Statistical physics of exchangeable sparse network ensembles

Ginestra Bianconi\textsuperscript{1,2}
\textsuperscript{1}School of Mathematical Sciences, Queen Mary University of London, London, E1 4NS, United Kingdom
\textsuperscript{2}The Alan Turing Institute, The British Library, London NW1 2DB, United Kingdom

Exchangeability is a desired statistical property of network ensembles requiring their invariance upon relabeling of the nodes. Here we propose a statistical physics framework and a Metropolis-Hastings algorithm defining exchangeable sparse network ensembles. The model generates networks with heterogeneous degree distributions by enforcing only global constraints while existing (non exchangeable) exponential random graphs enforce an extensive number of local constraints. The model can be generalized to networks with degree correlations and to generalized network structures.

Networks constitute the architecture of the vast majority of complex systems ranging from the brain to financial networks\textsuperscript{1,2}. Maximum entropy network ensembles\textsuperscript{3–18} and in general information theory frameworks\textsuperscript{19–21} are key to model and analyze such realistic networks, and can be used for a wide variety of applications. Due to the deep relation between information theory and statistical mechanics\textsuperscript{22} maximum entropy network ensembles can to large extent be treated as traditional statistical mechanics ensembles. Indeed recently it has been shown in Ref.\textsuperscript{7} that network ensembles can be distinguished between canonical and microcanonical network ensembles enforcing respectively soft and hard constraints. For instance Erdős and Rényi networks of $N$ nodes can enforce either a given total number of links $L$ (giving rise to the ensemble $G(N, L)$ ensemble) or a given expected number of links (giving rise to the ensemble $G(N, p)$ ensemble where $p$ is the probability that any two links are connected). Erdős and Rényi networks are certainly important, however in many applications it is observed that nodes have heterogeneous degree distribution\textsuperscript{23}, typically deviating from the Poisson degree distribution of ER networks in the sparse network regime. In order to treat network ensembles with heterogeneous degree distribution, exponential random graphs\textsuperscript{3,4,24} are considered instead. Exponential random graphs are canonical network ensembles enforcing a given expected degree sequence. While the Erdős Rényi impose a single global constraint such as the expected total number of links, exponential random graphs enforce an extensive number of local constraints each given by the expected degree of a single node of the network. This feature of exponential random graphs makes these ensembles significantly different from the Erdős Rényi ensembles. The first main difference is that these ensembles are not any more equivalent to their conjugated microcanonical network ensemble\textsuperscript{25} (the configuration model) which enforces a given degree sequence of the network. The second main difference is that these ensembles are not any more exchangeable. Exchangeability is a notion originally introduced by de Finetti\textsuperscript{26} whose theorem states that a sequence of random variables $X_1, X_2, \ldots$ is exchangeable if and only if there exists a random probability measure $\Theta$ such that the $X_i$ are conditionally identically independent variables given $\Theta$. This notion has been then extended to 2-arrays, (i.e. networks) for which the Aldous-Hoover theorem applies\textsuperscript{27,30}.

Exchangeability is a desired statistical property of network ensembles that ensure invariance of the model upon relabelling of the nodes. This property is therefore fundamental for using reliably the network model for sampling and for preserving privacy when processing real network data. In the dense network regime, graphons\textsuperscript{29,31} have been shown to be exchangeable and are known to allow a well defined infinite graph limit\textsuperscript{27,28}. Graphons are dense in the sense that they have a number of links of the same order of the number of nodes to the power two, i.e. $L = O(N^2)$. However this regime is seldom encountered in real networks. The mathematics literature has recently proposed several approaches to face the challenge of modelling exchangeable networks with a number of links that scales like $L = O(N^\alpha)$ with $0 < \alpha < 1$\textsuperscript{30,32,34} and to define ways to define the infinite network limit\textsuperscript{33} for such models. All these approaches are based on point processes on $\mathbb{R}^2$.

In this paper we propose a statistical physics approach to model sparse exchangeable network ensembles of a given number of nodes $N$ and a number of links that scales as $L = O(N)$. Therefore these network ensembles cover the scaling regime $L = O(N)$ which is important for the vast majority of applications. The exchangeable network ensembles are Hamiltonian and are not based on a point processes. These ensembles generate networks with given heterogeneous degree distribution $\rho(k)$ imposing only two global constraints: the energy (expressing the value of a global exchangeable Hamiltonian of the network ensemble) and the total number of links. The proposed exchangeable sparse network ensembles have the property that each link of the network has the same marginal probability, still the network display and heterogeneous degree distribution. Moreover the probability that two nodes are connected, when conditioned on their degrees reduces to the probability of the exponential random graph in the uncorrelated limit. Note that we do not make claims on the limit for $N \to \infty$, instead we take an equilibrium statistical mechanics approach and we consider $N$ finite but large. The model can be simulated with a constructive Metropolis algorithm and can be extended to network models with degree correlations, to directed, bipartite networks and to generalized network structures such as multiplex networks\textsuperscript{8,35} and simplicial complexes\textsuperscript{36–38}. 
Introduction to exchangeable networks - A network ensemble is exchangeable if the the probability $P(G)$ of a network $G = (V, E)$ is independent on the nodes labels, i.e.,

$$P(G) = P(\tilde{G}),$$

where $\tilde{G}$ is obtained from the network $G$ by permuting the nodes labels. It follows that in an exchangeable network ensemble the marginal probability $p_{ij}$ of a link between node $i$ and node $j$ must invariant under any permutation $\sigma$ of the node labels, i.e.

$$p_{ij} = p_{\sigma(i), \sigma(j)}.$$

In an exponential random graph ensemble with given expected degree sequence $k = \{k_1, k_2, \ldots, k_N\}$ with $k_i \leq K \ll K_S = \sqrt{(k)N}$ the marginal distribution $p_{ij}$ takes the well celebrated expression

$$p_{ij} = \frac{k_ik_j}{(k)N},$$

where $(k)N = \sum_{i=1}^{N} k_i = 2L$ is twice the expected total number of links of the network. This network ensemble is not exchangeable, unless the expected degree of each node is the same. Indeed the marginal probability $p_{ij}$ is not invariant upon permutation of the node labels if the expected degree distribution is heterogeneous. Only in the case in which the expected degree of each nodes is the same $k_i = \langle k \rangle$ we recover the exchangeable expression of the marginal probability of a sparse Poisson Erdős and Rényi network $p_{ij} = \langle k \rangle/N$.

In the following we will propose an exchangeable sparse network ensembles that imposes an heterogeneous expected degree distribution $p(k)$ enforcing only two global constraints: the energy of the ensemble and the total number of links. In this ensemble the marginal probability of a link between node $i$ and node $j$ is given by the exchangeable expression,

$$p_{ij} = \sum_{kk'} p(k)p(k') \frac{kk'}{(k)N} = \frac{(k)}{N}.$$  

In other words the marginal probability of any link is the same for every link, but when it is conditioned to the degree of the two linked nodes is given by the uncorrelated network expression

$$p_{ij|k_i=k,k_j=k'} = p(k,k') = \frac{kk'}{(k)N}.$$  

Therefore this expression has the same marginal of the Erdős and Rényi network but it can generate uncorrelated networks with heterogeneous degree distribution $p(k)$.

Exchangeable Sparse Networks-

Our goal is to construct sparse exchangeable network ensembles of $N$ nodes with degree distribution $p(k)$. Here by sparse we imply that these networks display a cutoff $K$ much smaller than the structural cutoff $K_S$, i.e. $k_i \leq K \ll K_S = \sqrt{(k)N}$ for all $i \in \{1, 2, 3, \ldots, N\}$ with $\langle k \rangle$ indicating the average over the $p(k)$ distribution, $\langle k \rangle = \sum_k kp(k)$. We assume that the nodes of the network can change their degree and we assign to each possible degree sequence of the network the probability

$$P(k) = \prod_{i=1}^{N} \left[p(k_i)\theta(K - k_i)\theta(k_i - m)\right].$$

where $\theta(x)$ indicates the Heaviside function with $\theta(x) = 1$ for $x \geq 0$ and $\theta(x) = 0$ otherwise. Therefore the probability of a degree sequence results from the product of the probability that each node has the observed degree. For keeping the model general we assume that the minimum degree of the network must be equal or greater than $m$. For instance if we want to impose a power-law degree distribution $p(k) = c \cdot k^{-\gamma}$ this allow us to impose a minimum degree $m \geq 1$ and to exclude isolated nodes for which $p(k)$ is not defined. In order to build an exchangeable network ensemble we need to define a probability $P(G)$ for any possible network $G = (V, E)$ in the ensemble described by the adjacency matrix $a$. In order to ensure sparsity we impose that the total number of links $L$ is fixed and given by $L = \langle k \rangle N/2$ and we impose that the probability of getting a degree sequence $k$ is $P(k)$. Since the number of networks $\mathcal{N}$ with given degree sequence $k$ can be expressed in terms of the entropy $\Sigma(k)$ of networks with degree sequence $k$ as $\mathcal{N} = \exp(\Sigma(k))$, the probability of each single network $G$ displaying a degree sequence $k$ is therefore taken to be

$$P(G) = P(k) e^{-\Sigma(k)} \delta\left(L, \sum_{i<j} a_{ij}\right),$$

where $\delta(x, y)$ indicates the Kronecker delta. For sparse networks the entropy $\Sigma(k)$ of networks with given degree sequence with $k_i \ll K_S$ obeys the Bender-Canfield formula

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

where in Eq. (7) and (8) we indicate with $k_i$ the degree of node $i$ given by $k_i = \sum_{j=1}^{N} a_{ij}$. Note that the sparse exchangeable network ensemble greatly differs from the network ensemble with given expected degree sequence because in the exchangeable ensemble the constraints are global and not local. Indeed the expression for $P(G)$ can be also given by

$$P(G) = e^{-H(G)} \delta\left(L, \sum_{i<j} a_{ij}\right) \theta\left(K - \max_{i=1,...,N} k_i\right) \theta\left(\min_{i=1,...,N} k_i - m\right),$$

with Hamiltonian $H(G)$ given by

$$H(G) = -\sum_{i=1}^{N} \ln p(k_i) + \Sigma(k).$$

Using Eq. (8) for $\Sigma(k)$ we can derive the explicit expression for $H(G)$:

$$H(G) = -\sum_{i=1}^{N} [p(k_i)k_i! + \ln((2L)!)]$$
The Hamiltonian $H(G)$ is clearly a global variable that depends on all the nodes of the network where each node is treated on equal footing. Therefore the expression of the Hamiltonian is clearly exchangeable as it is invariant upon a permutation of the node labels. In order to show that for this ensemble is exchangeable and that the marginal distribution is given by Eq. (4) we solve the model using saddle point method applied to a free-energy expressed in terms of a functional order parameter. In order to perform this calculation, let us write the probability $\mathcal{P}(G)$ as

$$\mathcal{P}(G) = \frac{1}{(2L)!^N} \sum_k \prod_{i=1}^N \left( \frac{1}{k!} \right) \delta \left( k_i, \sum_{j=1}^N a_{ij} \right) \delta \left( \sum_{j=1}^N a_{jj}, L \right),$$

where $\sum_k'$ indicates the sum over all possible degree sequences with a maximum degree equal or smaller than $K$ and a minimum degree greater or equal than $m$. Expressing the Kronecker deltas in Eq. (12) with their integral form

$$\delta(x, y) = \frac{1}{2\pi} \int_0^2 d\omega e^{i\omega(x-y)},$$

the partition function $Z$ can be expressed as

$$Z = \sum_a \mathcal{P}(G)e^{-\sum_{i=1}^N a_{ij}} = \frac{1}{(2L)!^N} \sum_a \int G(\lambda, \omega, k, h) \int d\lambda \int \frac{d\lambda}{2\pi} G(\lambda, \omega, k, h),$$

with

$$G(\lambda, \omega, k, h) = \sum_{i=1}^N [i\omega, k_i + \ln(k_i, p(k))] + i\lambda L$$

$$+ \frac{1}{2} \sum_{i,j} \ln(1 + e^{-i\lambda(i \rightarrow j - a_{ij} - h)}),$$

and $\mathcal{D}\omega = \prod_{i=1}^N [d\omega/(2\pi)]$. Let us now introduce the functional order parameter indicating the density of nodes with degree $k_i = k$ and with $\omega_i = \omega$.

$$c(\omega, k) = \frac{1}{N} \sum_{i=1}^N \delta(\omega - \omega_i) \delta(k, k_i).$$

By calculating the partition function $Z$ in the sparse regime (i.e. $K \ll K_N$) with the saddle point method (see SI) we can derive for the functional order parameter $c(\omega, k)$ when $h \rightarrow 0$ the expression

$$c(\omega, k) = \frac{1}{2\pi} \int d\omega p(k) e^{i\omega k + e^{-\omega}}.$$ 

This implies that the density of nodes of degree $k$ is given by

$$\int d\omega c(\omega, k) = p(k).$$

Therefore the degree of each node is fluctuating, but in the large network limit the degree distribution is given by $p(k)$ as desired. We are now in the position to evaluate the marginal probability $p_{ij}$ of a link between node $i$ and node $j$. A straightforward calculation (see SI) leads to the expression of the marginal probability $p_{ij} = \langle a_{ij} \rangle$ in terms of the functional order parameter $c(\omega, k)$ leading to

$$p_{ij} = \frac{1}{\langle k \rangle N} \sum_{m \leq k, n \leq k} \int_0^\pi d\omega \int_0^\pi d\omega' c(\omega, k) c(\omega', k') e^{-i\omega - i\omega'}. $$

From this expression, using Eq. (15) it follows immediately the expression for the marginal given by Eq. (4) leading to $p_{ij} = \langle k \rangle / N$ also if the marginal probability conditioned on the nodes degree (see SM for a detailed derivation) is given by Eq. (5). Therefore the marginal probability $p_{ij} = \langle k \rangle / N$ is the same for every node of the network and it is equal to the marginal probability in a Poisson Erdős Rényi network, but the degree distribution is $p(k)$, i.e. it can significantly differ from a Poisson distribution.

**Metropolis-Hastings algorithm** - The exchangeable ensemble of sparse networks can be obtained by implementing a simple Metropolis-Hastings algorithm using the network Hamiltonian given by Eq. (11). The Metropolis-Hastings algorithm for the exchangeable sparse networks is outlined below.

