPM) [84]. The idea is to maximize the Shannon entropy constraining not the expected values of the matrix marginals, but rather the expected value of each link weight. Similarly to the case of the DWCM, this leads to

$$q_{ij}^{\text{MECAPM}}(w) = (y_{ij})^w (1 - y_{ij}) \quad (21)$$

where  $y_{ij}$  is the Lagrange multiplier controlling for the weight of the link from  $i$  to  $j$ . This framework can be used for network reconstruction provided that the imposed expected weights depend only on the matrix marginals, which is the only information available on the system. This is naturally achieved by the MaxEnt recipe, hence:

$$\langle w_{ij} \rangle^{\text{MECAPM}} = \frac{y_{ij}}{1 - y_{ij}} \equiv w_{ij}^{\text{ME}}, \quad (22)$$

a position allowing for the Lagrange multipliers to be readily estimated as the link probabilities:

$$p_{ij}^{\text{MECAPM}} = y_{ij} = \frac{w_{ij}^{\text{ME}}}{1 + w_{ij}^{\text{ME}}}. \quad (23)$$

This algorithm falls into the category of dense reconstruction methods, since the MaxEnt weights are usually sufficiently large to induce  $p_{ij} \simeq 1 \forall i, j$ . And as the DWCM, the MECAPM algorithm produces a whole ensemble of networks.

### 3.2. Density-tunable reconstruction methods

The MaxEnt, the DWCM and the MECAPM methods suffer from the same limitation: the predicted configurations are often too dense to faithfully describe real-world networks. Therefore other reconstruction methods have been proposed. The rationale driving the algorithms described below is to produce configurations that are sparser than the ones obtained through the aforementioned algorithms.

#### 3.2.1. The IPF algorithm: generic formulation

As we have seen, the IPF algorithm basically acts by distributing the known marginals across the positive entries of the matrix. Hence, it requires that the position of the null entries is known in advance. This limitation is the reason why the method is often used in combination with other algorithms that estimate the positions of the zeros. Once these positions are known, the IPF algorithm adjusts the positive entries (typically initialized as MaxEnt estimates) to match the constraints.

Indeed, the freedom to choose the topological details turns out to be fundamental whenever an algorithm able to *generate* realistic configurations is needed. The general formulation of the IPF algorithm give us this freedom. Indeed, in order to account for either known or guessed subsets of entries (which do not necessarily need to be zero), it is enough to i) subtract them from the known marginals  $\{\hat{s}_i^{\text{out}}\}_{i=1}^N$  and  $\{\hat{s}_i^{\text{in}}\}_{i=1}^N$ , and ii) modify eqs. (14) and (15) by using these rescaled marginals and explicitly excluding known entries from the sums at the denominator [86]. In the most general case, the IPF estimation can be written as the infinite product

$$w_{ij}^{\text{IPF}} = \prod_{n=0}^{\infty} \frac{\hat{s}_i^{\text{out}}}{[s_i^{\text{out}}]^{(2n)}} w_{ij}^{(0)} \frac{\hat{s}_j^{\text{in}}}{[s_j^{\text{in}}]^{(2n+1)}} \quad (24)$$

where  $[s_i^{\text{out}}]^{(2n)} = \sum_{j(\neq i)} w_{ij}^{(2n)}$  and  $[s_i^{\text{in}}]^{(2n+1)} = \sum_{j(\neq i)} w_{ji}^{(2n+1)}$ .<sup>5</sup>

#### 3.2.2. The Drehmann & Tarashev approach

In the absence of a clear recipe to estimate the network density, a number of algorithms exploring the whole range  $[0, 1]$  of possible density values have been proposed. Drehmann & Tarashev devise a simple approach [88] to perturb the MaxEnt matrix and obtain sparse network, following three steps:

- choosing a random set of off-diagonal entries to be zero, thus manually setting a desired link density;

---

<sup>5</sup>The IPF algorithm is also known with the name of RAS algorithm, because the form of the solution devised in [87] is written as the product of three matrices whose symbols are  $\mathbf{R}$ ,  $\mathbf{A}$ ,  $\mathbf{S}$ . More specifically,  $\mathbf{A}^{(2n+1)} = \mathbf{R}^{(n+1)}\mathbf{A}^{(2n)}$  and  $\mathbf{A}^{(2n+2)} = \mathbf{A}^{(2n+1)}\mathbf{S}^{(n+1)}$ , with  $\mathbf{A}^{(0)}$  being the initial adjacency matrix and  $\mathbf{R}$ ,  $\mathbf{S}$  being the two diagonal matrices  $\mathbf{R}$ ,  $\mathbf{S}$  being the two diagonal matrices  $\mathbf{R}^{(n+1)} = \text{diag}(\hat{s}_i^{\text{out}}/[s_i^{\text{out}}]^{(2n)})$  and  $\mathbf{S}^{(n+1)} = \text{diag}(\hat{s}_i^{\text{in}}/[s_i^{\text{in}}]^{(2n+1)})$ .

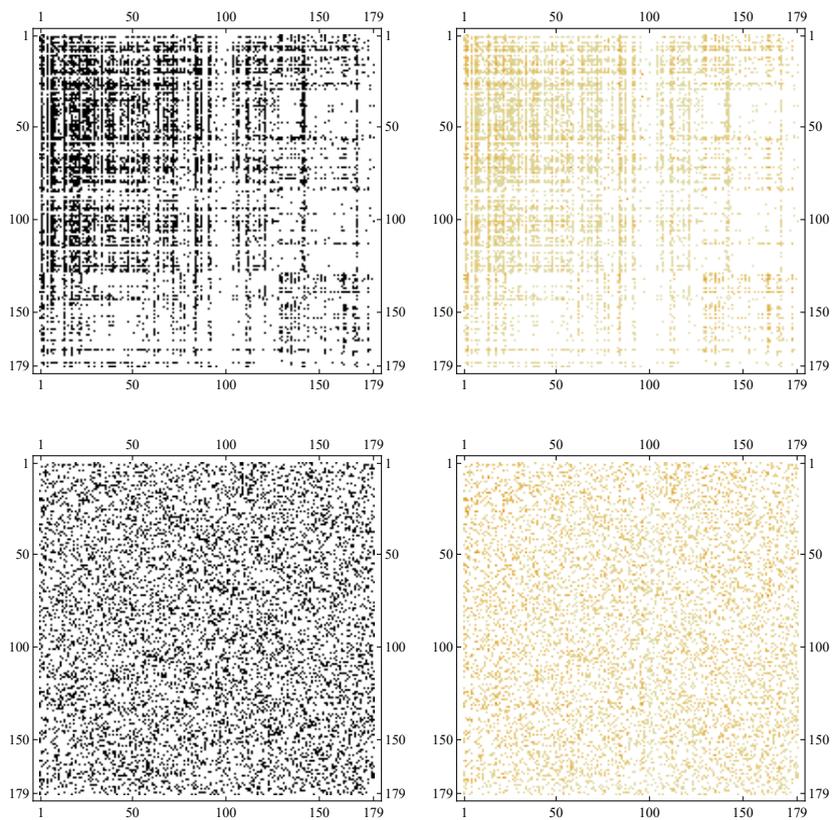


Figure 2: Comparison between the observed adjacency matrix of the eMID network in 2003 (top panels) and its reconstructed version according to the method by Drehmann & Tarashev described in section 3.2.2 (bottom panels). Left panels represent binary adjacency matrices with black/white denoting the presence/absence of connections, whereas, right panels represent weighted adjacency matrices with color intensity denoting the weight of connections.

- treating the remaining non-zero entries as random variables, uniformly distributed between zero and twice their MaxEnt estimated value:

$$w_{ij}^{\text{D-T}} \sim U(0, 2w_{ij}^{\text{ME}}) \quad (25)$$

(so that the expected value of weights under this distribution coincides with the MaxEnt estimates  $w_{ij}^{\text{ME}} = \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}} / \hat{W}$ );

- running the IPF algorithm to correctly restore the value of the marginals.

The value of the network density can be tuned to generate arbitrarily sparse configurations. Yet a drawback of the method lies in the completely random nature of the obtained topological structure(s) (see fig. 2).

### 3.2.3. The Mastromatteo, Zarinelli & Marsili approach

Another approach to generate reconstructed networks with an arbitrary density of links has been formulated in [79]. This method consists in sampling uniformly the set of network configurations that are compatible with the constraints defined by eqs. (5) and have a given value of the density. Two additional assumptions are made on the matrix  $\hat{\mathbf{W}}$ : i) the entries larger than a certain threshold  $\theta$  (supposed to be at most of order  $N$ ) are considered to be known—an hypothesis that in the case of financial networks meets the recent disclosure policies adopted for some markets [89]); ii) the unknown entries are rescaled by the threshold  $\theta$  itself, and thus become bounded in the range  $[0, 1]$ . Known entries can be completely omitted in the description of the method, and the focus can be kept only on the ensemble generated by the variability of the unknown entries.

The fundamental question the authors tackle is the following: given an arbitrary value of  $\rho$ , how many matrices exist that satisfy eqs. (5) and whose density matches the chosen one? Clearly, while the maximum density value allowing for (at least) one configuration to exist is  $\rho_{max} = 1$  (the MaxEnt algorithm always satisfies eqs. (5)), finding  $\rho_{min}$  for a given value of the constraints is non-trivial. The measure introduced in [79] to fairly sample the space of binary adjacency matrices that are *compatible* with a given  $\rho$  is  $P_0(\mathbf{A}|z) = \frac{1}{Z} z^{L(\mathbf{A})}$ , where  $Z$  is a normalization constant,  $L(\mathbf{A})$  is the number of links in  $\mathbf{A}$  and the parameter  $z$  sets the average density

$$\langle \rho \rangle = \sum_{\mathbf{A}} P_0(\mathbf{A}|z) \rho(\mathbf{A}) \quad (26)$$

to a desired value (notice that the sum runs over the compatible configurations only).

However, when  $\langle \rho \rangle < 1$  it is hard to analytically evaluate the sampling probability distribution  $P_0(\mathbf{A}|z)$ . Consequently, an approximated probability distribution is introduced:

$$P(\mathbf{A}|z) = \frac{e^{-\beta H(\mathbf{A})} z^{L(\mathbf{A})}}{Z(\beta, z)} \quad (27)$$

where  $H(\mathbf{A}) = \sum_i (\Theta[s_i^{\text{out}} - k_i^{\text{out}}] + \Theta[s_i^{\text{in}} - k_i^{\text{in}}])$ . The parameter  $z$  plays the role of the fugacity in statistical physics and fixes the mean value of the density over the *whole* set of adjacency matrices. Since the unknown entries of  $\mathbf{W}$  range between 0 and 1, the number of out-going (in-coming) links of each node in any compatible configuration with the

chosen density value is always larger than its out-strength (in-strength): this brings to the condition  $H(\mathbf{A}) = 0$ . Conversely, the probability of any infeasible configuration (*i.e.*, with  $H(\mathbf{A}) = 1$ ) to appear is suppressed by a coefficient  $e^{-\beta}$  which vanishes as  $\beta \rightarrow \infty$ . Since an analytical treatment of the distribution in eq. (27) is also infeasible, the authors implement a *message-passing algorithm* [79] to calculate the marginal link probabilities, which are then independently sampled to build a candidate binary network. And once the binary topology is obtained, weights are inferred via the IPF algorithm.

As the authors explicitly notice [79], being able to tune the network density implies being able to consider a whole range of structures characterized by (potentially) different degrees of robustness to, *e.g.*, financial shocks. However, as a result of the authors’ “compatibility analysis”, no allowable configurations can be found below a certain value of the network density. And while the MaxEnt method produces configurations believed to maximize the network robustness, sparser configurations may, on the other hand, provide lower bounds to it.

### 3.2.4. The Moussa & Cont approach

Another algorithm that combines the MaxEnt and the ERG frameworks is the one presented in [90], where the authors propose two different versions of their method—both similar to the entropy-based approaches described in the previous subsections. The major difference lies in the amount of available information required, which is now substantially larger: namely, the out- and in-degree distributions of the network, as well as the out- and in-strength distributions, all of them supposedly well-described by power-laws (whose parameters are tuned to reproduce stylized facts of financial networks). In a nutshell, the method generates a whole ensemble of  $R$  different binary network configurations (*i.e.*, the *prior* configurations), whose topological structure is then “adjusted” *a posteriori* to match the constraints represented by eqs. (5).

*The “exact” approach.* The first version of the method is designed to meet the constraints *exactly* in each configuration of the ensemble. Prior configurations are characterized by the same degree distributions but different topological structures (the algorithm used to generate the ensemble is the generalization of the preferential attachment algorithm to directed networks [91]). Once a binary configuration is generated, the IPF algorithm is used to assign weights to the realized links. The problem can thus be formally stated as that of determining, for each binary configuration in the ensemble, the set of weights  $\{w_{ij}^{\text{M-C}}\}$  such that

$$\min_{\mathbf{w}} \left\{ \sum_{\{a_{ij}=1\}} w_{ij} \ln \left( \frac{w_{ij}}{w_{ij}^{\text{ME}}} \right) \right\} = \sum_{\{a_{ij}=1\}} w_{ij}^{\text{M-C}} \ln \left( \frac{w_{ij}^{\text{M-C}}}{w_{ij}^{\text{ME}}} \right), \quad (28)$$

where the sum runs over the set of non-zero entries of any configuration in the ensemble (for all null entries we trivially have  $w_{ij}^{\text{M-C}} = 0$ ). Using a normalization condition that the entries to be estimated satisfy  $\sum_{\{a_{ij}=1\}} w_{ij} = 1$ , the solution to the optimization problem above can be formally written as

$$w_{ij}^{\text{M-C}} = \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}} \left( \frac{e^{\gamma_i + \delta_j}}{\sum_{\{a_{kl}=1\}} \hat{s}_k^{\text{out}} \hat{s}_k^{\text{in}} e^{\gamma_k + \delta_l}} \right), \quad (29)$$

where  $\{\gamma_i\}_{i=1}^N$  and  $\{\delta_i\}_{i=1}^N$  are, respectively, the Lagrange multipliers related to the constraints on nodes out- and in-strengths, found by solving  $\langle s_i^{\text{out}} \rangle = \hat{s}_i^{\text{out}}$  and  $\langle s_i^{\text{in}} \rangle = \hat{s}_i^{\text{in}}$   $\forall i$ .

*The “average” approach.* However, as the authors explicitly notice, dealing with too sparse matrices may prevent the IPF algorithm to converge, because a solution of the IPF algorithm exists and is unique if and only if the matrix is irreducible (*i.e.*, the network is strongly connected) [86]. Moreover, there is no guarantee that the numerical values of weights assigned by IPF are distributed according to some empirical observations (*e.g.*, following a power law) [90].

For this reason, a second version of the algorithm is proposed. The ensemble of configurations is now composed by weighted networks for which both degrees and weights distributions are specified. The probability distribution on this set of configurations, however, is not uniform: each configuration  $r = 1, \dots, R$  is, in fact, characterized by a statistical weight  $\nu_r \in (0, 1)$  such that the constraints represented by eqs. (5) are satisfied on average. The ensemble probability distribution  $\{\nu_r\}_{r=1}^R$  is derived by minimizing its KL divergence from the uniform distribution  $1/R$

$$\sum_{r=1}^R \nu_r \ln \left( \frac{\nu_r}{1/R} \right) \quad (30)$$

under the constraints  $\hat{s}_i^{\text{out}} = \sum_r \nu_r [s_i^{\text{out}}]^{(r)}$  and  $\hat{s}_i^{\text{in}} = \sum_r \nu_r [s_i^{\text{in}}]^{(r)}$   $\forall i$  (where  $\{[s_i^{\text{out}}]^{(r)}\}_{i=1}^N$  and  $\{[s_i^{\text{in}}]^{(r)}\}_{i=1}^N$  are the out- and in-strengths for the  $r^{\text{th}}$  configuration in the ensemble). Similar calculations to those used for eq. (29) lead to the analytical expression of the probability coefficients

$$\nu_r = \frac{e^{\sum_{i=1}^N \gamma_i [s_i^{\text{out}}]^{(r)} + \sum_{j=1}^N \delta_j [s_j^{\text{in}}]^{(r)}}}{Z(\vec{\gamma}, \vec{\delta})} \quad (31)$$

with  $Z(\vec{\gamma}, \vec{\delta}) = \sum_r e^{\sum_{i=1}^N \gamma_i [s_i^{\text{out}}]^{(r)} + \sum_{j=1}^N \delta_j [s_j^{\text{in}}]^{(r)}}$ . As usual, the numerical value of the Lagrange multipliers can be found by solving the constraints equations  $\langle s_i^{\text{out}} \rangle = \hat{s}_i^{\text{out}}$  and  $\langle s_i^{\text{in}} \rangle = \hat{s}_i^{\text{in}}$   $\forall i$ . The estimation of any quantity of interest is then carried out by computing the ensemble average of the quantity itself.

### 3.3. Exact density methods

The reconstruction algorithms described in the previous sections are attempts to circumvent the lack of information on the actual network density and, most importantly, to avoid predicting too dense configurations. The methods described below, instead, explicitly require the knowledge of the observed network density or, at least, the link density for a subset of nodes. This is because, as shown in [92, 81, 93], adding this piece of information can dramatically increase the performance of a reconstruction algorithm. We now introduce a series of algorithms of this kind, all strictly following the ERG formalism introduced in Section 2.2.

### 3.3.1. The density-corrected DWCM

The first example of ERG-based reconstruction method we have met in subsection 3.1.3 is the DWCM, obtained by constraining the out-strength and in-strength sequences. As we have seen, the DWCM-induced ensemble is basically characterized by fully-connected configurations, with a density of links which cannot be tuned independently from the distribution of weights (see eq. (20)). So the outcome of the DWCM is very close to that of the MaxEnt method, and indeed the DWCM can be seen a sort of stochastic generalization of MaxEnt. In order to overcome this limitation, the authors of [81] propose a *density-corrected* version of the DWCM, defined by constraining the total number of links  $L$  beside the out-strength and in-strength sequences:

$$H(\mathbf{W}|\vec{\gamma}, \vec{\delta}, \zeta) = \sum_{i=1}^N (\gamma_i s_i^{\text{out}} + \delta_i s_i^{\text{in}}) + \zeta L. \quad (32)$$

In analogy with the DWCM, also in this case links turn out to be statistically independent. When weights can assume only positive integer values, the weight probability distribution can be written as

$$q_{ij}^{\text{dcDWCM}}(w) = p_{ij}^{\text{dcDWCM}} (y_i^{\text{out}} y_j^{\text{in}})^{w-1} (1 - y_i^{\text{out}} y_j^{\text{in}}) \quad (33)$$

where

$$p_{ij}^{\text{dcDWCM}} = \frac{z y_i^{\text{out}} y_j^{\text{in}}}{1 + z y_i^{\text{out}} y_j^{\text{in}} - y_i^{\text{out}} y_j^{\text{in}}} \quad (34)$$

in which  $y_i^{\text{out}}$  and  $y_j^{\text{in}}$  are defined as in the DWCM and  $z = e^{-\zeta}$  (thus, for  $\zeta = 0$  we recover the standard DWCM). Notice that eq. (33) defines a composition of a single Bernoulli trial, controlling for the existence of a link between any two nodes  $i$  and  $j$ , and a geometric distribution for its weight, whose mean value reads

$$\langle w_{ij} \rangle^{\text{dcDWCM}} = \frac{p_{ij}^{\text{dcDWCM}}}{1 - y_i^{\text{out}} y_j^{\text{in}}}. \quad (35)$$

Finally the Lagrange multipliers are fixed as in the DWCM, while  $\zeta$  is obtained by  $\hat{L} = \langle L \rangle \equiv \sum_{i \neq j} p_{ij}^{\text{dcDWCM}}$ .

This method ideally refines the DWCM by explicitly adding to the recipe a piece of topological information. By doing so, the occurrence probability of a network in the ensemble still depends on the marginals, but also on the imposed number of links, hence very dense configurations become highly improbable.

### 3.3.2. The Directed Enhanced Configuration Model

Beyond the total number of links, also the degree heterogeneity can be explicitly taken into account. The Directed Enhanced Configuration Model (DECM) [94] is defined by:

$$H(\mathbf{W}|\vec{\alpha}, \vec{\beta}, \vec{\gamma}, \vec{\delta}) = \sum_{i=1}^N (\alpha_i k_i^{\text{out}} + \beta_i k_i^{\text{in}} + \gamma_i s_i^{\text{out}} + \delta_i s_i^{\text{in}}) \quad (36)$$

and encompasses many ERG-based models as special cases (for instance, the degree-corrected DWCM is obtained when  $\alpha_i = \beta_i = \zeta/2, \forall i$ ). The DECM probability distribution can be written as:

$$P(\mathbf{W}|\vec{\alpha}, \vec{\beta}, \vec{\gamma}, \vec{\delta}) = \prod_{i=1}^N \prod_{j(\neq i)=1}^N q_{ij}^{\text{DECM}}(w) \quad (37)$$

with

$$q_{ij}^{\text{DECM}}(w) = \begin{cases} 1 - p_{ij}^{\text{DECM}} & \text{if } w = 0, \\ p_{ij}^{\text{DECM}}(y_i^{\text{out}}y_j^{\text{in}})^{w-1}(1 - y_i^{\text{out}}y_j^{\text{in}}) & \text{if } w > 0 \end{cases} \quad (38)$$

and

$$p_{ij}^{\text{DECM}} = \frac{x_i^{\text{out}} x_j^{\text{in}} y_i^{\text{out}} y_j^{\text{in}}}{1 + x_i^{\text{out}} x_j^{\text{in}} y_i^{\text{out}} y_j^{\text{in}} - y_i^{\text{out}} y_j^{\text{in}}} \quad (39)$$

(where  $x_i^{\text{out}} = e^{-\alpha_i}$ ,  $x_i^{\text{in}} = e^{-\beta_i}$ ,  $y_i^{\text{out}} = e^{-\gamma_i}$  and  $y_i^{\text{in}} = e^{-\delta_i}$ ). Notice that from eq. (38) it is simple to evaluate the average value of the generic link weight as

$$\langle w_{ij} \rangle^{\text{DECM}} = \frac{p_{ij}^{\text{DECM}}}{1 - y_i^{\text{out}} y_j^{\text{in}}}. \quad (40)$$

Lagrange multipliers are as usual found by solving the corresponding  $4N$  equations derived from the likelihood-maximization principle:  $\forall i, \hat{k}_i^{\text{out}} = \langle k_i^{\text{out}} \rangle \equiv \sum_{j(\neq i)} p_{ij}^{\text{DECM}}$ ,  $\hat{k}_i^{\text{in}} = \langle k_i^{\text{in}} \rangle \equiv \sum_{j(\neq i)} p_{ji}^{\text{DECM}}$ ,  $\hat{s}_i^{\text{out}} = \langle s_i^{\text{out}} \rangle \equiv \sum_{j(\neq i)} \langle w_{ij} \rangle^{\text{DECM}}$ ,  $\hat{s}_i^{\text{in}} = \langle s_i^{\text{in}} \rangle \equiv \sum_{j(\neq i)} \langle w_{ji} \rangle^{\text{DECM}}$ .

As for the density-corrected DWCM, eq. (38) can be interpreted as a combination of a Bernoulli trial, with probability  $p_{ij}^{\text{DECM}}$ , and a drawing from a geometric distribution, with parameter  $y_i^{\text{out}}y_j^{\text{in}}$ ; in this case, however, the link probability depends also on the degrees of nodes  $i$  and  $j$ .

The DECM method has a simple interpretation when a specific economic system is considered, namely the World Trade Network (WTN). In economic terms, the two aforementioned processes respectively describe the tendency of a generic country  $i$  either to establish a new export towards country  $j$  (with probability  $p_{ij}$ ) or to reinforce an existing one (with probability  $y_i^{\text{out}}y_j^{\text{in}}$ , by rising the exchanged amount of goods of, so to say, “one unit” of trade). In order to understand which process is more probable, we can study the behavior of the ratio  $p_{ij}^{\text{DECM}}/(y_i^{\text{out}}y_j^{\text{in}})$ . Whenever this quantity is greater than 1, country  $i$  would probably establish a new export relation towards  $j$ , and at the same time experience a certain resistance to reinforce it; otherwise, country  $i$  would experience a certain resistance to start exporting to  $j$ , but once such relation were established it would be characterized by a relatively low “friction”, inducing the involved partners to strengthen it [95]. Note that the case  $p_{ij}^{\text{DECM}}/(y_i^{\text{out}}y_j^{\text{in}}) = 1$  implies reducing eq. (38) to eq. (18) of the DWCM. In other words, the DWCM does not discriminate between the first link and the subsequent ones, reducing  $q_{ij}(w)$  to a simple geometric distribution. As shown in [94], the DWCM fails in reproducing the observed properties of the WTN precisely because it cannot give the right importance to the very first link, which is treated as a simple unit of weight. This observation hints at the importance of the information encoded into nodes degrees, to be considered as a fundamental ingredient (together with nodes strengths) for a faithful reconstruction of real-world networks.

