Grand canonical ensembles of sparse networks and Bayesian inference

Ginestra Bianconi[1,2]

[1]*School of Mathematical Sciences, Queen Mary*
University of London, London, E1 4NS, United Kingdom

[2]*The Alan Turing Institute, The British Library, London NW1 2DB, United Kingdom*[*]

Abstract

Maximum entropy network ensembles have been very successful in modelling sparse network topologies and in solving challenging inference problems. However the sparse maximum entropy network models proposed so far have fixed number of nodes and are typically not exchangeable. Here we consider hierarchical models for exchangeable networks in the sparse limit, i.e. with the total number of links scaling linearly with the total number of nodes. The approach is grand canonical, i.e. the number of nodes of the network is not fixed a priori: it is finite but can be arbitrarily large. In this way the grand canonical network ensembles circumvent the difficulties in treating infinite sparse exchangeable networks which according to the Aldous-Hoover theorem must vanish. The approach can treat networks with given degree distribution or networks with given distribution of latent variables. When only a subgraph induced by a subset of nodes is known, this model allows a Bayesian estimation of the network size and the degree sequence (or the sequence of latent variables) of the entire network which can be used for network reconstruction.

---

[*] ginestra.bianconi@gmail.com

I. INTRODUCTION

Networks [1, 2] have the ability to capture the topology of complex systems ranging from the brain to financial networks. Network models are key to have reliable unbiased null models of the network and to explain emergent phenomena of network evolution. Network model can be classified in two major classes: equilibrium maximum entropy models [3–15] and growing network models [1, 16–18]. While growing network models have a number of nodes that increases in time, maximum entropy models are used so far only for treating networks of a given number of nodes N. In this paper we are interested in extending the realm of maximum entropy network models to networks of varying network size N.

Maximum entropy network ensembles are the least biased ensembles satisfying a given set of constraints. As such maximum entropy ensembles are widely used as null models and for network reconstruction starting from features associated to the nodes of the network. Given the profound relation between information theory and statistical mechanics [19, 20], maximum entropy network ensembles can be distinguished between microcanonical ensembles and canonical ensembles [3, 21, 22] similarly to the analogous distinction traditionally introduced in statistical mechanics for ensembles of particles. Microcanonical network ensembles are ensembles of networks of N nodes satisfying some hard constraints (such as the total number of links, or the given degree sequence). Canonical network ensembles instead are ensemble of networks of N nodes satisfying some soft constraints, (such as the expected total number of links or the expected degree sequence). The canonical ensembles with expected degree sequence can be also formulated as latent variable models where the latent variables can be associated to the nodes [5, 23].

Maximum entropy models have been very successful in solving challenging inference models [6, 8, 24–26], however they have the limitation that they only treat networks with a given fixed number of nodes N. Indeed in a number of scenarios, the number of nodes might not be fixed or might not be known. In this context an important problem is to compare networks of different network sizes. For instance in brain imaging one might choose a finer grid or a coarser grid of brain regions and an outstanding problem in machine learning is how to build neural networks that can generalize well when tested on network data with different network size than the network data in the training set [27, 28].

In order to have network ensembles that can treat networks of different size, here we introduce the grand canonical network ensembles in which the number of nodes can vary. A well-defined

grand-canonical network ensemble necessarily needs to be exchangeable [29], i.e. needs to be invariant under permutation of labels of the nodes of the network, so that removing or adding a node has an effect that is independent of the particular choice of the node added or removed.

The research on exchangeable networks is currently very vibrant. The graphon model [30] is the most well established exchangeable network model. However this model is dense, i.e. the number of links scales quadratically with the number of nodes while the vast majority of the network data is sparse with a total number of links scaling linearly with the network size. In other words most of the real world networks have constant average degree. However popular models for sparse networks such as the configuration model [31] and the exponential random graphs [4] are not exchangeable. In fact these models treat networks of labelled nodes with given degree or with given expected degree sequence. Therefore the network ensemble is not invariant under permutation of the node labels, except if all the degrees of all the expected degrees of the network are the same (for a more diffused discussion of why these networks are not exchangeable see discussion in Ref.[32]). Several works have been proposed exchangeable network models in the when the average degree of the network diverges sublinearly with the network size [33–38]. In Ref. [32] a framework able to model sparse exchangeable networks in the limit of constant degree, has been proposed. The model is very general and has been extended to treat generalized network structures including multiplex networks [39] and simplicial complexes [40]. However the model is well defined only for finite networks of large but finite number of nodes N as exchangeable sparse networks need to obey the Aldous-Hoover theorem [41, 42] according to which infinite sparse exchangeable networks must vanish. An alternative strategy for formulating exchangeable ensembles is to consider ensembles of unlabelled networks for which several results are already available [43].

Here we build on the recently proposed exchangeable sparse network ensembles [32] to formulate hierarchical grand-canonical ensembles of sparse networks. The proposed grand-canonical ensembles are hierarchical models [25, 44] with variable number of nodes N and with given degree distribution or alternatively given latent variable distributions. The grand canonical approach provides a way to circumvent the limitations imposed by the Aldous-Hoover theorem because in this framework one considers a mixture of network ensembles with finite but unspecified and arbitrary large network sizes. In this paper we define the grand-canonical ensembles and we characterize them with statistical mechanics methods, evaluating their entropy, the marginal probability of a link and proposing generative algorithms to sample networks from these ensembles. [Note that

the proposed grand canonical ensembles differ from the ensembles proposed in Refs. [45, 46], as in our case we consider networks with undetermined number of nodes, while in Refs.[45, 46] is the total sum of weights of weighted networks that is allowed to vary. From the statistical mechanics perspective our approach is fully classical while in Refs. [45, 46] networks ensembles are treated as quantum mechanical ensembles where the particles are associated to the links of the network and the adjacency matrix elements play the role of occupation numbers.]

Finally, we use the gran-canonical network ensembles to solve an inference problem. We consider a scenario in which the entire network has an unknown number of nodes, and we have only access to a subgraph induced by a subset of its nodes. In this hypothesis we use the grand-canonical network models to perform a Bayesian estimation of the true parameters of the network model (given by the network size and the degree sequence or the sequence of latent variables). This a posteriori estimate of the parameters can then be used to reconstruct the unknown part of the network.

II.   THE GRAND CANONICAL NETWORK ENSEMBLE WITH GIVEN DEGREE DISTRIBUTION

We consider the hierarchical grand canonical ensemble of exchangeable sparse simple networks where we associate to every network $G = (V, E)$ with $N = |V| > N_0$ nodes the probability

$$\mathbb{P}(G) = P(N)P(\mathbf{k}|N)P(G|\mathbf{k}, N) \tag{1}$$

where $P(N)$ indicates the probability that the network G has N nodes, $P(\mathbf{k}|N)$ indicates the conditional probability that the network has degree sequence $\mathbf{k}$ given that the network has N nodes, and $P(G|\mathbf{k}, N)$ indicates the probability of the network G with adjacency matrix $\mathbf{a}$ given that the network has N nodes and degree sequence $\mathbf{k}$ (see Figure 1 for a schematic representation of the model). To be specific we consider the following model giving rise to the hierarchical grand canonical ensemble of exchangeable simple models:

(1) *Drawing the total number of nodes N of the network.* Let us discuss suitable choices for the distribution of the number of nodes N with N greater or equal than some minimum number of nodes N_0. We indicate the distribution $P(N)$ as

$$P(N) = \pi(N), \quad \text{for } N \geq N_0. \tag{2}$$
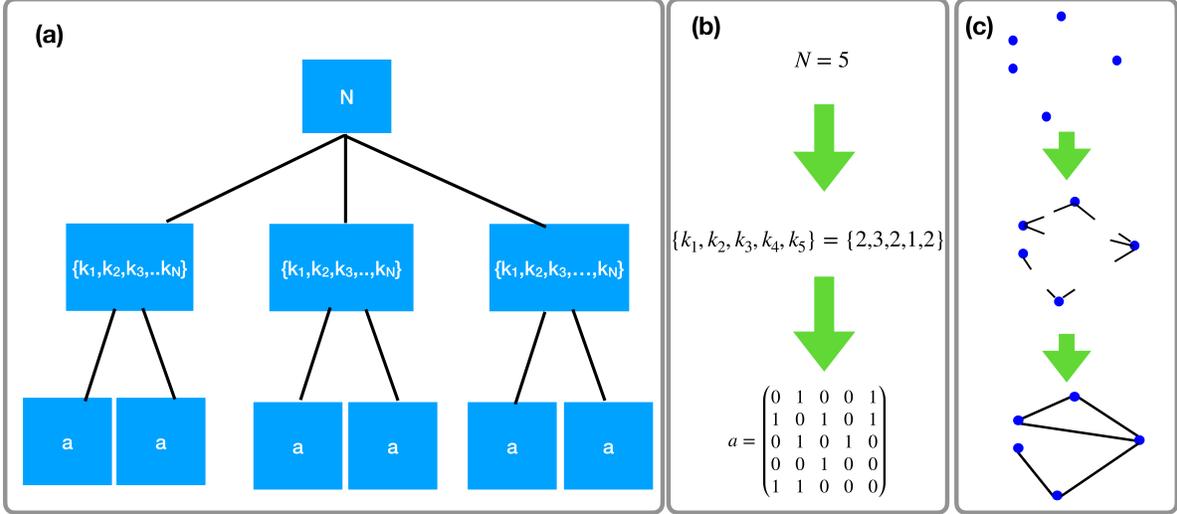
4

Figure 1. Schematic representation of the hierarchical grand canonical ensemble of exchangeable sparse simple networks. The proposed ensemble is a hierarchical model of networks in which first the total number of nodes N is drawn from a $P(N) = \pi(N)$ distribution, then a given degree sequence $\mathbf{k} = \{k_1, k_2, \ldots k_N\}$ is drawn from the distribution $P(\mathbf{k}|N)$ among all the degree sequence with the total number of nodes N; finally a network G with adjacency matrix $\mathbf{a}$ drawn from the distribution $P(G|\mathbf{k}, N)$ among all the networks with a given total number of nodes N and degree sequence $\mathbf{k}$. Panel (a) describes the hierarchical nature of the model, panel (b) provide an example of subsequent draw of the total number of nodes, the degree sequence and the adjacency matrix of the network, panel (c) is a visualization of the construction of a network according to the proposed model.

While a statistical mechanics approach would suggest to take a distribution $\pi(N)$ with a well defined mean value (such as the exponential distribution)

$$\pi(N) = C e^{-\mu N} \quad \text{for } N \geq N_0, \tag{3}$$

where C is a normalization constant and $\mu > 0$, in the context of network science it might actually be relevant to consider also broad distributions $\pi(N)$ such as power-law distributions

$$\pi(N) = D N^{-\nu} \quad \text{for } N \geq N_0, \tag{4}$$

where D is a normalization constant and $\nu > 1$.