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

2. Iterate the following steps until equilibration:
   (i) Choose randomly a link $\ell = (i, j)$ between node $i$ and $j$ and choose a pair of random nodes $(i', j')$ not connected by a link. By indicating with $a$ the adjacency matrix of the network we have $a_{ij} = 1$ and $a_{ij'} = 0$.
   (ii) Draw a random number $r$ from a uniform distribution in $[0, 1]$, i.e. $r \sim U(0, 1)$. Calculate the Hamiltonian $H_{mf}(a)$ for the network with adjacency matrix $a$ of the network and calculate the Hamiltonian $H_{mf}$ for the adjacency matrix $a'$ in which the link between nodes $(i, j)$ is removed and the link between the nodes $(i', j')$ is inserted instead. If $r > \min(1, e^{-\Delta H})$ where $\Delta H = H_{mf} - H_{mf}$ and if the move does not violate the conditions on the minimum and maximum degree of the network, set $a_{ij} \rightarrow 1 - a_{ij}$ and $a_{ij'} \rightarrow 1 - a_{ij'}$ and $a_{ij'} \rightarrow 1 - a_{ij'}$.

This algorithm can be used to generate exchangeable network ensembles with different degree distributions such as exponential distribution or power-law degree distribution (see Figure 1).

This approach can be directly extended to treat directed and bipartite networks and also generalized network structures such as multiplex networks and simplicial complexes (see SM). In the following we describe in more detail the extension to sparse networks with given degree correlations.

**Exchangeable network ensembles with degree correlations** - Here our aim is to construct a sparse exchangeable network ensemble of $N$ nodes in which each node has degree $k$ with
probability \( p(k) \) and every link between a node of degree \( k \) and a node of degree \( k' \) contributes to the probability of the network \( G \) by a factor \( Q(k, k') \) where \( Q(k, k') = Q(k') \) with \( Q(k, k') \) independent of \( N \). To this end, we follow a derivation similar to the one considered for the exchangeable uncorrelated network ensemble and we assign to each network \( G \) a probability \( \mathbb{P}(G) \) given by

\[
\mathbb{P}(G) = \prod_{i,j} [Q(k_i, k_j)]^{p(i,j)} \prod_{i=1}^{N} p(k_i) e^{-\Sigma k_i} \theta \left( L - \sum_{i,j} a_{ij} \right) \times \theta \left( K - \max_{i=1,...,N} k_i \right) \theta \left( m - \min_{i=1,...,N} k_i \right),
\]

where \( \Sigma(k) \) is the entropy of the ensemble of correlated networks with degree sequence \( k \). The entropy \( \Sigma(k) \) can be expressed as

\[
\Sigma(k) = \ln \left( \frac{2L!!}{K!!} \right) \prod_{i=1}^{N} \frac{[\gamma(k_i)]^{k_i}}{k_i!} + o(N)
\]

where \( \gamma(k) \) is a function that satisfies

\[
\gamma(k) = \frac{1}{\langle k \rangle} \sum_{k'} Q(k, k') p(k') \frac{k'}{\gamma(k')}. \quad (20)
\]

The probability \( \mathbb{P}(G) \) can be also written in Hamiltonian form as

\[
\mathbb{P}(G) = e^{-H(G)} \delta \left( L - \sum_{i,j} a_{ij} \right) \theta \left( K - \max_{i=1,...,N} k_i \right) \theta \left( m - \min_{i=1,...,N} k_i \right)
\]

where the (exchangeable) Hamiltonian \( H(G) \) is given by

\[
H(G) = - \sum_{i,j} a_{ij} \ln Q(k_i, k_j) - \sum_{i=1}^{N} \ln p(k_i) + \Sigma(k), \quad (21)
\]

where \( \Sigma(k) \) is given by Eq. (19). By studying this correlated network ensemble with statistical mechanics methods similar to the ones we have used in the case of the exchangeable uncorrelated network ensemble investigated earlier, we can show (see SM for details) that the degree distribution is \( p(k) \) and that the marginal probability of each link can be written as

\[
p_{ij} = \sum_{m \leq k \leq k, m \leq k'} p(k)p(k') = \frac{\langle k \rangle}{N}, \quad (22)
\]

with \( p(k, k') \) indicating the probability of a link between a node of degree \( k \) and a node of degree \( k' \), with

\[
p_{ij=k,k=k'} = p(k, k') = \frac{1}{\langle k \rangle N} Q(k, k') \frac{k k'}{\gamma(k) \gamma(k')} \quad (23)
\]

Note that for \( Q(k, k') = 1 \) it follows immediately from Eq. (20) that \( \gamma(k) = 1 \) and we recover the uncorrelated network ensemble. Eq. (22) and Eq. (23) clearly reveal the exchangeable nature of this ensemble as the marginal probability of a link is independent of the node labels. However the ensemble retains the ability to model sparse networks with arbitrary degree distribution and degree correlations.

**Conclusions** In conclusions, in this work we propose a statistical mechanics framework able to define sparse exchangeable network ensembles of a given number of nodes \( N \). Here by sparse we mean that the networks have a structural cutoff. This hypothesis is necessary for fully treating the model analytically but it can be removed as long as the entropy \( \Sigma(k) \) is known and numerically estimated for every possible degree sequence \( k \) of the network. The network ensemble can be generated by a simple Metropolis-Hastings algorithm. This statistical mechanics approach is based on enforcing two global constraints, such as the total number of links and the value of the exchangeable Hamiltonian of the ensemble. Despite every link has the same marginal probability the ensemble can generate networks with very heterogeneous degree distribution. This implies that in order to impose an heterogeneous degree distribution we do not need as for the exponential random graphs to impose an extensive number of local constraints but two global constraints are actually sufficient. This approach is here shown to be generalizable to networks with degree correlations, to directed and bipartite networks and to generalized network structures such as multilayer networks and simplicial complexes. This work provides a physical point of view for addressing the challenging problem of modelling exchangeable networks which has applications in a number of fields ranging from data analysis, to sampling of networks and has profound ramifications in mathematics. Therefore we hope that this work will stimulate further theoretical and applied research at the frontier between these disciplines.
SUPPLEMENTARY MATERIAL

EXCHANGEABLE ENSEMBLE OF SPARSE SIMPLE NETWORKS

Treatment of the exchangeable ensemble of uncorrelated networks

In this section our goal is to solve the partition function \( Z(h) \) (that for construction is expected for \( h = 0 \) to take the value \( Z(0) = 1 \)) for the exchangeable ensemble of simple networks using the saddle point equation deriving the expression of the functional order parameter \( c(\omega, k) \). The use start by recalling the expression given in the main text for the partition function \( Z(h) \) of this network ensemble,

\[
Z(h) = \sum_{G} \mathcal{F}(G) e^{-h \sum_{i} a^i} = \frac{1}{(2L)!^L} \sum_{\omega} \int \mathcal{D} \omega \int \frac{d\lambda}{2\pi} e^{G(\lambda, \omega, h)}, \tag{S-1}
\]

with

\[
G(\lambda, \omega, k, h) = \sum_{i=1}^{N} [i\omega k + \ln(p(k))] + i\lambda L + \frac{1}{2} \sum_{i,j} \ln(1 + e^{-i\omega k - i\omega i - k - h}),
\]

and with \( \mathcal{D} \omega = \prod_{i=1}^{N} d(\omega)/(2\pi) \). In Eq. (S-1) and in the following we use the notation \( \sum_{\lambda} \) to indicate the sum over all the possible values of the degree of each node \( i \) satisfying \( m \leq k_i \leq K \ll K = \sqrt{\langle k \rangle N} \). Let us now introduce the functional order parameter \( 37, 40, 41 \)

\[
c(\omega, k) = \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i) \delta(k, k_i), \tag{S-2}
\]

by enforcing its definition with a series of delta functions. Therefore, by assuming a discretization in \( \omega \) in intervals of size \( \Delta \omega \) we introduce for any value of \( (\omega, k) \) the term

\[
1 = \int dc(\omega, k) \delta \left( c(\omega, k) - \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i) \delta(k, k_i) \right) = \int d\hat{c}(\omega, k) d\hat{c}(\omega, k) \frac{e^{-2\pi i(\omega \hat{c}(\omega, k))}}{2\pi/(N\Delta\omega)} \exp \left[ i\Delta\omega \hat{c}(\omega, k) [Nc(\omega, k) - \sum_{i=1}^{N} \delta(\omega - \omega_i) \delta(k, k_i)] \right].
\]

After performing these operations, by imposing \( 2L = \langle k \rangle N \) where \( \langle k \rangle = \sum_{k} kp(k) \), the partition function reads in the limit \( \Delta \omega \to 0 \),

\[
Z(h) = \frac{1}{(2L)!^L} \sum_{\omega} \int \mathcal{D} c(\omega, k) \int \mathcal{D} \hat{c}(\omega, k) \int d\lambda \frac{d\lambda}{2\pi} e^{Nf(c(\omega, k), \hat{c}(\omega, k), h)} \tag{S-3}
\]

with

\[
f(\lambda, c(\omega, k), \hat{c}(\omega, k), h) = i \int d\omega \sum_{m \leq k \leq K} (\hat{c}(\omega, k)c(\omega, k) + i\lambda k/2 + \Psi + \ln \int d\omega \sum_{m \leq k \leq K} p(k) k! e^{i\omega k - i\omega k'}) \tag{S-4}
\]

where \( \Psi \) is given by

\[
\Psi = \frac{N}{2} \int d\omega \int d\omega' \sum_{m \leq k \leq K} c(\omega, k)c(\omega', k') \ln \left( 1 + e^{-i\omega - i\omega' - i\omega k} \right)
\]

and where \( \mathcal{D} c(\omega, k) \) is the functional measure \( \mathcal{D} c(\omega, k) = \lim_{\Delta \omega \to 0} \prod_{\omega} \prod_{k} [dc(\omega, k) \sqrt{N\Delta \omega}/(2\pi)] \) and similarly \( \mathcal{D} \hat{c}(\omega, k) = \lim_{\Delta \omega \to 0} \prod_{\omega} \prod_{k} [d\hat{c}(\omega, k) \sqrt{N\Delta \omega}/(2\pi)] \). Performing a Wick rotation in \( \lambda \) and assuming \( z/N = e^{i\nu} \) real and much smaller than one, i.e. \( z/N \ll 1 \) which is allowed in the sparse regime \( K \ll K_s \), we can linearize the logarithm and express \( \Psi \) as

\[
\Psi = \frac{1}{2} zv^2 e^h, \tag{S-5}
\]

with

\[
v = \int d\omega \sum_{m \leq k \leq K} c(\omega, k)e^{-i\omega}. \tag{S-6}
\]
The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of $f(\lambda, c(\omega, k), \hat{c}(\omega, k), h)$ with respect to $c(\omega, k), \hat{c}(\omega, k)$ and $\lambda$, obtaining for $h \to 0$,

$$-i\hat{c}(\omega, k) = zve^{-i\omega},$$

$$c(\omega, k) = \int \frac{d\omega}{2\pi} \sum_{m \leq K} p(k)ke^{i\omega}e^{-i\omega/2}e^{-i\omega/2},$$

$$zv^2 = \langle k \rangle.$$  \hspace{1cm} (S-7)

Let us first calculate the integral

$$\int \frac{d\omega}{2\pi} \sum_{m \leq K} p(k)ke^{i\omega}e^{-i\omega/2} = \int \frac{d\omega}{2\pi} \sum_{m \leq K} k!p(k)ke^{i\omega}e^{-i\omega/2}$$

where we have substituted the saddle point expression for $\hat{c}(\omega, k)$. This integral can be also written as

$$\int \frac{d\omega}{2\pi} \sum_{m \leq K} k!p(k)ke^{i\omega}e^{-i\omega/2} = \sum_{m \leq K} p(k)(zv)^k = \langle (zv)^k \rangle.$$  \hspace{1cm} (S-8)

Therefore $c(\omega, k)$ at the saddle point solution can be expressed as

$$c(\omega, k) = \frac{1}{2\pi} k!p(k)ke^{i\omega}e^{-i\omega/2}$$

With this expression, using a similar procedure we can express $\nu$ as

$$\nu = \int d\omega \sum_{m \leq K} c(\omega, k)e^{-i\omega} = \frac{1}{\langle (zv)^k \rangle} \sum_{m \leq K} kp(k)(zv)^{k-1}$$

Combing this equation with the third saddle point equation

$$zv^2 = \langle k \rangle,$$

it is immediate to show that $zv = 1$ is a solution with

$$z = \frac{1}{\langle k \rangle}, \quad \nu = \langle k \rangle.$$  \hspace{1cm} (S-9)

By inserting this expression in Eq. (S-10) we get Eq. (16), i.e.

$$c(\omega, k) = \frac{1}{2\pi} k!p(k)e^{i\omega}e^{-i\omega/2}.$$  \hspace{1cm} (S-10)

Calculating the partition function at the saddle point, we get $Z(h \to 0) = 1$.

**Calculation of the marginal probability of a link**

For calculating the marginal distribution $p_{ij}$ of a link between node $i$ and node $j$ in the exchangeable network ensemble we first note that given that the ensemble has an exchangeable Hamiltonian, the marginal probability of a link should be the same for every link of the network, i.e. $p_{ij} = \hat{p}$. In order to obtain $\hat{p}$ we can simply derive the free energy $F = Nf$ with $f$ given by Eq. (S-4) with respect to the auxiliary field $h$ obtaining

$$\frac{N(N-1)}{2} \hat{p} = - \frac{\partial(Nf)}{\partial h} \bigg|_{h=0} = - \frac{\partial(N\Psi)}{\partial h} \bigg|_{h=0} = \frac{N}{2} z \int d\omega \sum_{m \leq K} c(\omega, k)c(\omega', k')e^{-i\omega-i\omega'}$$

from which, inserting the saddle point value of $c(\omega, k)$ and $z$ we get for $N \gg 1$,

$$p_{ij} = \hat{p} = \sum_{m \leq K} \sum_{m \leq K} p(k)p(k') \frac{kk'}{(k)N} = \frac{\langle k \rangle}{N}.$$  \hspace{1cm} (S-11)
Expression of the marginal probability of a link conditioned on the degrees of its two endnodes

In this paragraph our goal is to derive the expression of the probability \( p_{ij; k, k'} = p(k, k') \) of a link between node \( i \) and node \( j \) in the exchangeable network ensemble conditioned on the degree of the two endnodes. The expression for \( p_{ij; k, k'} \) can be obtained by showing that the probability \( \pi_{ij} \) that node \( i \) is connected to node \( j \) in any network ensemble enforcing a given degree sequence \( k \) (the configuration model) is given by

\[
\pi_{ij} = \sum_{a} a_{ij} N \prod_{r=1}^{N} \delta \left( k_r - \sum_{s=1}^{N} a_{rs} \right) \delta \left( L - \sum_{r<s} a_{rs} \right) e^{-\Sigma(k)} = \frac{k_i k_j}{\langle k \rangle N} \quad (S-17)
\]

as long as the maximum degree of the network \( K \) is much smaller than the structural cutoff, i.e. \( K \ll K_S \). Since \( \pi_{ij} \) only depends on the degrees \( k_i \) and \( k_j \) of its two endnodes, in this ensemble the probability \( \pi(k, k') \) of any link between any two nodes of degree \( k \) and degree \( k' \), takes the expression,

\[
\pi_{ij; k, k'} = \pi(k, k') = \frac{k k'}{(k) N} \quad (S-18)
\]

as long as \( K \ll K_S \). The exchangeable network model is essentially an ensemble in which we can get very different degree distributions but each network \( G \) with a given distribution \( k \) is weighted by \( P(k) \exp[-\Sigma(k)] \). Therefore, the we can express \( p_{ij; k, k'} \) as

\[
p_{ij; k, k'} = p(k, k') = \frac{\sum_{h_{i,k} = k, h_{j,k} = k'} \prod_{r=1}^{N} p(k_r) \pi(k_i, k_j)}{p(k) p(k')} = \pi(k, k') = \frac{k k'}{(k) N} \quad (S-19)
\]