### 3.3.3. Simplifying the DECM: a two-step model

A simplified version of the DECM can be derived by noticing that the estimation of the topological structure of a network can be, in some circumstances, disentangled from the estimation of its weighted structure. This observation rests upon the evidence that the link probabilities of the DECM show a large (positive) correlation with the analogous probabilities of a much simpler ERG model, namely the Directed Binary Configuration Model (DBCM) obtained by constraining only the out- and in-degree sequences [95]. The DBCM is thus defined by the Hamiltonian

$$H(\mathbf{A}|\vec{\alpha}, \vec{\beta}) = \sum_{i=1}^N (\alpha_i k_i^{\text{out}} + \beta_i k_i^{\text{in}}), \quad (41)$$

which leads to the connection probability

$$p_{ij}^{\text{DBCM}} = \frac{x_i^{\text{out}} x_j^{\text{in}}}{1 + x_i^{\text{out}} x_j^{\text{in}}} \quad (42)$$

with  $x_i^{\text{out}} = e^{-\alpha_i}$  and  $x_i^{\text{in}} = e^{-\beta_i}$ , determined via the usual  $2N$  equations  $\hat{k}_i^{\text{out}} = \langle k_i^{\text{out}} \rangle \equiv \sum_{j(\neq i)=1}^N p_{ij}^{\text{DBCM}}$  and  $\hat{k}_i^{\text{in}} = \langle k_i^{\text{in}} \rangle \equiv \sum_{j(\neq i)=1}^N p_{ji}^{\text{DBCM}}$ ,  $\forall i$ .

Putting this expression into eqs. (38) and (40) leads to

$$q_{ij}^{2\text{s-DECM}}(w) = p_{ij}^{\text{DBCM}} (y_i^{\text{out}} y_j^{\text{in}})^{w-1} (1 - y_i^{\text{out}} y_j^{\text{in}}), \quad \langle w_{ij}^{2\text{s-DECM}} \rangle = \frac{p_{ij}^{\text{DBCM}}}{1 - y_i^{\text{out}} y_j^{\text{in}}}, \quad (43)$$

which defines a “two-step” version of the DECM, and whose Lagrange multipliers are found by imposing,  $\forall i$ ,  $\hat{k}_i^{\text{out}} = \langle k_i^{\text{out}} \rangle^{\text{DBCM}}$  and  $\hat{k}_i^{\text{in}} = \langle k_i^{\text{in}} \rangle^{\text{DBCM}}$  first and then  $\hat{s}_i^{\text{out}} = \langle s_i^{\text{out}} \rangle^{2\text{s-DECM}}$  and  $\hat{s}_i^{\text{in}} = \langle s_i^{\text{in}} \rangle^{2\text{s-DECM}}$ .

### 3.3.4. Fitness-induced Exponential Random Graphs

Despite the previous findings, we note that it is impossible to use either the DECM or its two-step version when the degrees of nodes are not known, which is unfortunately a rather common situation. Nevertheless, these cases can be treated by resorting to the *fitness* ansatz, which states that the link probability between any two nodes depends on non-topological features of the involved nodes, typical of the system under analysis. More precisely, it is assumed that the “activity” of each node  $i$  in the network is summed up by an “intrinsic” quantity, called *fitness* [52], which is directly related to the Lagrange multiplier  $x_i$  controlling the degree of that node through a monotone functional relation. Note that such relation between fitness values and degrees lies at the basis of the so-called *good-gets-richer* mechanism, according to which “better” nodes (those characterized by a higher fitness value) have more chances to “attract” connections [52].

For instance, in the case of the (undirected) WTN, where nodes represent countries and links represent trade relationships between them, a strong linear correlation can be observed between the Lagrange multipliers of nodes’ degrees and the Gross Domestic Product (GDP) values of the respective countries:  $x_i \simeq \sqrt{z} \text{GDP}_i \forall i$  [96, 57]. Consequently, the link probability between nodes  $i$  and  $j$  can be rewritten as

$$p_{ij}^{\text{UBCM}} \simeq \frac{z \text{GDP}_i \text{GDP}_j}{1 + z \text{GDP}_i \text{GDP}_j}, \quad (44)$$

where UBCM stands for Undirected Binary Configuration Model. Similar fitness ansatz have been successfully tested for financial networks, such as interbank markets [97, 98, 92] and shareholding networks [99, 100],

The validity of the fitness ansatz has profound implications on the kind of information that is necessary to have in order to accurately reconstruct a network, but in general leads to face the problem of finding node observables that are correlated with degrees. Remarkably, strengths often work well as fitnesses [101], a “stylized fact” implying that the DBCM Lagrange multipliers can be expressed as  $x_i^{\text{out}} = f(\hat{s}_i^{\text{out}})$  and  $x_i^{\text{in}} = f(\hat{s}_i^{\text{in}})$ . As the many empirical analyses of economic and financial systems mentioned above have pointed out, the functional form  $x_i^{\text{out}} = \sqrt{z}(\hat{s}_i^{\text{out}})^b$  and  $x_i^{\text{in}} = \sqrt{z}(\hat{s}_i^{\text{in}})^b$  with exponent  $b = 1$  is often accurate enough for all practical purposes.

*Estimating the degrees.* The above assumption of linear proportionality allows to estimate the unknown degrees in a straightforward way. Indeed, connection probabilities or the fitness-induced DBCM assume the form

$$p_{ij}^{\text{f-DBCM}} = \frac{z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}{1 + z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}, \quad (45)$$

so that the only variable left is the proportionality constant  $z$ .<sup>6</sup> The latter can be simply estimated provided that the total number of links  $\hat{L}$  of the empirical network is known. Imposing  $\hat{L} = \langle L \rangle$  in fact means solving only one equation

$$\hat{L} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N \frac{z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}{1 + z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}} \quad (46)$$

which has a single solution  $z > 0$  [98, 92]. Once  $z$  is found, the degrees of any node  $i$  in the network can be estimated as

$$\langle k_i^{\text{out}} \rangle^{\text{f-DBCM}} = \sum_{j(\neq i)=1}^N p_{ij}^{\text{f-DBCM}}, \quad \langle k_i^{\text{in}} \rangle^{\text{f-DBCM}} = \sum_{j(\neq i)=1}^N p_{ji}^{\text{f-DBCM}}. \quad (47)$$

An estimate of  $z$  can be obtained also using the information on the connectivity of a subset  $I$  of nodes, for instance the total number  $\hat{L}_I$  of links internal to  $I$ , or the degrees of all nodes belonging to  $I$  [102, 80]. In both cases, in fact, the likelihood-maximization principle leads to an equation similar-in-spirit to eq. (46). In the first case, we have

$$\hat{L}_I = \sum_{i \in I} \sum_{j(\neq i) \in I} \frac{z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}{1 + z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}, \quad (48)$$

while in the second case it is

$$\sum_{i \in I} (\hat{k}_i^{\text{out}} + \hat{k}_i^{\text{in}}) = \sum_{i \in I} \sum_{\substack{j=1 \\ (j \neq i)}}^N \left( \frac{z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}}{1 + z \hat{s}_i^{\text{out}} \hat{s}_j^{\text{in}}} + \frac{z \hat{s}_j^{\text{out}} \hat{s}_i^{\text{in}}}{1 + z \hat{s}_j^{\text{out}} \hat{s}_i^{\text{in}}} \right). \quad (49)$$

---

<sup>6</sup>Note that the MECAPM connection probabilities defined in eq. (23) are recovered here by the particular choice  $z = \hat{W}^{-1}$

As shown in [103], the way a specific subset of nodes is selected does matter. Whenever a faithful estimation of the network density is needed, nodes must be sampled according to a random selection scheme [80], any other procedure being biased towards larger or smaller density values.

### 3.3.5. Combining fitness-induced DBCM and IPF

An reconstruction method combining the fitness-induced ERG formalism and the IPF recipe is proposed in [104]. Here, the authors assume to know only the out- and in-strengths  $\{\hat{s}_i^{\text{out}}, \hat{s}_i^{\text{in}}\}_{i=1}^N$  and the total number of links  $\hat{L}$ . The algorithm then consists of two steps:

- the presence of a link between any two nodes  $i$  and  $j$  is estimated as in the fitness-induced DWCM, *i.e.*, via eq. (45) using as  $z$  the solution of eq. (46);
- the weights are placed on each generated binary configuration according to the IPF algorithm, and hence the constraints of eqs. (5) are always met exactly.

### 3.3.6. Combining fitness-induced DBCM and DECM

A more rigorous way to assign weights to the fitness-induced ERG formalism consists in solving “bootstrapped” version of the DECM [92]. More precisely, the system of equations to be solved to find the Lagrange multipliers  $\{x_i^{\text{out}}, x_i^{\text{in}}, y_i^{\text{out}}, y_i^{\text{in}}\}_{i=1}^N$  of the DECM becomes

$$\begin{cases} \langle k_i^{\text{out}} \rangle^{\text{f-DBCM}} &= \sum_{j(\neq i)=1}^N p_{ij}^{\text{DECM}} \\ \langle k_i^{\text{in}} \rangle^{\text{f-DBCM}} &= \sum_{j(\neq i)=1}^N p_{ji}^{\text{DECM}} \\ \hat{s}_i^{\text{out}} &= \sum_{j(\neq i)=1}^N p_{ij}^{\text{DECM}} (1 - y_i^{\text{out}} y_j^{\text{in}})^{-1} \\ \hat{s}_i^{\text{in}} &= \sum_{j(\neq i)=1}^N p_{ji}^{\text{DECM}} (1 - y_j^{\text{out}} y_i^{\text{in}})^{-1} \end{cases} \quad \forall i \quad (50)$$

where  $p_{ij}^{\text{DECM}}$  is defined in eq. (39) and  $\langle k_i^{\text{out}} \rangle^{\text{f-DBCM}}$  and  $\langle k_i^{\text{in}} \rangle^{\text{f-DBCM}}$  are the fitness-induced DBCM estimates defined by eqs. (47). The name *bootstrapped DECM* comes from the double role played by node out- and in-strengths, which are first used to estimate the degrees, and then imposed as complementary constraints.

### 3.3.7. The degree-corrected gravity model

Although the DECM (both in its original and “bootstrapped” versions) represents a very accurate reconstruction method [94, 92], its numerical resolution can represent a computationally-demanding task<sup>7</sup>. Building upon the MaxEnt recipe, economics provides the main inspiration for a simpler alternative. Indeed, although the MaxEnt method performs poorly in reproducing the topological structure of many real-world networks, the observed weights are nicely reproduced by eq. (13) [80, 81]. A straightforward way to both retain the explanatory power of the gravity model and avoid ending up with a complete network is provided by the heuristic recipe of the “degree-corrected gravity model” (dcGM) [98]:

$$w_{ij}^{\text{dcGM}} = \begin{cases} 0 & \text{with probability } 1 - p_{ij}^{\text{f-DBCM}} \\ w_{ij}^{\text{ME}} (p_{ij}^{\text{f-DBCM}})^{-1} & \text{with probability } p_{ij}^{\text{f-DBCM}} \end{cases} \quad \forall i \neq j. \quad (51)$$

<sup>7</sup>It should, however, be noticed that the estimation procedure leading to eqs. (47) has a regularizing effect on the values of degrees, which become smooth monotonic functions of the strengths [98]. This, in turn, may lead to a smaller computational effort to solve eqs. (50).

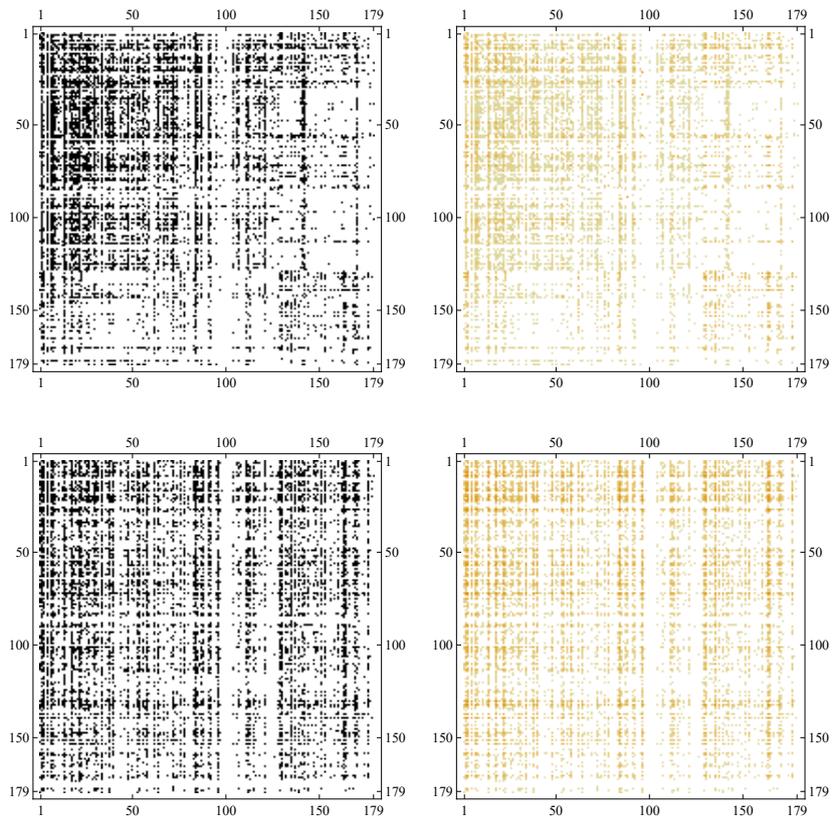


Figure 3: Comparison between the observed adjacency matrix of the eMID network in 2003 (top panels) and its reconstructed version according to the dcGM method described in section 3.3.7 (bottom panels). Left panels represent binary adjacency matrices with black/white denoting the presence/absence of connections, whereas, right panels represent weighted adjacency matrices with color intensity denoting the weight of connections.

This equation “corrects” the MaxEnt recipe by placing the weight  $w_{ij}^{\text{ME}}$  only with probability  $p_{ij}^{\text{f-DBCM}}$  (*i.e.*, conditional to the existence of the link), and rescaled in order to have  $\langle w_{ij}^{\text{dcGM}} \rangle = w_{ij}^{\text{ME}}$ . In this way, both the network marginals and the link density are correctly reproduced on average (see fig. 3). However, as for the original MaxEnt, this happens only if non-zero diagonal entries are considered as well. Otherwise, by restricting the sums over  $i \neq j$ , the expected values of the strengths obtained from the dcGM would require an extra-term to reproduce the observed marginals:  $\forall i$ , we would get  $\langle s_i^{\text{out}} \rangle^{\text{dcGM}} = \hat{s}_i^{\text{out}} - \hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}} / \hat{W}$  and  $\langle s_i^{\text{in}} \rangle^{\text{dcGM}} = \hat{s}_i^{\text{in}} - \hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}} / \hat{W}$ , and the missing term would be precisely the expected diagonal weight  $\langle w_{ii}^{\text{dcGM}} \rangle = \hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}} / \hat{W} \equiv w_{ii}^{\text{ME}}$ .

The authors in [80] devise a procedure, inspired by the IPF algorithm, to redistribute the diagonal terms across the non-diagonal entries of the network. Precisely, the correction term  $\delta w_{ij}^{(n)}$  to be added to the second line of eq. (51) at the  $n$ -th iteration of the IPF-like algorithm is defined as

$$\delta w_{ij}^{(n+1)} = \frac{\hat{s}_j^{\text{out}} \hat{s}_j^{\text{in}}}{\hat{W}} \left( \frac{\delta w_{ij}^{(n)}}{\sum_{k(\neq j)} \delta w_{kj}^{(n)}} \right), \quad \delta w_{ij}^{(n)} = \frac{\hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}}}{\hat{W}} \left( \frac{\delta w_{ij}^{(n-1)}}{\sum_{k(\neq i)} \delta w_{ik}^{(n-1)}} \right), \quad (52)$$

and where the algorithm is initialized at  $w_{ij}^{(0)} = 1 - \delta_{ij}$ . Once the asymptotic corrections  $\delta w_{ij}^{(\infty)}$  are determined, the heuristic recipe of eq. (51) is replaced by

$$w_{ij}^{\text{dcGM}} = \begin{cases} 0 & \text{with probability } 1 - p_{ij}^{\text{f-DBCM}} \\ (w_{ij}^{\text{ME}} + \delta w_{ij}^{(\infty)}) (p_{ij}^{\text{f-DBCM}})^{-1} & \text{with probability } p_{ij}^{\text{f-DBCM}} \end{cases} \quad \forall i \neq j. \quad (53)$$

For all practical purposes, a small number of iterations is often enough to achieve a satisfactory degree of accuracy. Here we explicitly report the functional form of the first three iterations:

$$\begin{aligned} w_{ij}^{(1)} &= \frac{\hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}}}{\hat{W}} \left( \frac{1}{N-1} \right); \\ w_{ij}^{(2)} &= \frac{\hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}}}{\hat{W}} \left( \frac{\hat{s}_j^{\text{out}} \hat{s}_j^{\text{in}}}{\sum_{l(\neq j)} \hat{s}_l^{\text{out}} \hat{s}_l^{\text{in}}} \right); \\ w_{ij}^{(3)} &= \frac{\hat{s}_i^{\text{out}} \hat{s}_i^{\text{in}}}{\hat{W}} \left( \frac{\hat{s}_j^{\text{out}} \hat{s}_j^{\text{in}}}{\sum_{l(\neq j)} \hat{s}_l^{\text{out}} \hat{s}_l^{\text{in}}} \right) \left( \frac{1}{\sum_{k(\neq i)} \frac{\hat{s}_k^{\text{out}} \hat{s}_k^{\text{in}}}{\sum_{m(\neq k)} \hat{s}_m^{\text{out}} \hat{s}_m^{\text{in}}}} \right). \end{aligned} \quad (54)$$

*A bipartite degree-corrected gravity model.* The degree-corrected gravity model can be straightforwardly extended to the case of bipartite (undirected) networks [100]. It is in fact enough to adapt eq. (46) and eq. (51) to the new problem setup. In particular, the equation to determine the unknown coefficient  $z$  relating the known and the expected total number of links is

$$\hat{L} = \sum_{i=1}^{N_1} \sum_{\alpha=1}^{N_2} \frac{z \hat{s}_i^{[1]} \hat{s}_\alpha^{[2]}}{1 + z \hat{s}_i^{[1]} \hat{s}_\alpha^{[2]}} \quad (55)$$

with  $N_1$  and  $N_2$  denoting the cardinality of the two layers of the network, and  $\{\hat{s}_i^{[1]}\}_{i=1}^{N_1}$  and  $\{\hat{s}_\alpha^{[2]}\}_{\alpha=1}^{N_2}$  indicating the *known* strength sequences of nodes belonging to the first and second layer, respectively. Notably, correction terms as the ones defined by eqs. (54) are no longer needed, since diagonal terms are now absent by definition.

### 3.3.8. Reconciling ERG and gravity models

Remarkably, the ERG framework allows the “economic” information defining gravity models to be translated into opportunely-defined fitnesses. Equation (44) provides a clear example. Another example is provided by the following definition

$$p_{ij}^{\text{GM}} = \frac{z \text{GDP}_i \text{GDP}_j e^{-\phi f(d_{ij})}}{1 + z \text{GDP}_i \text{GDP}_j e^{-\phi f(d_{ij})}} \quad (56)$$

where  $f(d_{ij})$  is an increasing function of the geographic distance  $d_{ij}$  between countries  $i$  and  $j$ . The simplest functional form  $f(d_{ij}) = d_{ij}$  comes from considering the Hamiltonian

$$H(\mathbf{W}|z, \phi) = - \sum_{i=1}^N (k_i \ln \text{GDP}_i) - L \ln z - F \phi \quad (57)$$

with  $F(\mathbf{A}) = \sum_i \sum_{j(\neq i)} a_{ij} d_{ij}$ . The latter term quantifies to what extent the topological structure of the network fills the geometric space in which the network itself is embedded [105]. As usual, the unknown parameters must be estimated by solving the equations  $\hat{L} = \langle L \rangle$  and  $\hat{F} = \langle F \rangle$ . In a sense, eq. (56) defines the closest network-based model to traditional gravity models.

The main difference between the approach proposed here and the traditional economic one becomes evident upon sketching the derivation of the so-called *zero-inflated gravity models* (ZIGM) [106]. In order to prevent this model from predicting a fully-connected network (the same limitation of the MaxEnt recipe), a probability coefficient reading

$$p_{ij}^{\text{ZIGM}} = \frac{1}{1 + e^{-\vec{\phi} \cdot \vec{C}_{ij}}} \quad (58)$$

is assumed to control for the presence of a link between any two nodes  $i$  and  $j$ . The vector  $\vec{\phi}$  of unknowns is estimated by considering the elements  $a_{ij}$  of the adjacency matrix to be the dependent variables, and the quantities usually employed to fit a gravity model (as the countries GDP, their geographic distances, etc.—*i.e.*, a whole vector  $\vec{C}$  of quantities for each pair of nodes) to play the role of explanatory variables. Both are then used to define the likelihood function for the actual network configuration  $\hat{\mathbf{G}}$ , which is as usual maximized with respect to  $\vec{\phi}$ . Once a matrix of probability coefficients is obtained, only the nodes pairs satisfying the condition  $p_{ij}^{\text{ZIGM}} \geq \hat{\rho}$  (where  $\hat{\rho}$  is the known density of links) are actually linked [106].

Zero-inflated gravity models and network-based gravity models differ in the amount of information needed to be fully specified. While the former require the whole adjacency matrix of a given (economic) network to be fully specified, the latter only require the knowledge of global (marginal) information. Remarkably, despite the much smaller amount of information needed, network-based gravity models perform much better than zero-inflated gravity models [106].

### 3.3.9. A remark on the ensemble methods

One of the reasons of the attractiveness of the “ensemble methods” lies in the possibility they offer to generate *different* topological structures that satisfy the *same* (weighted) constraints. This feature can be used to disentangle the impact of marginals such as the *balance sheets* of a financial system and of network structural details on the outcome of a dynamical process like the spreading of financial distress [93].

In order to generate realistic scenarios, however, some kind of topological information must be accessible. In the optimal case, the whole degree sequence of a real network would be available (and, thus, used as additional constrain beside the weighted marginals, via the DECM or its two-step version). Otherwise, if a more aggregate knowledge on the system is available (like the total number of links, or the degrees of a particular subset of nodes), an additional assumption is needed to make the best use out of such information. From the many attempts done so far, it seems that a preliminary estimation of the degree sequence (as in the bootstrapped DECM scheme) enhances the performance of a given reconstruction method, an evidence that explains the superiority of the algorithm in [98] with respect to, *e.g.*, the algorithm in [84]—although both are defined by exactly the same information [81]. From a quantitative point of view, providing a realistic estimate of the degrees of nodes from aggregate information implies having a good fitness ansatz and a realistic estimate of the whole network density. The latter requirements ultimately means having an estimate for the parameter  $z$  to be used in eq. (45).

### 3.4. Shannon-like approaches to reconstruction

The reconstruction algorithms presented in the previous subsections build on the constrained maximization of Shannon entropy, or closely-related functionals like the KL divergence. Shannon entropy is however only one out of many possible functionals that can be taken to extremes under the constraints representing the accessible information.

#### 3.4.1. Spectral entropy

Among Shannon-like functionals, entropic measures exist that are inspired by quantum physics. *Spectral entropy*, also known as *Von Neumann entropy*, deserves a special mention [107]. It is defined as

$$S^{\text{VN}} = \text{Tr}[\Xi \ln \Xi] = \sum_{m=1}^N \xi_m(\Xi) \ln \xi_m(\Xi) \quad (59)$$

*i.e.*, as the Shannon entropy of a probability distribution induced by the eigenvalues  $\{\xi_m\}_{m=1}^N$  of the matrix  $\Xi$ . In quantum physics, the density matrix  $\Xi$  describes a system that can be found in one of a set of pure states with different probabilities (precisely defined by the eigenvalues of  $\Xi$ ): in order to employ this concept in network theory, the density matrix has to be re-expressed in terms of network quantities. A network-based version of the density matrix satisfying the properties of positive semi-definiteness and trace unitarity has been defined as [108]<sup>8</sup>:

$$\Xi = \frac{e^{-\beta \mathbf{L}}}{Z}. \quad (60)$$

---

<sup>8</sup>Other proposals like  $\Xi = \frac{\mathbf{L}}{\text{Tr}[\mathbf{L}]}$  do not satisfy the (sub)additivity property [108].

where  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  is the Laplacian matrix ( $\mathbf{D}$  is a diagonal matrix of nodes degrees) with elements  $L_{ij} = k_i \delta_{ij} - a_{ij} \forall i, j$  and  $Z = \text{Tr}[e^{-\beta \mathbf{L}}]$ .

This approach ultimately boils down to the calculation of the divergence between the spectral density of an operator associated with the empirical graph and that of the corresponding operator associated with a graph model [108]. In principle this allows to optimize the parameters of the model using a sort of quantum analogue of the method described in section 2.2.

