

## Percolation transition and distribution of connected components in generalized random network ensembles

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2009 J. Phys. A: Math. Theor. 42 195007

(<http://iopscience.iop.org/1751-8121/42/19/195007>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.153

The article was downloaded on 03/06/2010 at 07:38

Please note that [terms and conditions apply](#).

# Percolation transition and distribution of connected components in generalized random network ensembles

Serena Bradde<sup>1,2</sup> and Ginestra Bianconi<sup>3</sup>

<sup>1</sup> International School for Advanced Studies, via Beirut 2/4, 34014, Trieste, Italy

<sup>2</sup> INFN, Via Valerio 2, Trieste, Italy

<sup>3</sup> The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014, Trieste, Italy

E-mail: [bradde@sissa.it](mailto:bradde@sissa.it) and [gbiancon@ictp.it](mailto:gbiancon@ictp.it)

Received 21 January 2009, in final form 27 February 2009

Published 23 April 2009

Online at [stacks.iop.org/JPhysA/42/195007](http://stacks.iop.org/JPhysA/42/195007)

## Abstract

In this work, we study the percolation transition and large deviation properties of generalized canonical network ensembles. This new type of random networks might have a very rich complex structure, including high heterogeneous degree sequences, non-trivial community structure or specific spatial dependence of the link probability for networks embedded in a metric space. We find the cluster distribution of the networks in these ensembles by mapping the problem to a fully connected Potts model with heterogeneous couplings. We show that the nature of the Potts model phase transition, linked to the birth of a giant component, has a crossover from second to first order when the number of critical colors  $q_c = 2$  in all the networks under study. These results shed light on the properties of dynamical processes defined on these network ensembles.

PACS numbers: 00.00, 20.00, 42.10

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

Recently the study of critical phenomena in complex networks has attracted a great deal of interest [1]. One of the main critical phenomena occurring in networks is the percolation transition which is a continuous structural phase transition that can be characterized by critical indices as a statistical mechanics second-order phase transition. This phase transition determines the robustness properties of complex networks [2–5] and the critical temperature of the Ising [6–8] and XY models [9, 10] on complex networks. Moreover, the onset of a percolating cluster determines a transition in between a phase in which small loops are suppressed and a phase in which the expectation value of small loops is positive in the limit of large network sizes [11].

The percolation phase transition in Erdős and Renyi networks is a classic subject of graph theory [12]. For this network ensemble the large deviation of the number of connected components (or clusters) has been characterized [13] by a mapping of the problem to a fully connected Potts model [14].

In uncorrelated complex networks, characterized by a non-Poisson degree distribution, the percolation transition depends on the second moment of the degree distribution [4, 5] and can show non-trivial critical exponents [1].

This phase transition has also been studied in directed networks [15] and in networks with degree–degree correlations [16].

In this paper, we study the percolation properties and the large deviation of the cluster distribution of the recently proposed generalized canonical random network ensembles [17, 18] with non-trivial degree distribution and an additional community structure or spatial structure. These network ensembles can be cast in the wide category of configuration or ‘hidden variable’ models extensively studied in the recent literature [2, 19–23]. The percolation properties and the large deviations of the cluster distribution in these ensembles are studied in this paper by mapping the problem to a fully connected Potts model with heterogeneous couplings. We find results in agreement with [24] where the Potts model formulation was first used for the study of the percolation properties of complex networks with heterogeneous degrees. In particular, our framework generalizes the results of [24] and can be applied in network ensembles with very diverse structure, not only network ensembles with heterogeneous degree distribution, but also network ensembles with an additional non-trivial community or spatial structure.

The paper is organized as follows. In section 2, we introduce the generalized canonical random ensembles. In section 3, we introduce the generating functions for the cluster distribution and we characterize its large deviations. In section 4, we relate the problem of finding the cluster distribution in a generalized canonical model, and their percolation transition, to the study of a fully connected Potts model with heterogeneous couplings. In section 5, we solve the fully connected Potts model with heterogeneous couplings and we find the percolation threshold and critical exponent  $\beta$  for the generalized canonical network ensembles. In section 6, we find the cluster distribution in the generalized canonical network ensembles. In section 7, we compare our theoretical predictions with simulation results. Finally, in section 8, we give the conclusions.