Let us now derive Eq. (S-18) for the ensemble in which we fix the degree sequence of the network (for the next examples of ensemble the derivation is similar and we will omit for space constraints). To this end we consider the partition function

\[
\tilde{Z}(h) = \sum_{a} \exp \left[ -\sum_{i<j} h_{i,k} a_{ij} \right] \prod_{r=1}^{N} \delta \left( k_r - \sum_{s=1}^{N} a_{rs} \right) \delta \left( L - \sum_{r<s} a_{rs} \right) e^{-\Sigma(k)}
\]

\[
\Sigma(k) = \ln \left( \frac{(2L)!^{N!}}{\prod_{k=1}^{N} k!} \right) + o(N) \quad (S-20)
\]

where in Eq. (S-20) and (S-21) we indicate with \( k_i \) the degree of node \( i \) given by \( k_i = \sum_{j=1}^{N} a_{ij} \). Expressing the Kronecker delta in Eq. (S-20) in integral form we get for the partition function \( \tilde{Z}(h) \) of this network ensemble,

\[
\tilde{Z}(h) = \frac{1}{(2L)!^{N!}} \sum_{a} \int d\omega \int \frac{d\lambda}{2\pi} e^{\tilde{G}(\lambda, \omega, k, h)},
\]

with

\[
\tilde{G}(\lambda, \omega, k, h) = \sum_{i=1}^{N} \left[ i \omega_i k_i + \ln(k_i! \right)] + i \omega L + \frac{1}{2} \sum_{ij} \ln \left( 1 + e^{-i\lambda - i \omega_i - i \omega_j - h_{ij}} \right),
\]

and with \( D\omega = \prod_{i=1}^{N} d\omega_i/(2\pi) \). In Eq. (S-22). By indicating with \( N_k \) the fraction of nodes with degree \( k \), let us introduce the functional order parameters \( c_k(\omega) \)

\[
c_k(\omega) = \frac{1}{N_k} \sum_{i=1}^{N} \delta(\omega - \omega_i) \delta(k, k_i), \quad (S-23)
\]

determining the fraction of nodes of degree \( k \) that are associated to \( \omega_i = \omega \). Therefore, by assuming a discretization in \( \omega \) in intervals of size \( \Delta \omega \) we introduce for any value of \( \omega \) and \( k \) the term

\[
1 = \int dc_k(\omega) \delta \left( c_k(\omega) - \frac{1}{N_k} \sum_{i=1}^{N} \delta(\omega - \omega_i) \delta(k, k_i) \right) = \frac{\int dc_k(\omega) dc_k(\omega)}{2\pi/(N_k \Delta \omega)} \exp \left[ i \Delta \omega \delta_k(\omega) (N_k c_k(\omega) - \sum_{i=1}^{N} \delta(\omega - \omega_i) \delta(k, k_i)) \right].
\]
After performing these operations, by imposing $2L = \langle k \rangle N$ where $\langle k \rangle = \sum_k kp(k)$, the partition function reads in the limit $\Delta \omega \to 0$,

$$
\tilde{Z}(h) = \frac{1}{(2L)!!} \int \mathcal{D} \xi_0(\omega) \int \mathcal{D} \xi_\lambda(\omega) \int \mathcal{D} \xi_\nu(\omega, k, h) e^{i \tilde{f}(\lambda, c_k(\omega), \tilde{c}_k(\omega), k, h)}
$$

(S-24)

with

$$
\tilde{f}(\lambda, c_k(\omega), \tilde{c}_k(\omega), k, h) = i \int d\omega \sum_{m \leq K} \tilde{P}(k) \tilde{c}_k^\dagger(\omega) c_k(\omega) + i\lambda(k)/2 + \Psi + \sum_{m \leq K} \tilde{P}(k) \ln \int \frac{d\omega}{2\pi} k! e^{iak - i\tilde{\omega}_k^\dagger(\omega)}
$$

(S-25)

where we have indicated with $\tilde{P}(k) = N_k/N$ and where $\Psi$ is given by

$$
\Psi = \frac{N}{2} \sum_{m \leq K, m \leq K} \tilde{P}(k) \tilde{P}(k') \int d\omega \int d\omega' c_k(\omega)c_k(\omega') \ln \left(1 + e^{-i\omega - i\omega' - h_k \nu} \right)
$$

and where $\mathcal{D} \xi_\nu(\omega)$ is the functional measure $\mathcal{D} \xi_\nu(\omega) = \lim_{\Delta \omega \to 0} \prod_\omega \int d\xi_\nu(\omega) \sqrt{N_\omega} \Delta \omega/(2\pi)$ and similarly $\mathcal{D} \xi_\lambda(\omega) = \lim_{\Delta \omega \to 0} \prod_\omega \int d\xi_\lambda(\omega) \sqrt{N_\omega} \Delta \omega/(2\pi)$. Performing a Wick rotation in $\lambda$ and assuming $\zeta/N = e^{-i\lambda}$ real and much smaller than one, i.e. $\zeta/N \ll 1$ which is allowed in the sparse regime $K \ll K_T$, we can linearize the logarithm and express $\Psi$ as

$$
\Psi = \frac{1}{2} \sum_{m \leq K} \tilde{P}(k) \tilde{P}(k') \int d\omega \int d\omega' c_k(\omega)c_k(\omega') e^{-h_k \nu},
$$

(S-26)

with

$$
\nu = \sum_{m \leq K} \tilde{P}(k) \int d\omega c_k(\omega) e^{-i\omega}.
$$

(S-27)

For later convenience let us also define $\nu$ as

$$
\nu = \sum_{m \leq K} \tilde{P}(k) \int d\omega c_k(\omega) e^{-i\omega}.
$$

(S-28)

The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of $f(\lambda, c_k(\omega), \tilde{c}_k(\omega), k, h)$ with respect to $c_k(\omega)$, $\tilde{c}_k(\omega)$ and $\lambda$, obtaining for $h_k \nu \to 0$,

$$
-i\tilde{c}_k(\omega) = \nu e^{i\omega},
$$

(S-29)

$$
c_k(\omega) = \frac{1}{2\pi} k! e^{iak - i\tilde{\omega}_k(\omega)},
$$

(S-30)

$$
\nu^2 = \langle k \rangle.
$$

(S-31)

Let us first calculate the integral

$$
\int \frac{d\omega}{2\pi} k! e^{-iak - i\tilde{\omega}_k(\omega)} = \int \frac{d\omega}{2\pi} k! e^{iak + \nu^2 e^{i\omega}}
$$

(S-32)

where we have substituted the saddle point expression for $\tilde{c}_k(\omega)$. This integral can be also written as

$$
\int \frac{d\omega}{2\pi} k! \sum_{n=0}^{\infty} \frac{(\nu^2)^n}{n!} e^{-i\omega n} = (\nu^2)^k \langle k \rangle.
$$

(S-33)

Therefore $c_k(\omega)$ at the saddle point solution can be expressed as

$$
c(\omega, k) = \frac{1}{2\pi} k! e^{iak + \nu^2 e^{i\omega}}
$$

(S-34)

With this expression, using a similar procedure we can express $\nu$ as

$$
\nu = \int \frac{d\omega}{2\pi} \sum_{m \leq K} \tilde{P}(k)c_k(\omega)e^{-i\omega} = \sum_{m \leq K} \frac{k \tilde{P}(k)}{\nu} = \frac{\langle k \rangle}{\nu}
$$

(S-35)

Therefore this equation reduces to the third saddle point equation

$$
\nu^2 = \langle k \rangle.
$$

(S-36)
it is immediate to show that \( z \nu = 1 \) is a solution with
\[
z = \frac{1}{\langle k \rangle}, \quad \nu = \langle k \rangle. \tag{S-35}\]

By inserting this expression in Eq. \( \text{S-32} \) we get Eq. \( \text{S-16} \), i.e.
\[
c_k(\omega) = \frac{1}{2\pi} k! e^{i\omega k} e^{-\nu}. \tag{S-36}\]
The marginal probability \( \pi(k, k') \) of a link between a node of degree \( k \) and a node of degree \( k' \) can be expressed as
\[
\pi(k, k') = \frac{\partial N f}{\partial \nabla_{k', \xi}} \bigg|_{h=0} \tag{S-37}
\]
leading to
\[
\pi(k, k') = \frac{z}{N} \int \! d\omega \int \! d\omega' c_k(\omega)c_{k'}(\omega') e^{-i\omega k - i\omega' k'} \tag{S-38}
\]
It follows that
\[
p(k, k') = \pi(k, k') = \frac{kk'}{\langle k \rangle \langle k' \rangle}. \tag{S-39}
\]

**EXCHANGEABLE ENSEMBLE OF SPARSE SIMPLE NETWORKS WITH DEGREE CORRELATIONS**

Treatment of the exchangeable ensemble of sparse correlated simple networks

In this section our goal is to treat the exchangeable ensemble of sparse correlated networks in which each node has degree \( k \) with probability \( p(k) \) and each link between a node of degree \( k \) and a node of degree \( k' \) contributes to the partition function by a term \( Q(k, k') = Q(k', k) \). Here we impose that the total number of links \( L = \langle k \rangle N / 2 \) and that the maximum degree of the network is below or equal to \( K \) and the minimum degree of the network is greater or equal to \( m \). For simplicity of notation we take the auxiliary field \( h = 0 \) from the beginning and we express the partition function \( Z \) of the exchangeable ensemble of sparse correlated networks as
\[
Z = \sum_a \sum_k \prod_{i<j} Q(k_i, k_j)^{a_{ij}} e^{-\frac{N}{\langle k \rangle}} \prod_{i=1}^N \delta \left( k_i, \sum_{j=1}^N a_{ij} \right) \delta \left( L, \sum_{i<j} a_{ij} \right), \tag{S-40}
\]
with the entropy \( \Sigma(k) \) given by
\[
\Sigma(k) = \ln \left( (2L)! \prod_{i=1}^N \frac{[\gamma(k)]^k}{k_i!} \right) + o(N) \tag{S-41}
\]
where \( \gamma(k) \) is determined by the self-consistent equation
\[
\gamma(k) = \frac{1}{\langle k \rangle} \sum_{m \leq k \leq K} Q(k, k') p(k') \frac{k'}{\gamma(k')} \tag{S-42}
\]
By expressing the Kronecker deltas in integral form
\[
\delta(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \! d\omega e^{i\omega(x-y)} \tag{S-43}
\]
we get
\[
Z = \frac{1}{(2L)!} \sum_k \int \! D\omega \int \! \frac{d\lambda}{2\pi} e^{G(\omega, \lambda, k)} \tag{S-44}
\]
where \( G(\omega, \lambda, k) \) is given by
\[
G(\omega, \lambda, k) = \sum_{i=1}^N \ln[\omega_i k_i + \ln(k_i! p(k_i)) - k_i \ln \gamma(k_i)] + i \lambda L + \frac{1}{2} \sum_{i<j} \ln \left( 1 + Q(k_i, k_j) e^{-i\omega_i - i\omega_j} \right). \tag{S-45}
\]
and where $D\omega = \prod_{i=1}^{N} [d\omega_i/(2\pi)]$. Let us now introduce the functional order parameter [37, 40, 41]

\[ c(\omega, k) = \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i)\delta(k, k_i), \]  

(S-46)

by enforcing its definition with a series of delta functions. Therefore, by assuming a discretization in $\omega$ in intervals of size $\Delta\omega$ we introduce for every $(\omega, k)$ the term

\[ 1 = \int dc(\omega, k) \delta \left( c(\omega, k) - \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i)\delta(k, k_i) \right) = \int dc(\omega, k) dc(\omega, k) \left| \frac{\Delta \omega}{2\pi} \right| \delta(\omega - \omega_i)\delta(k, k_i), \]  

(S-47)

After performing these operations, by imposing $2L = \langle k \rangle N$ where $\langle k \rangle = \sum_k kp(k)$, the partition function reads in the limit $\Delta\omega \to 0$

\[ Z = \frac{1}{(2L)!!} \sum_{k} \int Dc(\omega, k) \int D\hat{c}(\omega, k) \int \frac{d\lambda}{2\pi} e^{\lambda f(\lambda, c(\omega, k), \hat{c}(\omega, k))} \]  

(S-48)

with

\[ f(\lambda, c(\omega, k), \hat{c}(\omega, k)) = i \int d\omega \sum_{m \leq k} \hat{c}(\omega, k)c(\omega, k) + i\lambda(k)/2 + \Psi \ln \int d\omega \sum_{m \leq k} p(k) \frac{k!}{\gamma(k)!} e^{i\lambda k - i\lambda c(\omega, k)} \]  

(S-49)

where $\Psi$ is given by

\[ \Psi = \frac{N}{2} \int d\omega \int d\omega' \sum_{m \leq k, m' \leq k} c(\omega, k)c(\omega', k')Q(k, k') \ln \left( 1 + e^{-i\lambda c(\omega, k') - i\lambda c(\omega', k')} \right), \]  

(S-50)

where $Dc(\omega, k)$ and $D\hat{c}(\omega, k)$ have the same definition then in the simple uncorrelated case. Performing a Wick rotation in $\lambda$ and assuming $z/N = e^{-1/2}$ real and much smaller than one, i.e. $z/N \ll 1$ which is allowed in the sparse regime $K \ll K_s$, we can linearize the logarithm and express $\Psi$ as

\[ \Psi = \frac{z}{2} \int d\omega \int d\omega' \sum_{m \leq k, m' \leq k} c(\omega, k)c(\omega', k')Q(k, k')e^{-i\lambda c(\omega, k')} , \]  

(S-51)

The saddle point equations determining the value of the partition function read

\[ -i\dot{c}(\omega, k) = ze^{i\omega} \int d\omega' \sum_{m \leq k' \leq k} Q(k, k')c(\omega', k')e^{-i\lambda c'} \]  

\[ c(\omega, k) = \frac{1}{z} \int d\omega' \sum_{m \leq k' \leq k} p(k') \frac{k!}{\gamma(k')!} e^{i\lambda k' - i\lambda c(\omega, k')} \]  

\[ z \int d\omega \int d\omega' \sum_{m \leq k, m' \leq k} c(\omega, k)c(\omega', k')Q(k, k')e^{-i\lambda c(\omega, k')} = \langle k \rangle \]  