### 3.4.2. The Cressie-Read family of power divergences

A whole family of functionals to be extremized to reconstruct partially known networks, generalizing the usual Shannon entropy or Kullback-Leibler divergence (see subsection 3.1.2), is represented by the so-called *Cressie-Read power divergences*. The latter can be compactly written as

$$I(\mathbf{P}, \mathbf{Q}, \gamma) = \frac{1}{\gamma(\gamma+1)} \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \left[ \left( \frac{P(\mathbf{G})}{Q(\mathbf{G})} \right)^\gamma - 1 \right] \quad (61)$$

with the real parameter  $\gamma$  indexing the members of the family. Equation (61) generalizes the KL divergence and provides an alternative measure of “distance” between any two probability distributions  $\mathbf{P}$  and  $\mathbf{Q}$ . Notice that even if  $I(\mathbf{P}, \mathbf{Q}, \gamma)$  is not a proper metric distance for all values of  $\gamma$ , the properties it satisfies are nonetheless useful for quantifying the extent to which any two distributions differ [109]. More specifically,

- $I(\mathbf{P}, \mathbf{Q}, \gamma)$  is a continuous function of all its arguments  $\{P(\mathbf{G})\}_{\mathbf{G} \in \mathcal{G}}, \{Q(\mathbf{G})\}_{\mathbf{G} \in \mathcal{G}}$ ;
- $I(\mathbf{P}, \mathbf{Q}, \gamma) \geq 0$ , with equality if and only if  $P(\mathbf{G}) = Q(\mathbf{G}), \forall \mathbf{G}$ ;
- $I(\mathbf{P}, \mathbf{Q}, \gamma)$  is invariant under the addition of events with zero probability;
- $I(\mathbf{P}, \mathbf{Q}, \gamma)$  is log-additive<sup>9</sup>;
- the functionals indexed by values of the parameter  $\gamma \in (-1, 0)$  satisfy the triangle inequality;
- the only functional representing a proper metric distance (related to the Matusita distance) is the one characterized by  $\gamma = -1/2$ .

Since  $\mathbf{Q}$  is often intended as summarizing the prior information about the system, the prescription to search for the probability distribution  $\mathbf{P}$  which is maximally non-committal with respect to the missing information can be translated into the request of *minimizing* the divergence from  $\mathbf{Q}$  to  $\mathbf{P}$ . In the case constraints are present, this (first) optimization step leads to recover an expression for  $\mathbf{P}$  which depends on a vector of unknown Lagrange multipliers:

$$\min_{\mathbf{P}} \left\{ I(\mathbf{P}, \mathbf{Q}, \gamma) - \lambda_0 \left[ \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) - 1 \right] - \sum_{m=1}^M \lambda_m \left[ \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) C_m(\mathbf{G}) - \langle C_m \rangle \right] \right\}; \quad (62)$$

<sup>9</sup>The log-additivity property reads  $\ln[1 + \theta(\theta+1)I(\mathbf{P}, \mathbf{Q}, \gamma)] + \ln[1 + \theta(\theta+1)I(\mathbf{R}, \mathbf{S}, \gamma)] = \ln[1 + \theta(\theta+1)I(\mathbf{P} \times \mathbf{R}, \mathbf{Q} \times \mathbf{S}, \gamma)]$  with  $\mathbf{P} \times \mathbf{R}$  and  $\mathbf{Q} \times \mathbf{S}$  indicating the tensor product of the two involved probability distributions [110].

the second step of the whole procedure prescribes to substitute the recovered expression of  $\mathbf{P}$  into  $I$  itself and optimize  $I(\vec{\lambda})$  with respect to the unknown parameters  $\vec{\lambda}$  [109].

Equation (62) generalizes the two principles lying at the basis of the ERG formalism introduced in the previous sections. As  $\gamma$  varies, the functional describing the divergence between  $\mathbf{P}$  and  $\mathbf{Q}$  varies as well. Two noteworthy examples are retrieved by solving the following limits

$$\lim_{\gamma \rightarrow 0} I(\mathbf{P}, \mathbf{Q}, \gamma) = D_{\text{KL}}(\mathbf{P}||\mathbf{Q}) = \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln \left( \frac{P(\mathbf{G})}{Q(\mathbf{G})} \right), \quad (63)$$

$$\lim_{\gamma \rightarrow -1} I(\mathbf{P}, \mathbf{Q}, \gamma) = D_{\text{KL}}(\mathbf{Q}||\mathbf{P}) = \sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) \ln \left( \frac{Q(\mathbf{G})}{P(\mathbf{G})} \right) \quad (64)$$

*i.e.*, the KL divergence between  $\mathbf{P}$  and  $\mathbf{Q}$  and between  $\mathbf{Q}$  and  $\mathbf{P}$ . Whenever the maximally uninformative prior is adopted,  $Q(\mathbf{G}) = \frac{1}{|\mathcal{G}|} \forall \mathbf{G} \in \mathcal{G}$ , the well-known expressions

$$D_{\text{KL}}(\mathbf{P}||\mathbf{Q}) = \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln P(\mathbf{G}) + \ln |\mathcal{G}|, \quad (65)$$

$$D_{\text{KL}}(\mathbf{Q}||\mathbf{P}) = - \sum_{\mathbf{G} \in \mathcal{G}} \frac{\ln P(\mathbf{G})}{|\mathcal{G}|} + \ln |\mathcal{G}|, \quad (66)$$

are recovered, respectively defining (up to a sign) the Shannon entropy functional and the likelihood functional [109]. Notice that, for  $\gamma \rightarrow 0$ , minimizing  $I$  consistently translates into maximizing Shannon entropy, thus retrieving the procedure described previously.

In order to use the framework described above for reconstruction purposes, the most general problem of inferring the entries  $\{n_{ij}\}_{i=1\dots I, j=1\dots J}$  of a rectangular matrix by using the information provided by marginals  $n_{i\cdot} = \sum_j n_{ij} \forall i$  and  $n_{\cdot j} = \sum_i n_{ij} \forall j$  must be restated in probabilistic terms. As illustrated in table 1, upon introducing the variables  $p_{ij} = n_{ij}/n_{i\cdot} \forall i, j$ , and further dividing all entries by  $n$  (thus inducing the definitions  $x_i = n_{i\cdot}/n \forall i$  and  $y_j = n_{\cdot j}/n \forall j$ ), providing a numerical estimate of the table entries translates into estimating the entries of the matrix  $\mathbf{P}$  appearing within the set of linear equations

$$\mathbf{y} = \mathbf{xP}. \quad (67)$$

Indeed the entries of  $\mathbf{P}$  can be *formally* interpreted as probability coefficients, defined as fractions of marginals. This position, in turn, allows a problem formally analogous to the one stated in eq. 62 to be defined as

$$\min_{p_{jk}} \left\{ I(\{p_{jk}\}, \{q_{jk}\}, \gamma) - \sum_j \beta_j \left( \sum_k p_{jk} - 1 \right) - \sum_k \alpha_k \left( \sum_j p_{jk} x_j - y_k \right) \right\} \quad (68)$$

and a solution of the form  $I(\{p_{jk}\}, \{q_{jk}\}, \gamma) = \frac{1}{\gamma(\gamma+1)} \sum_j \sum_k p_{jk} \left[ \left( \frac{p_{jk}}{q_{jk}} \right)^\gamma - 1 \right]$  to be found. Notice that choosing  $\gamma \rightarrow 0$  and a maximally uninformative prior reduces to the

$n_{11}$	$n_{12}$	$n_{1\cdot}$	$\longrightarrow$	$p_{11}n_{1\cdot}$	$p_{12}n_{1\cdot}$	$n_{1\cdot}$	$\longrightarrow$	$p_{11}x_1$	$p_{12}x_1$	$x_1$
$n_{21}$	$n_{22}$	$n_{2\cdot}$		$p_{21}n_{2\cdot}$	$p_{22}n_{2\cdot}$	$n_{2\cdot}$		$p_{21}x_2$	$p_{22}x_2$	$x_2$
$n_{31}$	$n_{32}$	$n_{3\cdot}$		$p_{31}n_{3\cdot}$	$p_{32}n_{3\cdot}$	$n_{3\cdot}$		$p_{31}x_3$	$p_{32}x_3$	$x_3$
$n_{\cdot 1}$	$n_{\cdot 2}$	$n$		$n_{\cdot 1}$	$n_{\cdot 2}$	$n$		$y_1$	$y_2$	$1$

Table 1: Pictorial representation of the ill-posed inverse problem concerning the inference of the entries of a table, using only the information provided by marginals. Upon introducing the variables  $p_{ij} = n_{ij}/n_{i\cdot}$  and dividing the entries by  $n$  (whence the definitions  $x_i = n_{i\cdot}/n$ ,  $y_i = n_{\cdot i}/n$ ) a constrained-optimization problem naturally emerges [111].

usual exponential form coming from the minimization of the KL divergence<sup>10</sup>

$$p_{jk}^{(0)} = \frac{e^{\alpha_k x_j}}{\sum_{k'} e^{\alpha_{k'} x_j}}. \quad (69)$$

Other choices, instead, lead to coefficients described by different functional forms. As an example, adopting the functional induced by the choice  $\gamma \rightarrow -1$  leads to the expression

$$p_{jk}^{(-1)} = -\frac{1}{\alpha_k x_j + \beta_j}. \quad (70)$$

In general, for an arbitrary choice of the exponent  $\gamma$ , the functional form of the entries of  $\mathbf{P}$  induced by the functional  $I(\{p_{jk}\}, \{q_{jk}\}, \gamma)$  differs substantially from the “usual” exponential one shown in eq. (69). This, in turn, induces a reconstruction procedure whose performance is potentially very different from that of the KL-based approach.

### 3.4.3. Other entropic families

Just like the Cressie-Read functionals provide a generalization of the Kullback-Leibler divergence, generalizations of the Shannon entropy are provided by *Renyi entropies* [112] and *Tsallis entropies* [113]. These entropies depend on a free parameter and include Shannon entropy as a limiting case. More specifically, Renyi entropies are defined as

$$S_\alpha = \frac{\ln [\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G})^\alpha]}{1 - \alpha} \quad (71)$$

(with  $\alpha$  being a non-negative parameter, different from 1) and satisfy the additivity property<sup>11</sup>.

Tsallis entropies can be axiomatically defined upon generalizing the fourth Shannon-Khinchin axiom (see section 2). While this axiom unambiguously identifies Shannon entropy, substituting it with the requirement that  $S_q(W_{A+B}) = S_q(W_A) + S_q(W_{B|A}) +$

<sup>10</sup>This solution is formally equivalent to the MaxEnt one. However, since this approach is typically used to infer election percentages or estimate the purchases of a basket of commodities (*i.e.*, to reconstruct tables where zero entries are practically never observed), the evidence that non-zero marginals cannot induce zero entries does not constitute a problem [75].

<sup>11</sup>The additivity property reads  $S_\alpha(\mathbf{P} \times \mathbf{Q}) = S_\alpha(\mathbf{P}) + S_\alpha(\mathbf{Q})$  with  $\mathbf{P} \times \mathbf{Q}$  indicating the tensor product of the two involved probability distributions.

$(1 - q)S_q(W_A)S_q(W_{B|A})$  leads to the *only* functional that satisfies such a new set of axioms<sup>12</sup>:

$$S_q = \frac{1 - \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G})^q}{q - 1}. \quad (72)$$

Remarkably,  $S_q$  can be employed to define a non-extensive version of the ERG formalism, whose derivation proceeds along similar lines. For example, imposing only the normalization condition leads to the functional  $\mathcal{L}_q[P] = S_q - \lambda_0 [\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) - 1]$  which is maximized by the uniform distribution  $P(\mathbf{G}) = \frac{1}{|\mathcal{G}|} \forall \mathbf{G}$ . Imposing less trivial constraints, however, has not been attempted yet: as a consequence, a thorough comparison between the goodness of the reconstruction performances induced by extensive and non-extensive entropies is still missing.

### 3.5. Beyond Shannon entropy: alternative approaches to reconstruction

After having revised the existing Shannon-based and Shannon-like approaches to reconstruction, we now review algorithms that are not based on the maximization of Shannon-inspired functionals.

#### 3.5.1. The “copula” approach

The first “alternative” approach to entropy-based reconstruction is, actually, the closest one to traditional MaxEnt. The “copula” method, in fact, adopts the same philosophy and uses the entries of a given matrix to define the support of a probability distribution to be estimated; at the same time, however, it provides a more general solution to the problem.

The MaxEnt prescription represents the simplest method for estimating a bivariate probability distribution  $P_{xy}(X, Y)$ , given the two marginal distributions  $P_x(X)$  and  $P_y(Y)$ . Indeed, maximizing the Shannon entropy

$$S = - \sum_i \sum_j P_{xy}(X_i, Y_j) \ln P_{xy}(X_i, Y_j) \quad (73)$$

under the constraints represented by normalization  $\sum_i \sum_j P_{xy}(X_i, Y_j) = 1$  and the two marginal distributions  $P_x(X_i) = \sum_j P_{xy}(X_i, Y_j) \forall i$  and  $P_y(Y_j) = \sum_i P_{xy}(X_i, Y_j) \forall j$  leads precisely to the MaxEnt-like estimation

$$P_{xy}^{\text{MC}}(X_i, Y_j) = P_x(X_i)P_y(Y_j). \quad (74)$$

The recipe above, however, can be generalized by introducing the so-called *copula functions*. The rationale for employing copulas is provided by Sklar’s theorem, which states that every multivariate cumulative distribution function (CDF) can be expressed in terms of its marginal CDFs<sup>13</sup> (say,  $F_x(X)$ ,  $F_y(Y)$ , etc.) and a copula function  $\mathcal{C}$  which, as the name suggests, “couples” them:

$$F_{xy\dots}(X, Y \dots) = \mathcal{C}[F_x(X), F_y(Y) \dots]. \quad (75)$$

<sup>12</sup>The parameter  $q$  quantifies the degree of non-extensivity of such a functional.

<sup>13</sup>Sklar’s theorem requires the marginals to be continuous. Whenever discrete datasets are considered, the results described here can be thought as being applied to the kernel density estimations of the corresponding (discrete) CDFs.

In our case, the marginal CDFs are those of the constraints (defined by eqs. (5)) to be estimated from data. The choice of the particular copula function, on the other hand, is completely arbitrary<sup>14</sup>. The authors in [115] use the Gumbel copula, defined as

$$\mathcal{C}_{ij}^{\text{Gumbel}}(\theta) = e^{-\left[(-\ln F_x(s_i^{\text{out}}))^{\theta} + (-\ln F_y(s_j^{\text{in}}))^{\theta}\right]^{\frac{1}{\theta}}} \quad (76)$$

where the only parameter  $\theta$  quantifies the dependence between the marginals. Remarkably, the parameter estimation can be carried out by maximizing the likelihood-like function

$$\mathcal{L}(\hat{\mathbf{G}}|\theta) = \sum_{i=1}^N \sum_{j=1}^N \ln \mathcal{C}[F_x(\hat{s}_i^{\text{out}}), F_y(\hat{s}_j^{\text{in}}) \dots] \quad (77)$$

with respect to  $\theta$ . Once the model parameter has been estimated, a matrix whose entries are (interpreted as) probability coefficients is obtained. Finally, the IPF method is employed to readjust the sums along rows and columns and recover the observed marginals.

Note that if the so-called “independent” copula function, defined by  $\mathcal{C}[F_x(X), F_y(Y) \dots] = F_x(X)F_y(Y) \dots$ , is used, then the MaxEnt estimation is recovered. And as for MaxEnt, the copula approach cannot reproduce the topological structure of sparse networks [115].

### 3.5.2. A Bayesian approach to network reconstruction

The major difference between likelihood-based methods (as those described in the previous sections) and Bayesian methods lies in the role played by model parameters. Very broadly speaking, while likelihood-based methods provide a recipe for estimating the unknown parameters on the basis of the observations, Bayesian approaches treat the unknown parameters as (additional) random variables, described by properly-defined *prior* probability distributions, whose parameters (called *meta-parameters*) are chosen *a priori*.

The first example of this second kind of approach to network modeling is provided by the *fitness model* [52], resting upon the same ideas lying at the basis of the ERG approach:

- each node  $i$  is described by a hidden variables  $x_i$ , representing its “fitness”; generally speaking, this is a real numbers supposedly quantifying the importance of that node in the network, and is drawn from a given probability distribution  $\nu(x)$ ;
- any two nodes  $i$  and  $j$  establish a connection according to a coupling function  $f(x_i, x_j)$  that, for undirected networks, is symmetric in the hidden variables assigned to nodes  $i$  and  $j$ .

The main difference with respect to the ERG approach lies in the *a priori* choice of both the functional form of the coupling function  $f$  and the distribution  $\nu$  from which fitnesses are drawn. The fitness model can be straightforwardly implemented by adapting the discrete formulas derived within the ERG framework to the continuous case. For

---

<sup>14</sup>Notably, a maximum-entropy recipe for estimating copulas has been recently proposed [114].

instance, in the undirected case, the Bayesian derivation of the nodes degrees and of the total number of links reads

$$k_i = \sum_{j(\neq i)=1}^N f(x_i, x_j) \longrightarrow k_i(x) = (N-1) \int f(x, y) \nu(y) dy, \quad (78)$$

$$L = \sum_{i=1}^N \sum_{j(<i)=1}^N f(x_i, x_j) \longrightarrow L = \frac{N(N-1)}{2} \int \int f(x, y) \nu(x) \nu(y) dx dy, \quad (79)$$

where the integrations over the support of the distribution  $\nu$  are necessary to account for the fitness variability.

Remarkably, several combinations of  $f$  and  $\nu$  lead to recover power-law degree distributions. In particular, both the intuitive combination

$$f(x, y) = zxy, \quad \nu(x) \propto x^{-2} \quad (80)$$

and the highly non-trivial combination

$$f(x, y) = \Theta(x + y - z), \quad \nu(x) = e^{-x} \quad (81)$$

lead to  $P(k) \propto k^{-2}$ . This result points out that a number of topological properties believed to arise only as a consequence of microscopic dynamic processes (as the one described by the preferential attachment mechanism and its variants) can, instead, be replicated also via a static model [52]. In other words, whenever preferential attachment does not represent a plausible mechanism, it is reasonable to imagine that any two vertices establish a connection when the link creates a mutual benefit, depending on some intrinsic node property.

The aforementioned approach has been recently extended to account also for weights, through an algorithm which is not dissimilar in spirit from the DECM. More specifically, the model introduced in [116] is described by the following probability distribution for link weights:

$$q_{ij}^{\text{Bayes}}(w) = \begin{cases} 1 - p_{ij}^{\text{Bayes}} & \text{if } w = 0, \\ p_{ij}^{\text{Bayes}} \theta_{ij} e^{-\theta_{ij} w} & \text{if } w > 0. \end{cases} \quad (82)$$

(with  $\theta_{ij} > 0$ ). Hence, as in the DECM, distinct links are independent, and while the presence of the link is described by a Bernoulli trial, the value of its weight is set according to an exponential distribution. Note that the latter is the continuous version of the geometric distribution, which is obtained by ERG-based models upon assuming discrete weights. The philosophy of the fitness model is then encoded into the choice of functional forms reading

$$p_{ij}^{\text{Bayes}} = f(x_i + x_j), \quad \theta_{ij} = G_{\zeta, \eta}^{-1}(e^{-x_i}) + G_{\zeta, \eta}^{-1}(e^{-x_j}). \quad (83)$$

In these expressions,  $G_{\zeta, \eta}^{-1}$  is the quantile function of the Gamma distribution with positive shape and scale parameters  $\zeta$  and  $\eta$ , drawn from an a priori distribution  $\pi(\zeta, \eta)$ , while fitnesses are drawn from the distribution  $\nu(x) = e^{-x}$  and  $f$  is defined as in [116] such that the degree distribution exhibits a power law. An homogeneous version of the model

has been also introduced, defined by the choices  $p_{ij} = p \sim \text{Beta}(\alpha, \beta)$  and  $\theta_{ij} = \theta \sim \text{Gamma}(\gamma, \delta)$ .

Since this model induces an entire ensemble of networks, the expected value of the quantities of interest can be computed only after introducing a sampling procedure on the ensemble. The authors of [116] adopt a *Gibbs sampler* working as follows.

- The sampler is initialized with a matrix  $\mathbf{W}^{(0)}$ . When considering the homogeneous version, the initial matrix is generated via the Erdős-Rényi model whose only parameter is required to match the observed average degree. Since the initial configuration is required to satisfy the conditions  $s_i^{\text{out}}(\mathbf{W}^{(0)}) < \hat{s}_i^{\text{out}}$  and  $s_i^{\text{in}}(\mathbf{W}^{(0)}) < \hat{s}_i^{\text{in}} \forall i$ , the maximum-flow algorithm [117] is employed to obtain a matrix  $\mathbf{W}^{(1)}$  matching the constraints exactly.
- The whole ensemble of configurations is obtained by “perturbing”  $\mathbf{W}^{(1)}$ . Such perturbations generalize the *local rewiring algorithm* according to the following rules: 1) the dimension  $k$  of the sub-matrices to be updated is chosen and a set  $\kappa$  of  $k$  pairs of indices is selected; 2) the entries of  $\mathbf{W}^{(1)}$  are updated according to the rule

$$\mathbf{W}_{\kappa_i}^{(n+1)} = \mathbf{W}_{\kappa_i}^{(n)} + (-1)^{i+1} \Delta \quad (84)$$

with  $\kappa_i$  indicating the  $i$ -th pair of indices (*e.g.*, the  $r$ -th row and the  $c$ -th column) and  $\Delta \in [-\min_{i, \text{odd}} \mathbf{W}_{\kappa_i}^{(n)}, \min_{i, \text{even}} \mathbf{W}_{\kappa_i}^{(n)}]$ .

Such a sampling process does not leave unexplored regions of the space of configurations: the existence of a sequence of Gibbs moves allowing for a transition from any matrix compatible with the given constraints to any other is, in fact, guaranteed [116]). And although the algorithm allows one to generate networks characterized by different topological structures, every configuration satisfies the constraints defined by eqs. (5) exactly.

*An “empirical” Bayesian approach to network reconstruction.* The same authors of [116] have recently developed an “adjustable” version of their Bayesian approach [93]. In this model, the linking probability between any two nodes  $i$  and  $j$  is assumed to be  $p_{ij}^{\text{E-Bayes}} = \frac{zx_i x_j}{1+zx_i x_j}$ , with  $x_i = \hat{s}_i^{\text{out}} + \hat{s}_i^{\text{in}}$  (naturally, this position better models undirected networks).

### 3.5.3. A comment on the Bayesian approaches to reconstruction

Although the three aforementioned algorithms have been labeled as Bayesian, they share features of both likelihood-based and genuinely Bayesian methods. All of them are, in fact, characterized by the presence of one, or more, free parameters not to be drawn from *a priori* distributions, but to be estimated via properly-defined likelihood conditions.

For what concerns the model in [52], the only free parameter  $z$  is fixed by imposing the condition  $\hat{L} = \langle L \rangle$ , *i.e.*, by substituting either eq. (80) or eq. (81) into eq. (79) and solving the resulting equation for  $z$ . A main difference with the fitness-induced ERG formalism [98] remains in the way fitnesses are dealt with. In a sense, the fitness-induced

ERG formalism represents the likelihood-based analogue of the Bayesian approach discussed here: there, the information on the fitness distribution is completely ignored and just a point-estimation is carried out; here, the fitness variability is completely accounted for.

The model in [116], on the other hand, needs to be calibrated whenever used to reconstruct real-world networks. After assuming that many of the free parameters coincide, the authors are able to solve the equation  $\hat{W} = \langle W \rangle$ , which results in  $\theta_{ij} \equiv \theta = \sum_{ij} p_{ij}^{\text{Bayes}} / \hat{W}$ . Similarly, in [93] the free parameter  $z$  is “adjusted” in order to ensure that the expected density matches the observed one.

This discussion highlights the main limitation of using Bayesian approaches as reconstruction methods. Indeed, the freedom coming from treating model parameters as random variables does not necessarily help in reproducing the features of *specific* real-world configurations. As the authors in [116] explicitly recognize, Bayesian models need to be (at least partially) tuned in order to be used as reconstruction models. Whenever this step is missing, the arbitrariness in choosing the *a priori* distributions can be better employed to generate scenarios—useful, for instance, to obtain confidence intervals for stress tests outcomes.