## 2. Random network ensembles

In this section, we introduce the generalized random network ensembles described in [17, 18]. The generalized random network ensembles are an extension of the known  $G(N, M)$  and  $G(N, p)$  random network ensembles and are related to configuration and ‘hidden variable’ ensembles [2, 19–23].

### 2.1. The $G(N, M)$ and $G(N, p)$ random network ensembles

The mathematical literature has widely studied the properties of the  $G(N, M)$  and  $G(N, p)$  random network ensembles.

- A random network in the ensemble  $G(N, M)$  is a network having  $N$  nodes and  $M$  undirected links. If we indicate by  $a_{ij}$  the adjacency matrix of the network (with  $a_{ij} = 1$  if there is a link between node  $i$  and  $j$  and  $a_{ij} = 0$  otherwise), the probability that a network  $\mathcal{G}$ , associated with the adjacency matrix  $\mathbf{a}$ , belongs to the  $G(N, M)$  ensemble

is given by

$$P(\mathcal{G}) = \frac{1}{Z} \delta \left( M, \sum_{i < j} a_{ij} \right) \tag{1}$$

with

$$Z = \binom{N(N-1)}{2M} \tag{2}$$

and with  $\delta(\cdot)$  indicating the Kronecker delta. The probability of each link in this ensemble of networks is given by  $p = M/(N(N-1)/2)$ .

- A network in the  $G(N, p)$  ensemble is a network in which each possible pair of links is present with probability  $p$ . Therefore the probability of a specific network  $\mathcal{G}$  in this ensemble is equal to

$$P_C(\mathcal{G}) = \prod_{i < j} p^{a_{ij}} (1-p)^{1-a_{ij}}, \tag{3}$$

where  $a_{ij}$  is the adjacency matrix. In the  $G(N, p)$  ensemble the total number of links  $M$  is not fixed but is Poisson distributed with mean  $\langle M \rangle_{P_C(G)} = pN(N-1)/2$ .

The  $G(M, N)$  and the  $G(N, p)$  ensemble with  $p = M/(N(N-1)/2)$  are linked by a Legendre transform, and, in the asymptotic limit of  $N \rightarrow \infty$ , they share the same statistical properties.

### 2.2. Generalized random network ensembles

Recently a statistical mechanics approach has been proposed [17, 18] that is able to generalize the random network ensembles to network ensembles with much more complex structure including networks with highly heterogeneous degree sequences and non-trivial community structure or spatial dependence of the link probability. The statistical mechanics approach is able to describe both ‘microcanonical’ network ensembles (that satisfy hard structural constraints and generalize the  $G(N, M)$  random ensembles) and ‘canonical’ network ensembles (that satisfy the structural constraints when their properties are averaged over the whole ensemble and generalize the  $G(N, p)$  ensemble).

- The ‘microcanonical’ networks have to satisfy a series of hard constraints  $\vec{F}(\mathcal{G}) = \vec{C}$  and the probability of these networks is given by

$$P_{MC}(\mathcal{G}) = \frac{1}{Z} \delta(\vec{F}(\mathcal{G}) - \vec{C}), \tag{4}$$

with  $Z$  indicating the cardinality of the ensemble. The probability of each link  $p_{ij}$  is computed introducing some Lagrange multipliers [17, 18].

- The ‘canonical’ conjugated ensemble can be built starting from the probability of the links  $p_{ij}$  in the ‘microcanonical’ one. We assign to each network  $\mathcal{G}$  the probability

$$P_C(\mathcal{G}) = \prod_{i < j} p_{ij}^{a_{ij}} (1-p_{ij})^{1-a_{ij}}, \tag{5}$$

which generalizes (3) to heterogeneous networks. In the ‘canonical’ ensembles the structural constraints are satisfied on average

$$\overline{\vec{F}(\mathcal{G})} = \vec{C}. \tag{6}$$

Here and in the following we always indicate by  $\overline{\dots}$  the average over the ensemble probability  $P_C(\mathcal{G})$  given by (5) and by  $\langle \dots \rangle$  the average over all the nodes  $i = 1, \dots, N$ .

