

**Role of adjacency-matrix degeneracy in maximum-entropy-weighted network models**

O. Sagarra,\* C. J. Pérez Vicente, and A. Díaz-Guilera

*Departament de Física Fonamental, Universitat de Barcelona, 08028 Barcelona, Spain*

(Received 3 August 2015; published 30 November 2015)

Complex network null models based on entropy maximization are becoming a powerful tool to characterize and analyze data from real systems. However, it is not easy to extract good and unbiased information from these models: A proper understanding of the nature of the underlying events represented in them is crucial. In this paper we emphasize this fact stressing how an accurate counting of configurations compatible with given constraints is fundamental to build good null models for the case of networks with integer-valued adjacency matrices constructed from an aggregation of one or multiple layers. We show how different assumptions about the elements from which the networks are built give rise to distinctively different statistics, even when considering the same observables to match those of real data. We illustrate our findings by applying the formalism to three data sets using an open-source software package accompanying the present work and demonstrate how such differences are clearly seen when measuring network observables.

DOI: [10.1103/PhysRevE.92.052816](https://doi.org/10.1103/PhysRevE.92.052816)

PACS number(s): 89.75.Fb, 64.60.aq, 89.75.Hc, 89.65.—s

**I. INTRODUCTION**

Network science [1] is a prime example of the multiple uses that many tools and methodologies extracted from traditional physics can have when applied to a variety of transdisciplinary problems. The advent of the so-called Big Data Era given by the explosion of ICT technologies is providing researchers with unprecedented large data sets in a myriad of fields ranging from biology [2] to urban studies [3], including bibliometrics [4], to chemical sciences [5] or even history [6], to cite just a few. The current challenge is to extract knowledge and useful information from such enormous data sets. A standard approach is based on representing them as a graph. Network representation of data is especially useful due to its relative simplicity in terms of computational effort, visualization [7], and analysis. However, it presents some serious limitations, which force us to look for innovative methodological tools.

One way to go beyond simple network representation is to generate appropriate null models. They must be flexible and reliable enough for comparison with our original data in the search for statistically relevant patterns. In general, this is not a simple task; data processing is tricky and subtle in many situations and it may lead to wrong conclusions based on a poor understanding of the problem under study. A clever strategy to find efficient null models consists of generating randomized instances of a given network while keeping some quantities constant [7,8]. This can be done by algorithmic randomization of graphs [9] but such a procedure can be costly in terms of computational time (especially for dense data sets) and programming difficulty. Most importantly, most “rewiring techniques” do not always generate an unbiased sampling of networks [10].

A different approach to this problem has its roots in the analogy of networks with classical statistical mechanics systems [11–15], though it was originally proposed by sociologists and also by urban planners [16] under the name of exponential random graphs [17]. It is based on the idea of constructing an ensemble of networks with different probabilities of appear-

ance, which on average fulfill the considered constraints. The advantage of these methods is that they consider the possibility of having fluctuations (as usually happens in real data) and their derivation is completely analytical. Furthermore, such methods provide an easy way of rapidly simulating (and averaging) network instances belonging to a given ensemble. So far in the literature successful development of this kind of methodology has been performed for different types of monolayered networks [13,18–20], directed [14] and bipartite [21] structures, and stochastic block models [22] and some multiplex weighted networks [4].

Recently, there is growing interest in the study of more complex mathematical structures [23,24] to account for the inherent multilayered character of some network systems. This fact calls for the need to develop maximum entropy ensembles with a multilayered perspective [4], which will help in the analysis of real-world data sets. This is the main goal of this work. In this paper, we complement previous work on maximum-entropy-weighted networks by considering systems of aggregated multiplexes, where we have information about the layered structure of the system and the nature of their events, but—as usually happens for real data—we only have access to its accumulated structure (the sum of weights connecting two nodes in each layer for each pair of nodes). We show how the role of event and layer degeneracy induces important differences in the obtained statistics for each case, which recovers the monolayered studied cases when the number of layers is set to unity. We further show that, despite the statistics being different, all the cases considered are examples of maximum likelihood networks of the dual problem [25] but yield different expectations for network quantities, highlighting the need to choose an appropriate null model for each case study based on the weighted character of the networks.

In Sec. II we present the mathematical framework and calculations of degeneracy for maximum-entropy networks with arbitrary constraints. Section III extends the calculation to obtain explicit statistics for a very general case of constraints for the different cases considered. Finally, Sec. IV presents an application of the model for the particular case of fixed strengths to the analysis of three real-world data sets. Extended

\*osagarra@ub.edu

mathematical calculations are provided in the Appendixes, and details on the used data sets, measured quantities, and numerical methods in the Supplementary Material [26], including also a package for public use to apply the proposed models [27].

## II. MAXIMUM-ENTROPY-CONSTRAINED GRAND-CANONICAL NETWORK ENSEMBLES WITH INTEGER WEIGHTS

We consider a representation of a network of  $N$  nodes, based on an adjacency matrix  $\mathbf{T}$  composed of positive-integer-valued entries  $t_{ij} \in \mathbb{N}$ , which we call occupation numbers. Each of these entries accounts for the intensity of the interaction between any given pair of nodes  $i$  and  $j$  in the network, measured in terms of discrete events (which may be trips between locations in a mobility network or messages between users in social networks for instance). We study the case of directed networks with self-loops, albeit the undirected case follows from the derivation. Our objective is to fully determine the grand canonical ensemble of networks [11,20,28] which fulfill on average some  $Q + 1$  given constraints  $\{T \equiv \sum_{ij} t_{ij}, C_q(\mathbf{T})\}$ . The total number of events  $T = \sum_{ij} t_{ij}$  determines the sampling of the network and is the minimal required constraint to consider any ensemble under the present framework [20]. In this paper, we examine the problem where constraints take the form of linear combinations of functions of the individual occupation numbers  $t_{ij}$ :

$$C_q(\mathbf{T}) = \sum_{ij} f_q^{ij}(t_{ij}), \quad \forall q \in Q. \quad (1)$$

To completely determine an ensemble, it is not enough to specify the quantities we wish to fix (the constraints given by the original data); we must also define the statistical nature of the events allocated to the occupation numbers  $t_{ij}$ . In other words, we have to count all the network instances which give rise to the same particular configuration of the adjacency matrix  $\mathbf{T}$ . This degeneracy term  $D(\mathbf{T})$  depends solely on the specifics of the system one represents and counts the number of equivalent (micro) states that a particular unique realization of the adjacency matrix  $\mathbf{T}$  can describe.

Once a grand canonical ensemble is fully constructed, the probability of obtaining a particular configuration of occupation numbers  $\mathbf{T}$  reads

$$P(\{\theta_q\}, \mathbf{T}) = \mathcal{Z}^{-1} D(\mathbf{T}) e^{H(\{\theta_q\}, \mathbf{T})}, \quad (2)$$

$$H(\{\theta_q\}, \mathbf{T}) \equiv \theta_T T(\mathbf{T}) + \sum_q \theta_q C_q(\mathbf{T}).$$

The so-called grand-canonical partition function  $\mathcal{Z} = \sum_{\{\mathbf{T}\}} D(\mathbf{T}) e^{H(\{\theta_q\}, \mathbf{T})}$  must be summed considering all the possible configurations of the adjacency matrix  $\mathbf{T}$  one can consider, regardless of whether the proposed constraints are met. Such a probability with an exponential form [17] is obtained by maximizing the Shannon entropy  $S = \sum_{\{\mathbf{T}\}} P(\mathbf{T}) \ln P(\mathbf{T})$  associated with the ensemble while preserving the  $Q + 1$  constraints on average.