#### 3.5.4. The Montagna & Lux approach

Link probability coefficients can be also defined *ad hoc*, without any explicit derivation from a given optimization principle. For instance, the authors in [118] consider the following forms:

$$p_{ij}^{\text{M-L-1}} = d_1 (\hat{s}_i^{\text{out}})^\alpha (\hat{s}_j^{\text{out}})^\beta, \quad (85)$$

$$p_{ij}^{\text{M-L-2}} = d_2 (\hat{s}_i^{\text{out}} + \hat{s}_j^{\text{out}}), \quad (86)$$

$$p_{ij}^{\text{M-L-3}} = d_3 \Theta(\hat{s}_i^{\text{out}} + \hat{s}_j^{\text{out}} - z), \quad (87)$$

where the parameters  $d_1, d_2, d_3$  are used to adjust the density of the network. This model thus follows the philosophy of the fitness model, by assuming that any two nodes  $i$  and  $j$  are connected with a probability  $p_{ij}(\hat{s}_i^{\text{out}}, \hat{s}_j^{\text{out}})$ . An ensemble of network configurations is then generated according to

$$a_{ij} = \begin{cases} 0 & \text{with probability } 1 - p_{ij}^{\text{M-L-h}} \\ 1 & \text{with probability } p_{ij}^{\text{M-L-h}} \end{cases} \quad (88)$$

(with  $h = 1, 2, 3$  and with the additional rule of eliminating loops generated when  $a_{ij} = a_{ji} = 1$ ). Once a topological structure has been generated, the weights of the realized connections are set proportionally to the “size” of involved nodes as

$$w_{ij}^{\text{M-L-h}} = \hat{s}_i^{\text{out}} \left( \frac{p_{ij}^{\text{M-L-h}}}{\sum_{\{a_{ij}=1\}} p_{ij}^{\text{M-L-h}}} \right). \quad (89)$$

In the original paper [118], this approach is not used for network reconstruction, but rather to generate a financial interbank network assuming banks’ total interbank assets to be power-law distributed—a choice that also leads to power-law distributed degrees. Additionally, in the original paper total assets  $\hat{a}_i$  are used instead of total interbank assets  $\hat{s}_i^{\text{out}}$  as fitness values for each node  $i$  (see section 4.3.1). The formulation we present here is equivalent though, since a linear proportionality relation  $\hat{a}_i(1 - \theta) = \hat{s}_i^{\text{out}}$  is assumed [119, 118].

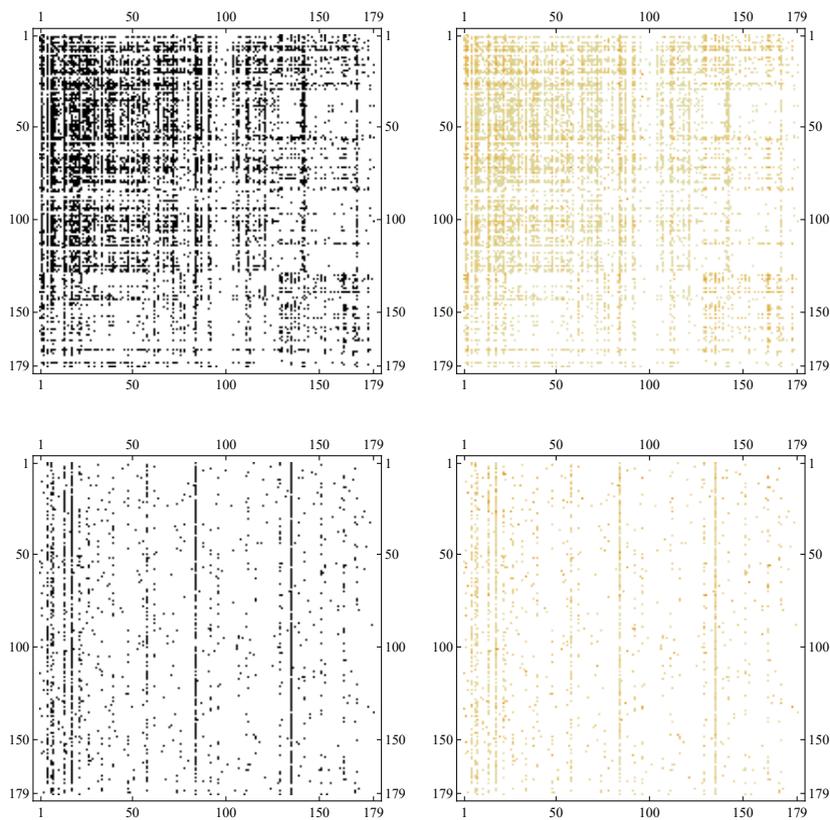


Figure 4: Comparison between the observed adjacency matrix of the eMID network in 2003 (top panels) and its reconstructed version according to the method by Halaj & Kok described in section 3.5.5 (bottom panels). Left panels represent binary adjacency matrices with black/white denoting the presence/absence of connections, whereas, right panels represent weighted adjacency matrices with color intensity denoting the weight of connections.

### 3.5.5. The probability map of Halaj & Kok

Halaj & Kok further define link probabilities assuming the additional information on nodes' *membership* to groups [120]. In the specific model implementation for interbank networks, each node (bank) belongs to a geographic area (*i.e.*, a country), and link probability are expressed as fraction of node out-strengths aggregated within countries (which are assumed to be known):

$$p_{ij}^{\text{H-K}} = \frac{\hat{w}_{g_i, g_j}}{\hat{w}_{g_i}}, \quad (90)$$

where  $\hat{w}_{g_i, g_j} = \sum_{i \in g_i} \sum_{j \in g_j} \hat{w}_{ij}$  is the total observed weight from area  $g_i$  to area  $g_j$  and  $\hat{w}_{g_i, \cdot} = \sum_{i \in g_i} \sum_j \hat{w}_{ij}$  is the total weight going out from area  $g_i$ .

Although this algorithm is similar in spirit to fitness models (and specifically to the *stochastic block-model*, see section 4.2.3), its formalism differs from that of ERG. Indeed, the network structure is determined according to the following steps:

- a pair of nodes is randomly drawn out of the set of possible ones;
- the link is realized according to the corresponding probability  $p_{ij}^{\text{H-K}}$ ;
- if the link is retained, a random number  $r_{ij} \in [0, 1]$  is generated, in order to determine the percentage of out-strength value of node  $i$ , and in-strength value of node  $j$ , assigned to the weight  $w_{ij}$ ;
- the residual magnitude of the out-strength of  $i$  and the in-strength of  $j$  is updated accordingly, *i.e.*,  $[s_i^{\text{out}}]^{(n)} = (1 - r_{ij}^{(n)})[s_i^{\text{out}}]^{(n-1)}$  and  $[s_j^{\text{in}}]^{(n)} = (1 - r_{ij}^{(n)})[s_j^{\text{in}}]^{(n-1)}$  (with  $n$  indicating the  $n$ -th iteration of the algorithm);
- the steps above are repeated until  $s_i^{\text{out}} \simeq \hat{s}_i^{\text{out}}$  and  $\hat{s}_i^{\text{in}} \simeq s_i^{\text{in}} \forall i$  (constraints may be satisfied only approximately, because of the purely numerical nature of the last step of the algorithm).

Besides requiring a substantial amount of additional information with respect to other reconstruction method, this approach treats all nodes belonging to the same geographic area as equivalent (the only variability being provided by the (random) allocation of fraction of weights): The structure of the sub-network linking any two geographic areas is random, and may not reflect its observed counterpart (see fig. 4).

### 3.5.6. The Minimum Density algorithm

As we have already stressed, the main reason for defining reconstruction algorithms alternative to MaxEnt is the densely-connected nature of the configurations predicted by eq. (13), which misrepresent the actual network structures (and may lead to underestimate the systemic risk). In order to overcome the intrinsic limitations of a complete network structure, the opposite approach of *minimizing* the link density while satisfying the observed constraints has been recently devised [121].

Contrarily to MaxEnt, which evenly shares the marginals across all connections, the Minimum Density (MD) algorithm allocates the marginals over the minimum possible number of links (see fig. 5). MD does not rest upon the maximization of Shannon entropy, but on an optimization principle based on minimizing the cost of maintaining connections. The algorithm, in fact, works as follows.

- Define deviations from the marginals to be satisfied

$$\Delta_{s_i^{\text{out}}}^{(n)} = \hat{s}_i^{\text{out}} - [s_i^{\text{out}}]^{(n)} = \hat{s}_i^{\text{out}} - \sum_{j(\neq i)=1}^N w_{ij}^{(n)}, \quad (91)$$

$$\Delta_{s_i^{\text{in}}}^{(n)} = \hat{s}_i^{\text{in}} - [s_i^{\text{in}}]^{(n)} = \hat{s}_i^{\text{in}} - \sum_{j(\neq i)=1}^N w_{ji}^{(n)} \quad (92)$$

where  $w^{(n)}$  is the matrix obtained at the  $n$ -th iteration of the MD algorithm.

- Choose a pair of nodes according to the probability coefficients

$$q_{ij}^{(n)} \propto \max \left\{ \frac{\Delta_{s_i^{\text{out}}}^{(n)}}{\Delta_{s_j^{\text{in}}}^{(n)}}, \frac{\Delta_{s_j^{\text{in}}}^{(n)}}{\Delta_{s_i^{\text{out}}}^{(n)}} \right\} \quad (93)$$

that privilege pairs of nodes where either the “out-strength deficit” of node  $i$  is large with respect to the “in-strength deficit” of node  $j$  or *vice-versa*.

- Link the two selected nodes via a connection weighting

$$w_{ij}^{(n)} = \min \{ \Delta_{s_i^{\text{out}}}^{(n)}, \Delta_{s_j^{\text{in}}}^{(n)} \}, \quad (94)$$

hence corresponding to the largest volume that this pair of nodes can exchange. This step, coupled with the previous one, ensures that each new link is assigned the maximum possible weight needed to satisfy either the marginal  $\hat{s}_i^{\text{out}}$  or the marginal  $\hat{s}_j^{\text{in}}$ . Moreover, as the authors explicitly notice, this updating rule also induces a disassortative topology, in order to reproduce a structural feature observed in many real-world networks.

- Decide if the proposed update must be retained, by evaluating the objective function

$$V(\mathbf{W}^{(n)}) = -cL^{(n)} - \sum_{i=1}^N \left[ \alpha_i \left( \Delta_{s_i^{\text{out}}}^{(n)} \right)^2 + \beta_i \left( \Delta_{s_i^{\text{in}}}^{(n)} \right)^2 \right], \quad (95)$$

with  $c$  quantifying the cost of establishing a link. If  $\Delta_V^{(n)} = V(\mathbf{W}^{(n)} + w_{ij}^{(n)}) > V(\mathbf{W}^{(n-1)})$  the link is retained (note that networks with lower densities have larger values of  $V$ ). If instead  $\Delta_V^{(n)} < 0$ , the likelihood of observing the resulting configuration is evaluated, *i.e.*, the proposed weight is retained with the Metropolis-Hasting probability  $P(\mathbf{W}^{(n)} + w_{ij}^{(n)})/P(\mathbf{W}^{(n-1)}) \propto e^{\Delta_V^{(n)}}$ .

- Repeat the aforementioned steps until all marginals have been allocated.

As the other probabilistic methods discussed so far, the MD method can be employed to generate a whole ensemble of networks, characterized by a value of link density close to the minimum possible one (close because of the variability introduced by the last step) but different topological structures. Note however that real networks with such a low density values are rarely observed. Thus, the main rationale behind the algorithm is not network reconstruction. Rather, the algorithm can be very useful to find an upper bound to systemic risk. As such, it can be successfully employed in conjunction with MaxEnt—which, instead, provides the lower bound, in order to obtain the interval in which the true value of systemic risk must lie. Additionally, in order to test for the supposed dependence of systemic risk on the network density, the authors also provide a more general rule for updating weights, by replacing eq. (94) with  $w_{ij}^{(n)} = \theta \min\{\Delta_{s_i^{\text{out}}}^{(n)}, \Delta_{s_j^{\text{in}}}^{(n)}\}$ : the parameter  $\theta \in [0, 1]$  is used to allocate percentages of marginals, thus relaxing the request of creating a network with minimum density. By tuning  $\theta$ , the whole range of link density values can be explored.

#### 4. Testing the network reconstruction

Now that we described a number of algorithms aiming at reconstructing a given network structure, we focus on some useful methods to *test* the effectiveness of the achieved reconstruction. In particular, three different kinds of indicators will be considered, of *statistical*, *topological* and *dynamical* nature.

##### 4.1. Statistical indicators

With this name, we refer to the entries of the so-called *confusion matrix*, the  $4 \times 4$  table whose elements are the number of *true positives*, *false negatives*, *true negatives* and *false positives*. The use of these indicators is justified upon considering that, from a purely topological perspective, a reconstruction method acts as a binary classifier “deciding” if a given pair of nodes should be linked or not, and whose performance can be thus evaluated in terms of the indices above [122].

To be more explicit, consider the problem of retrieving the position of both present and missing links of an observed binary matrix  $\hat{\mathbf{A}}$ . Denoting an individual network obtained by a given reconstruction method  $\mathbf{A}$ , for each node pair four different alternatives are possible: **a**)  $\hat{a}_{ij} = a_{ij} = 1$ : in this case, an observed link has been correctly predicted (we have a *true positive*); **b**)  $\hat{a}_{ij} = 1$  but  $a_{ij} = 0$ : in this case, an observed link has been incorrectly predicted as being missing (we have a *false negative*); **c**)  $\hat{a}_{ij} = a_{ij} = 0$ : in this case, a missing link has been correctly predicted (we have a *true negative*); **d**)  $\hat{a}_{ij} = 0$  but  $a_{ij} = 1$ : in this case, a missing link has been incorrectly predicted as being present (we have a *false positive*).

The total number of events within one of these four categories can be straightforwardly computed as follows:

$$TP = \mathbf{1}(\hat{\mathbf{A}} \circ \mathbf{A})\mathbf{1}^T, \quad (96)$$

$$FN = \mathbf{1}(\hat{\mathbf{A}} \circ (\mathbf{I} - \mathbf{A}))\mathbf{1}^T, \quad (97)$$

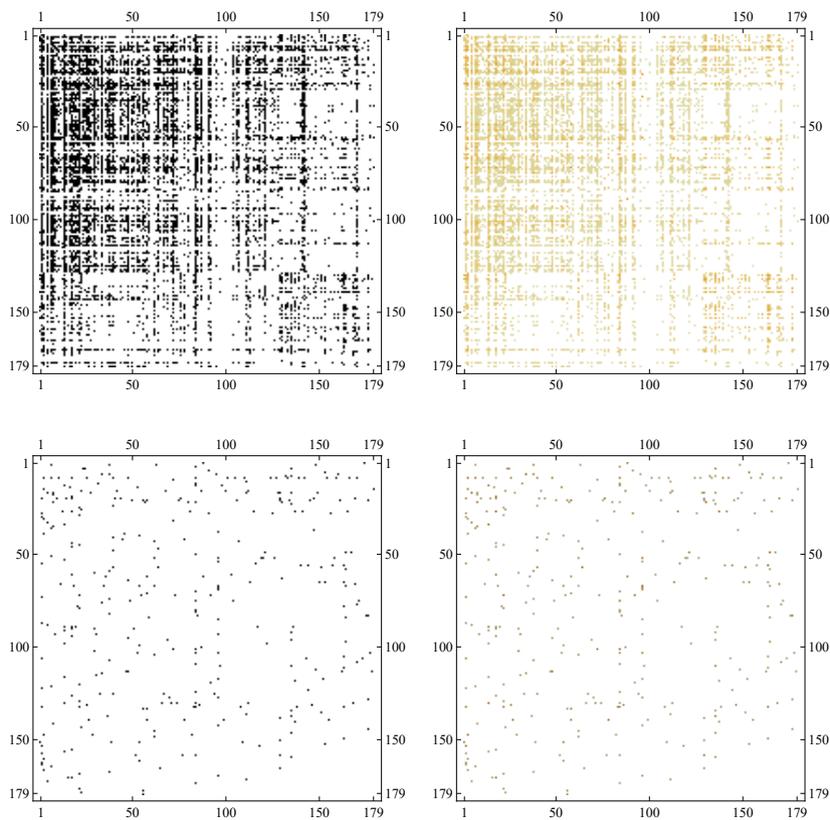


Figure 5: Comparison between the observed adjacency matrix of the eMID network in 2003 (top panels) and its reconstructed version according to the Minimum Density method described in section 3.5.6 (bottom panels). Left panels represent binary adjacency matrices with black/white denoting the presence/absence of connections, whereas, right panels represent weighted adjacency matrices with color intensity denoting the weight of connections.

$$TN = \mathbf{1}((\mathbf{I} - \hat{\mathbf{A}}) \circ (\mathbf{I} - \mathbf{A}))\mathbf{1}^T, \quad (98)$$

$$FP = \mathbf{1}((\mathbf{I} - \hat{\mathbf{A}}) \circ \mathbf{A})\mathbf{1}^T \quad (99)$$

where the symbol  $\circ$  indicates the element-wise product of two matrices,  $\mathbf{1} = (1, 1 \dots 1)$  is the  $N$ -th dimensional row-vector whose entries are all ones and  $\mathbf{I}$  is the  $N \times N$  matrix whose generic entry reads  $I_{ij} = 1 - \delta_{i,j}$ . For instance, the total number of true positives reads  $TP = \sum_i \sum_{j(\neq i)} \hat{a}_{ij} a_{ij}$ . Note that since these four indices sum up to the total number of nodes  $N$ , only three of them are independent. In particular, the last three ones can be compactly re-written in terms of  $TP$ . We have  $FN = \sum_i \sum_{j(\neq i)} \hat{a}_{ij} (1 - a_{ij}) = \hat{L} - TP$ ,  $TN = \sum_i \sum_{j(\neq i)} (1 - \hat{a}_{ij})(1 - a_{ij}) = N(N - 1) - L - \hat{L} + TP$  and  $FP = \sum_i \sum_{j(\neq i)} (1 - \hat{a}_{ij}) a_{ij} = L - TP$ , where  $\hat{L}$  and  $L$  are the total number of links in the observed and reconstructed network, respectively.

The four indices above provide absolute numbers which, by themselves, are of limited usefulness. This is the reason why the information provided by  $TP$ ,  $FN$ ,  $TN$  and  $FP$  is often combined to define “relative” indices. The first of them is the *sensitivity* (or *true positive rate*) [122], defined as

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{\hat{L}} \quad (100)$$

and quantifying the percentage of observed links that are correctly recovered. Note that for the performance a reconstruction method to be deemed satisfactory, the condition of a  $TPR$  value close to 1 is *necessary*, but not *sufficient*. Indeed, a method that overestimates the number of links achieves a high  $TPR$  value by construction (for a fully connected reconstructed network, it is  $TPR = 1$  by definition), but also lacks the ability to identify missing connections. The latter is quantified by the *specificity* (or *true negative rate*) [122], defined as

$$SPC = \frac{TN}{FP + TN} = \frac{TN}{N(N - 1) - \hat{L}} \quad (101)$$

*i.e.*, the percentage of observed missing links that are correctly recovered. The *false positive rate*  $FPR = 1 - SPC$  is the complementary index to  $SPC$  [122].

Thus, intuitively, any “good” reconstruction algorithm should be characterized by a large  $TPR$  value and a low  $FPR$  value (*i.e.*, the better its performance, the closer the  $TPR$  index to 1 and the  $FPR$  index to 0). This observation leads to the classical “graphical” way to visualize the performance of a classifier, by representing it as a point of coordinates  $(FPR, TPR)$  within the unit square of coordinates  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 1)$ ,  $(1, 0)$ . Any perfect classifier is thus to be found on the top-left corner of the square, whereas, a random classifier (predicting an equal number of present and missing links) lies on its main diagonal. The performance of a given reconstruction algorithm can then be quantified as the *area under the curve* (AUC) [122] identified by the three points of coordinates  $(0, 0)$ ,  $(FPR, TPR)$ ,  $(1, 1)$ . A perfect classifier is then characterized by an AUC of 1, a random classifier by an AUC equal to  $1/2$ , and in general for a non-random classifier it is  $1/2 < \text{AUC} \leq 1$ .

An alternative way of evaluating the performance of a reconstruction method is plotting its  $TPR$  versus a fourth index, its *precision* (or *positive predicted value*) [122]

$$PPV = \frac{TP}{TP + FP} = \frac{TP}{L} \quad (102)$$

which measures the percentage of correctly placed links with respect to the total number  $L$  of predicted links. In other words, the  $PPV$  index quantifies the “ability” of a given classifier to predict *only* true positives. Thus, and contrarily to the  $TPR$ , a large  $PPV$  cannot be trivially obtained by dense reconstruction methods.

Finally we consider an index measure the overall performance of a classifier in correctly placing both ones and zeros: the *accuracy* [123], defined as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{N(N - 1)}. \quad (103)$$

Whenever a reconstruction method defines an entire ensemble of candidate matrices, the above quantities have to be evaluated as averages over such ensemble. This can be done using, in eqs. (96-99), the average quantities  $\langle a_{ij} \rangle = p_{ij}$  and  $\langle \mathbf{A} \rangle = \mathbf{P}$  instead of those corresponding to a single realization  $a_{ij}$  and  $\mathbf{A}$ :

$$\langle TP \rangle = \mathbf{1}(\hat{\mathbf{A}} \circ \mathbf{P})\mathbf{1}^T, \quad (104)$$

$$\langle FN \rangle = \mathbf{1}(\hat{\mathbf{A}} \circ (\mathbf{I} - \mathbf{P}))\mathbf{1}^T, \quad (105)$$

$$\langle TN \rangle = \mathbf{1}((\mathbf{I} - \hat{\mathbf{A}}) \circ (\mathbf{I} - \mathbf{P}))\mathbf{1}^T, \quad (106)$$

$$\langle FP \rangle = \mathbf{1}((\mathbf{I} - \hat{\mathbf{A}}) \circ \mathbf{P})\mathbf{1}^T. \quad (107)$$

Using ensemble averages, the difference between dense and sparse reconstruction methods can be better discussed quantitatively. MaxEnt, which we take as the representative of dense reconstruction algorithms, satisfies a relation of the kind  $\langle a_{ij}^{\text{ME}} \rangle = p_{ij} = p \simeq 1 \forall i \neq j$ , leading to

$$\langle TP^{\text{ME}} \rangle \simeq \hat{L} \text{ and } \langle FP^{\text{ME}} \rangle \simeq N(N - 1) - \hat{L}. \quad (108)$$

As a consequence,  $\langle PPV^{\text{ME}} \rangle \simeq \hat{L}/N(N - 1)$ , *i.e.*, the power of the method coincides with the network density. In order to fully understand the importance of this result, we now consider the Directed Random Graph Model (DRGM), defined by the prescription  $p_{ij} = p = \hat{L}/N(N - 1) \forall i \neq j$ . We have  $\langle TPR^{\text{DRGM}} \rangle = \langle FPR^{\text{DRGM}} \rangle = p$  and, most importantly,  $\langle PPV^{\text{DRGM}} \rangle = p$ . In other words, a “random” classifier is not necessarily an algorithm guessing the (binary) value of each entry with probability 1/2: more generally, it is a reconstruction method defined by the DRGM recipe, and whose  $PPV$  represents a lower bound for any reconstruction algorithm. Notice that the MaxEnt method attains such value, thus confirming the weakness of its performance—unless very dense configurations are considered, since  $\langle ACC^{\text{ME}} \rangle = p$ .

Other indicators that have been extensively used to measure the goodness of a reconstruction algorithm are the Hamming distance  $H$ , the Jaccard similarity  $J$ , the cosine

similarity  $\vartheta$ , the Jensen-Shannon divergence  $JS$ , between  $\hat{\mathbf{A}}$  and  $\mathbf{A}$  [124]. Remarkably, whenever dealing with binary matrices, these indices can be rewritten in terms of the four basic quantities  $TP$ ,  $FN$ ,  $TN$ ,  $FP$ :

$$H = FN + FP, \quad (109)$$

$$J = \frac{TP}{FN + TN + FP}, \quad (110)$$

$$\vartheta = \frac{TP}{\hat{L}}, \quad (111)$$

$$JS = \frac{FN}{2\hat{L}} \ln(2\hat{L}) + \frac{FP}{2L} \ln(2L) - TP \left( \frac{\hat{L} + L}{2\hat{L}L} \right) \ln \left( \frac{\hat{L} + L}{2\hat{L}L} \right) - \frac{\ln(\hat{L}L)}{2} \quad (112)$$

(whenever a whole ensemble of configurations must be considered, the expressions above must be averaged accordingly).

Moving further, for testing the effectiveness of a given algorithm in reconstructing weights, a tempting possibility would be to simply extend some of the measures defined in the binary case to the weighted one. However, the non-binary nature of the entries makes it difficult to devise a best choice. Nevertheless, the most popular metric is the weighted counterpart of the cosine similarity [80, 124], reading

$$\vartheta_w = \frac{\mathbf{1}(\hat{\mathbf{W}} \circ \mathbf{W})\mathbf{1}^T}{\|\hat{\mathbf{W}}\|_2 \|\mathbf{W}\|_2} \quad (113)$$

with  $\|\dots\|$  indicating the  $L_2$  entry-wise matrix norm. In other words, these two matrices are treated as vectors of real numbers, whose overlap is approximated by a fictitious angle with values ranging from -1 indicating maximum dissimilarity to +1 meaning exact similarity, and with 0 indicating absence of correlation.

Additional indices are the  $L_1$  and the  $L_2$  entry-wise matrix distances [93], respectively defined as

$$\|\hat{\mathbf{W}} - \mathbf{W}\|_1 = \sum_{i=1}^N \sum_{j(\neq i)=1}^N |\hat{w}_{ij} - w_{ij}|, \quad (114)$$

$$\|\hat{\mathbf{W}} - \mathbf{W}\|_2 = \sqrt{\sum_{i=1}^N \sum_{j(\neq i)=1}^N (\hat{w}_{ij} - w_{ij})^2} \quad (115)$$

and the so-called ‘‘error measure’’ [115] reading

$$\epsilon = \frac{\sum_{i=1}^N \sum_{j=1}^N |\hat{w}_{ij} - w_{ij}|}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}}. \quad (116)$$

When ensemble methods are considered, eqs (113)-(116) can still be used, upon substituting  $\mathbf{W}$  with  $\langle \mathbf{W} \rangle$ . Notice, however, that the major drawback of norm-like quantities lies in the fact that they are unbounded, which makes it difficult to employ them for comparing different candidate matrices.