(S-52)

Let us define $\tilde{\gamma}(k)$ as

\[ \tilde{\gamma}(k) = z \int d\omega' \sum_{m \leq k \leq k} Q(k, k')c(\omega', k')e^{-i\lambda c'}. \]  

(S-53)

With this definition we have

\[ -i\dot{c}(\omega, k) = \tilde{\gamma}(k)e^{i\omega} \]  

(S-54)

Let us first calculate the integral

\[ \frac{1}{2\pi} \int d\omega \sum_{m \leq k \leq k} p(k) \frac{k!}{\gamma(k)!} e^{i\lambda k - i\lambda c(\omega, k)} = \frac{1}{2\pi} \int d\omega \sum_{m \leq k \leq k} \frac{k!}{\gamma(k)!} p(k)e^{i\lambda k + \tilde{\gamma}(k)c(\omega, k)} \]  

(S-55)

where we have substitute the saddle point expression for $\hat{c}(\omega, k)$. This integral can be also written as

\[ \int d\omega \sum_{m \leq k \leq k} \frac{k!}{\gamma(k)!} p(k)e^{i\lambda k + \tilde{\gamma}(k)c(\omega, k)} = \sum_{m \leq k \leq k} p(k) \left( \tilde{\gamma}(k) / \gamma(k) \right)^k. \]  

(S-56)
Let $w$ indicate the value of this integral, i.e.

$$w = \sum_{m \leq k \leq K} p(k) \frac{\tilde{\gamma}(k)}{\gamma(k)}.$$  \hspace{1cm} (S-57)

The functional order parameter $c(\omega, k)$ at the saddle point solution can be expressed as

$$c(\omega, k) = \frac{1}{2\pi w} \frac{k! p(k)}{[\gamma(k)]^k} e^{-i\omega \tilde{\gamma(k)}}.$$  \hspace{1cm} (S-58)

With this expression, using a similar procedure we can express $\tilde{\gamma}(k)$ as

$$\tilde{\gamma}(k) = z \int d\omega' \sum_{m \leq k \leq K} Q(k, k') c(\omega', k') e^{-i\omega'} = \frac{z}{w} \sum_{m \leq k \leq K} Q(k, k') p(k') \frac{k'}{\tilde{\gamma}(k')} \left(\frac{\tilde{\gamma}(k')}{\gamma(k')}\right)^k.$$  \hspace{1cm} (S-59)

Combining this equation with the third saddle point equation we get

$$z \int d\omega \int d\omega' \sum_{m \leq k \leq K} \sum_{m' \leq k' \leq K} Q(k, k') c(\omega, k) c(\omega', k') e^{-i\omega - i\omega'} = \frac{1}{\pi} \sum_{m \leq k \leq K} p(k) k \left(\frac{\tilde{\gamma}(k)}{\gamma(k)}\right)^k = \langle k \rangle.$$  \hspace{1cm} (S-60)

Given that $\gamma(k)$ is defined though the Eq. (S-42), it follows that

$$\tilde{\gamma}(k) = \gamma(k), \quad w = 1, \quad z = \frac{1}{\langle k \rangle}.$$  \hspace{1cm} (S-61)

Finally using Eqs. (S-194) we can derive the final expression for $c(\omega, k)$ given by

$$c(\omega, k) = \frac{1}{2\pi} \frac{k! p(k)}{[\gamma(k)]^k} e^{i\omega k + \gamma(k) e^{-i\omega}}.$$  \hspace{1cm} (S-62)

From this equation of the functional order parameter we can derive the marginal for each link of the network which is given by

$$p_{ij} = \frac{1}{N} \int d\omega \int d\omega' \sum_{m \leq k \leq K, m' \leq K} c(\omega, k) c(\omega', k') Q(k, k') e^{-i\omega - i\omega'},$$  \hspace{1cm} (S-63)
yielding,

$$p_{ij} = \sum_{m \leq k \leq K, m' \leq K} p(k) p(k') p(k, k').$$  \hspace{1cm} (S-64)

Here $p(k, k')$ indicates the probability of a link between node $i$ and node $j$ conditioned to the degree of the two nodes $k_i = k$ and $k_j = k'$, i.e.

$$p_{ij | k_i = k, k_j = k'} = p(k, k') = \frac{1}{\langle k \rangle N} Q(k, k') \frac{kk'}{\gamma(k) \gamma(k')}.$$  \hspace{1cm} (S-65)

Note that for $Q(k, k') = 1$ it follows that $\gamma(k) = 1$, and for $Q(k, k') = kk'$ it follows $\gamma(k) = k$ and hence in both cases we recover the exchangeable network ensemble of simple uncorrelated networks.

EXCHANGEABLE ENSEMBLE OF SPARSE DIRECTED NETWORKS

Directed exchangeable sparse network ensembles

In a directed network the adjacency matrix is not symmetric and for each node we can distinguish between in-degree and the out-degree. Assuming that $a_{ij} = 1$ indicates the presence of a link from node $i$ to node $j$, the in-degree and the out-degree of node $i$ can be expressed as

$$k_i^{\text{in}} = \sum_{j=1}^{N} a_{ji},$$
$$k_i^{\text{out}} = \sum_{j=1}^{N} a_{ij}.$$  \hspace{1cm} (S-66)
We assume that the in-degree and the out-degree have a maximum value equal or smaller than $K$ with $K \ll K_s = \sqrt{\langle k^m \rangle N}$ and that they a minimum value equal or larger than $m$. Here we define the exchangeable uncorrelated ensemble of directed networks with join degree distribution $p(k^m, k^{out})$ indicating the probability that a node has degree $k_i^m = k^m$ and $k_i^{out} = k^{out}$. This distribution is arbitrary, but needs to satisfy $\langle k^m \rangle = \langle k^{out} \rangle$. The exchangeable uncorrelated ensemble of directed networks assigns to each directed network $G$ the probability $\mathbb{P}(G)$ given by

$$\mathbb{P}(G) = \prod_{i=1}^{N} p(k_{i}^{in}, k_{i}^{out}) e^{-\Sigma(k_{i}^{in}, k_{i}^{out})} \sum_{a} \prod_{i,j} \delta \left( L_{i,j} \sum_{a_{ij}} a_{ij} \theta \left( K - \max_{\ell=1...N} k_{\ell}^{in} \right) \theta \left( K - \max_{\ell=1...N} k_{\ell}^{out} \right) \theta \left( \min_{\ell=1...N} k_{\ell}^{in} - m \right) \theta \left( \min_{\ell=1...N} k_{\ell}^{out} - m \right) \right)$$

(S-67)

where the entropy $\Sigma(k^{in}, k^{out})$ is given by

$$\Sigma(k) = \ln \left[ \frac{L!}{\prod_{i=1}^{N} [k_{i}^{in}, k_{i}^{out}]!} \right] + o(N).$$

(S-68)

The probability $\mathbb{P}(G)$ admits an Hamiltonian expression as

$$\mathbb{P}(G) = e^{-H(G)} \delta \left( L_{i,j} \sum_{a_{ij}} a_{ij} \theta \left( K - \max_{\ell=1...N} k_{\ell}^{in} \right) \theta \left( K - \max_{\ell=1...N} k_{\ell}^{out} \right) \theta \left( \min_{\ell=1...N} k_{\ell}^{in} - m \right) \theta \left( \min_{\ell=1...N} k_{\ell}^{out} - m \right) \right),$$

(S-69)

where the Hamiltonian $H(G)$ of this ensemble is given by

$$H(G) = -\sum_{i=1}^{N} \ln \left( p(k_{i}^{in}, k_{i}^{out}) k_{i}^{in} k_{i}^{out} \right) + \ln(L!).$$

(S-70)

In the following paragraph we will treat the statistical mechanics of this model showing that the density of nodes with in-degree $k^{in}$ and out-degree $k^{out}$ is given by the desired joint distribution $p(k^{in}, k^{out})$ although the marginal probability of each node is equal for each node and given by

$$p_{ij} = \sum_{k_{i}^{in}, k_{i}^{out}} p_{in}(k_{i}^{in}) p_{out}(k_{i}^{out}) \frac{k_{i}^{in} k_{i}^{out}}{\langle k^{in} \rangle N},$$

(S-71)

where

$$p_{in}(k_{i}^{in}) = \sum_{k_{i}^{out}} p(k_{i}^{in}, k_{i}^{out}),$$

$$p_{out}(k_{i}^{out}) = \sum_{k_{i}^{in}} p(k_{i}^{in}, k_{i}^{out}).$$

(S-72)

Note that although the marginal probability is the same for each node the marginal probability of a link conditioned on the degree of its two end node is not, i.e.

$$P_{i,j|k^{in}=k^{out}} = \frac{k_{i}^{in} k_{j}^{out}}{\langle k^{in} \rangle N}.$$

(S-73)

Derivation of the marginal probability

In this section our goal is the solve the partition function $Z$ for the exchangeable ensemble of directed networks using the saddle point equation the expression for the marginal probability of a link. For simplicity for this ensemble we put the auxiliary fields $h = 0$ from the beginning and we express the partition function $Z$ as

$$Z = \sum_{a} \sum_{k^{in}} \sum_{k^{out}} e^{-H(G)} \delta \left( L_{i,j} \sum_{a_{ij}} a_{ij} \prod_{i=1}^{N} \left[ \delta \left( k_{i}^{in} - \sum_{j=1}^{N} a_{ji} \right) \delta \left( k_{i}^{out} - \sum_{j=1}^{N} a_{ij} \right) \right] \right).$$

(S-74)

By expressing the Kronecker deltas in integral form

$$\delta(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} du e^{iu(x-y)}$$

(S-75)
we get

$$Z = \frac{1}{L!} \sum \int \sum \int \int D\omega \int D\hat{\omega} \int \frac{d\lambda}{2\pi} e^{iG(\lambda, \omega, \hat{\omega}, k, k')},$$  \hspace{1cm} (S-76)$$

with

$$G(\lambda, \omega, \hat{\omega}, k, k') = \sum_{i=1}^{N} \left[ i\omega \delta(k_{in} - k_{out}) + i\hat{\omega} \delta(k_{out} - k_{in}) + \ln(k_{in}^{\prime} k_{out}^{\prime} \{p(k_{in}', k_{out}')\}) + i\lambda L + \sum_{i,j} \ln(1 + e^{-ik_{in} \omega^{i} - i\hat{\omega}^{j}}) \right],$$  \hspace{1cm} (S-77)$$

and with \( D\omega = \prod_{i=1}^{N}[d\omega_{i}/(2\pi)] \) and \( D\hat{\omega} = \prod_{i=1}^{N}[d\hat{\omega}_{i}/(2\pi)] \). Let us now introduce the functional order parameter \( c(\hat{\omega}, \omega, k, k') \) by enforcing its definition with a series of delta functions by introducing the conjugated order parameter \( \hat{c}(\omega, \hat{\omega}, k, k') \) and by imposing \( L = (k^{in})N = (k^{out})N \) where \( (k^{in}) = \sum_{k^{in}} k^{in} p(k^{in}, k^{out}) \) and \( (k^{out}) = \sum_{k^{out}} k^{out} p(k^{in}, k^{out}) \), the partition function reads

$$Z = \frac{1}{L!} \sum \int \int \int Dc(\omega, \hat{\omega}, k, k') \int D\hat{c}(\omega, \hat{\omega}, k, k') \int D\omega \int D\hat{\omega} \int \frac{d\lambda}{2\pi} e^{i\lambda f}$$  \hspace{1cm} (S-79)$$

with

$$f = i \int d\omega \int d\hat{\omega} \sum_{m^{in}, k, m^{out}, K} \hat{c}(\omega, \hat{\omega}, k, k') c(\omega, \hat{\omega}, k, k') + i\lambda (k^{in}) + \Psi \ln \int d\omega \sum_{m^{in}, k, m^{out}, K} p(k^{in}, k^{out})! \exp \left[ i\omega k^{in} + i\hat{\omega} k^{out} - i\hat{c}(\omega, \hat{\omega}, k, k') \right]$$  \hspace{1cm} (S-80)$$

where \( \Psi \) is given by

$$\Psi = \frac{N}{2} \int d\omega \int d\omega' \sum_{m^{in}, k, m^{out}, K} \sum_{m^{out}, k, m^{out}, K} c(\omega, \hat{\omega}, k, k') c(\omega', \hat{\omega}', k^{in}, k^{out}) \ln \left( 1 + e^{-ik_{in} \omega^{i} - i\hat{\omega}^{j}} \right)$$

and where \( Dc(\omega, \hat{\omega}, k, k') \) and \( D\hat{c}(\omega, \hat{\omega}, k, k') \) are functional measures. Performing a Wick rotation in \( \lambda \) and assuming \( \omega, \hat{\omega} \) real and much smaller than one, i.e. \( \omega, \hat{\omega} \ll 1 \) which is allowed in the sparse regime \( K \ll K_{S} \), we can linearize the logarithm and express \( \Psi \) as

$$\Psi = zv\hat{v},$$  \hspace{1cm} (S-81)$$

with

$$v = \int d\omega \int d\hat{\omega} \sum_{m^{in}, k, m^{out}, K} c(\omega, \hat{\omega}, k, k') e^{-i\omega},$$

$$\hat{v} = \int d\omega \int d\hat{\omega} \sum_{m^{in}, k, m^{out}, K} c(\omega, \hat{\omega}, k, k') e^{-i\hat{\omega}}.$$  \hspace{1cm} (S-82)$$

The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of \( f(\lambda, c(\omega, k), \hat{c}(\omega, k)) \) with respect to \( c(\omega, \hat{\omega}, k, k') \), \( \hat{c}(\omega, \hat{\omega}, k, k') \) and \( \lambda \), obtaining

$$-i\hat{c}(\omega, \hat{\omega}, k, k') = zv e^{-i\omega} + zv e^{-i\hat{\omega}},$$

$$c(\omega, \hat{\omega}, k, k') = \frac{1}{(2\pi)^2} p(k^{in}, k^{out})! \exp \left[ i\omega k^{in} + i\hat{\omega} k^{out} \right] \frac{1}{\sqrt{\det \sum_{m^{in}, k, m^{out}, K} \sum_{m^{out}, k, m^{out}, K} p(k^{in}, k^{out})! \exp \left[ i\omega' k^{in} + i\hat{\omega}' k^{out} \right]}}.$$  \hspace{1cm} (S-83)$$