(2) *Drawing the degree sequence of the network.* In order to obtain a sparse exchangeable network ensemble with given degree distribution $p(k)$ having finite average degree $\langle k \rangle$, min-

5

imum allowed degree $\hat{m}$ and maximum allowed degree K we consider the following expression for the probability of a given degree sequence given the total number of nodes

$$P(\mathbf{k}|N) = \prod_{i=1}^{N} \left[p(k_i) \hat{\theta}(K - k_i) \theta(k_i - \hat{m}) \right] \delta \left(\sum_{i=1}^{N} k_i, \langle k \rangle N \right), \tag{5}$$

where $\hat{\theta}(x)$ indicates the Heaviside function $\hat{\theta}(x) = 1$ if $x \geq 0$ and $\hat{\theta}(x) = 0$ otherwise and where we used the notation $\langle k \rangle = \sum_k k p(k)$. In the following we will indicate with L the total number of links of the network given by $L = \langle k \rangle N/2$. Note that $P(\mathbf{k}|N)$ is independent of the labels of the nodes, i.e. all the degree sequences that can be obtained by a permutation of the node labels of a given degree sequence have the same probability $P(\mathbf{k}|N)$.

(3) *Drawing the adjacency matrix of the network.* The probability of a network G with adjacency matrix $\mathbf{a}$ given the total number of nodes N of the network and the degree sequence $\mathbf{k}$ is chosen in the least biased way by drawing the network from a uniform distribution, i.e. the conditional probability $P(G|\mathbf{k}, N)$ is equivalent to the probability of a network in the microcanonical ensemble. Therefore, by indicating with $\mathcal{N}(\mathbf{k}|N)$ the total number of networks with N nodes and degree sequence $\mathbf{k}$ and with $\Sigma_N(\mathbf{k}) = \ln \mathcal{N}(\mathbf{k}|N)$ the entropy of the ensemble we can express $P(G|\mathbf{k}, N)$ as

$$P(G|\mathbf{k}, N) = \frac{1}{\mathcal{N}(\mathbf{k}|N)} = e^{-\Sigma_N(\mathbf{k})} \tag{6}$$

Note that for sparse networks of $N \geq N_0$ nodes the entropy $\Sigma_N(\mathbf{k})$ obeys the Bender-Canfield formula as long as the network has a structural cutoff K_S, i.e. as long as $k_i \ll K_S = \sqrt{\langle k \rangle N_0}$ [3, 21, 22, 47]

$$\Sigma_N(\mathbf{k}) = \ln \left(\frac{(2L)!!}{\prod_{i=1}^{N} k_i!} \right) + o(N) \tag{7}$$

where in Eq. (7) we indicate with $\mathbf{k} = \{k_1, k_2, \ldots, k_N\}$ the degree sequence with k_i, the degree of node i, given by $k_i = \sum_{j=1}^{N} a_{ij}$.

It follows that the hierarchical grand canonical ensemble for exchangeable sparse networks can be cast into an Hamiltonian ensemble with probability $\mathbb{P}(G)$ given by

$$\mathbb{P}(G) = \frac{1}{Z} e^{-H(G)} \delta \left(\langle k \rangle N/2, \sum_{i<j} a_{ij} \right) \hat{\theta} \left(K - \max_{i=1}^{N} k_i \right) \hat{\theta} \left(\min_{i=1}^{N} k_i - \hat{m} \right), \tag{8}$$

6

with Hamiltonian $H(G)$ given by

$$H(G) = -\ln \pi(N) - \sum_{i=1}^{N} \ln \left[p(k_i)k_i! \delta \left(k_i, \sum_{j=1}^{N} a_{ij} \right) \right] + \ln((\langle k \rangle N)!!). \tag{9}$$

This Hamiltonian is global and is invariant under permutation of the node labels, therefore this hierarchical grand canonical ensemble is exchangeable. Indeed we have that the probability of a network $\mathbb{P}(G)$ given by Eq. (8) obeys

$$\mathbb{P}(G) = \mathbb{P}(\tilde{G}) \tag{10}$$

where $\tilde{G}$ is any network obtained from network G under a generic permutation σ of the labels of the nodes. Moreover we note that for $\pi(N) = \delta(N, \bar{N})$, i.e. when the network size is fixed this model reduces to the exchangeable model for sparse network ensemble proposed in Ref.[32].

III. THE GRAND CANONICAL NETWORK ENSEMBLE WITH GIVEN DISTRIBUTION OF THE LATENT VARIABLES

The grand canonical formalism can also be easily extended to treat network models with latent variables $\boldsymbol{\theta}$ associated to the nodes of the network $G = (V, E)$. Note that here and in the following we assume that the latent variables take discrete values. To this end we can consider the soft grand canonical hierarchical model associating to each network with $N = |V| > N_0$ nodes, latent variables $\boldsymbol{\theta}$ and adjacency matrix $\mathbf{a}$ the probability

$$P(G, \boldsymbol{\theta}, N) = P(N)P(\boldsymbol{\theta}|N)P(G|\boldsymbol{\theta}, N) \tag{11}$$

with

$$P(N) = \pi(N), \tag{12}$$

where $\pi(N)$ is an arbitrary prior on the number of nodes in the network defined for $N \geq N_0$. Typical examples of the distribution $\pi(N)$ are given by Eq. (3) and Eq. (4). The probability of the latent variables is chosen to be exchangeable and given by

$$P(\boldsymbol{\theta}|N) = \prod_{i=1}^{N} p(\theta_i) \tag{13}$$

where $p(\theta_i)$ is the probability distribution of each latent variable. The distribution $p(\theta)$ can be chosen arbitrarily, as long as the expectation of θ is finite. The probability of the network given

the network size and the latent variables is chosen to be derived by a Bernoulli variable for each link, with probability of observing a link between node i and node j conditioned on the value of their latent variables given by $p_N(\theta_i, \theta_j)$, i.e.

$$P(G|\boldsymbol{\theta}, N) = \prod_{i<j} p_N(\theta_i, \theta_j)^{a_{ij}} (1 - p_N(\theta_i, \theta_j))^{1-a_{ij}}. \tag{14}$$

To be concrete we consider the following expression for the probability $p_N(\theta_i, \theta_j)$ which is the general expression of the marginal probability of a link in canonical network ensembles (or equivalently exponential random graph models),

$$p_N(\theta_i, \theta_j) = \frac{\theta_i \theta_j / N}{1 + \theta_i \theta_j / N}. \tag{15}$$

The advantage of taking this expression for the probability $p_N(\theta_i, \theta_j)$ is that $p_N(\theta_i, \theta_j)$ is always smaller or equal to one for every value of the latent variables. Therefore in this model we do not need to impose a structural cutoff on the latent variables. In summary the grand canonical network ensemble with given latent variable distribution is a hierarchical network model in which given the network size and latent variables the network is drawn according to a canonical ensemble of networks. In this ensemble the probability of a network G can be written in Hamiltonian form as

$$\mathbb{P}(G) = \frac{1}{Z} e^{-H(G)} \tag{16}$$

with Hamiltonian $H(G)$ given by

$$H(G) = -\ln \pi(N) - \sum_{i=1}^{N} p_N(\theta_i) - \sum_{i<j} \left\{ a_{ij} \ln p_N(\theta_i, \theta_j) + (1 - a_{ij}) \ln[1 - p_N(\theta_i, \theta_j)] \right\}. \tag{17}$$

This Hamitonian is invariant under permutation of the node labels, therefore this model is exchangeable.

IV.   THE ENTROPY OF GRAND CANONICAL ENSEMBLES

In this paragraph we show that the entropy S [3, 48] of the two proposed grand canonical network ensembles, defined as

$$S = \sum_{G} \mathbb{P}(G) \ln \mathbb{P}(G), \tag{18}$$

can be decomposed into contributions that reflect the uncertainty related to an increasing number of hierarchical levels of the model. In order to show this results we discuss separately the entropy of the two proposed grand canonical ensembles.

A. Entropy of the grand canonical ensemble with given degree distribution

The entropy S of the ensemble fixing the degree distribution can be decomposed into the entropy of the model at different levels of the hierarchy according to the following expression,

$$S = S_{\pi(N)} + \langle S_{p(k)} \rangle_{\pi(N)} + \langle \Sigma_N(\mathbf{k}) \rangle_{\pi(N), p(k)} \tag{19}$$

where $S_{\pi(N)}$ is the entropy associated to the number of typical choices of the total number of nodes N, $\langle S_{p(k)} \rangle_{\pi(N)}$ is the entropy associated to the choice of the degree sequence averaged over the distribution $\pi(N)$ and $\langle \Sigma_N(\mathbf{k}) \rangle_{\pi(N), p(k)}$ is the average of the Gibbs entropy [3] of the networks with given degree sequence averaged over the distribution $\pi(N)$ and $P(\mathbf{k}|N)$. In other words we have

$$S_{\pi(N)} = -\sum_{N>N_0} \pi(N) \ln \pi(N),$$

$$\langle S_{p(k)} \rangle_{\pi(N)} = \sum_{N>N_0} \pi(N) \left[-N \sum_k p(k) \ln p(k) \right],$$

$$\langle \Sigma_N(\mathbf{k}) \rangle_{\pi(N), p(k)} = \sum_{N>N_0} \pi(N) \sum_{\mathbf{k}} P(\mathbf{k}|N) \Sigma_N(\mathbf{k}). \tag{20}$$

B. Entropy of the grand canonical ensemble with given latent variable distribution

Similarly to the previous case, it is easy to show that the entropy of the ensemble fixing the distribution of the latent variables can be decomposed into the entropy of the model at different levels of their hierarchy, according to the following expression

$$S = S_{\pi(N)} + \langle S_{p(\theta)} \rangle_{\pi(N)} + \langle S_N(\boldsymbol{\theta}) \rangle_{\pi(N), p(\theta)}, \tag{21}$$

where $S_{\pi(N)}$ is the entropy associated to the number of typical choices of the total number of nodes N, $\langle S_{p(\theta)} \rangle_{\pi(N)}$ is the entropy associated to the choice of the latent variable distribution averaged over the distribution $\pi(N)$ and $\langle S_N(\boldsymbol{\theta}) \rangle_{\pi(N), p(k)}$ is the average of the Shannon entropy [3] of the networks with given sequence of latent variables averaged over the distribution $\pi(N)$ and $P(\boldsymbol{\theta}|N)$. In other words we have

$$S_{\pi(N)} = -\sum_{N>N_0} \pi(N) \ln \pi(N),$$

$$\langle S_{p(\theta)} \rangle_{\pi(N)} = \sum_{N>N_0} \pi(N) \left[-N \sum_\theta p(\theta) \ln p(\theta) \right],$$

$$\langle S_N(\boldsymbol{\theta}) \rangle_{\pi(N), p(\theta)} = \sum_{N>N_0} \pi(N) \sum_\theta P(\boldsymbol{\theta}|N) S_N(\boldsymbol{\theta}), \tag{22}$$

where the Shannon entropy $S_N(\boldsymbol{\theta})$ of the network given the sequence of latent variables and the network size N can be expressed as

$$S_N(\boldsymbol{\theta}) = -\sum_{i<j} \left[p_N(\theta_i, \theta_j) \ln p_N(\theta_i, \theta_j) + (1 - p_N(\theta_i, \theta_j)) \ln(1 - p_N(\theta_i, \theta_j)) \right]. \tag{23}$$