Using Eq. (1) we reach

$$P(\mathbf{T}) = \mathcal{Z}^{-1} D(\mathbf{T}) \prod_{ij} z_T^{t_{ij}} z_{ij}(t_{ij}), \quad (3)$$

$$z_{ij}(t_{ij}) \equiv \prod_q e^{\theta_q f_q^{ij}(t_{ij})}, \quad z_T \equiv e^\theta.$$

Let us note that if the degeneracy term factorizes, i.e.,  $D(\mathbf{T}) \propto \prod_{ij} D_{ij}(t_{ij})$ , the partition function can be re-expressed as

$$\mathcal{Z} = \prod_{ij} \mathcal{Z}_{ij} = \prod_{ij} \sum_{t_{ij}=0}^{\infty} D_{ij}(t_{ij}) z_T^{t_{ij}} z_{ij}(t_{ij}), \quad (4)$$

$$P(\mathbf{T}) = \prod_{ij} p_{ij}(t_{ij}) = \prod_{ij} \frac{D_{ij}(t_{ij}) z_T^{t_{ij}} z_{ij}(t_{ij})}{\sum_{t'_{ij}=0}^{\infty} D_{ij}(t'_{ij}) z_T^{t'_{ij}} z_{ij}(t'_{ij})},$$

where the statistics of  $\mathbf{T}$  are formed by a set of independent random variables corresponding to the occupation numbers  $\{t_{ij}\}$ . Whenever one defines the degeneracy term and is able to sum the individual partition functions  $\mathcal{Z}_{ij}$ , then one gets the explicit statistics of the occupation numbers. The values of the Lagrange multipliers  $(\theta_T, \{\theta_q\})$  associated with the  $Q + 1$  constraints (which can also be understood as *a posteriori* hidden variables [29,30]) are obtained by solving the so-called saddle-point equations,

$$\hat{C}_q = \langle C(\mathbf{T}) \rangle = \sum_{ij} \langle f_q^{ij}(t_{ij}) \rangle = \sum_{ij} \sum_{t'_{ij}=0}^{\infty} p_{ij}(t'_{ij}) f_q^{ij}(t'_{ij}), \quad (5)$$

$$\hat{T} = \sum_{ij} \langle t_{ij} \rangle,$$

where  $\{\hat{C}_q\}$  are the values of the quantities one wants to keep fixed (on average) in the ensemble.

The degeneracy terms are in general subtle to compute and, to the best of our knowledge, are seldom considered in the literature. In order to calculate them, however, we need to make considerations about the type of networks under study and their elements. In this work, we consider systems composed of events that are either distinguishable or indistinguishable. Additionally, we study the general representation of an overlay multiplex network, which is obtained by aggregating several layers of a system into a single (integer-weighted) adjacency matrix [23,24]. Examples of such networks include aggregation of transportation layers [31], networks generated by accumulation of information over a certain time span such as origin-destination matrices [3], e-mail communications [32] or human contacts [33], and even an aggregation of trading activities in different sectors such as the World Trade Network (WTN) [34].

Thus the system under consideration is an aggregation of  $M$  network layers containing the same type of events: They can be either a group of layers composed of distinguishable (multiedge; ME) or indistinguishable (weighted; W) events or even an aggregation of binary (B) networks. The occupation numbers corresponding to layer  $m$  are denoted  $t_{ij}^m$ , but we only have *access* to information about their accumulated value through all the layers, i.e., the aggregated occupation numbers  $t_{ij} = \sum_m t_{ij}^m$ .

TABLE I. Degeneracy terms corresponding to the elements of the system and their layers in each case.

Network type	$D(\mathbf{T})_{\text{Events}}$	$D(\mathbf{T})_{\text{Layers}}$
Multidige (ME)	$\frac{T!}{\prod_{ij} (\sum_m t_{ij}^m)!}$	$\prod_{ij} \frac{(\sum_m t_{ij}^m)!}{\prod_m t_{ij}^m!} = \prod_{ij} M^{\sum_m t_{ij}^m}$
Weighted (W)	1	$\prod_{ij} \binom{M + \sum_m t_{ij}^m - 1}{\sum_m t_{ij}^m}$
Binary distinguishable (BD)	$T!$	$\prod_{ij} \binom{M}{\sum_m t_{ij}^m}$
Binary indistinguishable (BI)	1	$\prod_{ij} \binom{M}{\sum_m t_{ij}^m}$

Finally, the degeneracy term is the product of the multiplicity induced by the nature of the events times the nature of the layers (which, in the only possible real scenario, are always distinguishable):  $D(\mathbf{T}) = D(\mathbf{T})_{\text{Events}} D(\mathbf{T})_{\text{Layers}}$ . The latter term is computed (for each pair of nodes or state  $ij$ ) by counting the number of different groupings one can construct by splitting  $t_{ij} = \sum_m t_{ij}^m$  (distinguishable or indistinguishable) aggregated events into  $M$  layers respecting the occupation limitation of the considered events: either only one event per layer (B network) or an unrestricted number (W and ME networks).

The resulting degeneracy terms are listed in Table I (see details in Appendix D), in which one can see that, in principle, the event degeneracy term does not factorize for the distinguishable cases due to the presence of the variable  $T! = (\sum_{ij} t_{ij})!$ . One can nevertheless obtain an effective degeneracy term by substituting it with  $\hat{T}!$  (a constant) (as shown in Appendix D, where a complete discussion of the implications of this substitution for the different cases is provided), which leads to results fully equivalent to those obtained by performing the exact calculation for the ME case with constraints of the form (6). In doing so, two preliminary conclusions can be drawn. First, both the distinguishable and the indistinguishable binary cases will lead to the same statistics since their degeneracy term for events will be constant (hence in the remainder of the paper we omit the BD case). Second, in all cases the complete degeneracy terms will factorize into state  $ij$  independent terms, which means that the statistics of the aggregated occupation numbers will be state independent [Eq. (4)].

### III. LINEAR CONSTRAINTS ON AGGREGATED OCCUPATION NUMBERS

To go further in our derivation, we now consider the case where the constraints are linear functions of the aggregated occupation numbers:

$$f_q^{ij}(t_{ij}) = c_q^{ij} t_{ij} = c_q^{ij} \sum_m t_{ij}^m. \quad (6)$$

This case is very generic and includes networks with local constraints on nodes [35], community constraints [18], and generalized cost constraints such as distances [36]. The case where the constraints depend on both the binary projection of the occupation numbers and their values  $f_{ij}(t_{ij}) = c_q^{ij} t_{ij} +$

$\tilde{c}_q^{ij} \Theta(t_{ij})$  can be derived from the methodology developed here and is analyzed in Appendix A.

The individual partition functions can be summed:

$$\begin{aligned} Z_{ij} &= \sum_{t_{ij}} D_{ij}(t_{ij}) z_{ij}^{t_{ij}} \\ &= \begin{cases} \text{ME: } \sum_{t_{ij}=0}^{\infty} \frac{(M z_{ij})^{t_{ij}}}{t_{ij}!} = e^{M z_{ij}}. \\ \text{W: } \sum_{t_{ij}=0}^{\infty} \binom{M+t_{ij}-1}{t_{ij}} z_{ij}^{t_{ij}} = (1-z_{ij})^{-M}; & z_{ij} < 1. \\ \text{B: } \sum_{t_{ij}=0}^M \binom{M}{t_{ij}} z_{ij}^{t_{ij}} = (1+z_{ij})^M; & t_{ij} \leq M. \end{cases} \end{aligned} \quad (7)$$

In this case, we have redefined  $z_{ij} \equiv e^{\theta r} \prod_q e^{\theta_q c_q^{ij}}$  to ease notation. This leads to

$$\begin{aligned} p_{ij}^{\text{ME}}(t_{ij}) &= e^{-M z_{ij}} \frac{(M z_{ij})^{t_{ij}}}{t_{ij}!}, \\ p_{ij}^{\text{W}}(t_{ij}) &= \binom{M+t_{ij}-1}{t_{ij}} z_{ij}^{t_{ij}} (1-z_{ij})^M, \\ p_{ij}^{\text{B}}(t_{ij}) &= \binom{M}{t_{ij}} \left( \frac{z_{ij}}{1+z_{ij}} \right)^{t_{ij}} (1+z_{ij})^{-(M-t_{ij})}. \end{aligned} \quad (8)$$

And we recover well-known probability distributions: a Poisson distribution for the ME case [20] (independent of the number of layers  $M$ ), negative binomial for the W case (the geometric distribution [19] being a special case when  $M = 1$ ), and a binomial distribution for the aggregated B case (the Bernoulli distribution [11] being a special case for  $M = 1$ ).

The resulting statistics show some important features: On the one hand, one sees that although the degeneracy term changes for ME networks for the case of either a monolayer or a multilayer, the form of the obtained statistics does not. This means that it is not possible to distinguish an ME monolayered network from an aggregation of multiple ME layers belonging to an ensemble with the same constraints. On the other hand, the situation for the other cases changes: For multiplexes the resulting occupation numbers will have statistics different from those in the monoplex case. This has the implication that one could in principle discern the aggregated nature of a network by inspecting the accumulated edge statistics  $\{t_{ij}\}$ , provided that one has access to enough realizations of a system and that it belongs to the same ensemble [i.e., the system evolves according to some given, even if unknown, linear constraints [3] of the form of Eq. (6)].

TABLE II. First and second moments of the considered distributions, together with the relative fluctuations.