In this paper, we focus on generalized ‘canonical’ networks. Each node  $i$  in this ensemble is characterized by two discrete *hidden variables*  $\theta_i$  and  $\alpha_i$ . We consider in this paper the link probability given by

$$p_{ij} = \frac{\theta_i \theta_j W(\alpha_i, \alpha_j)}{1 + \theta_i \theta_j W(\alpha_i, \alpha_j)} \quad (7)$$

which is fully specified once the function  $W(\alpha, \alpha')$  is given. The link probability (7) corresponds to maximally entropic ensembles with given degree and structural constraints related to the hidden variables  $\alpha_i$  of the nodes [17, 18].

In the ensembles described by (7), the degree of each node  $k_i$  is a Poisson variable [23] with average

$$\bar{k}_i = \sum_{j \neq i} p_{ij}. \quad (8)$$

In the following, we specifically comment on some relevant limiting cases for the general distribution (7).

- *The  $G(N, p)$  ensemble.* If the values of the *hidden variables*  $\theta$ 's are equal, i.e.  $\theta_i = \theta \forall i$  and  $W(\alpha, \alpha') = 1 \forall \alpha, \alpha'$ , the probability of a link is given by

$$p_{ij} = p = \theta^2 / (1 + \theta^2). \quad (9)$$

The degree of each node is a Poisson variable with equal average  $\bar{k} = pN$ . Performing also the average  $\langle \cdot \rangle$  over all the nodes of the network we get

$$p = \frac{\langle \bar{k} \rangle}{N}. \quad (10)$$

We recover therefore the Erdős and Renyi ensembles  $G(N, p)$  by taking

$$\theta = \sqrt{\frac{\langle \bar{k} \rangle / N}{1 - \langle \bar{k} \rangle / N}} \simeq \sqrt{\langle \bar{k} \rangle / N}, \quad (11)$$

where the last expression is valid for sparse networks with  $\langle \bar{k} \rangle$  finite.

- *The configuration model.* If the linking probability  $p_{ij}$  of equation (7) depends only on  $\theta_i$  and  $\theta_j$  (i.e.  $W(\alpha, \alpha') = 1 \forall \alpha, \alpha'$ ), then

$$p_{ij} = \frac{\theta_i \theta_j}{1 + \theta_i \theta_j}. \quad (12)$$

This ensemble is the canonical version of the configuration model, each node  $i$  having a degree  $k_i$  distributed according to a Poisson variable with average

$$\bar{k}_i = \sum_{j \neq i} \frac{\theta_i \theta_j}{1 + \theta_i \theta_j}. \quad (13)$$

This ensemble has in general non-trivial degree correlations that disappear for  $\max_i(\theta_i) \ll 1$ . In this last case,  $\max_i(\theta_i) \ll 1$  the linking probability  $p_{ij}$  defined in equation (13) can be approximated as

$$p_{ij} = \theta_i \theta_j. \quad (14)$$

Therefore, in this limit, the networks of the ensemble are uncorrelated and there is a simple relation between the hidden variables  $\theta_i$  and the average degree  $\bar{k}_i$  of the node  $i$ , i.e.

$$\theta_i = \frac{\bar{k}_i}{\sqrt{\langle \bar{k} \rangle N}}. \quad (15)$$

Finally, we observe that if we use (15) the linking probability  $p_{ij}$  can be expressed in the well-known expression for uncorrelated networks

$$p_{ij} = \frac{\bar{k}_i \bar{k}_j}{\langle k \rangle N}. \quad (16)$$

• *Structured networks.* In the more general structured case, we have two possibilities:

- (i) The index  $\alpha = 1, \dots, A$  with  $A = \mathcal{O}(N^{1/2})$  can indicate the community of a node and the function  $W(\alpha, \alpha')$  can be an  $A \times A$  matrix. In this case the number of links  $L(\alpha, \alpha')$  between the community  $\alpha$  and the community  $\alpha'$  will be distributed according to a Poisson distribution with average

$$\overline{L(\alpha, \alpha')} = \sum_{i < j} p_{ij} \delta(\min(\alpha_i, \alpha_j), \alpha) \delta(\max(\alpha_i, \alpha_j), \alpha'). \quad (17)$$

- (ii) The index  $\alpha$  can indicate a position in a metric space which determines the link probability. In this case the function  $W(\alpha, \alpha')$  is a vector depending only on the metric distance  $d[\alpha, \alpha']$ , i.e.  $W = W(d[\alpha, \alpha'])$ .