V. MARGINAL PROBABILITY OF A LINK

A. The case of the the grand canonical ensemble with given degree distribution

The grand canonical ensemble of exchangeable sparse network ensembles is an ensemble in which the total number of nodes is not specified. If we consider the networks of this ensemble having a given number of nodes N, the model reduces to the exchangeable sparse network ensemble proposed in Ref.[32] whose marginal probability of a link (i, j) is given by

$$\tilde{p}_{ij} = \sum_k p(k) \sum_{k'} p(k') \frac{kk'}{\langle k \rangle N}. \tag{24}$$

Since the grand-canonical ensemble of sparse exchangeable networks with given degree distribution can be interpreted as a mixture of the exchangeable sparse models proposed in Ref. [32] with different size N, it is immediate to show that the marginal probability of a link between node i and node j in the grand canonical ensembles is given by the exchangeable expression,

$$p_{ij} = \sum_{N>N_0} \pi(N) \sum_{k,k'} p(k)p(k') \frac{kk'}{\langle k \rangle N} = \sum_{N>N_0} \pi(N) \frac{\langle k \rangle}{N}. \tag{25}$$

Moreover the probability that two nodes are connected given that they have degree k and k' is given by

$$p_{ij|k_i=k, k_j=k'} = p(k, k') = kk' \sum_{N>N_0} \frac{\pi(N)}{\langle k \rangle N}. \tag{26}$$

Finally the probability that two nodes are connected given that they have degree k and k' and the actual size of the network is N is given by the uncorrelated network expression

$$p_{ij|k_i=k, k_j=k', N} = p_N(k, k') = \frac{kk'}{\langle k \rangle N}. \tag{27}$$

From these expressions of the marginal probability of a link it is possible to appreciate how the hierarchical grand canonical ensemble of sparse exchangeable networks circumvents the difficulties arising form the Aldous-Hoover theorem without violating it. Indeed the marginal probability $p_N(k, k')$ of a link conditioned on the degrees of the two linked nodes and the number of nodes N

of the network vanishes in the limit $N \to \infty$, however if the number of nodes of the network is arbitrarily large but unknown the marginal probability of the link remains finite (as both p_{ij} and $p(k, k')$ are finite).
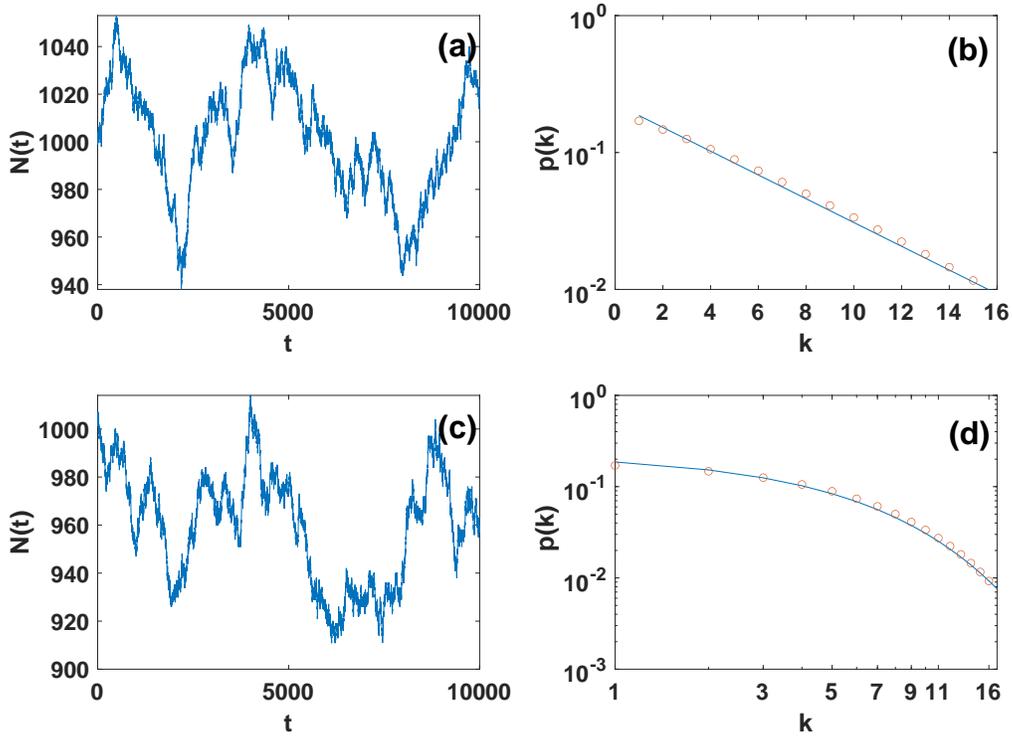

Figure 2. The number of nodes $N(t)$ at a function of time t in the Metropolis-Hastings simulation of an exponential networks (panel a) and networks with more general degree distribution (panel c) are shown together with the average degree distribution of the networks that is stable as the number of networks varies (symbols of panel (b) and (d)). The solid lines in panel(b) and panel (d) indicate the target degree distributions $p(k) = Ce^{-k/m}$ with $m = 5$ (for panel (b)) and $p(k) = C(3 + k)^{-\gamma}$ with $\gamma = 3.4$ (for panel (d)). The prior on the number of nodes is taken to be exponential $\pi(N) = Ce^{-N/\bar{N}}$ with $\bar{N} = 1000$ with $N_0 = 500$ and $K = 16$.

B. The case of the the grand canonical ensemble with given latent variable distribution

For the grand canonical ensemble with given latent variable distribution $p(\theta)$ we have that the marginal probability of a link is given by

$$p_{ij} = \sum_{N>N_0} \pi(N) \sum_{\theta,\theta'} p(\theta)p(\theta')p_N(\theta,\theta'). \tag{28}$$

The probability of the link given the latent variable of the nodes is given by

$$p(\theta,\theta') = \theta\theta' \sum_{N>N_0} \pi(N) \frac{1}{N + \theta\theta'}; \tag{29}$$

the probability of a link given the network size and the latent variables is given by

$$p_N(\theta,\theta') = \frac{\theta\theta'/N}{1 + \theta\theta'/N}. \tag{30}$$

As we discussed in the case of the grand canonical ensemble with given degree distribution also for the grand canonical ensemble with given latent variable distribution the grand canonical approach allows to circumvent the Aldous-Hoover theorem without violating it as the marginal probability of a link in an arbitrarily large network of unknown size is finite.

VI. GENERATING SINGLE INSTANCES OF GRAND-CANONICAL NETWORK ENSEMBLES

In this section we describe two algorithms to generate single instances of the proposed grand canonical ensembles. In particular we will discuss a Metropolis-Hastings ensemble to generate single instances of networks drawn from the grand canonical ensemble with given degree distribution and a Monte Carlo algorithm to generate single instances of networks drawn from the grand canonical ensemble with given distribution of latent variables.

A. Metropolis-Hastings algorithm for the grand-canonical ensemble with given degree distribution

The grand-canonical exchangeable ensemble of sparse networks can be obtained by implementing a Metropolis-Hastings algorithm using the network Hamiltonian given by Eq.(9).

(1) Start with a network of N nodes having exactly $L = \langle k \rangle N/2$ links and in which the minimum degree is greater of equal to $\hat{m}$ and the maximum degree is smaller or equal to K.

(2) Perform the Metropolis-Hastings algorithm for exchangeable sparse networks with N nodes (defined below);

(3) Propose to change the number of nodes to $N' = N + 1$ (addition of one node) or $N' = N - 1$ (removal of one node) with equal probability and accept the move with probability $\max(1, \pi(N')/\pi(N))$ as long as $N' > N_0$. If the move is accepted change the number of nodes adding or removing a node, set the number of links to $L = \langle k \rangle N/2$ and ensure that each node has minimal degree at least $\hat{m}$ and maximum degree less than K. In particular if a node is added ensure it has at least $\hat{m}$ links by rewiring randomly the existing links of the networks and adding a number of links so that the total number of links is the integer that better approximates $\langle k \rangle N/2$. Instead, if a node needs to be removed, choose a random node of the network remove it and rewire/remove links in order to enforce that the total number of links is the integer that better approximates $\langle k \rangle N/2$.

The Metropolis-Hastings algorithm for the exchangeable sparse networks with N nodes is the same algorithm used in Ref. [32] for exchangeable networks with finite size N and is indicated below.

(1) Start with a network of N nodes having exactly $L = \langle k \rangle N/2$ links and in which the minimum degree is greater of equal to $\hat{m}$ and the maximum degree is smaller or equal to K.

(2) Iterate the following steps until equilibration:

    (i) Let **a** be the adjacency matrix of the network;

    (i) Choose randomly a random link $\ell = (i, j)$ between node i and j and choose a pair of random nodes (i', j') not connected by a link.

    (ii) Let **a**′ be the adjacency matrix of the network in which the link (i, j) is removed and the link (i', j') is inserted instead. Draw a random number r from a uniform distribution in $[0, 1]$, i.e. $r \sim U(0, 1)$. If $r < \max(1, e^{-\Delta H})$ where $\Delta H = H(\mathbf{a}') - H(\mathbf{a})$ and if the move does not violate the conditions on the minimum and maximum degree of the network, replace **a** by **a**′.

The Metropolis-Hastings algorithm can be used to sample the space of networks with variable number of nodes and given (stable) degree distribution (see Figure 2).