Network type	$\langle t_{ij} \rangle$	$\sigma_{t_{ij}}^2$	$\frac{\sigma_{t_{ij}}^2}{\langle t_{ij} \rangle^2}$	Domain $z_{ij}$
ME	$M z_{ij}$	$M z_{ij}$	$(M z_{ij})^{-1}$	$[0, \infty)$
W	$M \frac{z_{ij}}{1-z_{ij}}$	$M \frac{z_{ij}}{(1-z_{ij})^2}$	$(M z_{ij})^{-1}$	$[0, 1)$
B	$M \frac{z_{ij}}{1+z_{ij}}$	$M \frac{z_{ij}}{(1+z_{ij})^2}$	$(M z_{ij})^{-1}$	$[0, \infty)$

Another important implication of the obtained statistics is the very different interpretations encoded in the  $z_{ij}$  values. This collection of values is related to the constraints originally imposed on the network ensemble through the set of Lagrange multipliers  $(\theta_T, \{\theta_q\})$  [Eqs. (3) and (5)] and can be understood as *a posteriori* measures related to the intensity of each node pair  $ij$ . These measures encode the correlations between nodes imposed by the constrained topology (note that for local constraints only at the level of nodes do we obtain a factorization  $\langle t_{ij} \rangle = M x_i y_j$ ). Table II lists the first two central moments of each distribution. For the ME case  $z_{ij}$  is directly mapped both to the average occupation of the considered link  $ij$ ,  $\langle t_{ij} \rangle$ , and to its (relative) importance in the network [20]. In all the other cases, however,  $z_{ij}$  relates to a probability of a set of events emerging from a given node, to be allocated to a link  $ij$ . Obviously, as  $\langle t_{ij} \rangle$  increases,  $z_{ij}$  increases in all cases, but not in the same linear way (in the W case, for instance,  $z_{ij}$  is bounded to a maximum value of 1). This means that while in all cases  $z_{ij}$  is related to the importance of a given link with respect to the others, the dependence in all non-ME cases is highly nonlinear. Finally, we can see that for a large number of layers  $M \gg T/N^2$ , the ensembles become equivalent to the ME, as the degeneracy term for (distinguishable) layers dominates the configuration space of the ensembles.

The different obtained statistics are highly relevant, as their marked differences point out a (regularly overlooked) problem: Different maximum-entropy ensembles yield very different statistics for the same considered constraints, and hence each data set needs to be analyzed carefully, since the process behind the formation of each network dictates its degeneracy term. Furthermore, all the obtained statistics are derived from a maximum-entropy methodology, and hence the values  $z_{ij}$  obtained from (5) are in all cases maximum-likelihood estimates of the probability that  $\mathbf{T}$  will belong to the set of models described by Eq. (8) (see Appendix B). Thus, any of the presented models will be a correct ensemble in a maximum-likelihood sense [25] for some given constraints, and the appropriate choice for each network representation depends on the system under study, in contrast to the interpretation given by [35].

This means that if one wants to assess the effects a given constraint has on a network constructed from real data, one needs to choose very carefully the appropriate null model to compare the data to. It is also worth pointing out that most of these ensembles are not equivalent to a manual rewiring of the network [37] (although one expects small differences; see Appendix C). However, maximum-entropy models allow for an analytical treatment of the problem and simplify the gen-

eration of network samples when the considered constraints are increasingly complicated (both at the coding level and at the computational one). This has many implications, including the possibility of computing  $p$  values, information-theoretic-related quantities such as ensemble entropies [13–15,18,22] and model likelihoods as well as efficient weighted network pruning algorithms [38,39]. Moreover, this procedure helps in the fast and simple generation of samples of networks with prescribed constraints.

The main difficulty with the *soft-constrained* maximum-entropy framework presented here for null model generation is the problem of solving the saddle-point equations, (5), associated with each ensemble. With the exception of some particular cases [37], these equations do not have an analytical solution and must be obtained numerically. In this case, the best approach is to maximize the associated loglikelihood of each model to a set of observations (constraints), yet the difficulty of each problem increases with the number of constraints since each fixed quantity has an associated variable to be solved. Considering the different statistics obtained in this paper, the most difficult case by far is the W one, since the condition that  $0 \leq z_{ij} < 1$  imposes a nonconvex condition in the domain of the loglikelihood function to maximize, while the others are in general easily solved using iterative balancing algorithms (see the Supplementary Material for an extended discussion and details on the numerical methods used [26]).

#### IV. APPLICATION TO REAL DATA: THE CASE OF FIXED STRENGTHS

To highlight the importance of considering an appropriate null model for the assessment of real data features, in this final section we consider the case of networks with a fixed strength sequence. Real networks usually display highly skewed node strength distributions, which have important effects on their observables. Hence, to correctly assess whether some observed feature in a data set can be explained solely by the strength distribution, it is crucial to choose an appropriate null model to compare the data to. This situation is especially important, for instance, with regard to community analysis through modularity maximization for weighted networks, because the modularity function to be optimized [18] needs as input a prediction from a null model with fixed strengths [ $Q \propto \sum_{ij} (\hat{t}_{ij} - \langle t_{ij} \rangle) \delta_{c_i, c_j}$ , where  $\{c\}$  are the community node labels associated with the optimal network partition]. For a directed model with fixed strengths, the constraints in Eq. (1) read  $(\theta_T$  is not needed because  $T = \sum_i s_i^{\text{in}} = \sum_j s_j^{\text{out}}$ )

$$\begin{aligned}
 s_q^{\text{out}} &= \sum_{ij} c_q^{ij} t_{ij} = \sum_{ij} \delta_{qi} t_{ij} = \sum_j t_{qj}, \quad \forall q = 1 \dots N; \\
 s_r^{\text{in}} &= \sum_{ij} c_r^{ij} t_{ij} = \sum_{ij} \delta_{jr} t_{ij} = \sum_i t_{ir}, \quad \forall r = 1 \dots N; \quad (9) \\
 z_{ij} &= \prod_q e^{\theta_q \delta_{iq}} \prod_r e^{\theta_r \delta_{ri}} = x_i y_j, \quad x_q \equiv e^{\theta_q}, \quad y_r \equiv e^{\theta_r}.
 \end{aligned}$$

So the resulting saddle-point equations, (5), are

$$\hat{s}_i^{\text{out}} = \langle s^{\text{out}} \rangle, \quad \hat{s}_i^{\text{in}} = \langle s^{\text{in}} \rangle, \quad i = 1 \dots N, \quad (10)$$



where  $\hat{s}$  denotes the numerical value found in the data for the random variable  $s$ , which, particularized to each case, reads

$$\begin{aligned} \text{ME} \quad & \begin{cases} \hat{s}_i^{\text{out}} = M x_i \sum_j y_j, \\ \hat{s}_j^{\text{in}} = M y_j \sum_i x_i; \end{cases} \\ \text{W} \quad & \begin{cases} \hat{s}_i^{\text{out}} = M x_i \sum_j \frac{y_j}{1-x_i y_j}, \\ \hat{s}_j^{\text{in}} = M y_j \sum_i \frac{x_i}{1-x_i y_j}; \end{cases} \\ \text{B} \quad & \begin{cases} \hat{s}_i^{\text{out}} = M x_i \sum_j \frac{y_j}{1+x_i y_j}, \\ \hat{s}_j^{\text{in}} = M y_j \sum_i \frac{x_i}{1+x_i y_j}. \end{cases} \end{aligned} \quad (11)$$

The ME case has an analytical solution [37], while the others must be solved computationally. The Supplementary Material [26] provides extended details about the network quantities computed, simulations, averaging and computational methods, and algorithms used in this section, which are available in the freely provided, open-source ODME package [27].

As real-world data sets we use a snapshot of the WTN, the OD matrix generated by taxi trips in Manhattan for the year 2011 [3,40], and the multiplex European Airline Transportation network [31]. The WTN has been widely studied in the literature and recently has been represented as an aggregated system of  $M \sim 100$  layers [41] representing

different types of commodities being traded. In this network, nodes represent countries and weights represent the amount of trade between them, measured in millions of U.S. dollars. In the OD taxi data set, which we construct as the aggregation of  $M = 365$  daily layer snapshots, each node represents an intersection and each weight the number of trips recorded between them [42]. Finally, in the airline network each node is an airport and weights correspond to the number of airlines providing direct connections between them, so the network is an aggregation of  $M = 37$  binary layers (one for each airline).