Let us first calculate the integral

$$\int d\omega \int d\hat{\omega} \sum_{m^{in}, k, m^{out}, K} p(k^{in}, k^{out})! \exp \left[ i\omega k^{in} + i\hat{\omega} k^{out} \right]$$  \hspace{1cm} (S-84)$$
where we have substituted the saddle point expression for \( \hat{c}(\omega, k) \). Using expanding the exponential and proceeding as in the simple uncorrelated case we get

\[
\int \frac{d\omega}{2\pi} \int \frac{d\hat{\omega}}{2\pi} \sum_{m=\infty}^{N} \sum_{k=\infty}^{M} p(k^{in}, k^{out}) \hat{p}(k^{in}, k^{out}) \exp \left[ i\omega k^{in} + i\hat{\omega} k^{out} - i\hat{c}(\omega, \hat{\omega}, k^{in}, k^{out}) \right] = \sum_{m=\infty}^{N} \sum_{k=\infty}^{M} \frac{p(k^{in}, k^{out}) k^{in}! k^{out}! \exp \left[ i\omega k^{in} + i\hat{\omega} k^{out} - i\hat{c}(\omega, \hat{\omega}, k^{in}, k^{out}) \right]}{\left( z\hat{\nu} k^{in} (z\nu) k^{out} \right)}.
\]

Therefore \( c(\omega, k) \) at the saddle point solution can be expressed as

\[
c(\omega, k) = \frac{1}{(2\pi)^2} k^{in} k^{out} p(k^{in}, k^{out}) e^{ik^{in}+i\hat{\omega}k^{out}} e^{-ig}
\]

(S-85)

With this expression, using a similar procedure we can express \( \nu \) as

\[
\nu = \left< z\hat{\nu} k^{in} (z\nu) k^{out} \right> \sum_{m=\infty}^{N} \sum_{k=\infty}^{M} k^{in} p(k^{in}, k^{out}) (z\hat{\nu})^{in-1} (z\nu)^{out}
\]

\[
\hat{y} = \left< z\hat{\nu} k^{in} (z\nu) k^{out} \right> \sum_{m=\infty}^{N} \sum_{k=\infty}^{M} k^{out} p(k^{in}, k^{out}) (z\hat{\nu})^{out} (z\nu)^{in-1}
\]

(S-87)

Combing this equation with the third saddle point equation

\[
zv \hat{y} = \left< k^{in} \right> = \left< k^{out} \right>
\]

(S-88)

it is immediate to show that \( zv = z\hat{y} = 1 \) is a solution with

\[
z = \frac{1}{<k^{in}>}, \quad \nu = \hat{y} = \left< k^{in} \right> = \left< k^{out} \right>
\]

(S-89)

By inserting this expression in Eq. (S-86) we get

\[
c(\omega, \hat{\omega}, k^{in}, k^{out}) = \frac{1}{(2\pi)^2} k^{in} k^{out} p(k^{in}, k^{out}) \exp \left[ i\omega k^{in} + i\hat{\omega} k^{out} + e^{-ig} \right].
\]

(S-90)

From this equation we can conclude that the networks of these ensemble have heterogeneous degree distribution, as the density of nodes of in-degree \( k^{in} \) and out-degree \( k^{out} \) is given the desired joint probability distribution, i.e.

\[
\int d\omega \int d\hat{\omega} c(\omega, \hat{\omega}, k^{in}, k^{out}) = p(k^{in}, k^{out}).
\]

(S-91)

However the marginal for each link is the same and given by Eq. (S-71) with the marginal probability of a link conditioned on the degrees of its two endnodes be given by Eq. (S-73).

**EXCHANGEABLE ENSEMBLE OF SPARSE BIPARTITE NETWORKS**

**General framework**

We consider exchangeable ensembles of bipartite networks formed by two set of nodes \( V \) and \( U \) with \( |V| = N \) and \( |U| = M \) with the condition

\[
M = \alpha N,
\]

(S-92)

with \( \alpha > 0 \) being a constant independent of \( N \). We indicate with \( i \) the nodes belonging to the set \( V \) and with \( \mu \) the nodes belonging to the set \( U \). The structure of the bipartite network is determined by the \( N \times M \) incidence matrix \( b_{\mu i} = 1 \) if there is a link between node \( i \) and node \( \mu \), otherwise \( b_{\mu i} = 0 \). The degree of the nodes in \( V \) and in \( U \) is determined from the incidence matrix \( b \) according to the following equations

\[
k_i = \sum_{\mu=1}^{M} b_{\mu i},
\]

\[
q_{\mu} = \sum_{i=1}^{N} b_{\mu i}.
\]

(S-93)
The exchangeable sparse bipartite network ensemble is an ensemble designed in order to obtain bipartite networks in which the nodes in $V$ have degree distribution $p(k)$ and the nodes in $U$ have degree distribution $\hat{p}(q)$. These distributions can be arbitrary but must obey $N(k) = M(q)$ which implies $\langle k \rangle = \alpha(q)$. Moreover these ensembles have fixed number of links $L = \langle k \rangle N$ and the degree $k$ ($q$) of the nodes in $V$ ($U$) has maximum smaller or equal to $K \ll K_{S} = \sqrt{\langle k \rangle N}$ ($K \ll K_{S} = \sqrt{\langle k \rangle N}$) and minimum degree greater or smaller than $m$ ($\hat{m}$). The probability $\mathbb{P}(G)$ for each bipartite network $G$ is taken to be

$$
\mathbb{P}(G) = \prod_{i=1}^{N} p(k_i) \prod_{\mu=1}^{M} \hat{p}(q_\mu) e^{-\sum_{k,q} \delta \left(L, \sum_{i,\mu} b_{i\mu} \right) \theta \left(K - \max_{i=1,...,N} k_i \right) \theta \left(\hat{K} - \max_{\mu=1,...,M} q_\mu \right) \theta \left(\min_{i=1,...,N} k_i - m \right) \theta \left(\min_{\mu=1,...,M} q_\mu - \hat{m} \right)} \quad (S-94)
$$

The entropy of this ensemble is given by

$$
\Sigma(k, q) = \ln \left[ \frac{L!}{\prod_{i=1}^{N} k_i! \prod_{\mu=1}^{M} q_\mu!} \right] + o(N) \quad (S-95)
$$

This ensemble is Hamiltonian as $\mathbb{P}(G)$ can be expressed as

$$
\mathbb{P}(G) = e^{-H(G)} \delta \left(L, \sum_{i,\mu} b_{i\mu} \right) \theta \left(K - \max_{i=1,...,N} k_i \right) \theta \left(\hat{K} - \max_{\mu=1,...,M} q_\mu \right) \theta \left(\min_{i=1,...,N} k_i - m \right) \theta \left(\min_{\mu=1,...,M} q_\mu - \hat{m} \right) \quad (S-96)
$$

with the Hamiltonian $H(G)$ given by

$$
H(G) = -\sum_{i=1}^{N} \ln \left( p(k_i) \right) - \sum_{\mu=1}^{M} \ln \left( \hat{p}(q_\mu) \right) + \ln(L!) \quad (S-97)
$$

This ensemble produces a network in which the nodes in $V$ have degree distribution $p(k)$ and the nodes in $U$ have degree distribution $\hat{p}(q)$ which can be heterogeneous also if the marginal of every link is the same and given by

$$
p_{i\mu} = \sum_{k,q} p(k) \hat{p}(q) \frac{kq}{(k)N} \quad (S-98)
$$

with

$$
p_{i\mu k_i=q_\mu=q} = \frac{kq}{(k)N} \quad (S-99)
$$

**Derivation of the marginal probability**

In this section our goal is the solve the partition function $Z$ for the exchangeable ensemble of bipartite networks using the saddle point equation the expression for the marginal probability of a link. The the partition function $Z$ of this network ensemble, where for simplicity we have put the auxiliary field $h = 0$ from the beginning is given by

$$
Z = \sum_{a} \sum_{k} \sum_{q} e^{-H(G)} \delta \left(L, \sum_{i,\mu} a_{i\mu} \right) \prod_{i=1}^{N} \delta \left(k_i - \sum_{\mu=1}^{M} b_{i\mu} \right) \prod_{\mu=1}^{M} \delta \left(q_\mu - \sum_{i=1}^{N} b_{i\mu} \right) \quad (S-100)
$$

By expressing the Kronecker deltas in integral form

$$
\delta(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\omega e^{ix\omega} \quad (S-101)
$$

we get

$$
Z = \sum_{a} \mathbb{P}(G) = \frac{1}{L!} \sum_{a} \sum_{k} \sum_{q} \int D\omega \int D\tilde{\omega} \int d\lambda \frac{G(\lambda, \omega, \tilde{\omega}, k, q)}{2\pi} \quad (S-102)
$$

with

$$
G(\lambda, \omega, \tilde{\omega}, k, q) = \sum_{i=1}^{N} \left[ i\omega_i k_i + \ln(k_i) p(k_i) \right] + \sum_{\mu=1}^{M} \left[ i\tilde{\omega}_\mu q_\mu + \ln(q_\mu) \hat{p}(q_\mu) \right] + i\lambda L + \sum_{i,\mu} \ln(1 + e^{-i\omega_i \omega - \tilde{\omega}_\mu}) \quad (S-103)
$$
and with $D\omega = \prod_{i=1}^{N}[d\omega_i/(2\pi)]$, and $D\hat{\omega} = \prod_{\mu=1}^{M}[d\hat{\omega}_\mu/(2\pi)]$. Let us now introduce the two functional order parameters \[ c(\omega, k) = \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i)\delta(k, k_i), \] \[ \sigma(\hat{\omega}, q) = \frac{1}{M} \sum_{\mu=1}^{M} \delta(\hat{\omega} - \hat{\omega}_\mu)\delta(q, q_\mu), \] (S-104)

by enforcing these definitions with a series of delta functions involving $c(\omega, k)$ and $\sigma(\hat{\omega}, q)$ and their conjugated order parameters $\hat{c}(\omega, k)$ and $\hat{\sigma}(\omega, k)$. Imposing also $L = \langle k \rangle N = \langle q \rangle M$ where $\langle k \rangle = \sum_i k p(k)$, $\langle q \rangle = \sum_q q \hat{p}(q)$, the partition function reads

$$ Z = \frac{1}{L!} \sum_{k} \sum_{q} f \left( \int D\hat{c}(\omega, k) \int Dc(\omega, k) \int D\hat{\sigma}(\hat{\omega}, q) \int D\sigma(\hat{\omega}, q) \int \frac{d\lambda}{2\pi} e^{Nf} \right) $$

(S-105)

with

$$ f = i \int d\omega \sum_{m \leq K} \hat{c}(\omega, k) c(\omega, k) + i\pi \int d\hat{\omega} \sum_{\hat{m} \leq \hat{K}} \hat{\sigma}(\hat{\omega}, q) \sigma(\hat{\omega}, q) + i\lambda \langle k \rangle $$

$$ + \Psi + \ln \int \frac{d\omega}{2\pi} \sum_{m \leq K} p(k) k! \exp \left[ i\omega k - i\hat{c}(\omega, k) \right] + \alpha \ln \int \frac{d\hat{\omega}}{2\pi} \sum_{\hat{m} \leq \hat{K}} \hat{p}(q) q! \exp \left[ i\hat{\omega} q - i\hat{\sigma}(\hat{\omega}, q) \right] $$

(S-106)

and $\Psi$ is given by

$$ \Psi = \frac{\alpha^N}{2} \int d\omega \sum_{m \leq K, \hat{m} \leq \hat{K}} c(\omega, k) \sigma(\hat{\omega}, q) \ln \left( 1 + e^{-i\omega - i\hat{\omega}} \right), $$

where $D\hat{c}(\omega, k)$, $Dc(\omega, k)$, $D\hat{\sigma}(\hat{\omega}, q)$ and $D\sigma(\hat{\omega}, q)$ are functional measures. Performing a Wick rotation in $\lambda$ and assuming $z/N = e^{-i\lambda}$ real and much smaller than one, i.e. $z/N \ll 1$ which is allowed in the sparse regime $K \ll K_S$, we can linearize the logarithm and express $\Psi$ as

$$ \Psi = z \alpha v \hat{v}, $$

(S-107)

with

$$ v = \int d\omega \sum_{m \leq K} c(\omega, k) e^{-i\omega}, $$

$$ \hat{v} = \int d\hat{\omega} \sum_{\hat{m} \leq \hat{K}} \sigma(\hat{\omega}, q) e^{-i\hat{\omega}}. $$

(S-108)

The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of $f(\lambda, c(\omega, k), \hat{c}(\omega, k))$ with respect to $c(\omega, \hat{\omega}, k^\text{in}, k^\text{out}), \hat{c}(\omega, \hat{\omega}, k^\text{in}, k^\text{out})$ and $\lambda$, obtaining

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

$$ c(\omega, k) = \int \frac{dk'}{2\pi} \sum_{m \leq K} p(k') k'! \exp \left[ i\omega k' - i\hat{c}(\omega, k') \right] $$

$$ -i\hat{\sigma}(\hat{\omega}, q) = z v e^{-i\hat{\omega}}, $$

$$ \sigma(\hat{\omega}, q) = \int \frac{dq'}{2\pi} \sum_{\hat{m} \leq \hat{K}} \hat{p}(q') q'! \exp \left[ i\hat{\omega} q' - i\hat{\sigma}(\hat{\omega}, q') \right]. $$

(S-109)

Let us first calculate the integrals

$$ \int \frac{d\omega}{2\pi} \sum_{m \leq K} p(k) k! \exp \left[ i\omega k - i\hat{c}(\omega, k) \right] = \sum_{m \leq K} p(k) (az\hat{v})^k = \langle (az\hat{v})^k \rangle $$

$$ \int \frac{d\hat{\omega}}{2\pi} \sum_{\hat{m} \leq \hat{K}} \hat{p}(q) q! \exp \left[ i\hat{\omega} q - i\hat{\sigma}(\hat{\omega}, q) \right] = \sum_{\hat{m} \leq \hat{K}} \hat{p}(q) (z\hat{v})^q = \langle (z\hat{v})^q \rangle $$

(S-110)
where we have substituted the saddle point expression for \( \hat{c}(\omega, k) \) and \( \hat{\sigma}(\hat{\omega}, q) \) and we have followed the same procedure as for calculating the corresponding integrals in the previous case. It follows that \( c(\omega, k) \) and \( \sigma(\hat{\omega}, q) \) at the saddle point solution can be expressed as

\[
\begin{align*}
c(\omega, k) &= \frac{1}{2\pi} k! p(k) e^{i\omega k + (\alpha z) e^{-i\omega}} \langle (\alpha z) e^{i\omega} \rangle, \\
\sigma(\hat{\omega}, q) &= \frac{\alpha}{2\pi} q! \hat{p}(q) e^{i\hat{\omega} q + (\alpha z) e^{-i\hat{\omega}}} \langle (\alpha z) e^{i\hat{\omega}} \rangle.
\end{align*}
\] (S-111)

With this expression, using a similar procedure as in the precedent integrals, we can express \( \nu \) as

\[
\nu = \frac{1}{\langle (\alpha z) e^{i\omega} \rangle} \sum_{m \leq K} kp(k)(\alpha z)^{k-1},
\]

\[
\hat{\nu} = \frac{1}{\langle (\alpha z) e^{i\hat{\omega}} \rangle} \sum_{m \leq K} q\hat{p}(q)(\alpha z)^{\hat{q}-1}.
\] (S-112)

Combining this equation with the third saddle point equation

\[
\alpha z \hat{v} = \langle k \rangle = \alpha \langle q \rangle,
\] (S-113)

it is immediate to show that \( nz = az \hat{v} = 1 \) is a solution with

\[
z = \frac{1}{\langle k \rangle}, \quad \nu = \langle k \rangle, \quad \hat{\nu} = \langle q \rangle.
\] (S-114)

By inserting this expression in Eq. (S-111) we get

\[
\begin{align*}
c(\omega, k) &= \frac{1}{2\pi} k! p(k) \exp \left[ i\omega k + e^{-i\omega} \right], \\
\sigma(\hat{\omega}, q) &= \frac{\alpha}{2\pi} q! \hat{p}(q) \exp \left[ i\hat{\omega} q + e^{-i\hat{\omega}} \right].
\end{align*}
\] (S-115)

From this equation we can conclude that the networks of these ensemble have heterogeneous degree distribution, as the density of nodes in \( V \) with degree \( k \) is given by \( p(k) \) while the density of nodes in \( U \) having degree \( q \) is given by \( \hat{p}(q) \), i.e.

\[
\int d\omega \int d\hat{\omega} \hat{c}(\omega, \hat{\omega}, k) = p(k)
\]

\[
\int d\omega \int d\hat{\omega} \hat{\sigma}(\omega, \hat{\omega}, q) = \hat{p}(q).
\] (S-116)

However the marginal for each link is the same and given by Eq. (S-98) with the marginal probability of a link conditioned on the degrees of its two endnodes be given by Eq. (S-99).