## 4.2. Topological indicators

The second set of indicators is represented by quantities of topological nature providing a “coarse-grained” description of the network under consideration, such as degree-degree correlations and the mesoscale community structure.

### 4.2.1. Testing weights reconstruction

The most straightforward way to compare observed weights with their corresponding estimates is to scatter-plot them. However in order to consistently compare only realized connections, it is preferable to scatter (realized) observed weights versus *conditional* weights

$$\langle w_{ij} | a_{ij} = 1 \rangle = \frac{\langle w_{ij} \rangle}{p_{ij}} \quad (117)$$

which to some extent encode the (available) structural information. This prescription is particularly useful to compare algorithms generating the same expected weights but predicting different topological structures (*e.g.*, MaxEnt and one of the exact density methods) [80].

### 4.2.2. Testing higher-order patterns reconstruction

Besides reconstructing link weights, a good reconstruction method is also expected to reproduce the higher-order trends characterizing the network  $\hat{\mathbf{G}}$  under observation. To this aim, the observed value of a generic quantity of interest  $X(\hat{\mathbf{G}})$  is usually compared with the corresponding prediction obtained by the reconstruction algorithm. Importantly, whenever dealing with ensemble methods, the entire set  $\mathcal{G}$  of configurations must be accounted for, whence the need to find statistical measures compactly describing all possible (alternative) outcomes. The most basic and useful choices are the ensemble average and standard deviation of  $X$  [58], namely

$$\langle X \rangle = \sum_{\mathbf{G} \in \mathcal{G}} X(\mathbf{G}) P(\mathbf{G}), \quad (118)$$

$$\sigma_X = \sqrt{\sum_{\mathbf{G} \in \mathcal{G}} (X(\mathbf{G}) - \langle X \rangle)^2 P(\mathbf{G})}. \quad (119)$$

The evaluation of eqs. (118) and (119) in principle requires the knowledge of the whole ensemble  $\mathcal{G}$ . Since listing all the configurations belonging of the ensemble is simply not feasible, analytical or numerical techniques can be used to tackle this problem.

In the first case, a simple remedy is provided by the *delta method* [125], based on the Taylor expansion of the observed value  $X(\mathbf{G})$  around the expected value of the variables it depends on:

$$X(\mathbf{G}) = X(\langle \mathbf{G} \rangle) + \sum_{i=1}^N \sum_{j(\neq i)=1}^N (g_{ij} - \langle g_{ij} \rangle) \left( \frac{\partial X}{\partial g_{ij}} \right)_{\mathbf{G}=\langle \mathbf{G} \rangle} + \dots \quad (120)$$

Equation (120) is a “tensorial” Taylor expansion, since each entry of the adjacency matrix  $\mathbf{G}$  is an independent random variable. By taking the expected value of both sides of eq.

(120), one recovers the delta method prescription to calculate the expected value of the quantity  $X$ , *i.e.*,

$$\langle X \rangle \simeq X(\langle \mathbf{G} \rangle). \quad (121)$$

The standard deviation  $\sigma_X$  is then estimated by inserting eqs. (120-121) into eq. (119):

$$\sigma_X \simeq \sqrt{\sum_{i=1}^N \sum_{j(\neq i)=1}^N \sum_{t=1}^N \sum_{s(\neq t)=1}^N \text{Cov}[g_{ij}, g_{ts}] \left( \frac{\partial X}{\partial g_{ij}} \frac{\partial X}{\partial g_{ts}} \right)_{\mathbf{G}=\langle \mathbf{G} \rangle}}. \quad (122)$$

Remarkably, eqs. (121) and (122) are exact in the case of linear constraints represented by degrees and strengths. Other examples of topological quantities whose ensemble averages and standard deviations can be computed exactly are the so-called *dyadic motifs*, defined by the expressions

$$N^{\leftrightarrow} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N a_{ij} a_{ji}, \quad (123)$$

$$N^{\rightarrow} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N a_{ij} (1 - a_{ji}), \quad (124)$$

$$N^{\nleftrightarrow} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N (1 - a_{ij})(1 - a_{ji}). \quad (125)$$

Upon considering that distinct dyads are independent, the expected value and standard deviations of the expressions above become

$$\langle N^{\leftrightarrow} \rangle = \sum_{i=1}^N \sum_{j(\neq i)=1}^N p_{ij} p_{ji}, \quad (126)$$

$$\langle N^{\rightarrow} \rangle = \sum_{i=1}^N \sum_{j(\neq i)=1}^N p_{ij} (1 - p_{ji}), \quad (127)$$

$$\langle N^{\nleftrightarrow} \rangle = \sum_{i=1}^N \sum_{j(\neq i)=1}^N (1 - p_{ij})(1 - p_{ji}) \quad (128)$$

$$\sigma_{N^{\leftrightarrow}} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N 2p_{ij} p_{ji} (1 - p_{ij} p_{ji}), \quad (129)$$

$$\sigma_{N^{\rightarrow}} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N p_{ij} (1 - p_{ji}) [1 - p_{ij} (1 - p_{ji}) - p_{ji} (1 - p_{ij})], \quad (130)$$

$$\sigma_{N^{\nleftrightarrow}} = \sum_{i=1}^N \sum_{j(\neq i)=1}^N 2(1 - p_{ij})(1 - p_{ji}) [1 - (1 - p_{ij})(1 - p_{ji})] \quad (131)$$

Other quantities, however, can be dealt with far less ease. As last resort, it is possible to proceed numerically by explicitly sampling  $\mathcal{G}$ . Once a (properly sampled) subset  $\tilde{\mathcal{G}}$  has been obtained, the ensemble average  $\langle X \rangle$  can be approximated by the arithmetic mean

$$\langle X \rangle \simeq \bar{X} = \sum_{\mathbf{G} \in \tilde{\mathcal{G}}} X(\mathbf{G})F(\mathbf{G}), \quad (132)$$

where  $P(\mathbf{G})$  is replaced by the sampling frequency  $F(\mathbf{G}) = \frac{N_{\mathbf{G}}}{|\tilde{\mathcal{G}}|}$ , and  $N_{\mathbf{G}}$  is the number of networks in the sample whose adjacency matrix is equal to  $\mathbf{G}$ . Analogously, the standard deviation  $\sigma_X$  becomes

$$\sigma_X \simeq \sqrt{\sum_{\mathbf{G} \in \tilde{\mathcal{G}}} (X(\mathbf{G}) - \bar{X})^2 F(\mathbf{G})}. \quad (133)$$

Finally, once an estimate for  $X$  (together with some measure about its uncertainty) has been obtained, the comparison between  $X(\hat{\mathbf{G}})$  and  $\langle X \rangle$  can be carried out by checking whether the observed value  $X(\hat{\mathbf{G}})$  lies within the region delimited by the values  $\langle X \rangle \pm z\sigma_X$ , with  $z$  set to determine a desired level of statistical significance. More compactly, this is expressed as the z-score<sup>15</sup>

$$z_X = \frac{X(\hat{\mathbf{G}}) - \langle X \rangle}{\sigma_X} \quad (134)$$

which measures the difference between the observed and the expected value in units of standard deviation. A z-score whose numerical value is close to zero then indicates that the chosen reconstruction algorithm generates an expected value of  $X$  that is close enough to the observed one: More generally, whenever  $|z| \leq z_{th}$  (with  $z_{th}$  usually being 1, 2 or 3), the discrepancy between the two values cannot be considered significant (with a confidence interval of 0.683, 0.954 and 0.997 respectively). Whenever dealing with the models defined within the ERG formalism, this further implies that the structure of the real network  $\hat{\mathbf{G}}$ , proxied by the quantity  $X$ , is completely explained by the (topological information encoded into the) imposed constraints. By contrast, if  $|z| > z_{th}$  the observed value  $X(\hat{\mathbf{G}})$  lies outside the chosen confidence interval: the structure of the observed network determining  $X$  is not completely explained by the specific constraints imposed, and further model specifications should be used (*i.e.*, additional or more complex constraints) [58, 126].

#### 4.2.3. What may and what may not be reconstructed

Even when a given method does not provide a good reconstruction of an observed network, it is nevertheless useful to understand what kind of information it can provide. To this end, let us consider the z-score again. Whenever  $z_X$  is significantly positive,  $X$  is said to be over-represented in  $\hat{\mathbf{G}}$ , meaning that the network under analysis shows a positive tendency towards it. For instance, chain-like motifs are significantly abundant (*i.e.*, they are found more often than expected) in food-webs. Analogously, whenever the

---

<sup>15</sup>The z-scores assumes a Gaussian distribution for the random variable under consideration. If deviations from this hypothesis are expected, a different statistical test should be employed.

z-score is significantly negative,  $X$  is said to be under-represented. Again in the case of food webs, loop-like motifs are significantly missing (*i.e.*, they are found less often than expected).

z-scores (and statistical tests in general) provide information also on ongoing structural changes of a given network. A particularly interesting issue concerns the detection of early-warning signals of upcoming critical events. As shown in [127], this can be done by computing  $z_X$  for each temporal snapshot of the considered system<sup>16</sup> and, then, plotting  $z_X(t)$  versus  $t$ . As far as the discrepancy between observed and expected values evolves “smoothly” from out-of-equilibrium to equilibrium states (or *vice-versa*), early-warning signals can be possibly detected [128, 129],

An important aspect of a network to be tested for statistical significance is, without doubts, its mesoscale organization into modules or communities. An approach similar to what presented in the previous subsection works as follows [130]. Suppose that we know the community organization of the network, characterized by  $\Lambda$  total intra-community links and  $\Pi$  intra-community pair of nodes. The probability that a random network with  $N$  nodes and  $L$  links has at least these values of  $\Lambda$  and  $\Pi$  derives from an urn model without reinsertion, and thus is given by the inverse cumulative hyper-geometric distribution

$$\Sigma = \sum_{l=\Lambda}^L \frac{\binom{\Pi}{l} \binom{N(N-1)-\Pi}{L-l}}{\binom{N(N-1)}{L}}. \quad (135)$$

Hence the smaller the value of  $\Sigma$ , known as *surprise*, the more significant the mesoscale organization of the considered network [130].

More refined benchmarks for mesoscale structures are provided by the Stochastic Block Model (SBM) [131] and its degree-corrected version (dcSBM) [132]. The effectiveness of these models in reproducing block structures in economic and financial networks has been investigated in [133]. In particular, these models allow to interpolate between two alternative kinds of partition structures: core-periphery and bipartite. In the context of interbank networks, a core-periphery structure indicates the existence of a set of core-banks acting as intermediaries between periphery-banks, whereas, a bipartite structure represents an intermediaries-free market, with banks trading (exclusively) according to their preferences for the counterparties [134].

Network configurations characterized by a structure of  $m$  blocks can be represented by an  $m \times m$  symmetric matrix—called the *affinity matrix*, whose entries represent the density of links within and between modules:

$$\mathcal{A} = \begin{pmatrix} \rho_{g_1 g_1} & \rho_{g_1 g_2} & \cdots & \rho_{g_1 g_m} \\ \rho_{g_1 g_2} & \rho_{g_2 g_2} & \cdots & \rho_{g_2 g_m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{g_1 g_m} & \rho_{g_2 g_m} & \cdots & \rho_{g_m g_m} \end{pmatrix}.$$

Using this representation, the SBM and dcSBM respectively assume that connection probability between any two nodes  $i$  and  $j$  assume the form

$$p_{ij}^{\text{SBM}} = \rho_{g_i g_j}, \quad p_{ij}^{\text{dcSBM}} = \rho_{g_i g_j} x_i x_j, \quad (136)$$

<sup>16</sup>In [127], the monitored quantity is precisely the abundance of dyadic motifs,  $X = N^{\leftrightarrow}, N^{\rightarrow}, N^{\leftarrow}$ .

where  $x_i$  is the parameter controlling for the degree of generic node  $i$ . Upon varying the model parameters, a whole range of different topologies can be generated. In the simple case of the dcSBM with only two blocks  $g_1$  and  $g_2$ , the authors of [133] impose a “background” bipartite-like structure with  $\rho_{g_1 g_2} > \rho_{g_1 g_1} = \rho_{g_2 g_2}$ , while progressively rising the degree heterogeneity of nodes belonging to  $g_1$  and  $g_2$  respectively. Upon running a belief-propagation (BP) algorithm [135, 136], the likelihood of the network

$$\mathcal{L}(\mathbf{A}|\vec{x}) = \ln \left[ \prod_{i=1}^N \prod_{j(<i)=1}^N p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}} \right] \quad (137)$$

highlights the transition from a purely bipartite structure to a purely core-periphery structure. The consistency check is done by comparing the numerical value of the likelihood function  $\mathcal{L}^{\text{BP}}$  with that of  $\mathcal{L}^{\text{SBM}}$  and  $\mathcal{L}^{\text{dcSBM}}$ .

This transition is not a surprise if the generative model is known. However, some care must be adopted when studying real-world networks. Indeed for the empirical interbank networks considered in [133], the kind of emerging mesoscale organization “depends” on the used benchmark. In particular, while the SBM-induced belief-propagation reveals a bipartite structure ( $\rho_{g_1 g_2} > \rho_{g_1 g_1} > \rho_{g_2 g_2}$ ) on daily data aggregation scales and a core-periphery structure ( $\rho_{g_1 g_1} > \rho_{g_1 g_2} > \rho_{g_2 g_2}$ ) on longer time scales, the dcSBM-induced belief-propagation always reveals a bipartite structure.

As also noticed elsewhere [137], this behavior is likely due to the tendency of SBM to detect homogeneous modules (*e.g.*, blocks of nodes with large degree and blocks of nodes with low degree [131]). As a consequence, the dcSBM should be preferred whenever the degree heterogeneity is strong. In the case of sparse networks, it is <sup>17</sup>

$$p_{ij}^{\text{dcSBM}} \simeq \hat{L}_{g_i g_j} \left( \frac{\hat{k}_i}{\hat{K}_{g_i}} \right) \left( \frac{\hat{k}_j}{\hat{K}_{g_j}} \right), \quad (138)$$

with  $\hat{L}_{g_i g_j}$  indicating the (observed) total number of links between blocks  $g_i$  and  $g_j$ , and  $\hat{K}_{g_i}$  indicating the (observed) total degree of nodes belonging to block  $g_i$ .

It is also worth mentioning that a block-wise extension of the Configuration Model can be directly defined by an Hamiltonian constraining the block-specific degree sequences [132]. This leads to link probability coefficients depending on the imposed block-structure. That is, the connection probability between node  $i$  belonging to block  $r$  and node  $j$  belonging to block  $s$  reads

$$p_{ij}^{rs} = \frac{x_i^{rs} x_j^{rs}}{1 + x_i^{rs} x_j^{rs}}, \quad (139)$$

where the Lagrange multipliers are now expressed in a tensorial form, and have to be numerically determined by solving block-specific likelihood conditions.

---

<sup>17</sup>This estimation can be obtained by maximizing the sparse-case likelihood function  $\mathcal{L}(\hat{\mathbf{A}}|\vec{\lambda}) \simeq \sum_i \sum_{j(<i)} (a_{ij} \ln p_{ij} - p_{ij})$  with  $p_{ij} = \rho_{g_i g_j} \lambda_i \lambda_j$  [131].

### 4.3. Dynamical indicators

The third family of indicators include metrics reflecting the outcome of diffusion processes over the network. Here in particular we deal with distress propagation across a financial network and the related issue of *systemic risk*, *i.e.*, the possibility that a local event triggers a global instability through a cascading effect. This issue received a lot of attention especially after the global financial crisis of 2007/2008. Since then, it was realized that the complex pattern of interconnections between financial institutions makes the system as a whole inherently fragile: those connections constitute the channels through which financial distress can spread which, eventually, lead to amplification effects like default cascades [60, 138, 139, 140, 141, 142, 143, 144].

Indeed, while *interconnectedness* implies *diversification* and as such helps reducing the individual risk, it also makes the system as a whole more vulnerable [145, 146, 147]. As a consequence, both researchers and regulators have started to pay attention to the structural features of financial systems [148, 119, 40, 149, 150], with the aim of properly estimating the *systemicness* (or *impact*) and the *vulnerability* of each bank. While the former represents the total loss induced on the system by the distress of that bank, the latter is the loss experienced by that bank when the whole system is under distress [140, 151].

The shortcoming of requiring data on individual exposures to obtain these indicators represents the very motivation for the use of effective network reconstruction techniques in finance. In particular, reconstruction methods can generate scenarios that are compatible with the available information and, as such, can be employed to test the resilience of both single institutions and the system as a whole. Clearly, the operative definition of any *dynamical (financial)* indicator depends on the underlying model of shock-propagation assumed. While the literature on the topic is extensive, here we only outline basic concepts and provide a few illustrative examples that, in our opinion, physicists can easily become familiar with. For an in-depth analysis of network-based systemic risk models, we remand to the recent reviews and books [152, 153, 154].

#### 4.3.1. Balance sheets and financial networks

The financial position of a given bank  $i$  is summed up by its balance sheet, which reports total assets  $a_i$  and liabilities  $l_i$  at a given date. Assets are resources with a positive economic value (such as loans, derivatives, stocks, bonds, options, real estates, etc.), whereas liabilities have a negative economic value (obligations, debits, customer deposits, accrued expenses, etc.). The net difference between assets and liabilities defines the bank equity

$$e_i = a_i - l_i \tag{140}$$

and the bank is said to be solvent as long as its equity is positive. Indeed, negative equity means insolvency, as the bank cannot pay back its liabilities even by selling all its assets. In the dedicated literature [77, 141, 119], insolvency is usually considered as a proxy for default, which in turn occurs when the bank actually fails to fulfill a legal obligation.

Let us now consider  $N$  banks and  $Q$  securities (stocks, bonds, options, etc...). The detailed composition of the balance sheet of  $i$  can be described, schematically, as follows. On the assets side we find the loans granted to other banks  $\{w_{ij}\}_{j=1}^N$ , the securities

constituting the investment portfolio  $\{\omega_{i\alpha}\}_{\alpha=1}^Q$  and other (*i.e.*, fixed and intangible) assets  $a_i^o$ :

$$a_i = \sum_{j(\neq i)=1}^N w_{ij} + \sum_{\alpha=1}^Q \omega_{i\alpha} + a_i^o. \quad (141)$$

On the liabilities side, there are the loans granted from other banks  $\{w_{ji}\}_{j=1}^N$ , as well as debts to outside parties  $l_i^o$ :

$$l_i = \sum_{j(\neq i)=1}^N w_{ji} + l_i^o. \quad (142)$$

Hence, financial networks naturally emerge from the interconnections between banks balance sheets. In particular, an interbank lending market is represented by a monopartite network of loans between banks, whereas an equity market is represented by the bipartite network between banks and owned securities. Financial shocks propagate across these networks according to three main mechanisms that we describe below.

#### 4.3.2. Counterparty risk and credit shocks

Bilateral exposures between banks expose them to what is perhaps the most intuitive channel of financial contagion: *counterparty risk*. Suppose that bank  $i$  undergoes significant losses and defaults, failing to meet its contractual obligations: this results in actual losses for the creditors of  $i$ , commonly labeled as *credit shocks* [140, 149]. In particular, bank  $j$  undergoes a loss equal to  $\varphi w_{ju}$ , where  $\varphi$  indicates the amount of loss given default<sup>18</sup>. Bank  $j$  can, in turn, default if this loss exceeds its equity  $e_j$ , originating a new wave of credit shocks.

Credit shocks have been extensively studied in literature (see, *e.g.*, [155, 156, 157]) Here we briefly describe the DebtRank model [158, 159, 160], whose peculiarity consists in allowing for credit shocks propagation also in absence of defaults, provided that balance sheets are deteriorated. Indeed, losses suffered by financial institutions from credit shocks are not only due to the actual default of counterparties, but also to the mark-to-market revaluation of obligations after the deterioration of counterparties creditworthiness (counterparties which are “closer” to default are less likely to pay back their debts at maturity). In particular, the Debt Rank assumes that relative changes of equity translate linearly into relative changes of asset values, resulting in an *impact* of bank  $i$  on bank  $j$  equal to  $\varphi w_{ji}/e_j$ . Individual banks losses are then obtained by iteratively spreading the individual banks distress levels weighted by the potential wealth affected.

Formally, the dynamics of the model consists of several rounds of shock propagation, hereafter index by  $t$ . The state of bank  $i$  at each  $t$  can be compactly described by the relative change of equity  $h_i(t) = 1 - e_i(t)/e_i(0)$ , which ranges between 0 and 1. By definition,  $h_i(t) = 0$  when no equity losses occurred for the bank,  $h_i(t) = 1$  when the bank defaults and  $0 < h_i(t) < 1$  for intermediate distress levels. Starting at  $t = 0$  from

---

<sup>18</sup>For uncollateralized market (mostly studied in the literature),  $\varphi = 1$ . If, instead, any central counterparty guarantees for interbank loans,  $\varphi = 0$  and, in principle, banks face no losses - and the risk goes to the counterparty.

$h_i(0) = 0, \forall i$ , the first-round losses at  $t = 1$  consist of exogenous shocks decreasing the equity of some banks:  $0 \leq h_i(1) \leq 1, \forall i$ . Later-round losses from subsequent credit shocks are then computed as:

$$h_i(t+1) = \min \left\{ 1, h_i(t) + \sum_{j \in \mathcal{A}(t)} \frac{\varphi w_{ij}}{e_i} [h_j(t) - h_j(t-1)] \right\} \quad (143)$$

where  $\mathcal{A}(t) = \{j : h_j(t-1) < 1\}$  is the set of banks that have not defaulted up to time  $t-1$  and, thus, can still spread their financial distress. The dynamics stops at convergence (say  $\tilde{t}$ ), *i.e.*, when no more banks can propagate their distress and  $\mathcal{A}(\tilde{t}) = \emptyset$ . Individual bank indicators are then computed over an appropriate ensemble of initial conditions of the dynamics. In particular, by denoting as  $h_j(\tilde{t}|i)$  the final relative equity change of  $j$  when the initial condition is the single default of bank  $i$  (*i.e.*,  $h_i(1) = 1$  and  $h_j(1) = 0 \forall j \neq i$ ), whose relative systemic importance is  $\nu_i = e_i(0) / \sum_j e_j(0)$ :

- the *impact* of bank  $i$  is the relative equity loss experienced by the system from the initial default of  $i$

$$I_i = \frac{\sum_{j(\neq i)=1}^N h_j(\tilde{t}|i) \nu_j}{1 - \nu_i}; \quad (144)$$

- the *vulnerability* of  $j$  is the relative equity loss for that bank averaged over the initial defaults of all other banks

$$V_i = \frac{\sum_{j(\neq i)=1}^N h_i(\tilde{t}|j)}{N - 1}. \quad (145)$$

In both cases, first-round losses caused by exogenous shocks are explicitly excluded to account for network effects only.

#### 4.3.3. Rollover risk and liquidity shocks

A more involving channel of financial contagion is related to *rollover risk*, faced by banks in need to refinance their debt which is about to mature with new debt [140, 161, 162, 163]. In periods of financial distress, diffuse worries on future losses and counterparty credit-worthiness can lead banks to adopt a micro-prudential *liquidity hoarding* policy by withdrawing liquidity from the market [164, 165, 166]. In this situation, banks which are short on liquidity may be unable to borrow all the needed money from the market and be forced to sell their illiquid assets. However, when assets sales are widespread, the market demand cannot cover for the supply: the market prices of illiquid assets decreases (a circumstance known as *fire sales*), resulting in effective losses for banks labeled as *liquidity shocks* [140, 149]. Note that fire sales spillovers may also originate by the leverage targeting policy adopted by banks [104, 151, 167] (that is, banks may respond to exogenous shocks by selling assets in order to maintain the desired level of debt over equity) and be exacerbated by indirect exposures between banks due to common assets holdings (see below). In any event, liquidity shocks do represent an important dimension of systemic risk, comparable to credit shocks [147] but traveling in the opposite direction.

Suppose that bank  $i$  suffers significant losses and defaults, thus stopping its liquidity provision to the market. Bank  $j$ , which would have rolled its debt over  $j$ , replenishes a

fraction  $(1 - \psi)$  of the lost funding with its liquid assets or from other sources<sup>19</sup> and the remaining fraction  $\psi$  by selling its illiquid assets. The latter, however trade at a discount, so that  $j$  must sell assets worth  $(1 + \chi)\psi w_{ij}$  in book value terms, corresponding to an overall loss of  $\chi\psi w_{ij}$ , where the parameter  $\chi$  sets the change in asset price<sup>20</sup>. Then, bank  $j$  also defaults if this loss exceeds its equity  $e_j$ , originating a new wave of liquidity shocks.