In all cases we consider directed networks, and throughout this paper we only show results in the outgoing direction, as the results in the incoming direction are qualitatively equal. Note that the aggregated B case cannot always be applied since the maximum number of events allocated per node pair cannot exceed the number of layers, and for the WTN data set this condition [ $\max(\{\hat{s}_i\}) \leq s^{\text{max}} = NM$ ] is violated for some nodes.

To analyze the difference between models, we compute ensemble expectations for different edge- and node-related properties suitably rescaled (fixing the original strength distribution of each data set) and then compare the obtained results with the real observed data features. The airline data set is very sparse and differences between models are not great, which points out the need for adequate sampling for a

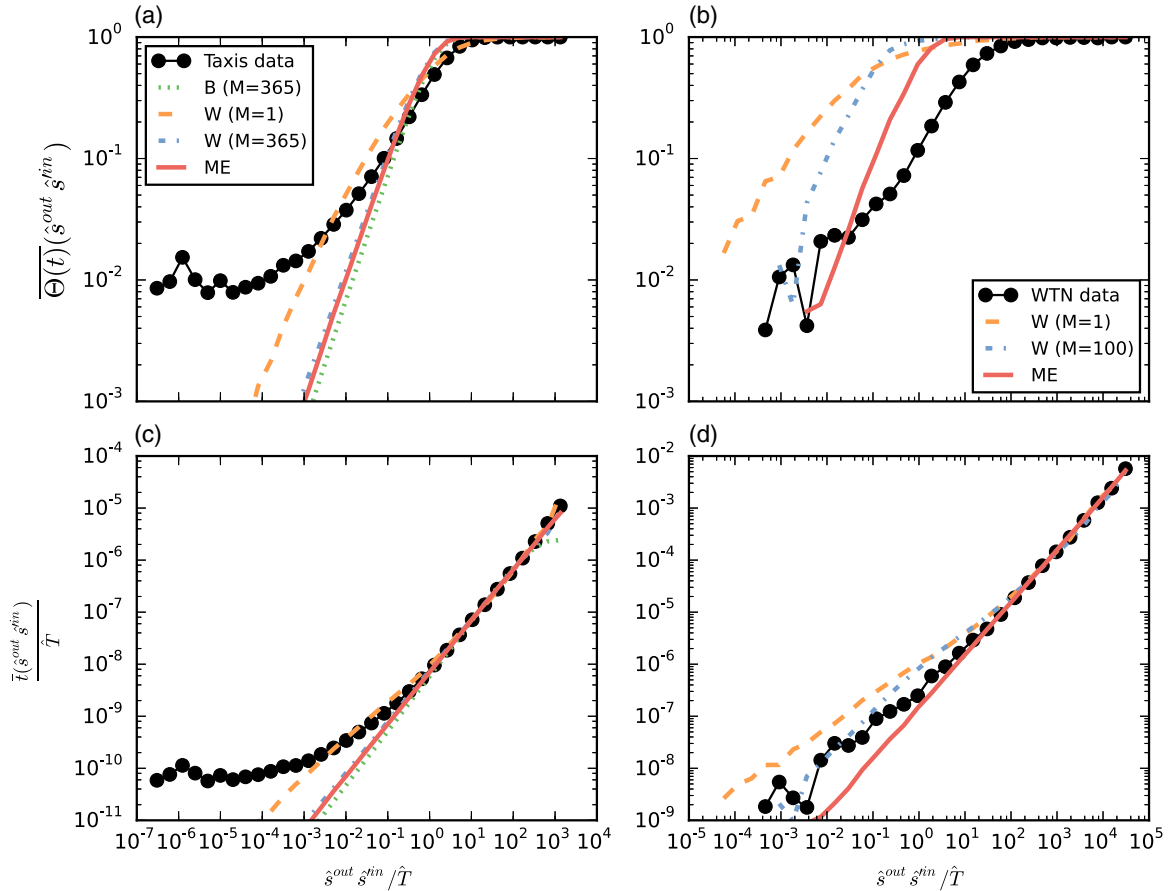


FIG. 1. (Color online) Node pair statistics. (a, b) Binary connection probability and (c, d) rescaled average edge weight as a function of the product of origin and destination node strength. Results averaged over  $r = 5 \times 10^2$  and  $r = 10^4$  realizations for the different models, respectively, with applied log-binning. The sudden increase in the binary pair-node connection probability is clearly shown for the W case.

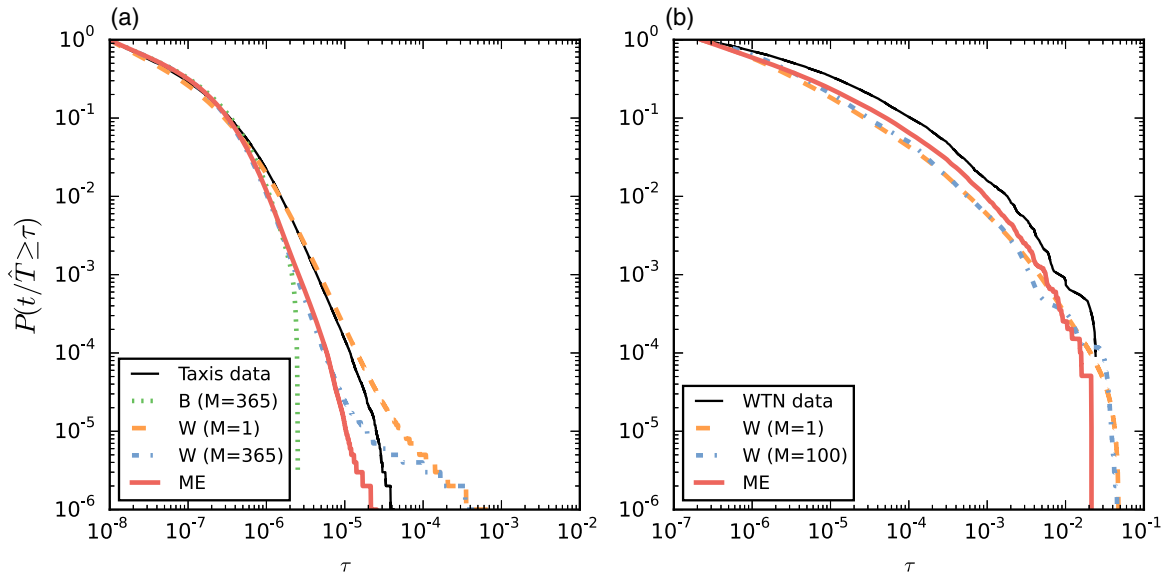


FIG. 2. (Color online) Weight statistics. Existing node pair weight complementary cumulative distribution for (a) the taxi and (b) the WTN data sets. The same conditions as for Fig. 1 apply. The presence of extremely high weights is shown in the tails of the distributions for both the W monolayer and the W multilayer case.