**EXCHANGEABLE ENSEMBLE OF SPARSE MULTIPLEX NETWORKS**

Sparse exchangeable multiplex networks

Exchangeable sparse multiplex networks can be also defined using a similar approach. To this end we can consider a multiplex network \( G = (G_1, G_2, \ldots, G_M) \) formed by \( M \) layer \( \alpha \in \{1, 2, \ldots, M\} \) each determined by an adjacency matrix \( A^{[\alpha]} \) \[^{[35]}\]. To keep the discussion simple we will assume that each adjacency matrix is undirected and unweighted. The degree \( k^{[\alpha]}_i \) of each node \( i \) in layer \( \alpha \in \{1, 2, \ldots, M\} \) is determined by the equation

\[
k^{[\alpha]}_i = \sum_{j=1}^{N} a^{[\alpha]}_{ij}.
\] (S-117)

An important feature of multiplex networks are multilinks \( m = (m^{[1]}, m^{[2]}, \ldots, m^{[M]}) \) (with \( m^{[\alpha]} \in (0, 1) \)) indicating the pattern of connection between any two nodes. For instance in a duplex network \( (M = 2) \) with two layers indicating mobile phone and email...
interaction a two nodes are connected by a multilink $(1, 0)$ if they only communicate with mobile phone, are connected by a multilink $(0, 1)$ if they only communicate via email and they are connected by a multilink $(1, 1)$ if they communicate both via mobile phone and email. In order to indicate if two nodes $i$ and $j$ are connected by a multilink $\vec{m}$ we can use the multi-adjacency matrices $A^m$ given by

$$A^m_{ij} = \prod_{\alpha=1}^{M} \left[ a^m_{ij}^\alpha m_\alpha + (1 - a^m_{ij}^\alpha)(1 - m_\alpha) \right].$$

(S-118)

Since any two nodes can be connected only by a single multilink we have

$$\sum_{\vec{m}} A^m_{ij} = 1.$$  \hspace{1cm} (S-119)

Having defined the multi-adjacency matrices, it is possible to introduce the definition of the multidegree $k^m_i$ as the sum of multilinks $\vec{m}$ incident to the node $i$ [8], i.e.

$$k^m_i = \sum_{j=1}^{N} A^m_{ij}.$$  \hspace{1cm} (S-120)

Using the approach described in this work we can either define exchangeable sparse multiplex networks in which each layer is independent of the other and has a given degree distribution (eventually dependent on the choice of the layer); or we can define exchangeable sparse multiplex networks in which the multidegree distribution is kept fixed.

The first case can be modelled by drawing each layer of the multiplex network independently from an exchangeable ensemble of uncorrelated simple networks. Given the simplicity of the approach here we neglect to treat this case in detail. The latter case can be modelled by an exchangeable multiplex network ensemble in which each nodes has a series of non trivial multidegrees of uncorrelated simple networks.

Having defined the multi-adjacency matrices, it is possible to introduce the definition of the multidegree $k^m_i$ as the sum of multilinks $\vec{m}$ incident to the node $i$ [8], i.e.

$$k^m_i = \sum_{j=1}^{N} A^m_{ij}.$$  \hspace{1cm} (S-120)

Using the approach described in this work we can either define exchangeable sparse multiplex networks in which each layer is independent of the other and has a given degree distribution (eventually dependent on the choice of the layer); or we can define exchangeable sparse multiplex networks in which the multidegree distribution is kept fixed.

The first case can be modelled by drawing each layer of the multiplex network independently from an exchangeable ensemble of uncorrelated simple networks. Given the simplicity of the approach here we neglect to treat this case in detail. The latter case can be modelled by an exchangeable multiplex network ensemble in which each nodes has a series of non trivial multidegrees $k^m_i$ with $\vec{m} \neq \vec{0}$ [e.g. $k^m_i = (k^{(1, 2)}_i, k^{(0, 1)}_i, k^{(1, 1)}_i)$ in the case of $M = 2$ layers] with multidegree distribution $\pi(k^m_i)$. Moreover we impose that in the network there are exactly $L^m = \langle k^m \rangle / N/2$ multilinks of type $\vec{m} \neq \vec{0}$ and that the multiplex is sparse, i.e. the multidegree $k^m_i$ has a minimum value greater or equal than $\hat{m}$ and a maximum value smaller or equal than $K^m_i$ with $K^m_i \ll K^m_{\max} = \sqrt{\langle k^m \rangle N}$. Therefore the ensemble is defined by associating to each multiplex network of $M$ layers $G = (G_1, G_2, \ldots, G_M)$ the probability

$$\mathbb{P}(G) = P\left(k^m_i\right) e^{-\Sigma\left(k^m_{ji}\right)} \prod_{\vec{m} \neq \vec{0}} \left[ \delta\left(L^m \sum_{i<j} A^m_{ij}\right) \theta\left(K^m_{\max} - \max_{i=1, \ldots, N} k^m_i\right) \theta\left(\min_{i=1, \ldots, N} k^m_i - \hat{m}\right) \right],$$  \hspace{1cm} (S-121)

where here $\{k^m\}$ indicates the sequence of all the non trivial multidegrees $\vec{m} \neq \vec{0}$ of every node $i$ of the multiplex network, and where the entropy $\Sigma\left(k^m_{ji}\right)$ is given by [8]

$$\Sigma\left(L^m_{ji}\right) = \ln \left( \prod_{\vec{m} \neq \vec{0}} \frac{(2L^m_{ji})!!}{\prod_{\vec{m} \neq \vec{0}} L^m_{ji}^{k^m_i}} \right) + o(N).$$  \hspace{1cm} (S-122)

Here $P\left(k^m_i\right)$ is given by the product of the probability that each nodes has multidegrees $k^m_i$

$$P\left(k^m_i\right) = \prod_{i=1}^{N} \pi\left(k^m_i\right).$$  \hspace{1cm} (S-123)

This exchangeable multiplex network ensemble is Hamiltonian as the probability $\mathbb{P}(G)$ can be written as

$$\mathbb{P}(G) = e^{-H(G)} \prod_{\vec{m} \neq \vec{0}} \left[ \delta\left(L^m \sum_{i<j} A^m_{ij}\right) \theta\left(K^m_{\max} - \max_{i=1, \ldots, N} k^m_i\right) \theta\left(\min_{i=1, \ldots, N} k^m_i - \hat{m}\right) \right],$$  \hspace{1cm} (S-124)

with Hamiltonian $H(G)$ given by

$$H(G) = - \sum_{i=1}^{N} \ln \left( \pi\left(k^m_i\right) \right) - \sum_{i=1}^{N} \sum_{\vec{m} \neq \vec{0}} k^m_{ji} + \sum_{\vec{m} \neq \vec{0}} (2L^m_{ji})!!.$$  \hspace{1cm} (S-125)
The marginal probability \( p_{ij} \) of observing a multilink \( \vec{m} \neq \vec{0} \) between node \( i \) and node \( j \) is given by

\[
p_{ij} = \langle A_{ij} \rangle = \sum_{k,k'} \pi(k) \pi_{k'}(k') p(k, k'),
\]

(S-126)

where the marginal probability \( p(k, k') \) of observing a multilink \( \vec{m} \neq \vec{0} \) between node \( i \) and node \( j \) conditioned on the multidegrees \( \vec{m} \) of the two nodes is given by

\[
p_{ij}^m(k, k') = \frac{k(k, k')}{k}.
\]

(S-127)

**Treatment of the exchangeable ensemble of sparse multiplex networks**

In this section our goal is to solve the partition function \( Z \) for the exchangeable ensemble of multiplex networks. The partition function \( Z \) of this multiplex network ensemble, is given by

\[
Z(h) = \sum_A \sum_{|k|} e^{-H(G)} e^{-\sum_i \sum_{a \neq b} h_{ab}^i k^i_a} \prod_{a \neq b} \delta \left( m^i_a, \sum_{i,j} A_{ij}^i \right) \prod_i \delta \left( k^i - \sum_j A^i_{ji} \right).
\]

(S-128)

Here and in the following we use the notation \( \sum_i \) to indicate the sum over all the possible values of the degree of each node \( i \) satisfying \( m^i_a \leq k^i_a \leq K^i_a = \sqrt{\langle k^i_a \rangle N} \). By expressing the Kronecker deltas in integral form

\[
\delta(x, y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\omega e^{i\omega(x-y)}
\]

we get for the partition function \( Z \) of this network ensemble,

\[
Z(h) = \frac{1}{\prod_{a \neq b}(2L^a)^!! \prod_A \sum_{|k|} \int D\omega \int \frac{dA}{2\pi} e^{G(A, \omega, k, h)},
\]

(S-130)

with

\[
G(\lambda, \omega, k^i_a, h) = \sum_{i=1}^{N} \left[ \sum_{a \neq b} h_{ab} k^i_a + \ln(k^i_a) \right] + i \sum_i A^i \ln L^i + \frac{1}{2} \sum_i \ln \left( 1 + \sum_{a \neq b} e^{-i\lambda k^i_a - i\omega k^i_a} \right),
\]

and with \( D\omega = \prod_{i=1}^{N} \prod_{a \neq b} \frac{d\omega^i}{(2\pi)} \). Let us now introduce the functional order parameter \( c(h) \)

\[
c(\omega, k) = \frac{1}{N} \sum_{i=1}^{N} \prod_{a \neq b} \delta(\omega^i - \omega^a) \delta(k^i_a, k^a_j),
\]

(S-131)

by enforcing its definition with a series of delta functions, and by imposing \( 2L^i = \langle k^i_a \rangle N \) where \( \langle k^i_a \rangle = \sum_{k^i_a} k^i_a p(k^i_a) \), we get

\[
Z(h) = \frac{1}{\prod_{a \neq b}(2L^a)^!! \prod_A \sum_{|k|} \int Dc(\omega, k) \int D\hat{c}(\omega, k) \int \frac{dA}{2\pi} e^{f(\lambda, c(h), \hat{c}(h), \omega, k, h)}
\]

(S-132)

with

\[
f(\lambda, c(h), \hat{c}(h), \omega, k, h) = i \int d\omega \sum_k c(h, k) c(\omega, k) + i \sum_{a \neq b} \hat{c}(h, k^a) / 2 + \Psi
\]

(S-133)

where \( W = 2^M - 1 \) indicates the number of non-trivial multilinks \( \vec{m} \neq \vec{0} \) and \( \Psi \) is given by

\[
\Psi = \frac{N}{2} \int d\omega \int d\omega' \sum_{k, k'} c(h, k) c(h', k') \ln \left( 1 + \sum_{a \neq b} e^{-i\lambda k^a - i\omega k^a - i\omega' k^a} \right).
\]
and where $Dc(\omega, k)$ and $D\hat{c}(\omega, k)$ are functional measures. Performing a Wick rotation in $\lambda$ and assuming $z^\mathbb{R}/N = e^{-i\omega t}$ real and much smaller than one, i.e. $z^\mathbb{R}/N \ll 1$ which is allowed in the sparse regime $K^\mathbb{R} \ll K^\mathbb{C}$, we can linearize the logarithm and express $\Psi$ as

$$\Psi = \frac{1}{2} \sum_{\alpha \in \mathbb{R}} z^\mathbb{R} [v(m)]^2 e^{-i\omega t},$$  \hspace{2cm} (S-134)

with

$$v(m) = \int d\omega \sum_k c(\omega, k)e^{-i\omega t}.$$  \hspace{2cm} (S-135)

The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of $f(\lambda, c(\omega, k), \hat{c}(\omega, k))$ with respect to $c(\omega, k)$, $\hat{c}(\omega, k)$ and $z^\mathbb{R}$, obtaining for $h^\mathbb{R} \to 0$,

$$-i\hat{c}(\omega, k) = \sum_{\alpha \in \mathbb{R}} z^\mathbb{R} v(m) \exp[-i\omega^\mathbb{R} - k^\mathbb{R}],$$

$$c(\omega, k) = \frac{1}{(2\pi)^W} \pi(k^\mathbb{R}) \prod_{\alpha \in \mathbb{R}} [k^\mathbb{R}! \exp[i\omega^\mathbb{R} k^\mathbb{R}]] \exp[-i\hat{c}(\omega, k)],$$

$$z^\mathbb{R}[v(m)]^2 = \langle k^\mathbb{R} \rangle.$$  \hspace{2cm} (S-136)

By proceeding like in the previous examples, we can perform the integral

$$\int \frac{d\omega}{(2\pi)^W} \sum_k c(\omega, k) = \prod_{\alpha \in \mathbb{R}} \pi(k^\mathbb{R}) \prod_{\alpha \in \mathbb{R}} [k^\mathbb{R}! \exp[i\omega^\mathbb{R} k^\mathbb{R}]] \exp[-i\hat{c}(\omega, k)] = \prod_{\alpha \in \mathbb{R}} \pi(k^\mathbb{R}) \prod_{\alpha \in \mathbb{R}} [z^\mathbb{R} v(m)]^{1/2} = w.$$  \hspace{2cm} (S-137)