As for credit shocks, liquidity shocks do propagate also in absence of defaults: equity losses experienced by a bank do imply not only a decreasing value of its obligations, but also a decreasing ability and willingness to lend money to the market. Thus, liquidity shocks can be smoothly incorporated into the DebtRank formalism when the network of interbank exposures is *annealed* (*i.e.*, when the dynamics of shock propagation is on the same time scale of contracts duration) [169]. By assuming that the ability of banks to lend money decreases proportionally to their equities, the *impact* of bank  $i$  on bank  $j$  reads  $\psi\chi w_{ij}/e_j$ , which sums to the term  $\varphi w_{ji}/e_j$  in eq. (143) to have a dynamical equation incorporating both credit and funding shocks. Financial indicators can then be computed as illustrated in the previous section.

#### 4.3.4. Overlapping portfolios and fire-sales spillovers

Beyond direct exposures, financial contagion can spread among banks through indirect exposures to commonly owned securities, namely *portfolio overlap* [129, 151, 161, 170, 171, 172]. Indeed, when the occurrence of financial distress triggers fire sales and prices start to fall, losses by banks with overlapping holdings become self-reinforcing and trigger further simultaneous sell orders, ultimately leading to downward spirals for asset prices.

Here we discuss a simple linear model of fire sales spillovers due to target leveraging by banks and driven by portfolio overlaps [151]. Upon defining  $\Omega_i = \sum_{\beta=1}^Q \omega_{i\beta}$  as the total portfolio size of bank  $i$  and  $\tilde{\omega}_{i\alpha} = \omega_{i\alpha}/\Omega_i$  as the weight of security  $\alpha$  within the portfolio of  $i$ , the model dynamics consist of two time steps. At  $t = 1$ , each bank  $i$  collects the return of its investments:

$$R_i(1) = \sum_{\alpha=1}^Q \tilde{\omega}_{i\alpha} f_{\alpha}(1) \quad (146)$$

where  $f_{\alpha}(1)$  denotes the net return of security  $\alpha$ . In order to simulate exogenous shocks,  $f_{\alpha}(1)$  is taken as a negative number so that  $R_i(1) < 0$ . Since the equity of  $i$  has now changed by  $\Omega_i R_i(1)$ , in order to return to the leverage target  $b_i = a_i/e_i$  the bank has to reallocate  $b_i \Omega_i R_i(1)$  assets on its balance sheet. To this end, it is assumed that banks reallocate assets proportionately to existing holdings, so that the net purchase of bank  $i$  on security  $\alpha$  is:

$$\phi_{i\alpha} = \tilde{\omega}_{i\alpha} b_i \Omega_i R_i(1). \quad (147)$$

<sup>19</sup>In periods of severe distress, exceptional monetary policies are usually implemented and central banks become lenders of last resort, corresponding to the case  $\psi = 0$ .

<sup>20</sup>To compute  $\chi$ , it is usually assumed that assets fire sales generate a linear impact on prices [119, 151, 168], so that the relative assets price change is proportional the aggregate amount of assets that need to be liquidated.

However, asset sales generate price impact (for simplicity, according to a linear model), so that the return of security  $\alpha$  is now:

$$f_\alpha(2) = \sum_{\beta=1}^Q L_{\alpha\beta} \sum_{j=1}^N \phi_{j\beta}, \quad (148)$$

where  $L_{\alpha\beta}$  is a generic entry of the matrix of price impact ratios. Note that if all securities are perfectly liquid (meaning that all elements of the matrix are zero), then price impact vanishes. The *illiquidity* of security  $\alpha$  is thus defined as  $\Lambda_\alpha = \sum_\beta L_{\alpha\beta}$ . Finally, the return of bank  $i$  at  $t = 2$  becomes:

$$R_i(2) = \sum_{\alpha=1}^Q \tilde{\omega}_{i\alpha} f_\alpha(2) \quad (149)$$

and in principle this process can be iterated multiple times.

Using this framework, individual bank indicators can be computed as follows (see also [80, 84]):

- the *impact* (or *systemicness*) of  $i$  is the contribution of that bank to the system equity wiped out by bank “de-leveraging” due to the initial shock

$$I_i = \left[ \sum_{\alpha=1}^Q \left( \sum_{j=1}^N \omega_{j\alpha} \right) \Lambda_\alpha \omega_{i\alpha} \right] \frac{b_i R_i(1)}{\sum_j e_j}; \quad (150)$$

- the (indirect) *vulnerability* of  $i$  is the impact of the initial shock on its equity through the de-leveraging of other banks

$$V_i = \frac{1 + b_i}{\Omega_i} \sum_{\alpha=1}^Q \Lambda_\alpha \omega_{i\alpha} \sum_{j=1}^N \omega_{j\alpha} b_j R_j(1). \quad (151)$$

Note that beyond balance sheet quantities and individual positions, the above expressions depend on securities illiquidity parameters, which are difficult to estimate. However, by assuming that fire sales in one security do not directly affect prices in other securities, the matrix of price impact ratios becomes diagonal and all illiquidity parameters become equal. In this special case, the ratios between reconstructed and empirical indicators assume a particularly simple form as, for instance [80],

$$\frac{\langle I_i \rangle}{\hat{I}_i} = \frac{\sum_{j=1}^N \sum_{\alpha=1}^Q \langle \omega_{i\alpha} \omega_{j\alpha} \rangle}{\sum_{j=1}^N \sum_{\alpha=1}^Q \hat{\omega}_{i\alpha} \hat{\omega}_{j\alpha}}. \quad (152)$$

## 5. Model selection criteria

Of course, each indicator described in the previous section can be used to assess only a specific aspect of the performance of a given reconstruction algorithm. Here we introduce more general information-based criteria, able to capture the overall performance of a reconstruction method.

### 5.1. The Likelihood Ratio Test

The very basic criterion to compare different reconstruction algorithms consists in comparing their likelihood functions. Since the likelihood represents the probability that the observed network is reproduced by the chosen model, the closer its value to 1, the better the model<sup>21</sup>.

However, since increasing the number of model parameters causes the likelihood function to increase as well, this basic criterion completely ignores the over-fitting issue, *i.e.*, the risk of introducing unnecessary parameters not providing any relevant information but weakening the overall predictive power of the model. Indeed, one of the desirable features of any model lies in the possibility to generalize/apply it on different systems. By tuning too many parameters over a single specific system may induce a model that is able to reproduce every detail of the system itself, without capturing more general and essential features potentially shared by similar systems. For this reason, a more refined criterion is needed, possibly discounting the number of parameters entering into the models definition.

The simplest choice is provided by the Likelihood Ratio Test (LRT), designed to compare *pairs* of *nested* models. This means that 1) only two models at a time can be compared, 2) the space of the parameters of one model must be a subspace of the parameters of the other model. This second requirement sheds light on the meaning of the test itself, which is intended to verify the *need* of enlarging the parameter space, *i.e.*, of adopting a more complex model to describe the observations. A concrete example is provided by the pair DECM and DWCM, defined by the Hamiltonians of eqs. (36) and (16), respectively. By switching off the DECM Lagrange multipliers controlling for the degrees (*i.e.*, setting  $x_i^{\text{out}} = x_i^{\text{in}} = 1 \forall i$ ), the likelihood function of the DECM reduces to the likelihood function of the DWCM:

$$p_{ij}^{\text{DECM}} = \frac{x_i^{\text{out}} x_j^{\text{in}} y_i^{\text{out}} y_j^{\text{in}}}{1 + x_i^{\text{out}} x_j^{\text{in}} y_i^{\text{out}} y_j^{\text{in}} - y_i^{\text{out}} y_j^{\text{in}}} \quad \longrightarrow \quad p_{ij}^{\text{DWCM}} = y_i^{\text{out}} y_j^{\text{in}}, \quad (153)$$

$$\langle w_{ij} \rangle^{\text{DECM}} = \frac{p_{ij}^{\text{DECM}}}{1 - y_i^{\text{out}} y_j^{\text{in}}} \quad \longrightarrow \quad \langle w_{ij} \rangle^{\text{DWCM}} = \frac{y_i^{\text{out}} y_j^{\text{in}}}{1 - y_i^{\text{out}} y_j^{\text{in}}}. \quad (154)$$

Provided that  $r_1$  is the model with the lower number of parameters (*i.e.*, the DWCM) and  $r_2$  is the model with the larger number of parameters (*i.e.*, the DECM), the LRT compares the quantity

$$D = 2\mathcal{L}_{r_2}(\hat{\mathbf{G}}|\vec{\lambda}_{(r_2)}) - 2\mathcal{L}_{r_1}(\hat{\mathbf{G}}|\vec{\lambda}_{(r_1)}) \quad (155)$$

to some properly-defined threshold value  $D_{th}$ . The latter is determined by the Wilks' theorem [173], stating that the probability distribution of  $D$  is approximately a chi-squared distribution with a number of degrees of freedom equal to  $|\vec{\lambda}_{(r_2)}| - |\vec{\lambda}_{(r_1)}|$ , *i.e.*, to the difference between the number of parameters of model  $r_2$  and model  $r_1$ .

---

<sup>21</sup>When considering log-likelihood functions, instead, the best algorithm is characterized by the closest value to zero.

### 5.2. The Akaike Information Criterion

In order to allow for the comparison of more than two models, the more refined Akaike Information Criterion (AIC) [174, 175, 176] can be used. Among a set of competing models, the best performing  $r$  is characterized by the largest value of

$$\text{AIC}_r = 2M_r - 2\mathcal{L}_r(\hat{\mathbf{G}}|\vec{\lambda}_{(r)}). \quad (156)$$

AIC is thus a model-specific index that is (proportional to) the difference between the number of parameters  $M_r$  of the model and its maximum log-likelihood. Adding the number of parameters to the log-likelihood function allows to get rid of the overfitting issue, and AIC represents an attempt of finding an optimal trade-off between explanatory power and simplicity.

Equation (156) provides the baseline for other similar criteria that have been subsequently defined. As an example, whenever the number  $n$  of empirical observations becomes too small with respect to the number of parameters (a rule of thumb being  $n/M_r < 40$  [174, 175]), the modified quantity

$$\text{AICc}_r = \text{AIC}_r + \frac{2M_r(M_r + 1)}{n - M_r - 1} \quad (157)$$

should be employed. AICc penalizes models with too many parameters even more severely than AIC; consistently, whenever  $n \gg M_r^2$ , AICc converges to AIC and eq. (156) is recovered.

### 5.3. The Bayesian Information Criterion

An alternative criterion to AICc is the *Bayesian Information Criterion* (BIC) [174, 175, 176]. The difference between the two lies in the functional form of the term to be added to the maximized likelihood. The BIC discounts not only the number of parameters but also the number of observations:

$$\text{BIC}_r = M_r \ln n - 2\mathcal{L}_r(\hat{\mathbf{G}}|\vec{\lambda}_{(r)}). \quad (158)$$

The extra term  $\ln n$  is believed to make BIC more restrictive than AIC, as the former tends to select models with a lower number of parameters than those selected by the latter [174, 175]. However, which criterion performs best, and under which conditions, is still a debated issue.

As a final comment, we would like to stress the general applicability of the aforementioned criteria. In fact, all of them can be extended to quantum-inspired entropic measures [108]. Additionally, although all these criteria are likelihood-based (*i.e.*, they can be used to compare only models defined by means of a likelihood function), they can be also employed to consistently compare probabilistic as well deterministic algorithms (it is enough to set the likelihood function of these algorithms to zero). In any case, despite their formal differences, all the described information-based criteria convey the same message: a “good” reconstruction algorithm is not only required to accurately fit the observed data but also to avoid over-fitting them, thus encompassing a good trade-off between accuracy and parsimony.

#### 5.4. A quick look at multimodel averaging

Beside individuating the best model within a basket of alternatives, AIC, AICc and BIC also allow to quantify the relative improvement brought by each model. This is achieved by computing the *Akaike* (or, equivalently, the *Bayesian*) *weights*, reading

$$w_r = \frac{e^{-\Delta_r/2}}{\sum_{s=1}^R e^{-\Delta_s/2}}, \quad (159)$$

where  $\Delta_r = \text{AIC}_r - \min\{\text{AIC}_s\}_{s=1}^R$  (or, in the BIC case,  $\Delta_r = \text{BIC}_r - \min\{\text{BIC}_s\}_{s=1}^R$ ), with  $R$  being the total number of competing models.

The Akaike (or Bayesian) weight of a certain model is usually interpreted as the probability that the corresponding model is the most appropriate one. In particular, models with  $\Delta \leq 2$  have substantial statistical support; models with  $4 \leq \Delta \leq 7$  have less support and models with  $\Delta > 10$  have essentially no support (remarkably, confidence intervals can also be also defined) [174, 175, 176]. Finally, in order to quantify how better a given model  $r_1$  is with respect to a competitor model  $r_2$ , the ratio  $w_{r_1}/w_{r_2}$  can be computed.

In table 2 we provide a sample test of this kind for the (undirected) ECM and WCM on several empirical networks. We see that, apart from the first two social networks, the ECM is always superior to the WCM, achieving unit probability (within machine precision). A closer inspection of the networks, for which the opposite result holds, reveals that these networks are (almost) fully connected. In these cases, the degree sequence represents a redundant constraint and therefore a model with less parameters is preferable. These results provide additional evidence that degrees convey an information which is not reducible to that of strengths in order to reconstruct a network with non-trivial topology.

<b>Networks</b>	$w_{\text{WCM}}^{\text{AIC}}$	$w_{\text{ECM}}^{\text{AIC}}$
Office social network	1	0
Research group social network	1	0
Fraternity social network	0	1
Maspalomas Lagoon food	0	1
Chesapeake Bay food web	0	1
Crystal River (control) food web	0	1
Crystal River food web	0	1
Michigan Lake food web	0	1
Mondego Estuary food web	0	1
Everglades Marshes food web	0	1
Italian Interbank network (in 1999)	0	1
World Trade Web (in 2000)	0	1

Table 2: Akaike weights for the WCM and ECM applied to reconstruct the empirical networks listed in the first column [94]. Except for the first two networks that are basically fully connected, the inclusion of degrees information is non-redundant, and the ECM drastically outperforms the WCM.

Finally, whenever no winning model emerges from these tests, Akaike (or Bayesian) weights can still be used to account for several estimates of the same parameter—say,  $\{\hat{\mu}\}$ . A prescription retaining the explanatory power of the models providing these estimates

reads

$$\bar{\hat{\mu}} = \sum_r w_r^{\text{AIC}} \hat{\mu}_{(r)}, \quad (160)$$

and consists in averaging the estimates themselves according to the “relevance” of the model they derive from (quantified by, *e.g.*, the Akaike weights). Equation (160) illustrates the concept of *multimodel average*: the estimation to be employed for reconstructing the network is, thus,  $\bar{\hat{\mu}}$ .

## 6. Conclusions and perspectives

Networks are increasingly pervasive in our life, hence network models and methods are becoming and will become more and more important in science and society. And whatever the development and technological level of this world will become, we shall always struggle (probably more and more) to get the information necessary to describing it. Already now, for systems as large as the WWW, it is essentially impossible to collect anything but partial information. We therefore believe that the knowledge of basic instruments needed to deal with partial information will be more and more needed in the future, paving the way for a novel use of ensemble methods in modern Statistical Physics.

The state of the art so far shows clearly that the performance of a reconstruction method crucially depends on several factors. In order to detecting the best method for the problem at hand, a great effort has been recently devoted to compare different algorithms on a large set of empirical networks. Such “horse-races” are intended to quantify the performance of the various algorithms with respect to the families of indicators introduced in section 4, paying particular attention to the topological ones. Below we briefly summarize the current state of the art presented above and the perspectives for future works.

### 6.1. Comparing different reconstruction algorithms on real-world networks

Although the rationale of the MaxEnt method is rooted into the empirical observation of financial systems (banks tend to maximize their diversification since, in case of distress propagation, a complete market is believed to be more robust than an incomplete market [138]), it is widely recognized that MaxEnt performs very poorly in reproducing the topological details of a given network. On the other hand, both the MaxEnt and the “copula” approach are strong performers in reproducing the observed weights, as indicated by the (weighted) cosine similarity [61].

However, real financial markets are sparse [177]. Thus, MaxEnt method must be complemented by a prescription able to reproduce the topological details of a given network structure. Although the density-corrected DWCM constraints both the marginals and the link density (thus reproducing both), it may fail in reproducing higher-order topological quantities [81]. A better performance is achieved by those algorithms estimating the topological details independently from the weights (*e.g.*, in a two-step fashion). For example, the fitness-induced ERG model described by eq. (46) not only reproduces (to a large extent) the in-degree and out-degree sequences of empirical systems [92], but is also characterized by a functional form of link probabilities guaranteeing that the observed

disassortative trends are correctly replicated [81]. Additionally, this method has been shown to satisfactorily reproduce the structural details of the same networks [80], thus outperforming other methods taking as input the same kind of information [81].

A more detailed analysis, focusing on statistical indicators, has been carried out in [61] where the indices defined by eqs. (103), (109), (110) and (111) have been used to rank different reconstruction algorithms. What emerges is that measures “emphasizing” the link structure between financial institutions favor methods that produce sparser networks, whereas measures that “emphasize” the magnitude of bilateral exposures favor methods that allocate exposures as evenly as possible. More specifically, both the fitness-induced ERG model and the Minimum-Density algorithm are strong performers according to the accuracy index, the Hamming distance and the Jaccard distance. The former, however, is “the clear winner among the ensemble methods [...] across all measures of interest” [61].

An additional round of comparisons has been carried out in [93], where the performance of the Bayesian approach(es) has been compared with the performance of the fitness-induced ERG model. To this aim, the authors have employed the “empirical” approach discussed in subsection 3.5.2, where parameters are tuned to match the actual network density. Both the “empirical” Bayesian approach and the fitness-induced ERG model are strong performers under the accuracy, the sensitivity and the specificity indices. This confirms that, as long as the binary network topology is concerned, the fitness-induced ERG model seems to represent the best algorithm available so far. For what concerns weights reconstruction, the fitness-induced ERG model achieves the best score under the  $L_1$  and  $L_2$  norms, but is outperformed by the Bayesian approach under the PTS index. The latter quantifies the probability of finding, within a numerically generated sample of networks, reconstructed weights whose magnitude lies within the 10% of the observed value. As noticed in [93], the reason of the success of the Bayesian approach lies in the fact that it allows one to generate configurations characterized by a whole range of different weights, whereas the fitness-induced ERG model assigns weights via a simple Bernoulli distribution, whence the smaller probability of finding values that satisfy the requirements above. It should, however, be noticed that the same successful result is achieved by using the Weighted Random Graph Model (which performs even better than the Bayesian method), pointing out that any algorithm which is sufficiently generic seems to be able to achieve a high score under the PTS index. In other words, the requirement to perform satisfactorily under this index does not seem to pertain to any algorithm specifically designed for reconstruction. Overall, the versatility of the Bayesian approach (*i.e.*, its capability of producing a large number of different topological structures and weights distributions) make it a good method for designing possible scenarios over which running stress tests, rather than for reconstructing a particular network configuration.

In [178], the authors adjust several methods described in the previous sections to deal with the reconstruction of bipartite bank-firm credit networks. The main difference between the original and the modified versions of such algorithms lies in the weights allocation step, realized through the IPF algorithm. Generally speaking, what the authors observe is that the best performance depends on the specific indicator and the level of aggregation. However, apart from the trivial result that the MaxEnt and the Minimum-Density methods achieve, respectively, the highest sensitivity and the highest specificity, the authors find that the considered variants of ERG models (*i.e.*, the methods inspired

Name	ME	Type	Category	Brief description	Sec.	Ref.
MaxEnt	✓	Dense	Deterministic	Maximizes Shannon entropy on network entries by constraining marginals	3.1.1	[76, 77]
IPF	✓	Tunable	Deterministic	Minimizes the KL divergence from MaxEnt	3.1.2	[86]
MECAPM	✓	Dense	Probabilistic	Constrains matrix entries to match, on average, MaxEnt values	3.1.4	[84]
Drehmann & Tarashev	✓	Tunable	Probabilistic	Randomly perturbs the MaxEnt reconstruction	3.2.2	[88]
Mastromatteo et al.	✓	Tunable	Probabilistic	Explores the space of network structures with the message-passing algorithm	3.2.3	[79]
Moussa & Cont	✓	Tunable	Probabilistic	Implements IPF on non-trivial topologies	3.2.4	[90]
Fitness-induced ERG	✓	Exact	Probabilistic	Uses the fitness ansatz to inform an exponential random graph model	3.3.4	[98, 92]
Copula approach	×	Dense	Deterministic	Generates a network via a copula function of the marginals	3.5.1	[115]
Gandy & Veraart	×	Tunable	Probabilistic	Implements an adjustable Bayesian reconstruction	3.5.2	[116]
Montagna & Lux	×	Tunable	Probabilistic	Assumes <i>ad-hoc</i> connection probabilities depending on marginals	3.5.4	[118]
Hałaj & Kok	×		Probabilistic	Uses external information to define a (geographical) probability map	3.5.5	[120]
Minimum-Density	×	Sparse	Probabilistic	Minimizes the network density while satisfying the marginals	3.5.6	[121]

Table 3: Overview of the reconstruction methods reviewed in the present work.

by [179] and [104]) “consistently perform best” [178] and “are able to reconstruct adjacency matrices and weighted networks relatively well, and they are capable to preserve the statistical properties of the actual network at all (data) aggregation levels” [178]. Interestingly, the performance of reconstruction methods in reproducing *dynamical (financial)* indicators is also tested. As the authors notice, even null models preserving degrees fail to accurately reproduce the actual level of systemic risk (defined as the probability of default of a bank [178]). However, the model inspired by [179] (followed by the model inspired by [104] and MaxEnt) “has the closest behavior to the actual network overall, while Minimum-Density shows an inconsistent performance across different aggregation levels” [178].

Generally speaking the fitness-induced ERG model performs well in replicating the binary topology because it provides a realistic estimate of the degrees in the network. Some studies have shown that methods that take as input local topological properties can outperform models that take as input more information of non-topological nature (*e.g.*, geographical distances in the case of the WTW) [180].

## 6.2. Is link density really needed?

All aforementioned “horse-races” point out the superior performance of the algorithms that can be calibrated to reproduce the observed link density over the ones which cannot: the network link density, in other words, constitutes a piece of information that *must* be taken into account, in order to achieve an accurate reconstruction.

However, as stressed by the authors in [93], the network density cannot be deduced from the marginals alone. As a consequence, it must be known from the beginning. If this is not the case (as it often happens to be), dealing with the issue of estimating link density is not always straightforward. Although the knowledge of just a network subgraph, in fact, often ensures that the actual link density is accurately estimated, such entries must be selected carefully. As recommended in [103], the random selection scheme should be employed, in order to avoid possible biases, which may arise when nodes are sampled according to some criterion. As the authors in [80] show, if (sets of) nodes were selected according to their total strength,  $\hat{s}_i^{\text{tot}} = \hat{s}_i^{\text{out}} + \hat{s}_i^{\text{in}}$ , the resulting density estimate would be strongly dependent on the specific subset value  $\sum_{i \in I} \hat{s}_i^{\text{tot}}$ . Nodes characterized by large total strengths, in fact, tend to cluster into densely-connected groups whereas nodes characterized by small total strengths tend to cluster into loosely-connected groups. The analysis in [80] thus suggests that a sampling-based reconstruction procedure should rest upon a “balanced” sampling of the nodes set, biased neither towards “core” nodes nor towards “peripheral” nodes.

In order to further show how relevant the role played by topological information can be in providing accurate estimates of structural quantities, let us compare the statistical fluctuations of weights estimates, output by the (bipartite versions of the) MECAPM and degree-corrected gravity model (the latter named “Enhanced” CAPM) [100]. Provided that

$$(\sigma_{w_{i\alpha}}^2)^{\text{MECAPM}} = w_{i\alpha}^{\text{ME}}(1 + w_{i\alpha}^{\text{ME}}) \quad (161)$$

and

$$(\sigma_{w_{i\alpha}}^2)_{w_{i\alpha}}^{\text{dcGM}} = (w_{i\alpha}^{\text{ME}})^2 \left[ \frac{1}{p_{i\alpha}} - 1 \right] \quad (162)$$

one finds that

$$\frac{(\sigma_{w_{i\alpha}}^2)^{\text{dcGM}}}{(\sigma_{w_{i\alpha}}^2)^{\text{MECAPM}}} \simeq \sqrt{\frac{1}{p_{i\alpha}} - 1}, \quad (163)$$

a ratio that is (strictly) smaller than 1 whenever  $p_{i\alpha} > 1/2$ . In other words, provided that the MaxEnt method satisfactorily estimates the observed weights, adding topological information helps reducing the error affecting these estimates by “shrinking” the ensemble over which higher-order properties must be estimated. As a consequence, the error accompanying the latter ones is reduced as well.

The example above also helps clarifying the role played by purely weighted information in the whole reconstruction process. Loosely speaking, we may say that the information encoded into nodes strengths is not *per se* of “lower quality” with respect to the information encoded into link density (or into nodes degrees). What emerges is rather that it should not be used to *directly* reconstruct a given network, but first to estimate nodes degrees, and only after be enforced as a complementary constraint.