13

B. Monte Carlo generation of grand canonical network ensemble with given latent variable distribution

A single instance of the grand canonical model with given latent variable distribution can be obtained by performing the following algorithm:

1. Draw the network size N from the $\pi(N)$ distribution;

2. Draw the latent variable θ_i of each node i independently from the latent variable distribution $p(\theta)$.

3. Draw each link (i, j) of the network with probability $p_N(\theta_i, \theta_j)$.

VII. BAYESIAN ESTIMATION OF THE NETWORK PARAMETERS GIVEN PARTIAL KNOWLEDGE OF THE NETWORK

In this section we will use the grand canonical network ensembles for calculating the posterior distribution of the network parameters given partial information of a network $G = (V, E)$. In particular let us assume that we only know the subgraph $\hat{G}(\hat{V}, \hat{E})$ induced by a set of nodes $\hat{V} \subset V$ of $\hat{N} = |\hat{V}|$ nodes and of adjacency matrix $\hat{\mathbf{a}}$ and we do not have access to the full network G with adjacency matrix $\mathbf{a}$. Without loss of generality let us label the nodes of the network in such a way that the labels i with $1 \leq i \leq \hat{N}$ indicate the nodes in $\tilde{V}$ (denote as sampled nodes) and the labels i with $i > \hat{N}$ indicate the nodes in $V \setminus \hat{V}$ (denoted also as unsampled or unknown nodes). We indicate with κ the degree sequence of the sampled network $\hat{G}$. Our goal is to make a Bayesian estimation of the network size N and the true network parameters given the observed subgraph $\hat{G}$. These a posteriori estimates of the true parameters of the network can then be used to reconstruct the unknown part of the network G.

A. Inferring the true parameters with the grand canonical ensemble with given degree distribution

In this paragraph we will use the grand canonical ensemble with given degree distribution to find the posterior probability distribution of the network parameters. For convenience we will indicate with k_i the true degree of the sampled nodes $1 \leq i \leq \hat{N}$ and we will indicate q_i the true
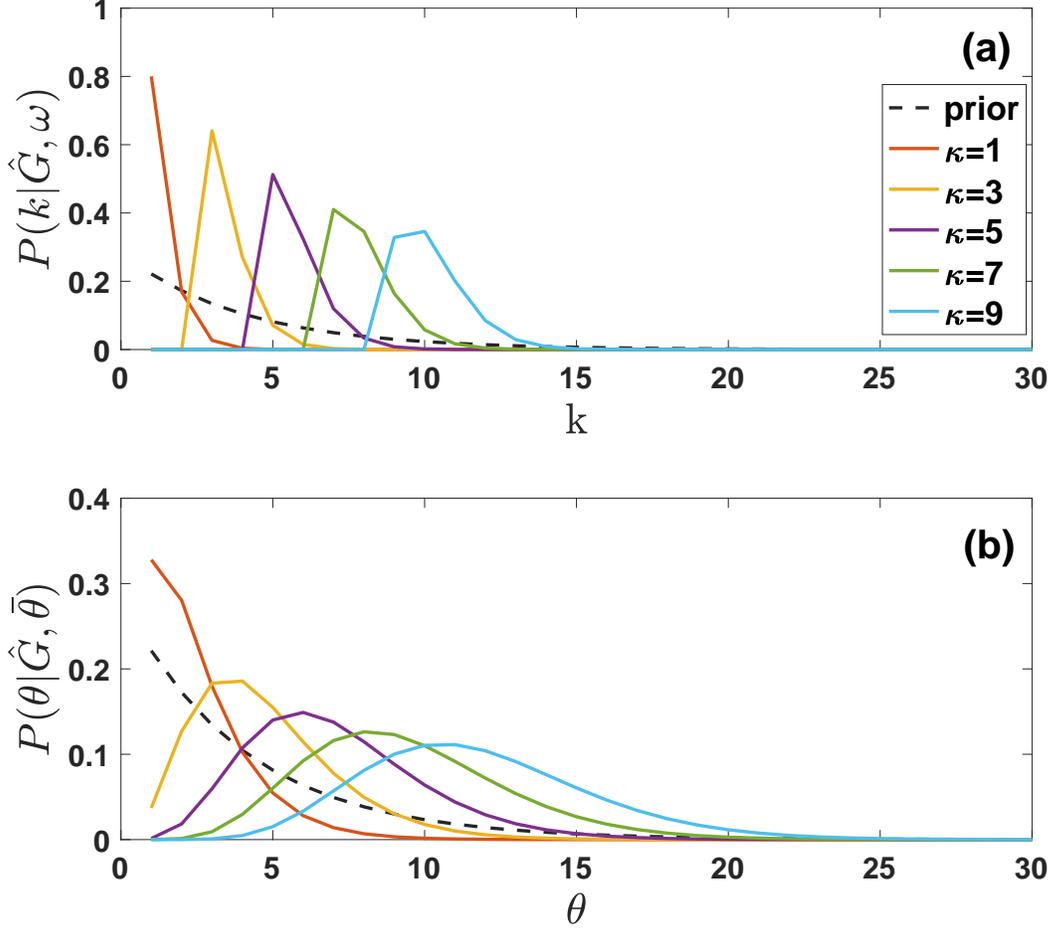
Figure 3. Marginal posterior probability for the true degree of sampled nodes (panel (a)) and for the true latent variable of sampled node (panel b). The posterior probability $P(k_i|\hat{G}, \omega)$ (panel (a)) of the true degree of a sampled nodes depends on the degree κ of the nodes in the sampled network $\hat{G}$ and is non-zero only for $k \geq \kappa$. The posterior probability $P(\theta|\hat{G}, \bar{\theta})$ of the latent variable of a sampled node can be non-zero on the entire range of θ values allowed by the prior. Here we have plotted $P(k_i|\hat{G}, \omega)$ and $P(\theta|\hat{G}, \bar{\theta})$ for different values of κ and we have chosen $\omega = 2$ and $\bar{\theta} = 0.6$. The dashed lines indicate the exponential prior on the degrees (panel (a)) and on the latent variables (panel (b)).

degree of the remaining unsampled $N - \hat{N}$ nodes $\hat{N} + 1 \leq i \leq N$. To this end, using the Bayes rule we get the following expression for the posterior distribution of the network parameters given the observed subgraph $\hat{G}$

$$P(N, \mathbf{k}, \mathbf{q}|\hat{G}) = \frac{P(N)P(\mathbf{k}, \mathbf{q}|N)P(\hat{G}|\mathbf{k}, \mathbf{q}, N)}{P(\hat{G})} \tag{31}$$
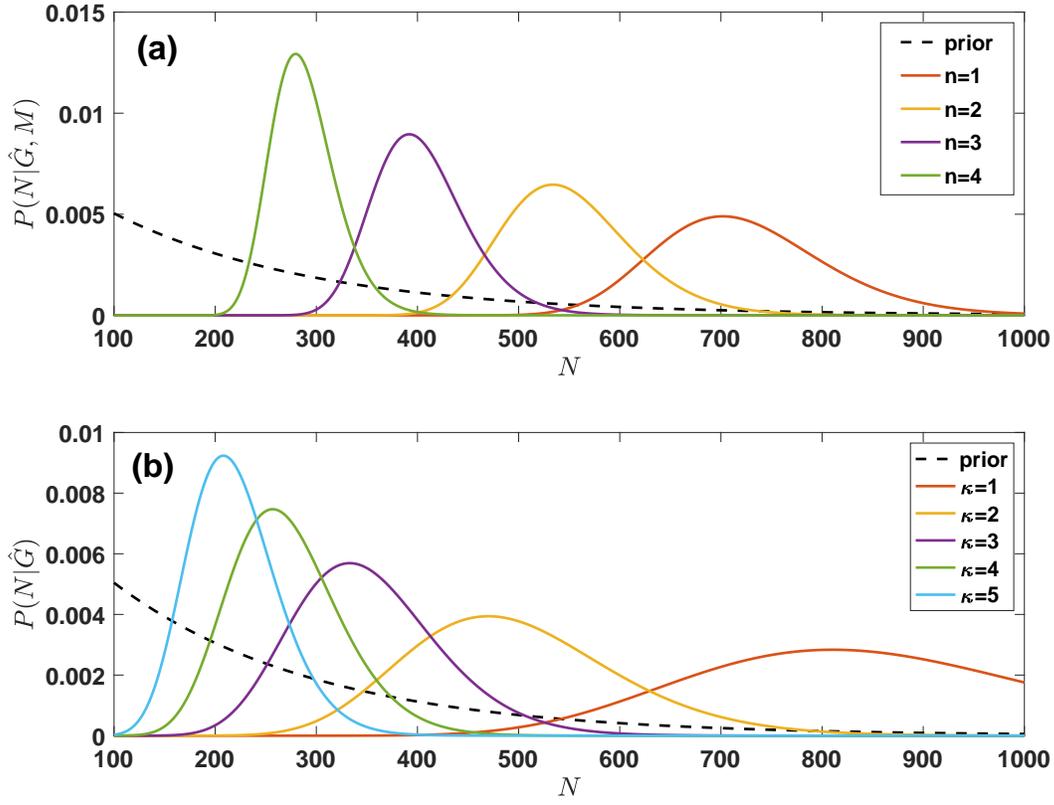
15

Figure 4. Marginal posterior probability for the true number of nodes in the grand canonical ensemble with given degree distribution (panel (a)) and in the grand canonical ensemble with given latent variable distribution (panel (b)). The posterior probability $P(N|\hat{G}, M)$ in panel (a) of the true number of nodes depends on the total number M of true but not observed links of the sampled nodes and on the total number of sampled links $\hat{L}$; the posterior probability $P(N|\hat{G})$ in panel (b) depends instead only on the degree κ of the nodes in the sampled network $\hat{G}$. We took $\hat{N} = 100$ and the priors given by $\pi(N) \propto e^{-N/N0}$, $p(k) \propto e^{-k/m}$, $p(\theta) \propto e^{-\theta/m}$ with $N0 = 200$, and $m = 7$. In panel (a) we have plotted $P(N|\hat{G}, M)$ for different values of $M = (\langle k \rangle - n)\hat{N}$ with $n = 1, 2, 3, 4$ and $\hat{L} = \hat{N}/2$; in panel (b) we have plotted $P(N|\hat{G})$ assuming that $\hat{G}$ is regular with all sampled nodes having sampled degree $\kappa = 1, 2, 3, 4, 5$. The dashed lines indicate the exponential prior $\pi(N)$ on the number of nodes.

where

$$P(N) = \pi(N),$$

$$P(\mathbf{k}, \mathbf{q}|N) = \prod_{i=1}^{\hat{N}} p(k_i) \prod_{i=1+\hat{N}}^{N} p(q_i),$$

$$P(\hat{G}|\mathbf{k}, \mathbf{q}, N) = e^{-\Delta_N \Sigma(\mathbf{k}, \mathbf{q}|\kappa)}, \tag{32}$$

with $\Delta_N \Sigma(\mathbf{k}, \mathbf{q}|\kappa)$ given by

$$\Delta_N \Sigma(\mathbf{k}, \mathbf{q}|\kappa) = \Sigma_N(\mathbf{k}, \mathbf{q}) - \hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\kappa). \tag{33}$$

Here $\Sigma_N \mathbf{k}, \mathbf{q})$ indicates the entropy of the network fo size N with degree sequence $[\mathbf{k}, \mathbf{q}]$ whose expression is given by the Bender-Canfield formula [3, 21, 22, 47] (Eq.(7)) which reads in this case

$$\Sigma_N(\mathbf{k}, \mathbf{q}) = (2L)!! \left[\prod_{i=1}^{\hat{N}} k_i! \prod_{i=1+\hat{N}}^{N} q_i! \right]^{-1}. \tag{34}$$

Moreover $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\kappa)$ indicates the logarithm of the number of networks of N nodes having $\hat{G}$ (with adjacency matrix $\hat{\mathbf{a}}$ and degree sequence κ) as induced subgraph between the $\hat{N}$ sampled nodes.