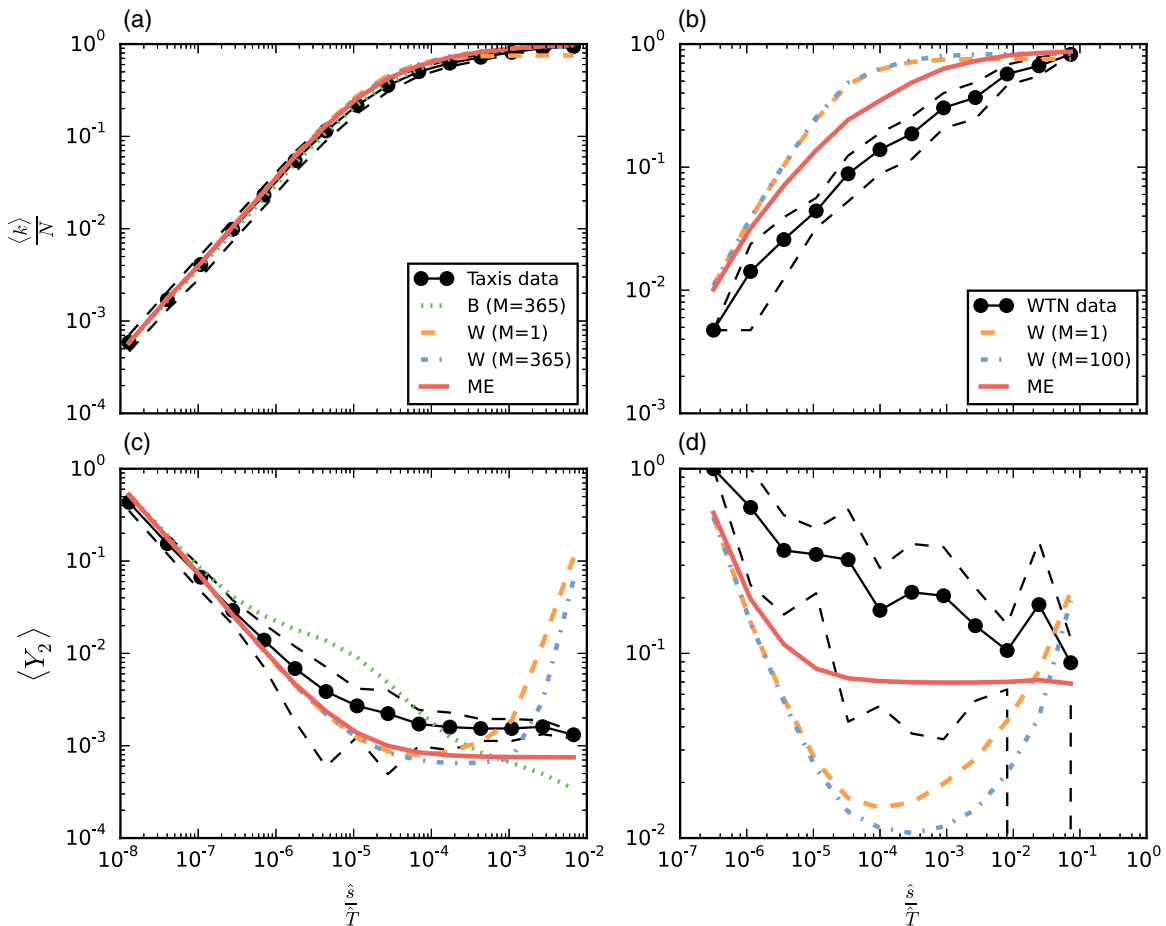


FIG. 3. (Color online) First-order node statistics. Rescaled (a, b) degree and (c, d) disparity for (a, c) the taxi and (b, d) the WTN data sets. The same conditions as for Fig. 1 apply. Dashed lines represent log-binned standard deviation ranges for the real data. The U-shaped disparity profile is clearly shown for the W cases, in sharp contrast with the monotonous behavior of both the real data and the ME model.

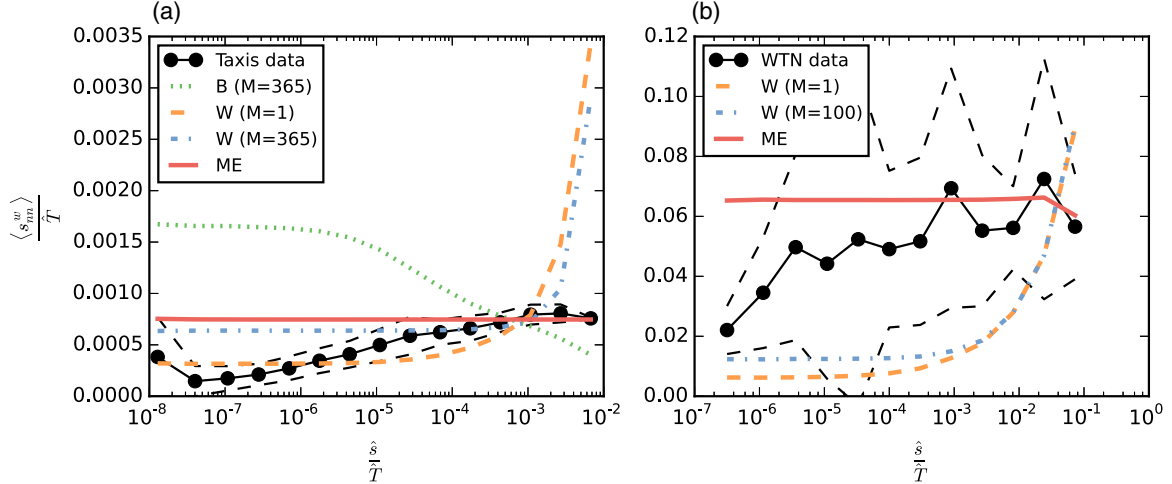


FIG. 4. (Color online) Second-order node statistics. Rescaled weighted average strength for (a) the taxi and (b) the WTN data sets. The same conditions as for Fig. 1 apply. Dashed lines represent log-binned standard deviation ranges for the real data. A sharp increase is clearly shown for high-strength nodes in the W cases.

successful analysis of weighted networks. Anyway, since it is by construction an aggregation of binary layers, the B case displays the most resemblance to the data, both qualitatively and quantitatively (see the Supplementary Material [26]).

The WTN and taxi data sets, in contrast, contain enough sampling for the wide differences between models to emerge. All cases have the same number of events  $\hat{T}$  on average, but they are not distributed among connections between nodes in the same way for the different models. Zero being the most probable value for the geometric distribution, for the W case with a single layer the connection probability initially increases distinctively faster than in all the other cases, leading to larger numbers of binary connections between low-strength nodes [Figs. 1(a) and 1(b)]. Yet the higher relative fluctuations of the geometric statistics also generate the extremely large maximum weights in the tail of the existing occupation number distribution (Fig. 2), which are concentrated in connections between high-strength nodes [Figs. 1(c) and 1(d)]. Since the total number of incoming and outgoing events at each node is fixed, this means that the W case has comparatively the lowest degrees for the most weighted nodes despite counting the larger number of binary connections,  $E = \sum_{ij} \Theta(t_{ij}) = \sum_i k_i^{\text{out}} = \sum_j k_j^{\text{in}}$ , as shown in Figs. 3(a) and 3(b).

These anomalies for low- and high-strength nodes, respectively, for the W case produce wild asymmetries in the allocation of weights per node, which can be studied by measuring their disparity,  $Y_2 = \sum_j t_{ij}^2 / (\sum_j t_{ij})^2$  [Figs. 3(c) and 3(d)], which quantifies how homogeneously distributed the weights emerging from each node are: It displays a U-shaped form, with both low- and high-strength nodes tending to very strongly concentrate their weights in a few connections. This nonmonotonic behavior is in strong contrast with that observed for the real data and usually for other data sets [4]. Concerning second-order node correlations, the outgoing weighted-average neighbor strength  $s_{nn}^w = \sum_j t_{ij} s_j^{\text{in}} / s_i^{\text{out}}$  (Fig. 4) again displays a large range of variation for the W case (with either one or more than one layer), in contrast with the slightly assortative profile of the real data, the uncorrelated

profile in the ME case, and the slight disassortative trend in the B case. The latter case is caused by the combination of two factors: The limitation on the maximum weight of the edges cannot compensate (with high weights connecting the nodes of higher strength) the tendency of large nodes to be connected to a macroscopic fraction of the network, which is dominated by low-strength nodes.

Obviously none of the null models used reproduce the real data, however, the goal in model construction is rather to assess the structural impact that a given constraint (in this case, a strength distribution) has on the network observables. In this sense, we show that different models provide very different insights into such impacts. In particular, since the airline data set is by construction an aggregated B network and the taxi data set an ME one (people riding in taxis are clearly distinguishable), the fact that the B and ME cases, respectively, lie closer to the real data comes as no surprise. The WTN case, however, is unclear: The modeling of trade transactions does not have a clearly defined nature, but if one assumes the WTN to be a multilayered network, its aggregated analysis should be performed using either the W case (with  $M > 1$  layers) or the ME case, which, again, are closer in both functional form and qualitative values to the real case (in contrast with [41], where the W model with a single layer is used).

## V. CONCLUSIONS

In this work we have shown the importance of considering the nature of the events one wishes to model when using an integer-valued network representation of a system. We have developed and solved a maximum-entropy framework for model generation applied to networks generated by aggregation [23,24] of multiplexes. We have shown how different considerations about the nature of the events generating the elements of the multiplex give rise to distinctively different node pair statistics. For the case where one wants to fix properties expressed as linear functions of the individual weights (and, optionally, their binary projection) in the network, we elegantly

recovered well-known statistics such as Poisson, binomial, negative binomial, geometric, and Bernoulli in each case.

We have, further, provided a practical example by focusing on the case of fixed strengths and applying the models to assess relevant features of three real-world data sets containing different types of weights, showing how the role of adjacency-matrix degeneracy plays a crucial role in model construction. To this end, we have considered the statistical nature of the obtained models as well as the weaknesses and strengths derived from their practical applications in real cases. Finally, we provide the open-source software package ODME [27], with which practitioners can apply the proposed models to other data sets.