### 6.3. Policy-making implications: the case of the OTC market

As mentioned in the Introduction, the more information is available about the interconnections shaping a given economic or financial network, the more effective a regulatory intervention can be whenever a systemic event is detected in the system. As a consequence, discovering which kind of information plays a major role in reconstructing a given economic or financial network is not only interesting from an academic point of view, but of paramount importance also for regulators and policy-makers, and has profound societal implications.

A currently debated topic is thus what kind of information should be disclosed by financial institutions. As an example, consider the case of the OTC market. In order to improve the transparency of this market, new reporting rules have been introduced in the aftermath of the crisis [89] establishing that, beside marginals, a certain number of entries of the adjacency matrix have to be made accessible (*i.e.*, the ones exceeding a given threshold). This constitutes a challenge that can be approached by employing (some of) the algorithms reviewed in the present work. For example, constraining single entries of the adjacency matrix can be easily implemented both within the ERG formalism and by slightly modifying the IPF algorithm. While, in the second case, it is enough to subtract the known entries from the marginals and redistribute what remains on the unconstrained entries, in the first case the additional constraints can be dealt with by keeping the corresponding Lagrange multipliers in the tensor form (the latter would be estimated by solving the equation  $\hat{w}_{ij} = \langle w_{ij} \rangle$  for each known entry).

In any case, and whatever algorithm is chosen, the adaptation of existing reconstruction methods to newly-disclosed kinds of information is not only important to determine the effectiveness of the new reporting rules, but can also lead to the design better data-sharing agreements.

## Appendix A. A combinatorial derivation of Shannon entropy

Since Shannon entropy is the quantity underlying many of the approaches reviewed in the present work, in what follows we present a very intuitive derivation of it.

Let us start by considering the set of sequences of binary symbols (0, 1) whose length is  $n$  (*i.e.*, the sequences like 01001001...). We can ask how many bits are needed in order to transmit any message of this particular kind. Upon considering that there are  $|\mathcal{M}| = 2^n$  messages satisfying the aforementioned properties, the answer is simply  $n = \log_2 |\mathcal{M}|$ , that is the length of the message itself.

Let us now consider a less trivial situation, by allowing our messages to be composed by a number of symbols larger than two (*e.g.*,  $R_1, R_2 \dots R_N$ ) and to be emitted by a source sequentially, according to the probability coefficients  $P_1, P_2 \dots P_N$  (with the latter satisfying the normalization condition  $\sum_i P_i = 1$ ). In this case, the set of admissible messages has cardinality  $|\mathcal{M}'| = N^n$ , and its generic element is composed by  $n_i$  symbols of the  $R_i$  kind. Since each symbol is emitted independently from the others, when  $n \rightarrow \infty$  we can apply the law of large numbers and say that the number of times  $n_i$  in which the outcome  $R_i$  is observed satisfies the limit  $n_i/n \rightarrow P_i$ . As a consequence, when  $n$  becomes sufficiently large, the sequences characterized by an occurrence of symbol  $R_i$  which is exactly  $P_i$  become overwhelmingly more likely to occur than the others (*i.e.*, a negligible amount of information is lost upon discarding sequences *not* satisfying  $n_i/n \simeq P_i$ ). We can thus restrict ourselves to consider the set of messages of length  $\sum_i n_i = n$  characterized by probability coefficients reading

$$P(R_{i_1} R_{i_2} \dots R_{i_n}) = P_1^{n_1} P_2^{n_2} \dots P_N^{n_N}. \quad (\text{A.1})$$

The messages above constitute a set of

$$|\mathcal{M}'| = \frac{n!}{n_1! n_2! \dots n_N!} \quad (\text{A.2})$$

equiprobable elements: as for the binary case considered at the beginning of this Appendix, the number of bits needed to transmit one of these messages can be estimated as  $I' = \log_2 |\mathcal{M}'|$ , that is

$$I' \simeq -n \sum_i P_i \log_2 P_i \quad (\text{A.3})$$

(upon using the Stirling approximation  $\ln(x!) \simeq x \ln x - x$ ). As a consequence, one can define the quantity

$$S = - \sum_i P_i \log_2 P_i \simeq \frac{I'}{n} \quad (\text{A.4})$$

as the Shannon entropy. In other words, Shannon entropy is proportional to the logarithm of the probability of a typical (long) sequence, divided by the number of symbols composing the sequence itself. Otherwise stated,  $S$  quantifies the (average) number of bits needed to transmit a typical sequence (*i.e.*, one of those belonging to  $\mathcal{M}'$ ).

The reasoning leading to eq. (A.4) can be restated more precisely by invoking the law of large numbers to gain insight into the asymptotic behavior of the i.i.d. variables describing our symbols  $R_1, R_2 \dots R_N$ :

$$\frac{1}{n} \log_2 \frac{1}{P(R_{i_1} R_{i_2} \dots R_{i_n})} = -\frac{1}{n} \log_2 \prod_i P_i(R_i) = -\frac{1}{n} \sum_i \log_2 P_i(R_i) \rightarrow S. \quad (\text{A.5})$$

As a consequence, the set of all possible messages is split in two: with high probability, the observed messages will belong to “a typical” subset, whose members are described by a coefficient approaching

$$P(R_{i_1} R_{i_2} \dots R_{i_n}) \simeq 2^{-nS} \quad (\text{A.6})$$

and inducing an uniform distribution over this set. This result is known as *asymptotic equipartition property* (AEP).

Although the early derivation of Shannon entropy rested upon the concept of *bit*, thus forcing the base of the logarithm to be 2, from a purely numerical viewpoint it is more convenient to make use of the natural logarithm (adopted in the present review). This means measuring information in *nats* (1 nat equals  $\log_2 e$  bits).

## Appendix B. Sketching a principled derivation of Shannon entropy

A possible derivation of Shannon entropy from the Shannon-Khinchin axioms is shown below (alternative proofs can be found in [71, 72, 73]). The fourth axiom implies that whenever a combination of two independent subsystems is considered, Shannon entropy reads  $S(W_A W_B) = S(W_A) + S(W_B)$ . Upon deriving it twice, the first time with respect to  $W_A$  and the second time with respect to  $W_B$ , the expression  $S'(W_A W_B) + W_A W_B S''(W_A W_B) = 0$  is obtained. Upon posing  $W_A W_B \equiv W$ , the expression above can be rearranged as

$$(W S'(W))' = 0, \quad (\text{B.1})$$

the latter derivative being taken with respect to  $W$ . The derived function is, thus, a constant. Solving this differential equation leads to the celebrated logarithmic functional form  $S(W) = k \ln W$ .

Let us now consider the case of dependent subsets. Upon posing  $W_{A+B} = W$ , it is enough to consider that the requirement  $S(W_{A+B}) = S(W_A) + S(W_{B|A})$  induces the definition of conditional entropy [181], as a (weighted) average of the number of configurations of subsystem  $B$  (say  $V_1, V_2, \dots$ ), the weights being provided by the fraction of configurations of subsystem  $A$  inducing them (*i.e.*,  $w_1 = \frac{V_1}{V}$ ,  $w_2 = \frac{V_2}{V} \dots$ ):

$$S(W_{A+B}) = S(W_A) + w_1 S(V_1) + w_2 S(V_2). \quad (\text{B.2})$$

Upon noticing that  $S(W_{A+B}) = k \ln V$ , we obtain the expression

$$S(W_A) = -k(w_1 \ln w_1) - k(w_2 \ln w_2) \dots \quad (\text{B.3})$$

### Appendix C. Two relevant properties of Shannon entropy

Some of the models reviewed in this work are defined within the ERG formalism. These are exponential distributions satisfying the so-called Gibbs property, *i.e.*, that of being maximally non-committal with respect to the missing information [73]. In fact, upon substituting

$$P(\mathbf{G}|\vec{\lambda}) = \frac{e^{-\sum_m \lambda_m C_m(\mathbf{G})}}{Z(\vec{\lambda})} \quad (\text{C.1})$$

within  $S = -\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln P(\mathbf{G})$ , we obtain

$$S_P = \sum_m \lambda_m \langle C_m \rangle + \ln Z(\vec{\lambda}). \quad (\text{C.2})$$

Let us now calculate the quantity  $S_P - S_Q$ , that is the difference between the Shannon entropy of an exponential distribution and the entropy of a generic distribution  $Q(\mathbf{G})$  which satisfies the same constraints as  $P(\mathbf{G})$  (*i.e.*,  $\sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) = 1$  and  $\sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) C_m(\mathbf{G}) = \langle C_m \rangle \forall m$ ):

$$\begin{aligned} S_P - S_Q &= \sum_m \lambda_m \langle C_m \rangle + \ln Z(\vec{\lambda}) + \sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) \ln Q(\mathbf{G}) = \\ &= \sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) \left[ \sum_m \lambda_m C_m + \ln Z(\vec{\lambda}) + \ln Q(\mathbf{G}) \right] = \\ &= \sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) [-\ln P(\mathbf{G}) + \ln Q(\mathbf{G})] = \\ &= \sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) \left[ -\ln \left( \frac{P(\mathbf{G})}{Q(\mathbf{G})} \right) \right] \geq \sum_{\mathbf{G} \in \mathcal{G}} Q(\mathbf{G}) \left[ 1 - \frac{P(\mathbf{G})}{Q(\mathbf{G})} \right] = \\ &= 1 - 1 = 0. \end{aligned} \quad (\text{C.3})$$

The result above shows that the entropy of an exponential distribution is larger than the entropy of any other distribution satisfying the same constraints (the equality is valid if and only if the probability coefficients of the two distributions coincide).

Unconstrained Shannon entropy instead attains its maximum in correspondence of the uniform distribution. In fact,

$$\begin{aligned} S_U - S_P &= -\sum_{\mathbf{G} \in \mathcal{G}} \frac{1}{|\mathcal{G}|} \ln \frac{1}{|\mathcal{G}|} + \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln P(\mathbf{G}) = \\ &= -\sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln \frac{1}{|\mathcal{G}|} - \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \ln \left( \frac{1}{P(\mathbf{G})} \right) = \\ &= \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \left[ -\ln \left( \frac{1}{|\mathcal{G}| P(\mathbf{G})} \right) \right] \geq \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) \left[ 1 - \frac{1}{|\mathcal{G}| P(\mathbf{G})} \right] = \\ &= 1 - 1 = 0. \end{aligned} \quad (\text{C.4})$$

This can be also verified by calculating the stationary point and the Hessian matrix of the functional

$$\mathcal{L}[P] = S - \lambda_0 \left[ \sum_{\mathbf{G} \in \mathcal{G}} P(\mathbf{G}) - 1 \right], \quad (\text{C.5})$$

whose generic entries read, respectively

$$\frac{\partial \mathcal{L}[P]}{\partial P(\mathbf{G})} = \frac{1}{|\mathcal{G}|} \quad ; \quad \frac{\partial^2 \mathcal{L}[P]}{\partial P(\mathbf{G}) \partial P(\mathbf{G}')} = -\frac{\delta_{\mathbf{G}\mathbf{G}'}}{P(\mathbf{G}) \ln 2}, \quad (\text{C.6})$$

and evaluating the second derivative in correspondence of the uniform distribution  $P(\mathbf{G}) = \frac{1}{|\mathcal{G}|}$ . Since the Hessian matrix is diagonal and its entries are strictly negative (for those distributions satisfying  $P(\mathbf{G}) > 0, \forall \mathbf{G}$ ), such a matrix is negative-definite: the stationary point of the Shannon entropy is a maximum.

#### Appendix D. A notable, continuous case

So far, we have considered distributions obtained by constraining only first moments. Let us now discuss the case in which the second moment is constrained as well. In order to do so, let us imagine we have a one-dimensional real variable  $x \in \mathbb{R}$  described by the probability density function  $p(x)$ . Upon rewriting our constraints as

$$1 = \int_{-\infty}^{+\infty} p(x) dx, \quad (\text{D.1})$$

$$\mu = \int_{-\infty}^{+\infty} x p(x) dx \quad (\text{D.2})$$

and

$$\mu_2 = \int_{-\infty}^{+\infty} x^2 p(x) dx \quad (\text{D.3})$$

we may ask what is the least-biased pdf that contains information on the mean value and the variance of  $x$ . Let us apply the entropy-maximization prescription to the continuous version of Shannon entropy, by defining the functional

$$\begin{aligned} \mathcal{L}[p] = & - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx - \lambda_0 \left[ \int_{-\infty}^{+\infty} p(x) dx - 1 \right] - \lambda_1 \left[ \int_{-\infty}^{+\infty} x p(x) dx - \mu \right] \\ & - \lambda_2 \left[ \int_{-\infty}^{+\infty} x^2 p(x) dx - \mu_2 \right] \end{aligned} \quad (\text{D.4})$$

with  $\lambda_0, \lambda_1, \lambda_2$  being the Lagrange multipliers corresponding to the three conditions (D.1), (D.2) and (D.3). Maximizing the functional above means looking for the function  $p(x)$  which makes the functional derivative  $\frac{\delta \mathcal{L}[p]}{\delta p(x)}$  vanish, *i.e.*,

$$\frac{\delta \mathcal{L}[p]}{\delta p(x)} = -\ln p(x) - 1 - \lambda_0 - \lambda_1 x - \lambda_2 x^2 = 0. \quad (\text{D.5})$$

The solution is

$$p(x) = e^{[-1-\lambda_0-\lambda_1x-\lambda_2x^2]} \quad (\text{D.6})$$

which is of course a Gaussian probability density function. The Lagrange multipliers can be found using the normalization condition  $e^{1+\lambda_0} = \int_{-\infty}^{+\infty} e^{-\lambda_2x^2-\lambda_1x} = \sqrt{\frac{\pi}{\lambda_2}} e^{\frac{\lambda_1^2}{4\lambda_2}}$  together with eqs. (D.2) and (D.3). Upon identifying  $\sigma^2 = \mu_2 - \mu^2$  one gets

$$p(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}. \quad (\text{D.7})$$

Finally, in order to verify that the Gaussian distribution is actually a maximum of the constrained entropy, it is sufficient to verify (in analogy with the discrete case presented above) that the second functional derivative is negative-definite as well. This is indeed true because  $\frac{\delta^2 \mathcal{L}[p]}{\delta p(x) \delta p(x')} = -\frac{\delta(x-x')}{p(x)}$ .

## Acknowledgements

TS, GC and GC acknowledge support from the EU projects CoeGSS (grant num. 676547) and SoBigData (grant num. 654024) and the FET project DOLFINS (grant num. 640772). DG acknowledges support from the Econophysics foundation (Stichting Econophysics, Leiden, the Netherlands). AG acknowledges support from the Italian PNR project CRISIS-Lab.

## References

### References

- [1] R. Albert, A.-L. Barabási, *Statistical mechanics of complex networks*, Reviews of Modern Physics **74** (1) (2002) 47–97. doi:10.1103/RevModPhys.74.47.
- [2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang, *Complex networks: Structure and dynamics*, Physics Reports **424** (4-5) (2006) 175–308. doi:10.1016/j.physrep.2005.10.009.
- [3] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology*, Oxford University Press, 2007.
- [4] S. C. Carlson, *Königsberg Bridge Problem*, Encyclopedia Britannica.
- [5] N. L. Biggs, E. K. Lloyd, R. J. Wilson, *Graph Theory 1736-1936*, Clarendon Press, Oxford, 1998.
- [6] J. L. Moreno, H. H. Jennings, *Who Shall Survive?: A New Approach to the Problem of Human Interrelations*, Nervous and mental disease monograph series, Nervous and mental disease publishing co., 1934.
- [7] J. L. Moreno, *Foundations of sociometry: An introduction*, Sociometry **4** (1) (1941) 15–35. doi:10.2307/2785363.
- [8] D. Lazer, A. Pentland, L. A. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M. Van Alstyne, *Computational social science*, Science **323** (5915) (2009) 721–723. doi:10.1126/science.1167742.
- [9] P. Erdős, A. Rényi, *On random graphs*, Publicationes Mathematicae Debrecen **6** (1959) 290–297.
- [10] P. Erdős, A. Rényi, *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960) 17–61.
- [11] B. Bollobás, *Graph Theory, An Introductory Course*, 1st Edition, Springer, 1979.
- [12] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.
- [13] M. Faloutsos, P. Faloutsos, C. Faloutsos, *On power-law relationships of the Internet topology*, SIGCOMM Proceeding (1999) 251–262doi:10.1.1.37.234.
- [14] G. Caldarelli, R. Marchetti, L. Pietronero, *The fractal properties of internet*, Europhysics Letters **52** (4) (2000) 386–391. doi:10.1209/epl/i2000-00450-8.
- [15] B. Huffaker, M. Fomenkov, D. Plummer, D. Moore, Claffy, K. Claffy, *Distance metrics in the Internet*, IEEE International Telecommunications Symposium, Sept. 2002 (2002) 200–202.
- [16] L. Page, S. Brin, R. Motwami, T. Winograd, R. Motwani, *The PageRank citation ranking: Bringing order to the web*, Unpublished work, Stanford InfoLab (1999).
- [17] J. M. Kleinberg, *Authoritative sources in a hyperlinked environment*, Journal of the ACM **46** (5) (1999) 604–632. doi:10.1145/324133.324140.
- [18] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, A. Arenas, *Self-similar community structure in a network of human interactions*, Physical Review E **68** (2003) 065103. doi:10.1103/PhysRevE.68.065103.
- [19] M. E. J. Newman, S. Forrest, J. Balthrop, *Email networks and the spread of computer viruses*, Physical Review E **66** (2002) 035101. doi:10.1103/PhysRevE.66.035101.
- [20] G. Caldarelli, F. Coccetti, P. De Los Rios, *Preferential exchange: Strengthening connections in complex networks*, Physical Review E **70** (2) (2004) 27102. doi:10.1103/PhysRevE.70.027102.
- [21] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, *Structure and tie strengths in mobile communication networks*, Proceedings of the National Academy of Sciences **104** (18) (2007) 7332–7336. doi:10.1073/pnas.0610245104.
- [22] N. Eagle, A. S. Pentland, D. Lazer, *Inferring friendship network structure by using mobile phone data*, Proceedings of the National Academy of Sciences **106** (36) (2009) 15274–15278. doi:10.1073/pnas.0900282106.

- [23] G. Miritello, E. Moro, R. Lara, *Dynamical strength of social ties in information spreading*, Physical Review E **83** (2011) 045102. doi:10.1103/PhysRevE.83.045102.
- [24] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, B. Bhattacharjee, *Measurement and analysis of online social networks*, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement, IMC '07, ACM, New York, NY, USA, 2007, pp. 29–42. doi:10.1145/1298306.1298311.
- [25] S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, *Network analysis in the social sciences*, Science **323** (5916) (2009) 892–895. doi:10.1126/science.1165821.
- [26] H. Kwak, C. Lee, H. Park, S. Moon, *What is Twitter, a social network or a news media?*, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 591–600. doi:10.1145/1772690.1772751.
- [27] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, W. Quattrociocchi, *The spreading of misinformation online*, Proceedings of the National Academy of Sciences **113** (2016) 554–559. doi:10.1073/pnas.1517441113.
- [28] O. Mason, M. Verwoerd, *Graph theory and networks in Biology*, IET Systems Biology **1** (2007) 89–119. doi:10.1049/iet-syb:20060038.
- [29] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen, C. von Mering, *STRING v10: protein-protein interaction networks, integrated over the tree of life*, Nucleic Acids Research **43** (D1) (2015) D447–D452. doi:10.1093/nar/gku1003.
- [30] G. C. K. W. Koh, P. Porras, B. Aranda, H. Hermjakob, S. E. Orchard, *Analyzing protein-protein interaction networks*, Journal of Proteome Research **11** (4) (2012) 2014–2031. doi:10.1021/pr201211w.
- [31] R. J. Williams, N. D. Martinez, *Simple rules yield complex food webs*, Nature **404** (6774) (2000) 180–183. doi:10.1038/35004572.
- [32] J. A. Dunne, R. J. Williams, N. D. Martinez, *Food-web structure and network theory: The role of connectance and size*, Proceedings of the National Academy of Sciences **99** (20) (2002) 12917–12922. doi:10.1073/pnas.192407699.
- [33] S. R. Proulx, D. E. Promislow, P. C. Phillips, **Network thinking in ecology and evolution**, Trends in Ecology & Evolution **20** (6) (2005) 345–353. doi:10.1016/j.tree.2005.04.004.
- [34] D. Garlaschelli, M. I. Loffredo, *Structure and evolution of the world trade network*, Physica A: Statistical Mechanics and its Applications **355** (1) (2005) 138–144. doi:10.1016/j.physa.2005.02.075.
- [35] C. A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann, A.-L. Barabási, R. Hausmann, *The product space conditions the development of nations*, Science **317** (2007) 482–487. doi:10.1126/science.1144581.
- [36] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D. R. White, *Economic networks: The new challenges*, Science **325** (5939) (2009) 422–425. doi:10.1126/science.1173644.
- [37] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, L. Pietronero, *A new metrics for countries' fitness and products' complexity*, Scientific Reports **2** (2012) 723. doi:10.1038/srep00723.
- [38] G. Cimini, A. Gabrielli, F. Sylos Labini, *The scientific competitiveness of nations*, PLoS ONE **9** (12) (2014) e113470. doi:10.1371/journal.pone.0113470.
- [39] E. Pugliese, G. Cimini, A. Patelli, A. Zaccaria, L. Pietronero, A. Gabrielli, *Unfolding the innovation system for the development of countries: co-evolution of Science, Technology and Production*, <https://arxiv.org/abs/1707.05146> (2017).
- [40] G. Iori, G. De Masi, O. V. Precup, G. Gabbi, G. Caldarelli, *A network analysis of the Italian overnight money market*, Journal of Economic Dynamics and Control **32** (1) (2008) 259–278. doi:10.1016/j.jedc.2007.01.032.
- [41] J. Glattfelder, S. Battiston, *Backbone of complex networks of corporations: The flow of control*, Physical Review E **80** (3) (2009) 036104. doi:10.1103/PhysRevE.80.036104.
- [42] I. Bordino, S. Battiston, G. Caldarelli, M. Cristelli, A. Ukkonen, I. Weber, *Web search queries can predict stock market volumes.*, PLoS ONE **7** (7) (2012) e40014. doi:10.1371/journal.pone.0040014.
- [43] A.-L. Barabási, *Scale-free networks: A decade and beyond*, Science **325** (5939) (2009) 412–413. doi:10.1126/science.1173299.
- [44] A.-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science **286** (5439) (1999) 509–512. doi:10.1126/science.286.5439.509.
- [45] D. J. Watts, S. H. Strogatz, *Collective dynamics of 'small-world' networks.*, Nature **393** (6684)