Moreover in Eq. (31) $P(\hat{G})$ indicates the evidence of the data given by

$$P(\hat{G}) = \sum_N \sum_{\mathbf{k}, \mathbf{q}} \pi(N) \prod_{i=1}^{\hat{N}} p(k_i) \prod_{i=1+\hat{N}}^{N} p(q_i) e^{-\Delta_N \Sigma(\mathbf{k}, \mathbf{q}|\kappa)}. \tag{35}$$

Calculating the entropy $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\kappa)$ using statistical mechanics methods including the use of a functional order parameter (see Appendix), we derive the following expression:

$$\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\kappa) = \ln \left[\frac{M!(Q-M)!!}{\prod_{i=1}^{\hat{N}}(k_i - \kappa_i)! \prod_{i=1+\hat{N}}^{N} q_i!} \binom{Q}{M} \delta(Q + M, 2L + 2\hat{L}) \right] \tag{36}$$

where M indicates the number of links between the sampled nodes and the unsampled nodes and Q indicates the sum over all the degrees of the unsampled nodes, i.e.

$$M = \sum_{i=1}^{\hat{N}} (k_i - \kappa_i),$$

$$Q = \sum_{i=1+\hat{N}}^{N} q_i, \tag{37}$$

where M and Q need to satisfy the constraint enforcing that the total number of true links is given by $L = \langle k \rangle N/2$. Therefore, indicating with $\hat{L} = \sum_{i=1}^{\hat{N}} \kappa_i/2$, we must impose

$$Q + M = 2L - 2\hat{L}. \tag{38}$$

17

The expression obtained for the entropy $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$ implies that the asymptotic expression for the number of networks with N nodes, degree sequence $[\mathbf{k}, \mathbf{q}]$ having $\hat{G}$ as a subgraph is given by (see Appendix for the derivation)

$$\mathcal{N}(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa}, N) = e^{\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})} = \frac{M!(Q - M)!!}{\prod_{i=1}^{\hat{N}}(k_i - \kappa_i)! \prod_{i=1+\hat{N}}^{N} q_i!} \binom{Q}{M} \delta(Q + M, 2L + 2\hat{L}). \tag{39}$$

This expression admits a simple combinatorial interpretation. In fact the networks with degree sequence $[\mathbf{k}, \mathbf{q}]$ having as subgraph $\hat{G}$ can be constructed by adding (unsampled) links to the graph $\hat{G}$. The unsampled part of the network can be constructed by assigning to each node i with $1 \leq i \leq \hat{N}$ a number of stubs given by $k_i - \kappa_i$ and to each node i with $i > \hat{N}$ a number of stubs given by q_i. The unsampled networks can then be obtained by matching the stubs pairwise with the constrains that the stubs of the first $\hat{N}$ nodes can be only matched with the stubs of the unsampled nodes $i > \hat{N}$. Therefore the reconstructed part of the network is formed by a bipartite network between the sampled and the unsample nodes with a number of links given by M and a simple network among the unsampled nodes with number of links given by $(Q - M)/2$. The number of matchings of the M links of the bipartite network is given by $M!$ the number of matching of the stubs of the simple network among unsampled nodes is $(Q - M)!!$. In order to get the number of distinct networks G with degree sequence $[\mathbf{k}, \mathbf{q}]$ having as subgraph $\hat{G}$ we need to divide by the number of permutations of the stubs belonging to the same nodes and we need to multiply by Q choose M indicating the number of ways in which we can choose the M stubs of the unsampled nodes to be matched with the stubs of the sampled nodes.

Given the expression for $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$ provided by Eq.(36), we can deduce the explicit expression for $\Delta_N\Sigma(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$:

$$\Delta_N\Sigma(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa}) = \ln\left[\prod_{i=1}^{\hat{N}} \frac{k_i!}{(k_i - \kappa_i)!} \frac{M!(Q - M)!!}{(\langle k \rangle N)!!} \binom{Q}{M} \delta(Q + M, 2L - 2\hat{L})\right]. \tag{40}$$

It follows that the describe Bayesian inference assigns a probability to the model parameters a probability

$$P(N, \mathbf{k}, \mathbf{q}|\hat{G}) \propto \pi(N) \prod_{i=1}^{\hat{N}} p(k_i) \prod_{i=1+\hat{N}+1}^{N} p(q_i) e^{-\Delta_N\Sigma(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})}, \tag{41}$$

with $\Delta_N\Sigma(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$ given by Eq. (40). From this expression, imposing with a delta function that $M = \sum_{i=1}^{\hat{N}}(k_i - \kappa_i)$, expressing the delta in integral form and using the saddle point to evaluate the

18

integral, we can calculate the marginal probability $P(k_i|\hat{G}, \omega)$ that a sampled node i with $1 \le i \le \hat{N}$ has true degree $k_i \ge \kappa_i$ given M and Q, i.e.

$$P(k_i|\hat{G}, \omega) \propto p(k_i)\frac{k_i!}{(k_i - \kappa_i)!}e^{-\omega k_i}\hat{\theta}(k_i - \kappa_i) \tag{42}$$

where ω is related to M by

$$M = \sum_{i=1}^{\hat{N}} \frac{\sum_k p(k)k\frac{k!}{(k-\kappa_i)!}e^{-\omega k}}{\sum_{k'} p(k')\frac{k'!}{(k'-\kappa_i)!}e^{-\omega k'}}. \tag{43}$$

In Figure 3 we show the difference between an exponential prior distribution $p(k)$ on the degree of the nodes and the posterior marginal probability of the true degree of the sampled nodes $P(k|\hat{G}, \omega)$ plotted for different values of the sampled degree κ of the same node. Finally, we can calculate the a posteriori probability $P(N|\hat{G}, M)$ that the real networks has N nodes, conditioned to M and to the sampled subrgraph $\hat{G}$. To this end we sum Eq. (41) over all the possible values of the degrees $\mathbf{k}$ and $\mathbf{q}$ such that Eqs.(37) are satisfied. Therefore, by inserting Eq. (40) into Eq.(41), enforcing Eqs.(37) with Kronecker deltas and integrating over all the possible values of $\mathbf{k}$ and $\mathbf{q}$ we get

$$P(N|\hat{G}, M) \propto \pi(N)\theta(N - \hat{N})C_{M,N}I^{(k)}(M)I^{(q)}(M, N), \tag{44}$$

where

$$C_{M,N} = \frac{M!(Q - M)!!}{(\langle k \rangle N)!!}\binom{Q}{M},$$

$$I^{(k)}(M) = \sum_{\mathbf{k}}\left[\prod_{i=1}^{\hat{N}} p(k_i)\frac{k_i!}{(k_i - \kappa_i)!}\delta\left(M, \sum_{i=1}^{\hat{N}} k_i\right)\right],$$

$$I^{(q)}(M, N) = \sum_{\mathbf{q}}\left[\prod_{i=1+\hat{N}}^{N} p(q_i)\delta\left(Q, \sum_{i=1+\hat{N}}^{N} q_i\right)\right], \tag{45}$$

where $Q = \langle k \rangle N - 2\hat{L} - M$. By expressing the Kronecker deltas in an integral form according to the expression

$$\delta(x, y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega(x-y)}, \tag{46}$$

performing a Wick rotation and evaluating the integrals at the saddle point, we can express $I^{(k)}(M)$ and $I^{(q)}(M, N)$ as

$$I^{(k)}(M) = \frac{1}{2\pi}\left[\prod_{i=1}^{\hat{N}}\sum_{k>\kappa_i} p(k)\frac{k!}{(k - \kappa_i)!}e^{-\omega^\star k}\right]e^{\omega^\star M},$$

$$I^{(q)}(M, N) = \frac{1}{2\pi}\left[\sum_q p(q)e^{-\bar{\omega}^\star q}\right]^{N-\hat{N}}e^{\bar{\omega}^\star Q}, \tag{47}$$

19

with $\omega^\star$ and $\bar{\omega}^\star$ fixed by the saddle point equations

$$M = \sum_{i=1}^{\hat{N}} \frac{\sum_{k>\kappa_i} p(k)\frac{k!}{(k-\kappa_i)!} k \, e^{-\omega^\star k}}{\sum_{k>\kappa_i} p(k)\frac{k!}{(k-\kappa_i)!} e^{-\omega^\star k}},$$

$$Q = (N - \hat{N})\frac{\sum_q p(q)q \, e^{-\bar{\omega}^\star q}}{\sum_q p(q)e^{-\bar{\omega}^\star q}}. \tag{48}$$

In Figure 4 we display the marginal a posteriori distribution $P(N|\hat{G}, M)$ as function of M demonstrating that the sampled network can modify significantly the prior assumptions on the total number of nodes in the network.