The insights derived from this paper can open the door to the objective identification of truly multiplex structures (except in one case where it has been shown to be impossible) by inspection of the statistics of their edges, provided that several instances of a network belonging to the same ensemble are available. The take-home message of this work is that, in order to perform a meaningful analysis of a given network, a practitioner needs to be able to select an appropriate null model, which depends not only on the endogenous constraints one considers but also on the very nature of the process one is modeling. Our work provides researchers with a range of maximum-entropy (and maximum-likelihood) models to choose from, covering a wide spectrum of possibilities for the case of weighted networks. Each of these models is not wrong or even right in the general case, despite yielding very different predictions for the same sets of constraints, but just more or less appropriate, depending on the problem one wants to study.

### ACKNOWLEDGMENTS

We thank A. Allard, A. Fernández, F. Font-Clos, and R. Toral for useful comments and suggestions. This work was partially supported by the LASAGNE (Contract No. 318132) and MULTIPLEX (Contract No. 317532) EU projects, the Spanish MINECO (FIS2012-38266), and the Generalitat de Catalunya (2014SGR608). O.S. acknowledges financial support from Generalitat de Catalunya (FI program) and the Spanish MINECO (FPU program).

### APPENDIX A: BINARY CONSTRAINTS

We develop here the general case where the constraints have the general form  $f_q^{ij}(t_{ij}) = \tilde{c}_q^{ij} \Theta(t_{ij}) + c_q^{ij} t_{ij}$ . The general derivation remains essentially unchanged from the linear case (see text), with a slight modification in the calculation of the explicit partition function,

$$\begin{aligned} Z_{ij} &= \sum_{t_{ij}=0}^{\infty} D_{ij}(t_{ij}) \tilde{z}_{ij}^{\Theta(t_{ij})} z_{ij}^{t_{ij}} \\ &= \tilde{z}_{ij} \left( \sum_{t_{ij}=0}^{\infty} D_{ij}(t_{ij}) \tilde{z}_{ij}^{t_{ij}} - D_{ij}(0) \right) + D_{ij}(0), \\ z_{ij} &\equiv e^{\theta \tau} \prod_q e^{\theta_q c_q^{ij}}, \quad \tilde{z}_{ij} \equiv \prod_q e^{\theta_q \tilde{c}_q^{ij}}, \end{aligned} \quad (\text{A1})$$

where  $\{z_{ij}\}$  have been redefined and the constraint on the total number of events  $T = \sum_{ij} t_{ij}$  is introduced in their redefinition. This yields

$$p_{ij}(t_{ij}) = \frac{D_{ij}(t_{ij}) \tilde{z}_{ij}^{\Theta(t_{ij})} z_{ij}^{t_{ij}}}{\tilde{z}_{ij} \left( \sum_{t'_{ij}=0}^{\infty} D_{ij}(t'_{ij}) \tilde{z}_{ij}^{t'_{ij}} - D_{ij}(0) \right) + D_{ij}(0)}, \quad (\text{A2})$$

which corresponds to the zero-inflated versions of the previous statistics recovered in the case  $f_{ij} = c_q^{ij} t_{ij}$ , that is, asymmetric statistics where the probability of the first occurrence is different from the rest. Note that, in this very general case, one can always set  $\{c_q^{ij} = 0 \forall ij\}$  to include the case where only binary constraints are considered. Explicitly, for the different statistics, we have the following.

ME (zero-inflated Poisson):

$$p_{ij}(t_{ij}) = \frac{(M z_{ij})^{t_{ij}}}{t_{ij}!} \frac{\tilde{z}_{ij}^{\Theta(t_{ij})}}{\tilde{z}_{ij} (e^{M z_{ij}} - 1) + 1}.$$

W (zero-inflated negative binomial):

$$p_{ij}(t_{ij}) = \binom{M + t_{ij} - 1}{t_{ij}} z_{ij}^{t_{ij}} \frac{\tilde{z}_{ij}^{\Theta(t_{ij})}}{\tilde{z}_{ij} ((1 - z_{ij})^{-M} - 1) + 1}.$$

B (zero-inflated binomial):

$$p_{ij}(t_{ij}) = \binom{M}{t_{ij}} z_{ij}^{t_{ij}} \frac{\tilde{z}_{ij}^{\Theta(t_{ij})}}{\tilde{z}_{ij} ((1 + z_{ij})^M - 1) + 1}.$$

We can then compute the binary connection statistics:

$$\langle \Theta(t_{ij}) \rangle = 1 - p_{ij}(0) = \begin{cases} \text{ME: } \frac{\tilde{z}_{ij} (e^{M z_{ij}} - 1)}{\tilde{z}_{ij} (e^{M z_{ij}} - 1) + 1} \\ \text{W: } \frac{\tilde{z}_{ij} ((1 - z_{ij})^{-M} - 1)}{\tilde{z}_{ij} ((1 - z_{ij})^{-M} - 1) + 1} \\ \text{B: } \frac{\tilde{z}_{ij} ((1 + z_{ij})^M - 1)}{\tilde{z}_{ij} ((1 + z_{ij})^M - 1) + 1} \end{cases} \quad (\text{A3})$$

Note how the binary projection in all cases corresponds to Bernoulli statistics. Regarding the occupation number statistics, one has explicitly

$$\begin{aligned} \langle t_{ij}^{\pm} \rangle &\equiv \langle t_{ij} | t_{ij} > 0 \rangle = \begin{cases} \text{ME: } M \frac{z_{ij}}{1 - e^{-M z_{ij}}} \\ \text{W: } M \frac{z_{ij}}{1 - z_{ij}} \frac{1}{(1 - (1 - z_{ij})^M)} \\ \text{B: } M \frac{z_{ij}}{1 + z_{ij}} \frac{1}{(1 - (1 + z_{ij})^{-M})} \end{cases} \\ \langle t_{ij} \rangle &= \langle \Theta(t_{ij}) \rangle \langle t_{ij}^{\pm} \rangle = \begin{cases} \text{ME: } M z_{ij} \frac{e^{M z_{ij}}}{1 + \tilde{z}_{ij} (e^{M z_{ij}} - 1)} \\ \text{W: } M \frac{z_{ij}}{1 - z_{ij}} \frac{1}{(1 - z_{ij})^M + \tilde{z}_{ij} (1 - (1 - z_{ij})^M)} \\ \text{B: } M \frac{z_{ij}}{1 + z_{ij}} \frac{1}{(1 + z_{ij})^{-M} + \tilde{z}_{ij} (1 - (1 + z_{ij})^{-M})} \end{cases} \end{aligned} \quad (\text{A4})$$

And we observe a clear nontrivial relation between binary statistics and weights, which leads to important correlations between degrees and strengths in networks belonging to these ensembles [20] which are also present in real data [43].

### APPENDIX B: MAXIMUM-LIKELIHOOD DISTRIBUTIONS

The probability distributions derived in this paper for networks belonging to the different described ensembles



fulfill the maximum-likelihood principle for networks [25]. Indeed, the constraint-point equations in (5) can be understood as the equations resulting from maximizing the likelihood of the inverse problem of finding the set of values for the Lagrange multipliers  $(\theta_T, \{\theta_q\})$  that maximize the likelihood of the observed adjacency matrix  $\hat{\mathbf{T}}$  to belong to each of the described models. In other words, defining the loglikelihood function of a network by

$$\begin{aligned}\mathcal{L}(\theta_T, \{\theta_q\} | \hat{\mathbf{T}}) &= \ln \left( \prod_{ij} p_{ij}(\theta_T, \{\theta_q\} | \hat{t}_{ij}) \right) \\ &= \sum_{ij} \ln p_{ij}(\theta_T, \{\theta_q\} | \hat{t}_{ij})\end{aligned}$$

and maximizing this expression with respect to  $(\theta_T, \{\theta_q\})$ , one is led to Eq. (5). Explicitly,

$$\begin{aligned}\partial_{\theta_q} \mathcal{L} &= \sum_{ij} (c_q^{ij} \hat{t}_{ij} + \tilde{c}_q^{ij} \Theta(\hat{t}_{ij}) - c_q^{ij} \langle t_{ij} \rangle - \tilde{c}_q^{ij} \langle \Theta(\hat{t}_{ij}) \rangle) \\ &= \hat{C}_q - \langle C_q \rangle(\{\theta_q\}) \equiv \Delta C_q, \\ \partial_{\theta_q, \theta_q} \mathcal{L} &= - \sum_{ij} (c_q^{ij} c_q^{ij} \sigma_{t_{ij}}^2 + \tilde{c}_q^{ij} \tilde{c}_q^{ij} \sigma_{\Theta(\hat{t}_{ij})}^2) = -\sigma_{C_q, C_q}^2(\{\theta_q\}),\end{aligned}\tag{B1}$$

which, at the critical points, lead to  $\Delta C_q = 0 \forall q$  and the condition of a maximum with respect to all the variables is fulfilled (we also note that the problem in this form is concave). We thus see that the initial statement of the paper is confirmed: It is not enough to specify the constraints to fully define a maximum-entropy ensemble; one needs also to state the nature of its elements, since any maximum-entropy ensemble will lead to a maximum-likelihood description of a data set. There is not a ‘‘correct’’ ensemble to fix a given constraint: just one that best describes the system that is being represented.