- (1998) 440–442. doi:10.1038/30918.
- [46] G. Szabó, M. Alava, J. Kertész, *Clustering in Complex Networks*, Vol. **650**, Springer, 2004, pp. 139–162.
- [47] P. Bonacich, *Power and centrality : A family of measures*, American Journal of Sociology **92** (5) (1987) 1170–1182. doi:10.1086/228631.
- [48] M. E. J. Newman, *Assortative mixing in networks*, Physical Review Letters **2** (4) (2002) 1–5. doi:10.1103/PhysRevLett.89.208701.
- [49] M. Almeida-Neto, P. Guimarães, P. R. Guimarães, R. D. Loyola, W. Ulrich, *A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement*, Oikos **117** (8) 1227–1239. doi:10.1111/j.0030-1299.2008.16644.x.
- [50] J. Bascompte, P. Jordano, C. J. Melián, J. M. Olesen, *The nested assembly of plant-animal mutualistic networks*, Proceedings of the National Academy of Sciences **100** (16) (2003) 9383–9387. doi:10.1073/pnas.1633576100.
- [51] S. Jonhson, V. Domínguez-García, M. A. Muñoz, *Factors determining nestedness in complex networks*, PLoS ONE **8** (9) (2013) e74025. doi:10.1371/journal.pone.0074025.
- [52] G. Caldarelli, A. Capocci, P. De Los Rios, M. Muñoz, *Scale-free networks from varying vertex intrinsic fitness*, Physical Review Letters **89** (25) (2002) 258702. doi:10.1103/physrevlett.89.258702.
- [53] J. Leskovec, L. Backstrom, R. Kumar, A. Tomkins, *Microscopic evolution of social networks*, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '08, ACM, New York, NY, USA, 2008, pp. 462–470. doi:10.1145/1401890.1401948.
- [54] M. Medo, G. Cimini, S. Gualdi, *Temporal effects in the growth of networks*, Physical Review Letters **107** (2011) 238701. doi:10.1103/PhysRevLett.107.238701.
- [55] J. Park, M. E. J. Newman, *The statistical mechanics of networks*, Physical Review E **70** (2004) 66117. doi:10.1103/PhysRevE.70.066117.
- [56] G. Bianconi, *The entropy of randomized network ensembles*, Europhysics Letters **81** (2) (2008) 28005. doi:10.1209/0295-5075/81/28005.
- [57] D. Garlaschelli, M. I. Loffredo, *Maximum likelihood: Extracting unbiased information from complex networks*, Physical Review E **78** (1) (2008) 015101. doi:10.1103/PhysRevE.78.015101.
- [58] T. Squartini, D. Garlaschelli, *Analytical maximum-likelihood method to detect patterns in real networks*, New Journal of Physics **13** (2011) 083001. doi:10.1088/1367-2630/13/8/083001.
- [59] A. Fronczak, *Exponential Random Graph Models*, Springer-Verlag New York, 2014, pp. 500–517.
- [60] S. Battiston, G. Caldarelli, R. May, T. Roukny, J. Stiglitz, *The price of complexity in financial networks*, Proceedings of the National Academy of Sciences **110** (2016) 10031–10036. doi:10.1073/pnas.1521573113.
- [61] K. Anand, I. van Lelyveld, A. Banai, S. Friedrich, R. Garratt, G. H. Iaj, J. Figue, I. Hansen, S. M. Jaramillo, H. Lee, J. L. Molina-Borboa, S. Nobili, S. Rajan, D. Salakhova, T. C. Silva, L. Silvestri, S. R. S. de Souza, *The missing links: A global study on uncovering financial network structures from partial data*, Journal of Financial Stability **35** (2018) 107–119. doi:10.1016/j.jfs.2017.05.012.
- [62] R. Guimerà, M. Sales-Pardo, *Missing and spurious interactions and the reconstruction of complex networks*, Proceedings of the National Academy of Sciences **106** (52) (2009) 22073–22078. doi:10.1073/pnas.0908366106.
- [63] L. Lü, T. Zhou, *Link prediction in complex networks: A survey*, Physica A: Statistical Mechanics and its Applications **390** (2011) 1150–1170. doi:10.1016/j.physa.2010.11.027.
- [64] C. Song, S. Havlin, H. A. Makse, *Self-similarity of complex networks.*, Nature **433** (7024) (2005) 392–395. doi:10.1038/nature03248.
- [65] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2006.
- [66] R. Hanel, S. Thurner, M. Gell-Mann, *How multiplicity determines entropy and the derivation of the maximum entropy principle for complex systems*, Proceedings of the National Academy of Sciences **111** (19) (2014) 6905–6910. doi:10.1073/pnas.1406071111.
- [67] S. Pressé, K. Ghosh, J. Lee, K. A. Dill, *Principles of maximum entropy and maximum caliber in statistical physics*, Review of Modern Physics **85** (3) (2013) 1115–1141. doi:10.1103/RevModPhys.85.1115.
- [68] R. Dewar, *Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states*, Journal of Physics A: Mathematical and General **36** (3) (2003) 631. doi:10.1088/0305-4470/36/3/303.
- [69] A. D. Wissner-Gross, C. E. Freer, *Causal entropic forces*, Physical Review Letters **110** (16) (2013)

168702. doi:10.1103/PhysRevLett.110.168702.
- [70] K. Anand, G. Bianconi, *Entropy measures for complex networks: toward an information theory of complex topologies*, Physical Review E **80** (2009) 045102(R). doi:10.1103/PhysRevE.80.045102.
- [71] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948) 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
- [72] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948) 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x.
- [73] E. T. Jaynes, *Information theory and statistical mechanics*, Physical Review **106** (4) (1957) 620–630. doi:10.1103/PhysRev.106.620.
- [74] A. Lesne, *Shannon entropy: A rigorous notion at the crossroads between probability, information theory, dynamical systems and statistical physics*, Mathematical Structures in Computer Science **24** (3). doi:10.1017/S0960129512000783.
- [75] T. Squartini, E. Ser-Giacomi, D. Garlaschelli, G. Judge, *Information recovery in behavioral networks*, PLoS ONE **10** (5) (2015) e0125077. doi:10.1371/journal.pone.0125077.
- [76] S. Wells, *Financial interlinkages in the United Kingdom's interbank market and the risk of contagion*, Working Paper 230, Bank of England (2004).
- [77] C. Upper, *Simulation methods to assess the danger of contagion in interbank markets*, Journal of Financial Stability **7** (3) (2011) 111–125. doi:10.1016/j.jfs.2010.12.001.
- [78] P. E. Mistrulli, *Assessing financial contagion in the interbank market: Maximum entropy versus observed interbank lending patterns*, Journal of Banking & Finance **35** (5) (2011) 1114–1127. doi:10.1016/j.jbankfin.2010.09.018.
- [79] I. Mastromatteo, E. Zarinelli, M. Marsili, *Reconstruction of financial networks for robust estimation of systemic risk*, Journal of Statistical Mechanics: Theory and Experiment **2012** (03) (2012) P03011. doi:10.1088/1742-5468/2012/03/P03011.
- [80] T. Squartini, G. Cimini, A. Gabrielli, D. Garlaschelli, *Network reconstruction via density sampling*, Applied Network Science **2** (1) (2017) 3. doi:10.1007/s41109-017-0021-8.
- [81] P. Mazzarisi, F. Lillo, *Methods for reconstructing interbank networks from limited information: A comparison*, Springer International Publishing, 2017, pp. 201–215. doi:10.1007/978-3-319-47705-3\_15.
- [82] T. Squartini, D. Garlaschelli, *Jan Tinbergen's legacy for economic networks: From the gravity model to quantum statistics*, 2013, pp. 161–186. doi:10.1007/978-3-319-00023-7.
- [83] W. F. Sharpe, *Capital asset prices: A theory of market equilibrium under conditions of risk*, The Journal of Finance **19** (3) (1964) 425–442. doi:10.2307/2977928.
- [84] D. Di Gangi, F. Lillo, D. Pirino, *Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction*, <http://arxiv.org/abs/1509.00607> (2016).
- [85] S. Kullback, R. A. Leibler, *On information and sufficiency*, The Annals of Mathematical Statistics **22** (1) (1951) 79–86. doi:10.1214/aoms/1177729694.
- [86] M. Bacharach, *Estimating nonnegative matrices from marginal data*, International Economic Review **6** (3) (1965) 294–310. doi:10.2307/2525582.
- [87] Y. M. Bishop, S. E. Fienberg, P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, Springer-Verlag New York, 2007. doi:10.1007/978-0-387-72806-3.
- [88] M. Drehmann, N. Tarashev, *Measuring the systemic importance of interconnected banks*, Journal of Financial Intermediation **22** (4) (2013) 586–607. doi:10.1016/j.jfi.2013.08.001.
- [89] *ESMA Trade Reporting*, <https://www.esma.europa.eu/policy-rules/post-trading/trade-reporting>.
- [90] A. Moussa, *Contagion and systemic risk in financial networks*, Ph.D. thesis, Columbia University PhD Thesis (2011).
- [91] B. Bollobás, C. Borgs, J. Chayes, O. Riordan, *Directed scale-free graphs*, in: Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, 2003, pp. 132–139.
- [92] G. Cimini, T. Squartini, A. Gabrielli, D. Garlaschelli, *Estimating topological properties of weighted networks from limited information*, Physical Review E **92** (4) (2015) 040802. doi:10.1103/PhysRevE.92.040802.
- [93] A. Gandy, L. A. M. Veraart, *Adjustable network reconstruction with applications to CDS exposures*, <https://ssrn.com/abstract=2895754> (2017).
- [94] R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli, *Enhanced reconstruction of weighted networks from strengths and degrees*, New Journal of Physics **16** (4) (2014) 043022. doi:10.1088/1367-2630/16/4/043022.
- [95] A. Almog, T. Squartini, D. Garlaschelli, *The double role of GDP in shaping the structure of the*

- International Trade Network*, <http://arxiv.org/abs/1512.02454> (2015).
- [96] D. Garlaschelli, M. Loffredo, *Fitness-dependent topological properties of the World Trade Web*, Physical Review Letters **93** (18) (2004) 1–4. doi:10.1103/PhysRevLett.93.188701.
- [97] G. De Masi, G. Iori, G. Caldarelli, G. D. Masi, *Fitness model for the Italian interbank money market*, Physical Review E **74** (6) (2006) 066112. doi:10.1103/PhysRevE.74.066112.
- [98] G. Cimini, T. Squartini, D. Garlaschelli, A. Gabrielli, *Systemic risk analysis on reconstructed economic and financial networks.*, Scientific Reports **5** (2015) 15758. doi:10.1038/srep15758.
- [99] D. Garlaschelli, S. Battiston, M. Castri, V. D. Servedio, G. Caldarelli, *The scale-free topology of market investments*, Physica A: Statistical and Theoretical Physics **350** (2-4) (2005) 491–499. doi:doi:10.1016/j.physa.2004.11.040.
- [100] T. Squartini, A. Almog, G. Caldarelli, I. van Lelyveld, D. Garlaschelli, G. Cimini, *Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks*, Physical Review E **96** (2017) 032315. doi:10.1103/PhysRevE.96.032315.
- [101] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, *The architecture of complex weighted networks*, Proceedings of the National Academy of Sciences **101** (11) (2004) 3747–3752. doi:10.1073/pnas.0400087101.
- [102] N. Musmeci, S. Battiston, G. Caldarelli, M. Puliga, A. Gabrielli, *Bootstrapping topology and systemic risk of complex network using the fitness model*, Journal of Statistical Physics **151** (2013) 720–734. doi:10.1007/s10955-013-0720-1.
- [103] N. Blagus, L. Šubelj, M. Bajec, *Empirical comparison of network sampling techniques*, <https://arxiv.org/abs/1506.02449> (2015).
- [104] S. Battiston, G. Caldarelli, M. D’Errico, S. Gurciullo, *Leveraging the network: a stress-test framework based on DebtRank*, Statistics and Risk Modeling **33** (2016) 117–138. doi:10.1515/strm-2015-0005.
- [105] F. Ruzzenenti, F. Picciolo, R. Basosi, D. Garlaschelli, *Spatial effects in real networks: Measures, null models, and applications*, Physical Review E **86** (6) (2012) 66110. doi:10.1103/PhysRevE.86.066110.
- [106] M. Duenas, G. Fagiolo, *Modeling the International-Trade Network: A gravity approach*, Journal of Economic Interaction and Coordination **8** (1) (2013) 155–178. doi:10.1007/s11403-013-0108-y.
- [107] K. Anand, G. Bianconi, S. Severini, *Shannon and von Neumann entropy of random networks with heterogeneous expected degree*, Physical Review E **83** (2011) 036109. doi:10.1103/PhysRevE.83.036109.
- [108] M. De Domenico, J. Biamonte, *Spectral entropies as information-theoretic tools for complex network comparison*, Physical Review X **6** (4) (2016) 41062. doi:10.1103/PhysRevX.6.041062.
- [109] G. G. Judge, R. C. Mittelhammer, *An Information Theoretic Approach to Econometrics*, Cambridge University Press, 2011.
- [110] N. Cressie, T. Read, *Multinomial goodness-of-fit tests*, Journal of the Royal Statistical Society. Series B (Methodological) **46** (3) (1984) 440–464.
- [111] W. K. T. Cho, G. Judge, *An information theoretic approach to network tomography*, Applied Economics Letters **22** (1) (2015) 1–6. doi:10.1080/13504851.2013.866199.
- [112] A. Rényi, *On Measures of Entropy and Information*, in: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1961, pp. 547–561.
- [113] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, Journal of Statistical Physics **52** (1) (1988) 479–487. doi:10.1007/BF01016429.
- [114] J. Piantadosi, P. Howlett, J. Borwein, *Copulas with maximum entropy*, Optimization Letters **6** (1) (2012) 99–125. doi:10.1007/s11590-010-0254-2.
- [115] P. Baral, J. P. Figue, *Estimation of bilateral exposures - A Copula approach*, Mimeo, Indiana University, University of Porto (2012).
- [116] A. Gandy, L. A. M. Veraart, *A Bayesian methodology for systemic risk assessment in financial networks*, Management Science (articles in advances) (2016) 4428–4446. doi:10.1287/mnsc.2016.2546.
- [117] A. Schrijver, *On the history of the transportation and maximum flow problems*, Mathematical Programming **91** (3) (2002) 437. doi:10.1007/s101070100259.
- [118] M. Montagna, T. Lux, *Contagion risk in the interbank market: A probabilistic approach to cope with incomplete structural information*, Quantitative Finance **17** (1) (2017) 101–120. doi:10.1080/14697688.2016.1178855.
- [119] E. Nier, J. Yang, T. Yorulmazer, A. Alentorn, *Network models and financial stability*, Journal of Economic Dynamics and Control **31** (6) (2007) 2033–2060. doi:doi:10.1016/j.jedc.2007.01.014.

- [120] G. Halaj, C. Kok, *Assessing interbank contagion using simulated networks*, Computational Management Science **10** (2) (2013) 157–186. doi:10.1007/s10287-013-0168-4.
- [121] K. Anand, B. Craig, G. von Peter, *Filling in the blanks: Network structure and interbank contagion*, Quantitative Finance **15** (4) (2015) 625–636. doi:10.1080/14697688.2014.968195.
- [122] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognition Letters **27** (2006) 861–874. doi:10.1016/j.patrec.2005.10.010.
- [123] C. E. Metz, *Basic principles of ROC analysis*, Seminars in Nuclear Medicine **8** (4) (1978) 283–298. doi:10.1016/S0001-2998(78)80014-2.
- [124] J. Wang, H. T. Shen, J. Song, J. Ji, *Hashing for similarity search: A survey*, <https://arxiv.org/abs/1408.2927> (2014).
- [125] G. Casella, R. L. Berger, *Statistical Inference*, Vol. 2, Duxbury Pacific Grove, CA, 2002.
- [126] D. T. Luu, T. Lux, B. Yanovski, *Structural correlations in the Italian overnight money market: An analysis based on network configuration models*, Entropy **19** (6) (2017) 259. doi:10.3390/e19060259.
- [127] T. Squartini, I. van Lelyveld, D. Garlaschelli, *Early-warning signals of topological collapse in interbank networks*, Scientific reports **3** (2013) 3357. doi:10.1038/srep03357.
- [128] F. Saracco, R. Di Clemente, A. Gabrielli, T. Squartini, *Detecting early signs of the 2007–2008 crisis in the world trade*, Scientific Reports **6** (2016) 30286. doi:10.1038/srep30286.
- [129] S. Gualdi, G. Cimini, K. Primicerio, R. D. Clemente, D. Challet, *Statistically validated network of portfolio overlaps and systemic risk*, Scientific Reports **6** (2016) 39467. doi:10.1038/srep39467.
- [130] C. Nicolini, A. Bifone, *Modular structure of brain functional networks: breaking the resolution limit by Surprise*, Scientific Reports **6** (2016) 19250. doi:10.1038/srep19250.
- [131] B. Karrer, M. E. J. Newman, *Stochastic blockmodels and community structure in networks*, Physical Review E **83** (1) (2011) 16107. doi:10.1103/PhysRevE.83.016107.
- [132] P. Fronczak, A. Fronczak, M. Bujok, *Exponential random graph models for networks with community structure*, Physical Review E **88** (3) (2013) 32810. doi:10.1103/PhysRevE.88.032810.
- [133] P. Barucca, F. Lillo, *Disentangling bipartite and core-periphery structure in financial networks*, Chaos, Solitons & Fractals **88** (2016) 244–253. doi:10.1016/j.chaos.2016.02.004.
- [134] B. Craig, G. von Peter, *Interbank tiering and money center banks*, Journal of Financial Intermediation **23** (3) (2014) 322–347. doi:http://dx.doi.org/10.1016/j.jfi.2014.02.003.
- [135] A. Decelle, F. Krzakala, C. Moore, L. Zdeborová, *Inference and phase transitions in the detection of modules in sparse networks*, Physical Review Letters **107** (6) (2011) 65701. doi:10.1103/PhysRevLett.107.065701.
- [136] X. Zhang, T. Martin, M. Newman, *Identification of core-periphery structure in networks*, Physical Review E **91** (2014) 0321803. doi:10.1103/PhysRevE.91.032803.
- [137] X. Yan, C. R. Shalizi, J. E. Jnesen, F. Krzakala, C. Moore, L. Zdeborova, P. Zhang, Y. Zhu, *Model selection for degree-corrected block models*, Journal of Statistical Mechanics 2014 (2014) P05007. doi:10.1088/1742-5468/2014/05/P05007.
- [138] F. Allen, D. Gale, *Financial contagion*, Journal of Political Economy **108** (1) (2000) 1–33. doi:10.1086/262109.
- [139] M. K. Brunnermeier, *Deciphering the liquidity and credit crunch 2007–2008*, Journal of Economic Perspectives **23** (1) (2009) 77–100. doi:10.1257/jep.23.1.77.
- [140] J. A. Chan-Lau, M. Espinosa, K. Giesecke, J. A. Solé, *Assessing the systemic implications of financial linkages*, Global financial stability report (chapter 2), IMF Monetary and Capital Markets Development (2009).
- [141] P. Gai, S. Kapadia, *Contagion in financial networks*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **466** (2120) (2010) 2401–2423. doi:10.1098/rspa.2009.0410.
- [142] A. G. Haldane, R. M. May, *Systemic risk in banking ecosystems*, Nature **469** (7330) (2011) 351–355. doi:10.1038/nature09659.
- [143] S. Battiston, J. D. Farmer, A. Flache, D. Garlaschelli, A. G. Haldane, H. Heesterbeek, C. Hommes, C. Jaeger, R. May, M. Scheffer, *Complexity theory and financial regulation*, Science **351** (6275) (2016) 818–819. doi:10.1126/science.aad0299.
- [144] M. Bardoscia, S. Battiston, F. Caccioli, G. Caldarelli, *Pathways towards instability in financial networks*, Nature Communications **8** (2017) 14416. doi:10.1038/ncomms14416.
- [145] N. Beale, D. G. Rand, H. Battey, K. Croxson, R. M. May, M. A. Nowak, *Individual versus systemic risk and the Regulator’s Dilemma*, Proceedings of the National Academy of Sciences **108** (31) (2011) 12647–52. doi:10.1073/pnas.1105882108.
- [146] F. Corsi, S. Marmi, F. Lillo, *When micro prudence increases macro risk: The destabilizing effects*

- of financial innovation, leverage, and diversification, *Operations Research* **64** (5) (2016) 1073–1088. doi:10.1287/opre.2015.1464.
- [147] P. Glasserman, H. P. Young, *How likely is contagion in financial networks?*, *Journal of Banking and Finance* **50** (2015) 383–399. doi:10.1016/j.jbankfin.2014.02.006.
- [148] M. Boss, H. Elsinger, M. Summer, S. Thurner, *Network topology of the interbank market*, *Quantitative Finance* **4** (6) (2004) 677–684. doi:10.1080/14697680400020325.
- [149] A. Krause, S. Giansante, *Interbank lending and the spread of bank failures: A network model of systemic risk*, *Journal of Economic Behavior & Organization* **83** (3) (2012) 583–608.
- [150] C.-P. Georg, *The effect of the interbank network structure on contagion and common shocks*, *Journal of Banking & Finance* **37** (7) (2013) 2216–2228. doi:10.1016/j.jbankfin.2013.02.032.
- [151] R. Greenwood, A. Landier, D. Thesmar, *Vulnerable banks*, *Journal of Financial Economics* **115** (3) (2015) 471–485. doi:10.1016/j.jfineco.2014.11.006.
- [152] D. Acemoglu, A. Ozdaglar, A. Tahbaz-Salehi, *Systemic risk and stability in financial networks*, *American Economic Review* **105** (2) (2015) 564. doi:10.1257/aer.20130456.
- [153] S. Bougheas, A. Kirman, *Complex Financial Networks and Systemic Risk: A Review*, Springer International Publishing, 2015, pp. 115–139. doi:10.1007/978-3-319-12805-4\_6.
- [154] T. R. Hurd, *Contagion! Systemic Risk in Financial Networks*, SpringerBriefs in Quantitative Finance, Springer International Publishing, 2016. doi:10.1007/978-3-319-33930-6.
- [155] L. Eisenberg, T. H. Noe, *Systemic risk in financial systems*, *Management Science* **47** (2) (2001) 236–249. doi:10.1287/mnsc.47.2.236.9835.
- [156] C. H. Furfine, *Interbank exposures: Quantifying the risk of contagion*, *Journal of Money, Credit and Banking* **35** (1) (2003) 111–128. doi:10.1353/mcb.2003.0004.
- [157] L. C. G. Rogers, L. A. M. Veraart, *Failure and rescue in an interbank network*, *Management Science* **59** (4) (2013) 882–898. doi:10.1287/mnsc.1120.1569.
- [158] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, G. Caldarelli, *DebtRank: Too central to fail? Financial networks, the FED and systemic risk.*, *Scientific Reports* **2** (2012) 541. doi:10.1038/srep00541.
- [159] M. Bardoscia, S. Battiston, F. Caccioli, G. Caldarelli, *DebtRank: A microscopic foundation for shock propagation.*, *PLoS ONE* **10** (6) (2015) e0130406. doi:10.1371/journal.pone.0130406.
- [160] P. Barucca, M. Bardoscia, F. Caccioli, M. D’Errico, G. Visentin, G. Caldarelli, S. Battiston, *Network valuation in financial systems*.
- [161] R. Cifuentes, G. Ferrucci, H. S. Shin, *Liquidity risk and contagion*, *Journal of the European Economic Association* **3** (2-3) (2005) 556–566. doi:10.1162/jeea.2005.3.2-3.556.
- [162] S. Kapadia, M. Drehmann, J. Elliott, G. Sterne, *Liquidity risk, cash flow constraints, and systemic feedbacks*, University of Chicago Press, 2012, pp. 29–61. doi:10.7208/chicago/9780226921969.003.0003.
- [163] K. Anand, P. Gai, M. Marsili, *Rollover risk, network structure and systemic financial crises*, *Journal of Economic Dynamics and Control* **36** (8) (2012) 1088–1100. doi:10.1016/j.jedc.2012.03.005.
- [164] D. Gale, T. Yorulmazer, *Liquidity hoarding*, *Theoretical Economics* **8** (2) (2013) 291–324. doi:10.3982/te1064.
- [165] M. K. Brunnermeier, L. H. Pedersen, *Market liquidity and funding liquidity*, *Review of Financial Studies* **22** (6) (2009) 2201–2238. doi:10.1093/rfs/hhn098.
- [166] V. V. Acharya, O. Merrouche, *Precautionary hoarding of liquidity and interbank markets: Evidence from the subprime crisis*, *Review of Finance* **17** (1) (2013) 107–160. doi:10.1093/rof/rfs022.
- [167] T. Adrian, H. S. Shin, *Liquidity and leverage*, *Journal of Financial Intermediation* **19** (3) (2010) 418–437. doi:10.1016/j.jfi.2008.12.002.
- [168] P. Feldhütter, *The same bond at different prices: Identifying search frictions and selling pressures*, *Review of Financial Studies* **25** (4) (2012) 1155–1206. doi:10.1093/rfs/hhr093.
- [169] G. Cimini, M. Serri, *Entangling credit and funding shocks in interbank markets*, *PLoS ONE* **11** (8) (2016) e0161642. doi:10.1371/journal.pone.0161642.
- [170] A. Shleifer, R. Vishny, *Fire sales in finance and macroeconomics*, *Journal of Economic Perspectives* **25** (1) (2011) 29–48. doi:10.1257/jep.25.1.29.
- [171] F. Caccioli, M. Shrestha, C. Moore, J. D. Farmer, *Stability analysis of financial contagion due to overlapping portfolios*, *Journal of Banking & Finance* **46** (2014) 233–245. doi:10.1016/j.jbankfin.2014.05.021.
- [172] R. Cont, L. Wagalath, *Fire sales forensics: Measuring endogenous risk*, *Mathematical Finance* **26** (4) (2016) 835–866. doi:10.1111/mafi.12071.
- [173] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*,

- The Annals of Mathematical Statistics **9** (1) (1938) 60–62. doi:10.1214/aoms/1177732360.
- [174] K. P. Burnham, D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, Springer-Verlag New York, 2002. doi:10.1007/b97636.
- [175] K. P. Burnham, D. R. Anderson, *Multimodel inference*, Sociological Methods & Research **33** (2) (2004) 261–304. doi:10.1177/0049124104268644.
- [176] H. Akaike, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control **19** (6) (1974) 716–723. doi:10.1109/TAC.1974.1100705.
- [177] P. Gai, A. Haldane, S. Kapadia, *Complexity, concentration and contagion*, Journal of Monetary Economics **58** (5) (2011) 453–470. doi:10.1016/j.jmoneco.2011.05.005.
- [178] A. Ramadiah, F. Caccioli, D. Fricke, *Reconstructing and stress testing credit networks*, <https://dx.doi.org/10.2139/ssrn.3084543> (2017).
- [179] F. Saracco, R. Di Clemente, A. Gabrielli, T. Squartini, *Randomizing bipartite networks: The case of the World Trade Web*, Scientific Reports **5** (2015) 10595. doi:doi:10.1038/srep10595.
- [180] F. Picciolo, T. Squartini, F. Ruzzenenti, R. Basosi, D. Garlaschelli, *The role of distances in the World Trade Web*, in: 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems, 2012, pp. 784–792. doi:10.1109/SITIS.2012.118.
- [181] A. I. Khintchine, *Mathematical Foundations of Information Theory*, Dover Publications (New York), 1957.