B.   Inferring the true parameters with the grand canonical ensemble with given latent variable distribution

In this section we treat the problem of Bayesian estimation of the parameters of the true network G given the sampled network $\hat{G}$ using the grand canonical model with given latent variable distribution. Let us indicate with θ_i the latent variables of the sampled nodes $1 \le i \le \hat{N}$ and with ϕ_i the latent variables of the unsampled nodes $i > \hat{N}$. Using Bayes rule we have

$$P(N, \boldsymbol{\theta}, \boldsymbol{\phi}|\hat{G}) = \frac{P(N)P(\boldsymbol{\theta}, \boldsymbol{\phi}|N)P(\hat{G}|\boldsymbol{\theta}, \boldsymbol{\phi}, N)}{P(\hat{G})}, \tag{49}$$

where $P(\hat{G}|\boldsymbol{\theta}, \boldsymbol{\phi}, N)$ is independent of ϕ, i.e. $P(\hat{G}|\boldsymbol{\theta}, \boldsymbol{\phi}, N) = P(\hat{G}|\boldsymbol{\theta}, N)$ and where

$$P(N) = \pi(N),$$

$$P(\boldsymbol{\theta}, \boldsymbol{\phi}|N) = \prod_{i=1}^{\hat{N}} p(\theta_i) \prod_{i=1+\hat{N}}^{N} p(\phi_i),$$

$$P(\hat{G}|\boldsymbol{\theta}, N) = \prod_{i<j|i,j\in\hat{V}} p_N(\theta_i, \theta_j)^{\hat{a}_{ij}}(1 - p_N(\theta_i, \theta_j))^{1-\hat{a}_{ij}} \tag{50}$$

with $p_N(\theta_i, \theta_j)$ given by Eq. (15) and with $\hat{\mathbf{a}}$ indicating the adjacency matrix of the sampled subgraph $\hat{G}$. In Eq. (49) $P(\hat{G})$ indicates the evidence of the data given by

$$P(\hat{G}) = \sum_N \sum_{\boldsymbol{\theta}} \pi(N) \prod_{i=1}^{\hat{N}} p(\theta_i)P(\hat{G}|\boldsymbol{\theta}, N). \tag{51}$$

Since, as we have observed previously, $P(\hat{G}|\boldsymbol{\theta}, \boldsymbol{\phi}, N)$ is independent of ϕ the Bayesian estimation of the parameters ϕ reduces simply to the prior in this case. Therefore we focus here only on the Bayesian estimate of the latent variables $\boldsymbol{\theta}$, i.e. we consider

$$P(N, \boldsymbol{\theta}|\hat{G}) = \frac{P(N)P(\boldsymbol{\theta}|N)P(\hat{G}|\boldsymbol{\theta}, N)}{P(\hat{G})}, \tag{52}$$

20

with $P(N), P(\hat{G}|\boldsymbol{\theta}, N), P(\hat{G})$ having the same definition as above and

$$P(\boldsymbol{\theta}|N) = \prod_{i=1}^{\hat{N}} p(\theta_i). \tag{53}$$

Using the explicit expression of $p_N(\theta_i, \theta_j)$ given by Eq. (15), we can express the likelihood $P(\hat{G}|\boldsymbol{\theta}, N)$ of the sampled network as

$$P(\hat{G}|\boldsymbol{\theta}, N) = \prod_{i=1}^{\hat{N}} \theta_i^{\kappa_i} \prod_{i<j|i,j\in\hat{V}} \left(1 + \frac{\theta_i \theta_j}{N}\right)^{-1} \frac{1}{N^{\hat{L}}}, \tag{54}$$

where $\hat{L}$ is the number of links of the sampled network $\hat{G}$. In the limit $N \gg 1$ we can approximate this expression as

$$P(\hat{G}|\boldsymbol{\theta}, N) \simeq \int d\bar{\theta} \prod_{i=1}^{\hat{N}} \left[\theta_i^{\kappa_i} e^{-\theta_i \bar{\theta}/2}\right] \frac{1}{N^{\hat{L}}} \delta\left(\bar{\theta}, \sum_{j=1}^{\hat{N}} \theta_j/N\right) \tag{55}$$

With this approximation we get that the posterior probability $P(N, \boldsymbol{\theta}|\hat{G})$ is given by

$$P(N, \boldsymbol{\theta}|\hat{G}) \propto \pi(N) \frac{1}{N^{\hat{L}}} \int d\bar{\theta} \prod_{i=1}^{\hat{N}} \left[p(\theta_i)\theta_i^{\kappa_i} e^{-\theta_i \bar{\theta}/2}\right] \delta\left(\bar{\theta}, \sum_{j=1}^{\hat{N}} \theta_j/N\right). \tag{56}$$

Calculating the marginal posterior probability of a single latent variable conditional of $\bar{\theta}$ we get

$$P(\theta_i|\hat{G}, \bar{\theta}) = p(\theta_i)\theta_i^{\kappa_i} e^{-\theta_i \bar{\theta}/2}. \tag{57}$$

In Figure 3 we show the difference between an exponential prior distribution $p(\theta)$ on the latent variables of the nodes and the posterior marginal probability of the true latent variables of the sampled nodes $P(\theta|\hat{G}, \bar{\theta})$ plotted for different values of the sampled degree κ of the same node.

Stating from Eq. (56) we can also calculate the posterior distribution $P(N|\hat{G})$ of the true number of nodes $N > \hat{N}$. To this end we express the delta function in an integral form and we sum over all possible latent variables $\boldsymbol{\theta}$, obtaining

$$P(N|\hat{G}) \propto \pi(N)\theta(N - \hat{N}) \frac{1}{N^{\hat{L}-1}} \frac{1}{2\pi} \int d\bar{\theta} d\omega e^{iN\omega\bar{\theta}} I^{(\theta)}(\omega, \bar{\theta}) \tag{58}$$

where $I^{(\theta)}(\omega, \bar{\theta})$ is given by

$$I^{(\theta)} = \prod_{i=1}^{\hat{N}} \left(\sum_{\theta} p(\theta)\theta^{\kappa_i} e^{-\theta(\bar{\theta}/2 - i\omega)}\right). \tag{59}$$

The integrals in Eq. (58) can be calculated at the saddle point getting

$$P(N|\hat{G}) \propto \pi(N)\theta(N - \hat{N}) \frac{1}{N^{\hat{L}-1}} e^{N \frac{(\bar{\theta}^\star)^2}{2}} \left[\prod_{i=1}^{\hat{N}} \left(\sum_{\theta} p(\theta)\theta^{\kappa_i} e^{-\theta\bar{\theta}^\star}\right)\right] \tag{60}$$

where

$$\bar{\theta}^\star = \frac{1}{N} \sum_{i=1}^{\hat{N}} \frac{\sum_\theta p(\theta) \theta^{\kappa_i+1} e^{-\theta\bar{\theta}^\star}}{\sum_\theta p(\theta) \theta^{\kappa_i} e^{-\theta\bar{\theta}^\star}}. \tag{61}$$

In Figure 4 we display the marginal a posteriori distribution $P(N|\hat{G})$ on the true number of nodes in the simplified assumption in which $\hat{G}$ is regular and all degree κ are the same demonstrating that the sampled network can modify significantly the prior assumptions on the total number of nodes in the network.

VIII.   CONCLUSIONS

In this paper we have proposed grand canonical network ensembles formed by networks of varying number of nodes. The grand canonical network ensembles we have introduced are both sparse and exchangeable, i.e. have a finite average degree and are invariant under permutation of the node labels. The grand canonical ensembles are hierarchical network models in which first the network size is selected, then the degree sequence (or the sequence of latent variables) and finally the network adjacency matrix is selected. The model circumvents the difficulties imposed by the Aldous-Hoover theorem that states that exchangeable infinite sparse network ensembles vanish, as the network is a mixture of finite networks, although the networks can have an arbitrarily large network size. Here we show how the grand-canonical ensembles can be used to perform a Bayesian estimation of the network parameters when only partial information about the network structures is known. This a posteriori estimation of the network parameters can then be used for network reconstruction.

The grand canonical framework for sparse exchangeable network ensembles is here described for the case simple networks but has the potential to be extended to generalized network structures including directed, bipartite networks, multiplex networks and simplicial complexes following the lines outlined in Ref.[32].

In conclusion we hope that this work, proposing hierarchical grand canonical network ensembles able to treat networks of different size and relating network theory to statistical mechanics will stimulate further results of mathematicians, physicists, and computer scientists working in

network science and related machine learning problems.

[1] Albert-László Barabási, *Network Science* (Cambridge University Press, 2016).

[2] EJ Newman, Mark, *Networks: An introduction* (Oxford University Press, 2010).

[3] Kartik Anand and Ginestra Bianconi, "Entropy measures for networks: Toward an information theory of complex topologies," Physical Review E **80**, 045102 (2009).

[4] Juyong Park and Mark EJ Newman, "Statistical mechanics of networks," Physical Review E **70**, 066117 (2004).

[5] Ginestra Bianconi, "Information theory of spatial network ensembles," in *Handbook on Entropy, Complexity and Spatial Dynamics* (Edward Elgar Publishing, 2021).

[6] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli, "The statistical physics of real-world networks," Nature Reviews Physics **1**, 58–71 (2019).

[7] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná, "Hyperbolic geometry of complex networks," Physical Review E **82**, 036106 (2010).

[8] Chiara Orsini, Marija M Dankulov, Pol Colomer-de Simón, Almerima Jamakovic, Priya Mahadevan, Amin Vahdat, Kevin E Bassler, Zoltán Toroczkai, Marián Boguná, Guido Caldarelli, *et al.*, "Quantifying randomness in real networks," Nature communications **6**, 1–10 (2015).

[9] Tiago P Peixoto, "Entropy of stochastic blockmodel ensembles," Physical Review E **85**, 056122 (2012).

[10] Filippo Radicchi, Dmitri Krioukov, Harrison Hartle, and Ginestra Bianconi, "Classical information theory of networks," Journal of Physics: Complexity **1**, 025001 (2020).

[11] Pedro Pessoa, Felipe Xavier Costa, and Ariel Caticha, "Entropic dynamics on gibbs statistical manifolds," Entropy **23**, 494 (2021).

[12] Hyunju Kim, Charo I Del Genio, Kevin E Bassler, and Zoltán Toroczkai, "Constructing and sampling directed graphs with given degree sequences," New Journal of Physics **14**, 023012 (2012).

[13] Charo I Del Genio, Hyunju Kim, Zoltán Toroczkai, and Kevin E Bassler, "Efficient and exact sampling of simple graphs with given arbitrary degree sequence," PloS one **5**, e10012 (2010).

[14] Anthony CC Coolen, Alessia Annibale, and Ekaterina Roberts, *Generating random networks and graphs* (Oxford university press, 2017).

[15] Kevin E Bassler, Charo I Del Genio, Péter L Erdős, István Miklós, and Zoltán Toroczkai, "Exact

sampling of graphs with prescribed degree correlations," New Journal of Physics **17**, 083052 (2015).

[16] Albert-László Barabási and Réka Albert, "Emergence of scaling in random networks," Science **286**, 509–512 (1999).

[17] Sergei N Dorogovtsev, Sergei N Dorogovtsev, and Jose FF Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford university press, 2003).

[18] Shubha R Kharel, Tamás R Mezei, Sukhwan Chung, Péter L Erdős, and Zoltan Toroczkai, "Degree-preserving network growth," Nature Physics , 1–7 (2021).

[19] Edwin T Jaynes, "Information theory and statistical mechanics," Physical Review **106**, 620 (1957).

[20] Kerson Huang, *Introduction to statistical physics* (Chapman and Hall/CRC, 2009).

[21] Kartik Anand and Ginestra Bianconi, "Gibbs entropy of network ensembles by cavity methods," Physical Review E **82**, 011116 (2010).

[22] Ginestra Bianconi, Anthony CC Coolen, and Conrad J Perez Vicente, "Entropies of complex networks with hierarchically constrained topologies," Physical Review E **78**, 016114 (2008).