### APPENDIX C: ENSEMBLE FLUCTUATIONS

If the maximum-entropy description provided here is to be successful, then the fluctuations of the obtained networks need to be bounded and have well-defined statistics. In particular, we require that the associated entropy of the distribution and the statistics of the constraints possess finite first and second moments and that their relative fluctuations around the average values need to be small in the limit of large sampling  $\hat{T}$ . Explicitly, we have

$$\begin{aligned}\frac{\sigma_{C_q}^2}{\langle C_q \rangle^2} &= \frac{\sum_{ij} (c_q^{ij})^2 \langle t_{ij} \rangle}{(\sum_{ij} c_q^{ij} \langle t_{ij} \rangle)^2} + \frac{a}{M} \frac{\sum_{ij} (c_q^{ij})^2 \langle t_{ij} \rangle^2}{(\sum_{ij} c_q^{ij} \langle t_{ij} \rangle)^2} \\ &= \frac{\sum_{ij} (c_q^{ij})^2 \langle t_{ij} \rangle}{(\sum_{ij} c_q^{ij} \langle t_{ij} \rangle)^2} + \frac{a}{M} \frac{1}{1 + \alpha_q}, \\ \alpha_q &\equiv \frac{\sum_{ij, kl | k \neq i, l \neq j} c_q^{ij} c_q^{kl} \langle t_{ij} \rangle \langle t_{kl} \rangle}{\sum_{ij} (c_q^{ij})^2 \langle t_{ij} \rangle^2},\end{aligned}\tag{C1}$$

where  $a = 0$  for the ME case,  $a = 1$  for the W case, and  $a = -1$  for the B case. We thus see that the fluctuations only disappear for large sampling in the linear case given by Eq. (6) for the ME description. By construction, the constraints are extensive for the occupation numbers  $t_{ij}$ , thus when the number

of events  $T$  increases, their average value  $\langle C_q \rangle$  must also increase [20]; yet only for the ME case do we have  $\langle t_{ij} \rangle \propto \hat{T}$ , and thus only in this case do relative fluctuations decay as  $\hat{T}^{-1}$ . Otherwise, the maximally random allocation of events will be made as homogeneous as possible among the states while preserving the constraints, hence  $\alpha_q$  will in general be a large number [the denominator in the sum has  $L$  terms, while the numerator has  $L(L-1)$ ,  $L$  being the number of available node pairs for the allocation] and relative fluctuations will be bounded and  $\mathcal{O}(M^{-1})$ . For similar reasons, one expects the first term to vanish for large sampling. For a very large number of layers, the ensembles become equivalent to the ME case, and fluctuations vanish in the large sampling limit [37].

For the case where any binary constraint is additionally imposed (Appendix A), the relative fluctuations of the binary structure dominate the statistics in the large sampling limit, and despite being bounded, these never vanish [20].

Concerning the associated Gibbs-Shannon entropy of the ensembles, since the occupation number statistics are independent, we have the random variable  $\ln P(\mathbf{T}) = \ln \prod_{ij} p_{ij}(t_{ij}) = \sum \ln p_{ij}(t_{ij})$  (associated with a given network instance), which is the sum of independent contributions that, in all cases studied (Poisson, negative binomial, and binomial), have well-defined first and second moments when averaged over the ensemble. Hence, the  $\ln P(\mathbf{T})$  statistics will be Gaussian, and no extreme outliers are expected. This indicates that the total average number of possible network instances compatible with a given set of constraints is a well-defined quantity, and one can define a typical network structure representing the ensemble (unlike other studied cases in the literature [44]).

### APPENDIX D: CALCULATION OF DEGENERACY TERMS

#### 1. Layer degeneracy

For each state  $ij$  of the possible  $L(N) = N^2$  node pairs [ $N(N-1)$  if not accepting self-loops], one needs to consider the process of allocating  $t_{ij}$  events in  $M$  possible distinguishable levels. For the W case this corresponds to the *urn problem* of placing  $t_{ij}$  identical balls in  $M$  distinguishable urns. For the B case one faces the problem of selecting groups of  $t_{ij} \leq M$  urns out of a set of  $M$  urns, and finally, for the ME case one must count how to place  $t_{ij}$  distinguishable balls in  $M$  distinguishable urns. These problems are well known and their solution leads to the third column in Table I, with the product over  $ij$  representing the fact that the allocation among the layers for each node pair is independent.

#### 2. Event degeneracy

For this calculation one only needs to take into account the distinguishable case (otherwise there is no degeneracy). This case, however is controversial to analyze. The correct counting of configurations in a grand-canonical ensemble is an issue spanning more than a century (see [45] and references therein for details and extended discussion), ever since Gibbs used it to establish the relation in classical statistical mechanics between the canonical and the grand-canonical ensembles of an ideal gas.

Grand-canonical ensembles of networks can be faced in many ways. The usual view is to imagine a collection of  $\mathcal{N}$  copies of a system in which to distribute  $F$  events in such a way that there are  $\hat{T} = F/\mathcal{N}$  events on average in each copy [46]. In this framework, the probability of obtaining a particular copy with  $T = \sum_{ij} t_{ij}$  events and a set of constraints  $\{C_q(\mathbf{T})\}$  reads

$$P(\mathbf{T}) \propto e^{\theta_T T} e^{\sum_q \theta_q C_q(\mathbf{T})}. \quad (\text{D1})$$

The prior expression is, however, related to the probability of obtaining a given configuration of  $\mathbf{T}$ , regardless whether or not it is unique (several configurations can give rise to the same  $\mathbf{T}$ ). For the case of distinguishable events, there are  $\binom{F}{T}$  different ways of obtaining the same number of events  $T$  among the set of copies and  $T!/\prod_{ij} t_{ij}!$  ways of distributing them to obtain a given adjacency matrix  $\mathbf{T}$ , hence one must consider an additional term in expression (D1),

$$\mathcal{D}(\mathbf{T})_{\text{Events}} = \binom{F}{T} \frac{T!}{\prod_{ij} t_{ij}!}. \quad (\text{D2})$$

For the case with linear constraints of the form of Eq. (6), the system partition function  $\mathcal{Z}$  reads ( $z_{ij} \equiv e^{\theta_T} \prod_{ij} e^{\theta_q c_{ij}^q}$ )

$$\begin{aligned} \mathcal{Z} &= \sum_{\{\mathbf{T}\}} \mathcal{D}(\mathbf{T}) \prod_{ij} z_{ij}^{t_{ij}} \\ &= \sum_{T=0}^F \binom{F}{T} \sum_{\{\mathbf{T} | \sum_{ij} t_{ij} = T\}} \frac{T!}{\prod_{ij} t_{ij}!} \prod_{ij} (M z_{ij})^{t_{ij}} \\ &= \sum_{T=0}^F \binom{F}{T} \left( M \sum_{ij} z_{ij} \right)^T = \left( 1 + M \sum_{ij} z_{ij} \right)^F. \end{aligned} \quad (\text{D3})$$

If we add the strong condition that the ensemble average number of events has to be equal to  $\hat{T}$ , a scaling of  $M \sum_{ij} z_{ij}$  on the total number of events  $F$  distributed among the copies  $\mathcal{N}$  is made apparent:

$$\begin{aligned} \langle T \rangle &= \partial_{\theta_T} \ln \mathcal{Z} = \sum_{ij} \langle t_{ij} \rangle = F \frac{M \sum_{ij} z_{ij}}{1 + M \sum_{ij} z_{ij}} \\ &= \hat{T} \Rightarrow M \sum_{ij} z_{ij} = \frac{\hat{T}/F}{1 - \hat{T}/F}. \end{aligned} \quad (\text{D4})$$