Therefore $c(\omega, k)$ at the saddle point solution can be expressed as

$$c(\omega, k) = \frac{1}{(2\pi)^W} \pi(k^\mathbb{R}) \prod_{\alpha \in \mathbb{R}} \exp[i\omega^\mathbb{R} k^\mathbb{R} + z^\mathbb{R} v(m)].$$  \hspace{2cm} (S-138)

With this expression, using a similar procedure we can express $v(m)$ as

$$v(m) = \int d\omega \sum_k c(\omega, k)e^{-i\omega t} = \frac{1}{(2\pi)^W} \pi(k^\mathbb{R}) k^\mathbb{R}! \sum_{\alpha \in \mathbb{R}} \pi(k^\mathbb{R}) \prod_{\alpha \in \mathbb{R}} [z^\mathbb{R} v(m)]^{1/2} = \frac{1}{(2\pi)^W} \pi(k^\mathbb{R}) k^\mathbb{R}! \prod_{\alpha \in \mathbb{R}} [z^\mathbb{R} v(m)]^{1/2}.$$  \hspace{2cm} (S-139)

Combing this equation with the third saddle point equation it is immediate to show that $z^\mathbb{R} v(m) = 1$ is a solution with

$$z^\mathbb{R} = \frac{1}{\langle k^\mathbb{R} \rangle}, \quad v(m) = \langle k^\mathbb{R} \rangle = w = 1.$$  \hspace{2cm} (S-140)

By inserting this expression in Eq. (S-138) we get

$$c(\omega, k) = \frac{1}{2\pi} \pi(k^\mathbb{R}) \prod_{\alpha \in \mathbb{R}} [k^\mathbb{R}! \exp[i\omega^\mathbb{R} k^\mathbb{R} + e^{-i\omega t}]].$$  \hspace{2cm} (S-141)

From this expression, by proceeding like in the simple network case, we can derive that each node of the network have multidegrees $k^\mathbb{R}$ with a probability $\pi(k^\mathbb{R})$ and that the marginal probability of multilinks are given by Eq. (S-126) and Eq. (S-127).

**EXCHANGEABLE ENSEMBLE OF SPARSE SIMPLICIAL COMPLEXES**

Uncorrelated exchangeable ensembles of simplicial complexes

The proposed approach to construct exchangeable network ensembles can be also immediately applied to generate and ensemble of exchangeable pure simplicial complexes of dimension $d$ with given generalized degrees of the nodes (for a background on simplicial complex and their ensembles see [35, 37]). We consider an ensemble of $d$-dimensional simplicial complexes $K$ whose structure is determined by the adjacency tensor $A$ of elements $a_{ij} = 1$ if the $d$-dimensional simplex $\alpha = [i_0, i_1, \ldots, i_d]$ belongs
to the simplicial complex, and with \( a_{ij} = 0 \) otherwise. The generalized degree \( k_i \) of the generic node \( i \) indicates the number of \( d \)-dimensional simplices incident to the node \( i \) and it can be expressed in terms of the adjacency tensor as
\[
k_i = \sum_{\alpha \in K} a_{\alpha} = \sum_{i_1 < i_2 < \ldots < i_d} a_{i_1i_2\ldots i_d}.
\] (S-142)

The ensemble of \( d \)-dimensional simplices with given generalized degree sequence \( \mathbf{k} \) has been studied in Ref.\cite{37}. Here we consider the exchangeable ensemble of uncorrelated \( d \)-dimensional simplicial complexes. We indicate with \( P(\mathbf{k}) \) the probability assigned to observing a generalized degree sequence \( \mathbf{k} \), with
\[
P(\mathbf{k}) = \prod_{i=1}^{N} \left[ p(k_i) \delta(K_i - k_i) \delta(m - k_i) \right].
\] (S-143)

Therefore the probability of the generalized degree sequence \( \mathbf{k} \) factorizes in the product of the probability \( p(k_i) \) that each node \( i \) has a generalized degree \( k_i \) equal or greater than \( m \). Moreover we consider that the simplicial complexes are sparse, i.e. they have a structural cutoff \cite{37}
\[
K_s = \left( \frac{\langle (k) N \rangle^d}{d!} \right)^{1/(d+1)}.
\] (S-144)

This implies that the generalized degree of the nodes \( k_i \) have a maximum value \( K \ll K_s \). Moreover we assume that each node as a generalized degree equal or greater than \( m \). This ensemble is generated by associating to each simplicial complex \( \mathcal{K} \) the probability \( \mathbb{P}(\mathcal{K}) \) given by
\[
\mathbb{P}(\mathcal{K}) = P(\mathbf{k}) e^{-\Sigma(\mathbf{k})} \left( S, \sum_{\alpha \in \mathcal{K}} a_{\alpha} \right)
\] (S-145)

where \( S = \langle k \rangle N / (d + 1) \) indicates the number of simplices in the simplicial complex and where \( \Sigma(\mathbf{k}) \) is the entropy of the ensemble with generalized degree sequence \( \mathbf{k} \). In presence of the structural cutoff, the entropy \( \Sigma(\mathbf{k}) \) of \( d \)-dimensional simplicial complexes with generalized degree sequence \( \mathbf{k} \) is given by \cite{37}
\[
\Sigma(\mathbf{k}) = \ln \left( \frac{\prod_{i=1}^{N} (k_i) ! (d !)^{\langle k \rangle N / (d + 1)}}{((k) N) !} \right) + o(N).
\] (S-146)

It follows that the exchangeable ensemble of \( d \)-dimensional simplicial complexes can be obtained by considering the Hamiltonian simplicial complex ensemble
\[
\mathbb{P}(\mathcal{K}) = e^{-H(\mathcal{G})} e^{-\Sigma(\mathbf{k})} \left( S, \sum_{\alpha \in \mathcal{K}} a_{\alpha} \right) \left( K_s - \max_{i=1,2,\ldots,N} k_i \right) \theta \left( \min_{i=1,2,\ldots,N} k_i - m \right).
\] (S-147)

with Hamiltonian \( H(\mathcal{G}) \) given by
\[
H(\mathcal{G}) = - \sum_{i=1}^{N} \ln p(k_i) + \Sigma(\mathbf{k}).
\] (S-148)

This ensemble is exchangeable and the marginal probability for each simplex \( \alpha \) is given by (see following paragraph for the derivation)
\[
p_{\alpha} = \sum_{|k_0,k_1,\ldots,k_d|} \left[ \prod_{r=0}^{d} p(k_r) \right] (k_0, k_1, \ldots, k_d) = d! \langle k \rangle^d N^d.
\] (S-149)

where the marginal probability \( p(k_0, k_1, \ldots, k_d) = p(\alpha = [i_0, i_1, \ldots, i_d] \mid k_i = k_i) \) of a simplex \( \alpha = [i_0, i_1, \ldots, i_d] \) conditioned on the value of the generalized degrees of its nodes \( k_i = k_i \) is given by the uncorrelated expression \cite{37}
\[
p(\alpha = [i_0, i_1, \ldots, i_d] \mid k_i = k_i) = p(k_0, k_1, \ldots, k_d) = d! \prod_{r=0}^{d} k_r \langle (k) N \rangle^d.
\] (S-150)
Derivation of the marginal probability of a simplex in the uncorrelated exchangeable simplicial complex ensembles

In this section our goal is the solve the partition function $Z(h)$ (that for construction is expected to take the value $Z(h = 0) = 1$) for the exchangeable ensemble of uncorrelated simplicial complexes. The us start by defying the for the partition function $Z(h)$ of this simplicial complex ensemble as

$$Z(h) = \sum_{a} \mathbb{P}(K)e^{-h\sum_{\omega}a_{\omega}} = \frac{[d!]^N/[((k)N)!]}{[(k)!(d+1)]} \sum_{a} \sum_{k} \int D\omega \int \frac{d\lambda}{2\pi} G(\lambda,\omega, h),$$

with

$$G(\lambda, \omega, h) = \sum_{i=1}^{N} (\omega_{ik} + \ln(k_{i}!p(k_{i}))) + i\lambda(k)/(d+1) + \sum_{\omega \in K} \ln \left(1 + e^{-\omega_{i}a_{\omega_{i}}(\omega_{\lambda} - \omega_{\lambda} + h)}\right),$$

and with $D\omega = \prod_{i=1}^{N} [d\omega_{i}]/(2\pi)$. In Eq. \[S-151\] and in the following we use the notation $\Sigma_{k}'$ to indicate the sum over all the possible values of the generalized degree of each node $i$ satisfying $m \leq k_{i} \leq K \ll K_{S}$. Let us now introduce the functional order parameter \[37\] \[40\] \[41\]

$$c(\omega, k) = \frac{1}{N} \sum_{i=1}^{N} \delta(\omega_{i} - \omega_{i}) \delta(k_{i}),$$

by enforcing its definition with a series of delta functions. By assuming a discretization in $\omega$ in intervals of size $\Delta \omega$ we then introduce for each choice of $(\omega, k)$ the term

$$1 = \int dc(\omega, k)\delta \left( c(\omega, k) - \frac{1}{N} \sum_{i=1}^{N} \delta(\omega_{i} - \omega_{i}) \delta(k_{i}) \right) = \int \frac{d\tilde{c}(\omega, k)dc(\omega, k)}{2\pi/(N\Delta\omega)} \exp \left[ i\Delta \omega \tilde{c}(\omega, k) [Nc(\omega, k) - \sum_{i=1}^{N} \delta(\omega_{i} - \omega_{i}) \delta(k_{i})] \right].$$

After performing these operations, by imposing $(d + 1)S = \langle k \rangle N$ where $\langle k \rangle = \sum_{i} kp(k_{i})$, the partition function reads in the limit $\Delta \omega \rightarrow 0$,

$$Z(h) = \frac{[d!]^N/[((k)N)!]}{[(k)!(d+1)]} \sum_{k} \int Dc(\omega, k) \int D\tilde{c}(\omega, k) \int \frac{d\lambda}{2\pi} e^{Nf(\lambda, c(\omega, k), \tilde{c}(\omega, k), h)}$$

with

$$f(\lambda, c(\omega, k), \tilde{c}(\omega, k), h) = i \int d\omega \sum_{k} \tilde{c}(\omega, k)c(\omega, k) + i\lambda(k)/(d+1) + \Psi + \ln \int \frac{d\omega}{2\pi} \sum_{m \leq k \leq K} p(k)ke^{i\tilde{c}(\omega, k)}$$

where $\Psi$ for $K \ll K_{S}$ can be approximated by

$$\Psi = \frac{N^d}{(d+1)!} e^{-h - i\lambda} \int_{r=0}^{d} \sum_{m \leq k \leq K} \int d\omega_{r} c(\omega_{r}, k_{r}) e^{-i\omega_{r}}$$

and where $Dc(\omega, k)$ is the functional measure $Dc(\omega, k) = \lim_{\Delta \omega \rightarrow 0} \prod_{\omega} \prod_{k} [dc(\omega, k) \sqrt{N\Delta\omega}/(2\pi)]$ and similarly $D\tilde{c}(\omega, k) = \lim_{\Delta \omega \rightarrow 0} \prod_{\omega} \prod_{k} [d\tilde{c}(\omega, k) \sqrt{N\Delta\omega}/(2\pi)]$. Performing a Wick rotation in $\lambda$ and assuming $z/N^d = e^{-i\nu}$ real and much smaller than one, i.e. $z/N^d \ll 1$ which is allowed in the sparse regime $K \ll K_{S}$, we can linearize the logarithm and express $\Psi$ as

$$\Psi = \frac{1}{(d+1)!} e^{z^{d+1} e^{-h}},$$

with

$$\nu = \int d\omega \sum_{m \leq k \leq K} c(\omega, k) e^{-i\omega}.$$

The saddle point equations determining the value of the partition function can be obtained by performing the (functional) derivative of $f(\lambda, c(\omega, k), \tilde{c}(\omega, k), h)$ with respect to $c(\omega, k)$, $\tilde{c}(\omega, k)$ and $\lambda$, obtaining for $h \rightarrow 0$

$$-i\tilde{c}(\omega, k) = \frac{z}{d!} e^{-i\omega},$$

$$c(\omega, k) = \frac{1}{2\pi} \int \frac{d\theta}{\sum_{m \leq k \leq K} \theta} \frac{p(k)ke^{i\tilde{c}(\omega, k)}}{p(k)ke^{i\tilde{c}(\omega, k)}},$$

$$\frac{z}{d!} = \langle k \rangle.$$
Let us first calculate the integral
\[
\int \frac{d\omega}{2\pi} \sum_{m \leq k \leq K} p(k) k! e^{-i\omega k - \tilde{\omega}(\omega, k)} = \int \frac{d\omega}{2\pi} \sum_{m \leq k \leq K} k! p(k) e^{i\omega k + (z\nu/d)!} e^{-i\omega}
\]  
(S-158)
where we have substituted the saddle point expression for \( \tilde{\omega}(\omega, k) \). This integral can be also written as
\[
\int \frac{d\omega}{2\pi} \sum_{m \leq k \leq K} k! p(k) e^{i\omega k} \sum_{h=0}^{\infty} \frac{(z\nu/d)!}{h!} e^{-i\omega h} = \sum_{m \leq k \leq K} p(k) \left( \frac{z}{d!} \nu^h \right)^k = \left( \left( \frac{z}{d!} \nu^h \right)^k \right).
\]  
(S-159)
Therefore \( c(\omega, k) \) at the saddle point solution can be expressed as
\[
c(\omega, k) = \frac{1}{2\pi} k! p(k) \exp \left[ i\omega k + (z\nu/d)! e^{-i\omega} \right].
\]  
(S-160)
With this expression, using a similar procedure we can express \( \nu \) as
\[
\nu = \int d\omega \sum_{k \leq K} c(\omega, k) e^{-i\omega} = \frac{1}{\langle (z\nu/d)! \rangle} \sum_{k \leq K} k p(k) (z\nu/d)!^{k-1}
\]  
(S-161)
Combing this equation with the third saddle point equation
\[
\frac{z}{d!} \nu^{d+1} = \langle k \rangle,
\]  
(S-162)
it is immediate to show that \( z\nu/d! = 1 \) is a solution with
\[
z = \frac{d!}{\langle k \rangle d!}, \quad \nu = \langle k \rangle.
\]  
(S-163)
By inserting this expression in Eq. (S-160) we get
\[
c(\omega, k) = \frac{1}{2\pi} k! p(k) e^{i\omega k + e^{-i\omega}}.
\]  
(S-164)
Calculating the partition function at the saddle point, we get \( Z(h \to 0) = 1 \). For calculating the marginal distribution \( p_\alpha \) of a simplex \( \alpha \) in the exchangeable network ensemble we first note that given that the ensemble has an exchangeable Hamiltonian, the marginal probability of a simplex should be the same for every simplex of the simplicial complex, i.e. \( p_\alpha = \tilde{p} \). In order to obtain \( \tilde{p} \) we can simply derive the free energy \( F = NF \) with \( f \) given by Eq. (S-154) with respect to the auxiliary field \( h \) obtaining
\[
\left( \begin{array}{c} N \\ d + 1 \end{array} \right) \tilde{p} = - \frac{\partial (N f)}{\partial h} \bigg|_{h=0} = - \frac{\partial (N \Psi)}{\partial h} \bigg|_{h=0} = \frac{N}{(d + 1)!} z \left[ \int d\omega \sum_{m \leq k \leq K} c(\omega, k) e^{-i\omega} \right]^{d+1}
\]  
(S-165)
from which, by approximating the binomial
\[
\left( \begin{array}{c} N \\ d + 1 \end{array} \right) \approx \frac{N^{d+1}}{(d + 1)!}
\]  
(S-166)
for \( N \gg 1 \) and \( d \) finite, and inserting the saddle point value of \( c(\omega, k) \) and \( z \) we get for \( N \gg 1 \),
\[
p_\alpha = \tilde{p} = \sum_{\langle k_i \rangle | m \leq k_i \leq K} \left[ \prod_{r=0}^{d} p(k_r) \right] \langle k_0 k_1, \ldots, k_d \rangle.
\]  
(S-167)
with
\[
p_{\alpha = \{i_0, i_1, \ldots, i_d\} | k_0 k_1 = k_0} = \frac{d! \prod_{r=0}^{d} k_r}{\langle k \rangle N^d}.
\]  
(S-168)
Correlated exchangeable simplicial complex ensemble