[23] Guido Caldarelli, Andrea Capocci, Paolo De Los Rios, and Miguel A Munoz, "Scale-free networks from varying vertex intrinsic fitness," Physical review letters **89**, 258702 (2002).

[24] Ginestra Bianconi, Paolo Pin, and Matteo Marsili, "Assessing the relevance of node features for network structure," Proceedings of the National Academy of Sciences **106**, 11433–11438 (2009).

[25] Edo M Airoldi, David Blei, Stephen Fienberg, and Eric Xing, "Mixed membership stochastic block-models," Advances in neural information processing systems **21** (2008).

[26] Arsham Ghavasieh, Carlo Nicolini, and Manlio De Domenico, "Statistical physics of complex information dynamics," Physical Review E **102**, 052304 (2020).

[27] Beatrice Bevilacqua, Yangze Zhou, and Bruno Ribeiro, "Size-invariant graph representations for graph classification extrapolations," in *International Conference on Machine Learning* (PMLR, 2021) pp. 837–851.

[28] Leonardo Cotta, Christopher Morris, and Bruno Ribeiro, "Reconstruction for powerful graph representations," Advances in Neural Information Processing Systems **34** (2021).

[29] Bruno De Finetti, "Funzione caratteristica di un fenomeno aleatorio, vol. 4 of 6," (1931).

[30] László Lovász, *Large networks and graph limits*, Vol. 60 (American Mathematical Soc., 2012).

[31] Fan Chung and Linyuan Lu, "The average distances in random graphs with given expected degrees," Proceedings of the National Academy of Sciences **99**, 15879–15882 (2002).

[32] Ginestra Bianconi, "Statistical physics of exchangeable sparse simple networks, multiplex networks,

and simplicial complexes," Physical Review E **105**, 034310 (2022).

[33] François Caron and Emily B Fox, "Sparse graphs using exchangeable random measures," Journal of the Royal Statistical Society. Series B, Statistical Methodology **79**, 1295 (2017).

[34] Christian Borgs, Jennifer T Chayes, Henry Cohn, and Nina Holden, "Sparse exchangeable graphs and their limits via graphon processes," arXiv preprint arXiv:1601.07134 (2016).

[35] Victor Veitch and Daniel M Roy, "The class of random graphs arising from exchangeable random measures," arXiv preprint arXiv:1512.03099 (2015).

[36] Victor Veitch and Daniel M Roy, "Sampling and estimation for (sparse) exchangeable graphs," The Annals of Statistics **47**, 3274–3299 (2019).

[37] Christian Borgs, Jennifer T Chayes, and Adam Smith, "Private graphon estimation for sparse graphs," arXiv preprint arXiv:1506.06162 (2015).

[38] Christian Borgs, Jennifer Chayes, Adam Smith, and Ilias Zadik, "Revealing network structure, confidentially: Improved rates for node-private graphon estimation," in *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, 2018) pp. 533–543.

[39] Ginestra Bianconi, *Multilayer networks: structure and function* (Oxford University Press, 2018).

[40] Ginestra Bianconi, *Higher-order networks: an introduction to simplicial complexes* (Cambridge University Press, 2021).

[41] David J Aldous, "Representations for partially exchangeable arrays of random variables," Journal of Multivariate Analysis **11**, 581–598 (1981).

[42] Douglas N Hoover, "Relations on probability spaces and arrays of random variables," Preprint, Institute for Advanced Study, Princeton, NJ **2**, 275 (1979).

[43] Jeremy Paton, Harrison Hartle, Jakob Stepanyants, Pim van der Hoorn, and Dmitri Krioukov, "Entropy of labeled versus unlabeled networks," arXiv preprint arXiv:2204.08508 (2022).

[44] Tiago P Peixoto, "Hierarchical block structures and high-resolution model selection in large networks," Physical Review X **4**, 011047 (2014).

[45] Andrea Gabrielli, Rossana Mastrandrea, Guido Caldarelli, and Giulio Cimini, "Grand canonical ensemble of weighted networks," Physical Review E **99**, 030301 (2019).

[46] Mika J Straka, Guido Caldarelli, and Fabio Saracco, "Grand canonical validation of the bipartite international trade network," Physical Review E **96**, 022306 (2017).

[47] Edward A Bender and E Rodney Canfield, "The asymptotic number of labeled graphs with given degree sequences," Journal of Combinatorial Theory, Series A **24**, 296–307 (1978).

[48] Ginestra Bianconi, "Entropy of network ensembles," Physical Review E **79**, 036114 (2009).

[49] Owen T Courtney and Ginestra Bianconi, "Generalized network structures: The configuration model and the canonical ensemble of simplicial complexes," Physical Review E **93**, 062311 (2016).

[50] Rémi Monasson and Riccardo Zecchina, "Statistical mechanics of the random k-satisfiability model," Physical Review E **56**, 1357 (1997).

APPENDIX: DERIVATION OF $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$

In this Appendix our goal is to derive the asymptotic expression of $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$ in the limit of large network size of the sampled network $\hat{N} \gg 1$, and of the true network $N = (1 + \alpha)\hat{N} \gg 1$ with $\alpha > 0$.

Let us assume that the sampled subgraph G is the network between the sampled nodes $1 \leq i \leq \hat{N}$ and has adjacency matrix $\hat{\mathbf{a}}$. The true network is instead formed by N nodes with adjacency matrix $\mathbf{a}$. We assume that $\mathbf{a}$ has the block structure given by

$$\mathbf{a} = \begin{pmatrix} \hat{\mathbf{a}} & \mathbf{b} \\ \mathbf{b}^\top & \tilde{\mathbf{a}} \end{pmatrix}, \tag{62}$$

where $\mathbf{b}$ indicates the $\hat{N} \times \alpha\hat{N}$ matrix between sampled nodes and the unsampled node and $\tilde{\mathbf{a}}$ indicates tha $(\alpha\hat{N}) \times (\alpha\hat{N})$ adjacency matrix among the unsampled nodes. As we have mentioned in the main text $\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})$ is the logarithm of the number $\mathcal{N}(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa}, N)$ of networks (or adjacency matrices $\mathbf{a}$) with degree sequence $[\mathbf{k}, \mathbf{q}]$ and admitting as a subgraph $\hat{G}$ having sampled degree sequence $\boldsymbol{\kappa}$. In statistical mechanics we also call $\mathcal{N}(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa}, N)$ the partition function of its corresponding statistical mechanics network model, and we indicate it by Z. In terms of the matrices $\mathbf{b}$ and $\tilde{\mathbf{a}}$ the partition function $Z = \mathcal{N}(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa}) = \exp\left(\hat{\Sigma}_N(\mathbf{k}, \mathbf{q}|\boldsymbol{\kappa})\right)$ can be written as

$$Z = \sum_{\mathbf{b}, \tilde{\mathbf{a}}} \prod_{i=1}^{\hat{N}} \delta\left(k_i - \sum_{j=1}^{N} a_{ij}\right) \prod_{i=1+\hat{N}}^{N} \delta\left(q_i - \sum_{j=1}^{N} a_{ij}\right) \delta\left(2L - \sum_{i=1}^{N} k_i\right) \tag{63}$$

Expressing the Kronecker deltas in the integral form and performing the sum over the elements of the matrices $\mathbf{b}$ and $\tilde{\mathbf{a}}$ we obtain

$$Z = \int \mathcal{D}\omega \int \mathcal{D}\tilde{\omega} \int \frac{d\lambda}{2\pi} e^{G(\omega, \tilde{\omega}, \lambda)} \tag{64}$$

with

$$G(\boldsymbol{\omega}, \tilde{\boldsymbol{\omega}}, \lambda) = \sum_{i=1}^{\hat{N}} [i\omega_i(k_i - \kappa_i)] + \sum_{i=1+\hat{N}}^{N} [i\tilde{\omega}_i q_i] + \sum_{i=1}^{\hat{N}} \sum_{j=1}^{\hat{N}} \ln(1 + e^{-i\omega_i - i\tilde{\omega}_j - i\lambda})$$

$$+ \frac{1}{2} \sum_{i=\hat{N}+1}^{N} \sum_{j=\hat{N}+1}^{N} \ln(1 + e^{-i\tilde{\omega}_i - i\tilde{\omega}_j - i\lambda}) + i\lambda(L - \hat{L}), \tag{65}$$

and with $\mathcal{D}\omega = \prod_{i=1}^{\hat{N}} [d\omega_i/(2\pi)]$ and $\mathcal{D}\tilde{\omega} = \prod_{i=1+\hat{N}}^{N} [d\tilde{\omega}_i/(2\pi)]$. Let us now introduce the functional order parameters [22, 49, 50]

$$c_{\kappa,k}(\omega) = \frac{1}{\hat{N}\hat{P}(\kappa, k)} \sum_{i=1}^{\hat{N}} \delta(\omega - \omega_i)\delta(k, k_i)\delta(\kappa, \kappa_i),$$

$$\rho_q(\tilde{\omega}) = \frac{1}{\alpha\hat{N}\tilde{P}(q)} \sum_{i=1+\hat{N}}^{N} \delta(\tilde{\omega} - \tilde{\omega}_i)\delta(q, q_i), \tag{66}$$

where $\hat{P}(k, \kappa)$ is the fraction of sampled nodes with degree κ in the sampled network and total inferred degree k; $\tilde{P}(q)$ is the fraction of unsampled nodes with degree q. Moreover we have indicated with $L = \langle k \rangle N/2$ and with $\hat{L} = \sum_{i=1}^{\hat{N}} \kappa_i/2$. By enforcing the definition of the order parameters with a series of delta functions we obtain

$$1 = \int dc_{\kappa,k}(\omega)\delta\left(c_{\kappa,k}(\omega) - \frac{1}{\hat{N}\hat{P}(\kappa, k)} \sum_{i=1}^{\hat{N}} \delta(\omega - \omega_i)\delta(k, k_i)\delta(\kappa, \kappa_i)\right)$$

$$= \int \frac{d\hat{c}_{\kappa,k}(\omega)dc_{\kappa,k}(\omega)}{2\pi/(\hat{N}\hat{P}(\kappa, k)\Delta\omega)} \exp\left[i\Delta\omega\hat{c}_{\kappa,k}(\omega)[\hat{N}\hat{P}(\kappa, k)c_{\kappa,k}(\omega) - \sum_{i=1}^{\hat{N}} \delta(\omega - \omega_i)\delta(k, k_i)\delta(\kappa, \kappa_i)]\right].$$

$$1 = \int d\rho_q(\tilde{\omega})\delta\left(\rho_q(\tilde{\omega}) - \frac{1}{\alpha\hat{N}\tilde{P}(q)} \sum_{i=1+\hat{N}}^{N} \delta(\tilde{\omega} - \tilde{\omega}_i)\delta(q, q_i)\right)$$

$$= \int \frac{d\hat{\rho}_q(\tilde{\omega})d\rho_q(\tilde{\omega})}{2\pi/(\alpha\hat{N}\tilde{P}(q)\Delta\tilde{\omega})} \exp\left[i\Delta\tilde{\omega}\hat{\rho}_q(\tilde{\omega})[\alpha\hat{N}\tilde{P}(q)\rho_q(\tilde{\omega}) - \sum_{i=1+\hat{N}}^{N} \delta(\tilde{\omega} - \tilde{\omega}_i)\delta(q, q_i)]\right].$$