Wrapping together the previous expressions and considering that the number of copies is arbitrary, one can imagine the

limit where it goes to  $\infty$  keeping  $\hat{T}$  constant. This amounts to considering  $F$  infinitely large too,

$$\mathcal{Z} = \lim_{F \rightarrow \infty} \left( 1 + \frac{\hat{T}/F}{1 - \hat{T}/F} \right)^F = e^{\hat{T}} = \prod_{ij} e^{\langle t_{ij} \rangle} = \prod_{ij} \mathcal{Z}_{ij}, \quad (\text{D5})$$

which leads to a factorizable partition function of the form of Eq. (4) which does not depend on the number of copies of the system; neither does its associated probability,

$$\begin{aligned} P(\mathbf{T}) &= \lim_{F \rightarrow \infty} \frac{\mathcal{D}(\mathbf{T}) \prod_{ij} z_{ij}^{t_{ij}}}{\mathcal{Z}} = \lim_{F \rightarrow \infty} \frac{\binom{F}{T} \frac{T!}{\prod_{ij} t_{ij}!} \prod_{ij} \left( \frac{\langle t_{ij} \rangle}{(F-T)} \right)^{t_{ij}}}{\left( 1 - \frac{\hat{T}}{F} \right)^{-F}} \\ &= \prod_{ij} \frac{\langle t_{ij} \rangle^{t_{ij}}}{t_{ij}!} e^{-T} \lim_{F \rightarrow \infty} \left( \frac{F - \hat{T}}{F - T} \right)^{F-T} \\ &= \prod_{ij} \frac{\langle t_{ij} \rangle^{t_{ij}}}{t_{ij}!} e^{-T} e^{-\hat{T}+T} = \prod_{ij} p_{ij}^{\text{ME}}(t_{ij}). \end{aligned} \quad (\text{D6})$$

We have thus reached the same independent Poisson probabilities as obtained by taking an effective factorizable degeneracy term,  $\hat{T}! \prod_{ij} (M^{t_{ij}}/t_{ij}!)$  [see Eq. (8)].

These results are in accordance with previous works [20] where the complete equivalence between canonical ensembles [ensembles with soft linear constraints as in Eq. (6) but with  $T = \hat{T}$  fixed for every network in the ensemble] and microcanonical ensembles (ensembles where all constraints are exactly fulfilled) of ME networks was proven. The equivalence between microcanonical ensembles and grand-canonical ensembles in Poisson form has also been validated by simulations for the case where strengths are fixed [37].

For nonlinear cases (such as the case with binary constraints or the B case with distinguishable events), the effective degeneracy term is an approximation, since the complete calculation using partial sums where  $T$  is exactly fixed cannot be performed. Approximating  $T!$  by  $\hat{T}!$  amounts to considering that the possible fluctuations of the macroscopic variable  $T$  are caused by the state-independent fluctuations of the microscopic structure given by  $\{t_{ij}\}$ . This, however, leads to the same statistics emerging from the leading-order terms in  $\hat{T}$  of the system partition function computed using a microcanonical formalism (see [20]).

[1] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2010).  
[2] M. Á. Serrano, M. Boguñá, and F. Sagués, *Mol. Biosyst.* **8**, 843 (2012).  
[3] O. Sagarra, M. Szell, P. Santi, A. Díaz-Guilera, and C. Ratti, *PLoS ONE* **10**, e0134508 (2015).  
[4] G. Menichetti, D. Remondini, and G. Bianconi, *Phys. Rev. E* **90**, 062817 (2014).  
[5] R. Guimera and L. A. N. Amaral, *Nature* **433**, 895 (2005).

[6] M. Schich, C. Song, Y.-Y. Ahn, A. Mirsky, M. Martino, A.-L. Barabási, and D. Helbing, *Science* **345**, 558 (2014).  
[7] P. Colomer-de-Simón, M. Á. Serrano, M. G. Beiró, J. I. Alvarez-Hamelin, and M. Boguñá, *Sci. Rep.* **3**, 2517 (2013).  
[8] G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondragón, and G. Bianconi, *PLoS One* **9**, e97857 (2014).  
[9] M. Molloy and B. Reed, *Random Struct. Alg.* **6**, 161 (1995).  
[10] A. Coolen, A. De Martino, and A. Annibale, *J. Stat. Phys.* **136**, 1035 (2009).

- [11] J. Park and M. E. J. Newman, *Phys. Rev. E* **70**, 066117 (2004).
- [12] G. Bianconi, A. C. C. Coolen, and C. J. Perez Vicente, *Phys. Rev. E* **78**, 016114 (2008).
- [13] A. Annibale, A. Coolen, L. Fernandes, F. Fraternali, and J. Kleinjung, *J. Phys. A: Math. Gen.* **42**, 485001 (2009).
- [14] E. Roberts, T. Schlitt, and A. Coolen, *J. Phys. A* **44**, 275002 (2011).
- [15] E. Roberts, A. Annibale, and A. Coolen, in *J. Phys. Conf. Ser.* **410**, 012097 (2013).
- [16] A. G. Wilson, *Transp. Research* **1**, 253 (1967).
- [17] P. W. Holland and S. Leinhardt, *JASA* **76**, 33 (1981).
- [18] G. Bianconi, P. Pin, and M. Marsili, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11433 (2009).
- [19] D. Garlaschelli and M. Loffredo, *Phys. Rev. Lett.* **102**, 038701 (2009).
- [20] O. Sagarra, C. J. Pérez Vicente, and A. Díaz-Guilera, *Phys. Rev. E* **88**, 062806 (2013).
- [21] E. S. Roberts and A. C. C. Coolen, *J. Phys. A* **47**, 435101 (2014).
- [22] T. P. Peixoto, *Phys. Rev. E* **85**, 056122 (2012).
- [23] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, *Phys. Rev. X* **3**, 041022 (2013).
- [24] E. Cozzo, G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, [arXiv:1504.05567](https://arxiv.org/abs/1504.05567).
- [25] D. Garlaschelli and M. I. Loffredo, *Phys. Rev. E* **78**, 015101 (2008).
- [26] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevE.92.052816> for details on the data sets used, measured network quantities, and numerical methods used.
- [27] O. Sagarra, ODME: Origin-Destination Multi-Edge Network Package (2014); [https://github.com/osagarra/ODME\\_lite](https://github.com/osagarra/ODME_lite).
- [28] G. Bianconi, *Europhys. Lett.* **81**, 28005 (2008).
- [29] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **68**, 036112 (2003).
- [30] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, *Phys. Rev. Lett.* **89**, 258702 (2002).
- [31] A. Cardillo, J. Gómez-Gardenes, M. Zanin, M. Romance, D. Papo, F. del Pozo, and S. Boccaletti, *Sci. Rep.* **3**, 1344 (2013).
- [32] R. Guimerà, A. Arenas, A. Díaz-Guilera, and F. Giralt, *Phys. Rev. E* **66**, 026704 (2002).
- [33] L. Gauvin, A. Panisson, C. Cattuto, and A. Barrat, *Sci. Rep.* **3**, 3099 (2013).
- [34] T. Squartini, G. Fagiolo, and D. Garlaschelli, *Phys. Rev. E* **84**, 046117 (2011).
- [35] T. Squartini and D. Garlaschelli, *New J. Phys.* **13**, 083001 (2011).
- [36] A. Halu, S. Mukherjee, and G. Bianconi, *Phys. Rev. E* **89**, 012806 (2014).
- [37] O. Sagarra, F. Font-Clos, C. J. Pérez-Vicente, and A. Díaz-Guilera, *Europhys. Lett.* **107**, 38002 (2014).
- [38] N. Dianati, [arXiv:1503.04085v3](https://arxiv.org/abs/1503.04085v3).
- [39] M. Á. Serrano, M. Boguñá, and A. Vespignani, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6483 (2009).
- [40] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13290 (2014).
- [41] R. Mastrandrea, T. Squartini, G. Fagiolo, and D. Garlaschelli, *Phys. Rev. E* **90**, 062804 (2014).
- [42] O. Sagarra, Master's thesis, Universitat Politècnica de Catalunya (2011).
- [43] M. A. Serrano, M. Boguna, and R. Pastor-Satorras, *Phys. Rev. E* **74**, 055101 (2006).
- [44] K. Anand, D. Krioukov, and G. Bianconi, *Phys. Rev. E* **89**, 062807 (2014).
- [45] R. H. Swendsen, *Am. J. Phys.* **83**, 545 (2015).
- [46] R. K. Pathria, *Statistical Mechanics*, 2nd ed. (Butterworth-Heinemann, London, 1996).