The final example of exchangeable ensemble is the ensemble of sparse correlated $d$-dimensional simplicial complexes in which each node has generalized degree $k$ with probability $p(k)$ and each $d$-simplex between $d + 1$ nodes of generalized degrees $\mathbf{k}_a = (k_0, k_1, \ldots, k_r)$ contributes to the probability of the simplicial complex by a term $Q(k_0, k_1, \ldots, k_d) = Q(\mathbf{k}_a)$ where $Q(\mathbf{k}_a)$ is invariant under any permutation of its arguments. Here we impose that the total number of $d$-simplices is $S = kN/(d + 1)$ and that the maximum generalized degree of the simplicial complex is below or equal to $K$ ensuring sparsity and the minimum generalized degree of the simplicial complex is greater or equal to $m$. To this end, we assign to each simplicial complex $\mathcal{K}$ a probability $P(\mathcal{K})$ given by

$$
P(\mathcal{K}) = \prod_{a \in \mathcal{K}} [Q(\mathbf{k}_a)]^{a_0} \prod_{i=1}^{N} p(k_i) e^{-\Sigma(k)} \left( S, \sum_{a \in \mathcal{K}} a_0 \right) \left( K - \max_{i=1 \ldots N} k_i \right) \left( m - \min_{i=1 \ldots N} k_i \right),
$$

where $\Sigma(k)$ is the entropy of the ensemble of correlated networks with degree sequence $\mathbf{k}$ that can be expressed as

$$
\Sigma(k) = \ln \left( ((\langle k \rangle N)!)^{d/(d+1)} ((\gamma(k_i))^{k_i/k_r}) \right) + o(N)
$$

where $\gamma(k)$ is defined self-consistently by the equation

$$
\gamma(k) = \frac{1}{\langle k \rangle !} \sum_{k_1, k_2, \ldots, k_d} Q(k, k_1, k_2, \ldots, k_d) \prod_{r=1}^{d} \left( \frac{p(k_r) k_r}{\gamma(k_r)} \right).
$$

The marginal probability of this ensemble is given by the exchangeable expression

$$
p_a = \sum_{k_0, k_1, \ldots, k_d} \left( \prod_{r=0}^{d} \frac{1}{N} \right) p(k_0, k_1, \ldots, k_d)
$$

with $p(k_0, k_1, \ldots, k_d)$ expressing the marginal probability of a simplex connecting $d + 1$ nodes with degrees $(k_0, k_1, \ldots, k_d)$,

$$
p(k_0, k_1, \ldots, k_d) = \frac{d!}{((\langle k \rangle N)!^d)} \sum_{k_0, k_1, k_2, \ldots, k_d} Q(k_0, k_1, k_2, \ldots, k_d) \prod_{r=0}^{d} \left( \frac{k_r}{\gamma(k_r)} \right)
$$

Derivation of the marginal probability of a simplex in the correlated sparse exchangeable ensemble of simplicial complexes

The partition function of the exchangeable ensemble of sparse correlated simplicial complexes can be written as

$$
Z(h) = \sum_{\mathbf{k}} e^{-h \Sigma(\mathbf{k})} \prod_{a \in \mathcal{K}} \left( \prod_{i=1}^{N} Q(k_i, k_{i_1}, \ldots, k_{i_d})^{a_0} e^{-\Sigma(\mathbf{k})} \left( k_i, \sum_{i_1 < i_2 < \ldots < i_d} a_{i_1, i_2, \ldots, i_d} \right) \delta(S, \sum_{a \in \mathcal{K}} a_0) \right),
$$

with the entropy $\Sigma(\mathbf{k})$ given by

$$
\Sigma(\mathbf{k}) = \ln \left( ((\langle k \rangle N)!)^{d/(d+1)} \right) + \ln \left( \prod_{i=1}^{N} \frac{[\gamma(k_i)]^{k_i/k_r}}{k_r!} \right) + o(N)
$$

where $\gamma(k)$ is determined by the self-consistent equation Eq. (25). By expressing the Kronecker deltas in integral form

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

we get

$$
Z(h) = \frac{(d!)^{-(\langle k \rangle N)/(d+1)}}{((\langle k \rangle N)!)^{d/(d+1)}} \sum_{\mathbf{k}} \int d\omega \int \frac{dA}{2\pi} e^{G(\omega, k, h)}
$$

where $G(\omega, \lambda, \mathbf{k})$ is given by

$$
G(\omega, \lambda, \mathbf{k}, h) = \sum_{i=1}^{N} \left[ i \omega k_i + \ln(k_i! p(k_i)) - k_i \ln(\gamma(k_i)) + i \lambda(k_i)/(d + 1) + \sum_{a \in \mathcal{K}} \ln \left( 1 + Q(k_0, k_1, \ldots, k_d) e^{-i \lambda \Sigma \omega_r - k_r} \right) \right] + \int d\omega \int \frac{dA}{2\pi} e^{G(\omega, \lambda, \mathbf{k}, h)}
$$

S-169

S-170

S-171

S-172

S-173

S-174

S-175

S-176

S-177

S-178
and where $D\omega = \prod_{n=1}^{N}[d\omega_n/(2\pi)]$. Let us now introduce the functional order parameter \[c(\omega, k) = \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i)\delta(k, k_i),\] (S-179)

by enforcing its definition with a series of delta functions introducing for each choice of $(\omega, k)$ the term

\[1 = \int dc(\omega, k) \delta\left(c(\omega, k) - \frac{1}{N} \sum_{i=1}^{N} \delta(\omega - \omega_i)\delta(k, k_i)\right) = \int \frac{d\tilde{c}(\omega, k)dc(\omega, k)}{2\pi/(N\Delta\omega)} e^{\Delta\omega \tilde{c}(\omega, k)/[Nc(\omega, k) - \sum_{i=1}^{N} \delta(\omega - \omega_i)\delta(k, k_i)]}.\] (S-180)

After performing these operations, by imposing $(d + 1)S = \langle k \rangle N$ where $\langle k \rangle = \sum p(k)$, the partition function reads

\[Z = \frac{1}{N} \sum_k Dc(\omega, k) D\tilde{c}(\omega, k) \int \frac{da}{2\pi} e^{\lambda(1,c(\omega, k),\tilde{c}(\omega, k), h)}\] (S-181)

with

\[f(\lambda, c(\omega, k), \tilde{c}(\omega, k), h) = i \int d\omega \sum_k \tilde{c}(\omega, k)c(\omega, k) + i\lambda(k)/(d + 1) + \Psi + \ln \int d\omega \sum_{m \leq K} p(k) \frac{k!}{[\gamma(k)]^k} e^{i\omega k - \tilde{c}(\omega, k)}\] (S-182)

where $\Psi$ is given by

\[\Psi = \frac{N^d}{(d + 1)!} \sum_{k=\langle k_i \rangle, \ldots, \langle k_d \rangle, m \leq K} \int d\omega \prod_{r} c(\omega_r, k_r)Q(k_0, k_1, \ldots, k_d) \ln \left(1 + e^{-ik^1 - \tilde{c}(\omega, k)}\right).\] (S-183)

where $Dc(\omega, k)$ and $D\tilde{c}(\omega, k)$ are functional measures. Performing a Wick rotation in $\lambda$ and assuming $z/N^d = e^{-ik}$ real and much smaller than one, i.e. $z/N^d \ll 1$ which is allowed in the sparse regime $K \ll K_s$, we can linearize the logarithm and express $\Psi$ as

\[\Psi = \frac{z}{(d + 1)!} \sum_{k=\langle k_i \rangle, \ldots, \langle k_d \rangle, m \leq K} \prod_{r=0}^{d} v(k_r),\] (S-184)

where

\[v(k) = \int d\omega c(\omega, k)e^{-i\omega}.\] (S-185)

The saddle point equations determining the value of the partition function read for $h \to 0$

\[-ic(\omega, k) = e^{-i\tilde{c}(\omega, k)}(k)\]

\[c(\omega, k) = \frac{1}{2\pi} \sum_{m \leq K} p(k) \frac{k!}{[\gamma(k)]^k} e^{iak - \tilde{c}(\omega, k)}\]

\[\Psi = \frac{\langle k \rangle}{d + 1}\] (S-186)

where $\tilde{c}(\omega, k)$ as

\[\tilde{c}(\omega, k) = \frac{z}{d!} \sum_{k=\langle k_0 \rangle, \ldots, \langle k_d \rangle, m \leq K} \prod_{r=1}^{d} \left[\int d\omega c(\omega_r, k_r)e^{-i\omega}\right].\] (S-187)

Let us first calculate the integral

\[\frac{1}{2\pi} \int d\omega \sum_{m \leq K} p(k) \frac{k!}{[\gamma(k)]^k} e^{iak - \tilde{c}(\omega, k)} = \frac{1}{2\pi} \int d\omega \sum_{m \leq K} \frac{k!}{[\gamma(k)]^k} p(k) e^{iak + \tilde{c}(\omega, k)}\] (S-188)

where we have substitute the saddle point expression for $\tilde{c}(\omega, k)$. This integral can be also written as

\[\int d\omega \sum_{m \leq K} \frac{k!}{[\gamma(k)]^k} p(k) e^{iak} \sum_{h=0}^{\infty} \left(\frac{\tilde{c}(\omega, k)}{h!}\right)^k e^{-ih\omega} = \sum_{m \leq K} p(k) \left(\frac{\tilde{c}(\omega, k)}{\gamma(k)}\right)^k.\] (S-189)
Let $w$ indicate the value of this integral, i.e.
\[
    w = \sum_{m \leq k \leq K} p(k) \left( \frac{\tilde{y}(k)}{\gamma(k)} \right)^k .
\]
\begin{equation}
    (S-190)
\end{equation}

The functional order parameter $c(\omega, k)$ at the saddle point solution can be expressed as
\[
    c(\omega, k) = \frac{k! p(k)}{2\pi \gamma(k)^d} e^{iak + \omega \gamma(k) k} .
\]
\begin{equation}
    (S-191)
\end{equation}

With this expression, using a similar procedure we can express $\tilde{y}(k)$ as
\[
    \tilde{y}(k) = \frac{z}{d! w d} \sum_{k_0 \leq k_1 \leq \ldots \leq k_d \leq K} Q(k_0, k_1, \ldots, k_d) \prod_{r=1}^d \left[ p(k_r) \frac{k_r}{\tilde{y}(k_r)} \left( \frac{\tilde{y}(k_r)}{\gamma(k_r)} \right)^k_r \right] .
\]
\begin{equation}
    (S-192)
\end{equation}

Combing this equation with the third saddle point equation we get
\[
    \Psi = \frac{1}{w} \sum_{m \leq k \leq K} \left[ p(k) k \left( \frac{\tilde{y}(k)}{\gamma(k)} \right)^k \right] = (k) .
\]
\begin{equation}
    (S-193)
\end{equation}

Given that $\gamma(k)$ is defined though the Eq. (S-171), it follows that $\tilde{y}(k) = \gamma(k)$, $w = 1$, $z = \frac{d!}{(k)^d}$. Finally using Eqs. (S-194) we can derive the final expression for $c(\omega, k)$ given by
\[
    c(\omega, k) = \frac{k! p(k)}{2\pi \gamma(k)^d} e^{iak + \omega \gamma(k) k} .
\]
\begin{equation}
    (S-195)
\end{equation}

From this equation of the functional order parameter we can derive the marginal for each link of the network which is given by
\[
    p_a = \frac{d!}{((k)^d) N^d} \sum_{k_0 \leq k_1 \leq \ldots \leq k_d \leq K} Q(k_0, k_1, \ldots, k_d) \prod_{r=1}^d \left[ \int d\omega c(\omega, k_r) e^{-i\omega} \right] .
\]
\begin{equation}
    (S-196)
\end{equation}

yielding,
\[
    p_a = \sum_{k_0 \leq k_1 \leq \ldots \leq k_d \leq K} \prod_{r=0}^d \left[ p(k_r) \right] p(k_0, k_1, \ldots, k_r) .
\]
\begin{equation}
    (S-197)
\end{equation}

Here $P(k_0, k_1, \ldots, k_d)$ indicates the probability of a link between node $i$ and node $j$ conditioned to the degree of the two nodes $k_i = k$ and $k_j = k'$, i.e.
\[
    p_{a=[k_i, k_1, \ldots, k_d]} = \frac{d!}{((k)^d) N^d} Q(k_0, k_1, \ldots, k_d) \prod_{r=0}^d \left[ \frac{k_r}{\gamma(k_r)} \right] .
\]
\begin{equation}
    (S-198)
\end{equation}

Note that for $Q(k_0, k_1, \ldots, k_d) = 1$ it follows that $\gamma(k) = 1$, and for $Q(k_0, k_1, \ldots, k_d) = \prod_{r=0}^d k_r$ it follows $\gamma(k) = k$ and hence in both cases we recover the exchangeable ensembles of uncorrelated simplicial complexes.