For structured networks with a generic distribution of  $\theta$ 's and a non-trivial function of  $W(\alpha, \alpha')$  we can consider the limit when the  $[\max_i(\theta)]^2 [\max_{\alpha, \alpha'} W(\alpha, \alpha')] \ll 1$ . In this limit the linking probability  $p_{ij}$  given by equation (7) reduces to the simple form

$$p_{ij} = \theta_i \theta_j W(\alpha_i, \alpha_j) \quad (18)$$

and we have

$$\theta_i = \frac{\bar{k}_i}{\sum_{j \neq i} \theta_j W(\alpha_i, \alpha_j)} \simeq \frac{\bar{k}_i}{\mathcal{N}_{\alpha_i}} \quad (19)$$

with  $\mathcal{N}_{\alpha_i} = \sum_{j \neq i} \theta_j W(\alpha_i, \alpha_j)$ .

### 3. Large deviation of the cluster distribution

The number of connected components  $C(\mathcal{G})$  or ‘clusters’ of a network  $\mathcal{G}$  gives direct information on the topological structure of the network and their percolating properties. Indeed if  $C(\mathcal{G})/N$  is small there are few large connected components while in the opposite case the network is divided into a huge number of small clusters. In the limit of large network sizes  $N \rightarrow \infty$  each canonical generalized network ensemble will be characterized by a typical value of the number of clusters  $C^*(N)$ . The typical distribution of clusters gives the percolating properties of the networks belonging to the ensemble and will be able to characterize the critical exponents of the percolation phase transition. Moreover, different network realizations  $\mathcal{G}$  of a generalized canonical ensemble will have a number of clusters  $C(\mathcal{G})$  which is subject to large deviations with respect to the typical value  $C^*(N)$ .

Given the probability of a network  $P_c(\mathcal{G})$  in the canonical generalized random ensembles, as defined in equation (5), we can define the probability density  $\hat{P}(C)$  of generating a random network  $\mathcal{G}$  in this ensemble with  $C$  clusters as in the following:

$$\hat{P}(C) = \sum_{\mathcal{G}} P_c(\mathcal{G}) \delta(C, C(\mathcal{G})). \quad (20)$$

In the thermodynamic limit,  $N \rightarrow \infty$ , the probability  $\hat{P}(C)$  is centered at some typical value  $C^*$  and decays extremely fast away from  $C^*$  in the large network limit. Let us indicate by  $c = C/N$  the number of connected components per vertex, the typical value of this quantity

converges in the thermodynamic limit to a size independent value  $c^*$ . Therefore, in order to characterize  $\hat{P}(C)$  in the thermodynamic limit, we consider the function  $\omega(c)$  defined as

$$\omega(c) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln \hat{P}(C) \tag{21}$$

implying clearly  $\omega(c) \leq 0$  for all  $c \in [0, 1]$  and  $\omega(c^*) = 0$ .

Finally we introduce the generating function  $Y(q)$  of the cluster probability  $\hat{P}(C)$

$$Y(q) = \sum_c \hat{P}(C) q^C = \sum_{\mathcal{G}(\mathbf{a})} \prod_{ij} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} q^{C(\mathcal{G})}, \tag{22}$$

where in the last expression we have used equation (5) defining the generalized random network ensembles. We characterize the asymptotic limit of the cluster generating function  $Y(q)$  by  $\phi(q)$  defined as

$$\phi(q) = \lim_{N \rightarrow \infty} \frac{1}{N} \ln Y(q). \tag{23}$$

From equation (22) we obtain, with a saddle point calculation, that the conjugated Legendre transform of the quantity  $\omega(c)$  can be expressed in terms of  $\phi(q)$  according to the relation

$$\omega(c) = \min_q [\phi(q) - c \log q]. \tag{24}$$

The cluster distribution is therefore fully characterized in the asymptotic limit if we know the function  $\phi(q)$ .

#### 4. The fully connected heterogeneous Potts model and the percolation transition of the generalized random network ensembles

In this section, we will reduce the problem of finding the cluster distribution in generalized canonical random network ensembles to the study of mean-field Potts models with heterogeneous couplings. We will prove that  $\phi(q)$ , given by (23), has a formal relation with the free energy of the mean field Potts model with heterogeneous couplings, after a suitable analytic continuation. This relation generalizes the known connection between the fully connected Potts model and the generating function of the cluster distribution of a random  $G(N, p)$  network [13, 14].