After inserting these expressions into the partition function in the limit $\Delta\omega \to 0$, indicating with $\sum'$ the sum over the allowed degree range we obtain

$$Z = \sum_{\boldsymbol{\kappa}}' \sum_{\mathbf{k}}' \sum_{\mathbf{q}}' \int \prod_{\kappa,k} \mathcal{D}c_{\kappa,k}(\omega) \int \prod_{\kappa,k} \mathcal{D}\hat{c}_{\kappa,k}(\omega) \int \prod_{q} \mathcal{D}\rho_q(\tilde{\omega}) \int \prod_{q} \mathcal{D}\hat{\rho}_q(\tilde{\omega}) \int \frac{d\lambda}{2\pi} e^{\hat{N}f}$$

with $f = f(c(\omega, k), \hat{c}(\omega, k), \rho(\tilde{\omega}, q), \hat{\rho}(\tilde{\omega}, q), \lambda, h)$ given by

$$f = \sum_{\hat{m} \leq \kappa \leq K} \sum_{\kappa \leq k \leq K} \hat{P}(\kappa, k)\mathrm{i} \int d\omega \hat{c}_{\kappa,k}(\omega)c_{\kappa,k}(\omega) + \alpha\mathrm{i} \int d\omega \sum_{\hat{m} \leq q \leq K} \tilde{P}(q)\hat{\rho}_q(\tilde{\omega})\rho_q(\tilde{\omega})$$

$$+\mathrm{i}\lambda(L - \hat{L})/\hat{N} + \Psi + \sum_{\hat{m} \leq \kappa \leq K} \sum_{\kappa \leq k \leq K} \hat{P}(\kappa, k) \ln \int \frac{d\omega}{2\pi} e^{\mathrm{i}\omega(k-\kappa)-\mathrm{i}\hat{c}_{\kappa,k}(\omega,k)}$$

$$+\alpha \sum_{\hat{m} \leq q \leq K} \tilde{P}(q) \ln \int \frac{d\tilde{\omega}}{2\pi} e^{\mathrm{i}\tilde{\omega}q - \mathrm{i}\hat{\rho}_q(\tilde{\omega})}, \tag{67}$$

where Ψ is given by

$$\Psi = \frac{\alpha^2 \hat{N}}{2} \sum_{\hat{m} \leq q \leq K, \hat{m} \leq q' \leq K} \tilde{P}(q)\tilde{P}(q') \int d\omega \int d\tilde{\omega}' \rho_q(\tilde{\omega})\rho_{q'}(\tilde{\omega}') \ln \left(1 + e^{-\mathrm{i}\tilde{\omega}-\mathrm{i}\tilde{\omega}'-\mathrm{i}\lambda} \right)$$

$$+\alpha\hat{N} \sum_{\hat{m} \leq \kappa \leq K} \sum_{\kappa \leq k \leq K} \hat{P}(\kappa, k) \sum_{\hat{m} \leq q \leq K} \tilde{P}(q) \int d\omega \int d\tilde{\omega} c_{\kappa,k}(\omega)\rho_q(\tilde{\omega}) \ln \left(1 + e^{-\mathrm{i}\omega-\mathrm{i}\tilde{\omega}-\mathrm{i}\lambda} \right),$$

and where the functional measures are defined as

$$\mathcal{D}c_{\kappa,k}(\omega) = \lim_{\Delta\omega \to 0} \prod_\omega [dc_{k\kappa}(\omega) \sqrt{\hat{N}\hat{P}(\kappa, k)\Delta\omega/(2\pi)}]$$

$$\mathcal{D}\hat{c}_{\kappa,k}(\omega) = \lim_{\Delta\omega \to 0} \prod_\omega [d\hat{c}_{\kappa,k}(\omega) \sqrt{\hat{N}\hat{P}(\kappa, k)\Delta\omega/(2\pi)}],$$

$$\mathcal{D}\rho_q(\tilde{\omega}) = \lim_{\Delta\tilde{\omega} \to 0} \prod_{\tilde{\omega}} [d\rho_q(\tilde{\omega}) \sqrt{\hat{N}\alpha\tilde{P}(q)\Delta\tilde{\omega}/(2\pi)}],$$

$$\mathcal{D}\hat{\rho}_q(\tilde{\omega}) = \lim_{\Delta\tilde{\omega} \to 0} \prod_{\tilde{\omega}} [d\hat{\rho}_q(\tilde{\omega}) \sqrt{\hat{N}\alpha\tilde{P}(q)\Delta\tilde{\omega}/(2\pi)}]. \tag{68}$$

By putting

$$e^{-\mathrm{i}\lambda} = \frac{z}{\hat{N}}, \tag{69}$$

and performing a Wick rotation in λ and assuming $z/\hat{N} = e^{-\mathrm{i}\lambda}$ real and much smaller than one, i.e. $z/\hat{N} \ll 1$ which is allowed in the sparse regime, we can linearize the logarithm and express Ψ as

$$\Psi = z\alpha\nu\left(\frac{1}{2}\alpha\nu + \hat{\nu}\right), \tag{70}$$

with

$$\nu = \sum_{\hat{m} \leq q \leq K} \tilde{P}(q) \int d\tilde{\omega}\rho_q(\tilde{\omega})e^{-\mathrm{i}\tilde{\omega}}.$$

$$\hat{\nu} = \sum_{\hat{m} \leq \kappa \leq K} \sum_{\hat{k}, \leq k \leq K} \hat{P}(\kappa, k) \int d\omega c_{\kappa,k}(\omega)e^{-\mathrm{i}\omega}. \tag{71}$$

The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of f with respect to the functional order parameters, obtaining

$$-i\hat{c}_{\kappa,k}(\omega) = z\alpha v e^{-i\omega},$$

$$-i\hat{\rho}_q(\omega) = z(\alpha v + \hat{v})e^{-i\tilde{\omega}},$$

$$c_{\kappa,k}(\omega) = \hat{P}(\kappa,k)\frac{\frac{1}{2\pi}e^{i\omega(k-\kappa)-i\hat{c}_{\kappa,k}(\omega)}}{\int\frac{d\omega'}{2\pi}e^{i\omega'(k-\kappa)-i\hat{c}_{\kappa,k}(\omega')}},$$

$$\rho_q(\tilde{\omega}) = \tilde{P}(q)\frac{\frac{1}{2\pi}e^{i\tilde{\omega}q-i\hat{\rho}_q(\tilde{\omega})}}{\int\frac{d\tilde{\omega}'}{2\pi}e^{i\tilde{\omega}'q-i\hat{\rho}_q(\tilde{\omega}')}},$$

$$2\frac{L-\hat{L}}{\hat{N}} = z\alpha v\,(\alpha v + 2\hat{v})\,. \tag{72}$$

Let us first calculate the integrals

$$I_{\kappa,k} = \int\frac{d\omega}{2\pi}e^{-i\omega(k-\kappa)-i\hat{c}_{\kappa,k}(\omega)} = \frac{1}{(k-\kappa)!}(z\alpha v)^{k-\kappa},$$

$$I_q = \int\frac{d\tilde{\omega}}{2\pi}e^{-i\tilde{\omega}q-i\hat{\rho}_q(\tilde{\omega})} = \frac{1}{q!}[z(\alpha v + \hat{v})]^q, \tag{73}$$

Using these expressions for the integral we can write the functional order parameters as

$$c_{\kappa,k}(\omega) = \hat{P}(\kappa,k)\frac{1}{2\pi}\frac{e^{i\omega(k-\kappa)+(z\alpha v)e^{-i\omega}}}{I_{\kappa,k}},$$

$$\rho_q(\tilde{\omega}) = \tilde{P}(q)\frac{1}{2\pi}\frac{e^{i\tilde{\omega}q+[zv(\alpha v+\hat{v})]e^{-i\tilde{\omega}}}}{I_q}. \tag{74}$$

With this expression, using a similar procedure we can express v as

$$\hat{v} = \int d\omega \sum_{\hat{m}\le\kappa\le K}\sum_{\kappa\le k\le K} c_{\kappa,k}(\omega)e^{-i\omega} = \sum_{\hat{k}\le k\le K}\hat{P}(\kappa,k)(k-\kappa)(\alpha z v)^{-1}.$$

$$v = \int d\tilde{\omega}\sum_{\hat{m}\le q\le K}\rho_q(\tilde{\omega})e^{-i\tilde{\omega}} = \sum_{\hat{m}\le q\le K}\tilde{P}(q)q[z(\alpha v + \hat{v})]^{-1}. \tag{75}$$

Combing these equations with the last saddle point equation it is immediate to show that z, v and $\hat{v}$ are given by

$$z = 1,$$

$$\alpha v = \sqrt{(Q-M)/\hat{N}},$$

$$\hat{v} = \frac{M/\hat{N}}{\sqrt{(Q-M)/\hat{N}}}. \tag{76}$$

with

$$2L - 2\hat{L} = M + Q. \tag{77}$$

Calculating the free energy $\hat{N}f$ at the saddle point, we get

$$\hat{N}f = -\frac{1}{2}(Q - M) - M + (L - \hat{L}) \ln \hat{N} + \hat{N} \sum_{\hat{m} \leq \kappa \leq K} \sum_{\hat{k} \leq k \leq K} \hat{P}(\kappa, k) \ln \frac{(\alpha\nu)^{k-\kappa}}{(k - \kappa)!}$$

$$+\alpha\hat{N} \sum_{\hat{m} \leq q \leq K} \tilde{P}(q) \ln \frac{[\alpha\nu + \hat{\nu}]^q}{q!}, \tag{78}$$

which leads to the following asymptotic expression for $Z = \mathcal{N}(\mathbf{k}, \mathbf{q} | \kappa, N) = \exp\left(\hat{\Sigma}_N(\mathbf{k}, \mathbf{q} | \kappa)\right)$

$$Z = \mathcal{N}(\mathbf{k}, \mathbf{q} | \kappa, N) \simeq \frac{M!(Q - M)!!}{\prod_{i=1}^{\hat{N}} k_i! \prod_{i=1+\hat{N}}^{N} q_i!} \binom{Q}{M}. \tag{79}$$