In order to present the results of the paper in a self-contained way we describe here the cluster expansion of the fully connected Potts model. The Potts model is a well-known statistical mechanical problem [25] describing  $N$  classical degrees of freedom  $\sigma_i$  associated with the nodes  $i = 1, \dots, N$  of a given network. Each variable  $\sigma_i$  can take  $q$  different values, namely  $\sigma_i = 0, \dots, q - 1$ , and is coupled to all the other degrees of freedom  $\sigma_j$  by means of a two-body interaction of strength  $J_{ij}$ . This interaction favors configurations where all the nodes in the network have the same value of  $\sigma$ . Thus the energy reads

$$E[\{\sigma\}] = - \sum_{i < j} \delta(\sigma_i, \sigma_j) J_{ij} - h \sum_{\sigma} u_{\sigma} \sum_i \delta(\sigma, \sigma_i), \tag{25}$$

where we assume that all the couplings are positive,  $J_{ij} > 0$  and that the first sum in (25) runs over all the pairs of nodes of a fully connected network. Moreover, we take the auxiliary field  $hu_{\sigma}$  parallel to the direction  $\sigma$ . The partition function of the model is

$$Z = \sum_{\{\sigma_i=0, \dots, q-1\}} \exp(-\beta E[\{\sigma\}]), \tag{26}$$

where  $\beta$  is the inverse temperature and the summation runs over all  $q^N$  spin configurations. In order to map the Potts model to the cluster structure of the generalized random network ensembles, we expand the partition function  $Z$  following the article [14]:

$$Z[\{v_{ij}\}, h] = \sum_{\sigma} \prod_{i < j} [1 + v_{ij} \delta(\sigma_i, \sigma_j)] e^{\beta h \sum_{\sigma} u_{\sigma} \sum_i \delta(\sigma_i, \sigma)}, \quad (27)$$

where we have defined

$$v_{ij} = e^{\beta J_{ij}} - 1. \quad (28)$$

Expanding equation (27) we obtain

$$Z[\{v_{ij}\}, h] = \sum_{\sigma} e^{\beta h \sum_{\sigma} u_{\sigma} \sum_i \delta(\sigma_i, \sigma)} \left[ 1 + \sum_{ij} v_{ij} \delta(\sigma_i, \sigma_j) + \sum_{i < j, k < l (ij) \neq (kl)} v_{ij} v_{kl} \delta(\sigma_i, \sigma_j) \delta(\sigma_k, \sigma_l) + \dots \right]. \quad (29)$$

Each term in the expansion (29) corresponds to a possible network  $\mathcal{G}$  formed by a subset  $E(\mathcal{G})$  of edges on the  $N$  complete network. Each contribution from a network  $\mathcal{G}$  is weighted by the probability  $\prod_{ij \in E(\mathcal{G})} v_{ij}$  and the sum is made over all possible networks  $\mathcal{G}$  of  $N$  nodes. Using this expansion, after performing the sum over the configurations  $\{\sigma_i = 0, \dots, q - 1\}$ , we can write the partition function reported in (27), in the form:

$$Z[\{v_{ij}\}, h] = \sum_{\mathcal{G}} \prod_{ij \in E(\mathcal{G})} v_{ij} \prod_{n=0}^{C(\mathcal{G})-1} \left( \sum_{\sigma} e^{\beta h u_{\sigma} S_n} \right), \quad (30)$$

with  $E(\mathcal{G})$  given by the set of all edges in  $\mathcal{G}$ ,  $C(\mathcal{G})$  given by the number of connected components in the network and  $S_n$  denotes the size of the  $n$ th component. From the previous equation it follows that in the absence of the external field

$$Z[\{v_{ij}\}, h = 0] = \sum_{\mathcal{G}} \prod_{ij \in E(\mathcal{G})} v_{ij} q^{C(\mathcal{G})}. \quad (31)$$

By comparing the definition of the cluster generating function (22) and expression (31) for the partition function of the Potts model, we observe that the two functions can be related by the following simple expression:

$$Y(q) = \prod_{i < j} (1 - p_{ij}) Z[\{v_{ij} = p_{ij}(1 - p_{ij})^{-1}\}, h = 0] \quad (32)$$

and the associated logarithmic function reads

$$\phi(q) = \sum_{i < j} \ln(1 - p_{ij}) - f[\{v_{ij}\}], \quad (33)$$

where  $v_{ij} = p_{ij}(1 - p_{ij})^{-1}$  and  $f$  is defined at the null external field  $h = 0$ . In the high temperature limit  $\beta \rightarrow 0$  the couplings  $J_{ij}$  given by (28) are linked to the edge probability by means of equation (28) so that

$$v_{ij} = \frac{p_{ij}}{(1 - p_{ij})} \sim \beta J_{ij}. \quad (34)$$

Therefore in order to find the cluster generating function we can simply solve the fully connected Potts model with heterogeneous couplings. Any assumption on the network ensemble will have a direct counterpart on the structure of the couplings in the Potts model.

We will solve the model in this framework, specializing the results for the cases of our interest (7). Using equation (34) we obtain

$$v_{ij} = \beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) = \theta_i \theta_j W(\alpha_i, \alpha_j). \quad (35)$$

In the various different cases under study the function  $J(\theta_i, \theta_j, \alpha_i, \alpha_j)$  takes different values.

- *The  $G(N, p)$  ensemble.* For the characterization of the cluster distribution of a Poisson random network in the  $G(N, p)$  ensemble with  $p = \langle k \rangle / N$  we take

$$\beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) \simeq \frac{\langle \bar{k} \rangle}{N} \quad (36)$$

for all pairs  $i, j$ .

- *The configuration model.* For the characterization of the cluster distribution in the configuration model we take

$$\beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) = \theta_i \theta_j. \quad (37)$$

In the case of an uncorrelated network we have  $\theta_i \ll 1$  and we can express the hidden variables  $\theta_i$  in terms of the expected average degree  $\bar{k}_i$ , as  $\theta_i = \bar{k}_i / \sqrt{\langle \bar{k} \rangle} N$ . Consequently the couplings of the Potts model take the form

$$\beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) = \frac{\bar{k}_i \bar{k}_j}{\langle \bar{k} \rangle N}. \quad (38)$$

- For the characterization of the cluster distribution in the structured network ensemble with a community structure or spatial dependence on the embedding geometric space, we have

$$\beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) = \theta_i \theta_j W(\alpha_i, \alpha_j). \quad (39)$$

In the case in which  $(\max_i \theta_i)^2 [\max_{\alpha, \alpha'} W(\alpha, \alpha')] \ll 1$  the previous equation simplifies

$$\beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) = \bar{k}_i \bar{k}_j \frac{W(\alpha_i, \alpha_j)}{N_{\alpha_i} N_{\alpha_j}}. \quad (40)$$

For  $q \rightarrow 1$  the properties of the partition function (31) are in correspondence with the percolation properties [14] of the generalized canonical network ensembles with linking probabilities  $p_{ij}$  given by (7). We will sketch the proof following [26]. It is straightforward that in the limit  $q \rightarrow 1$ , the partition function  $Z[\{v_{ij}\}, h] = \prod_{i < j} (1 - p_{ij})^{-1}$  so that

$$f_1 = \lim_{\substack{N \rightarrow \infty \\ q \rightarrow 1}} \frac{\ln Z[\{v_{ij}\}, h] + \sum_{i < j} \ln(1 - p_{ij})}{N(q - 1)} = \lim_{N \rightarrow \infty} \frac{\partial \ln Z[\{v_{ij}\}, h]}{N \partial q} \Big|_{q=1}. \quad (41)$$

We could choose the parameter  $u_\sigma = \delta(\sigma, 0)$ , so that the external field favors the  $\sigma = 0$  state, the partition function reported in (30) simplifies

$$Z[\{v_{ij}\}, h] = \sum_{\mathcal{G}} \prod_{ij \in E(\mathcal{G})} v_{ij} \prod_{n=0}^{C(\mathcal{G})-1} (q - 1 + e^{h S_n}). \quad (42)$$

Using the fact that  $\sum_n f(S_n) = \sum_S \sum_n \delta(S_n - S) f(S) = \sum_S C(S) f(S)$ , where  $S$  is the number of nodes in the same cluster and  $C(S)$  is the number of clusters with  $S$  nodes, we obtain that the previous equation becomes

$$Z[\{v_{ij}\}, h] = \sum_{\mathcal{G}} \prod_{ij \in E(\mathcal{G})} v_{ij} e^{\sum_S C(S) \ln(q - 1 + e^{h S})}. \quad (43)$$

Performing the summation over the graphs with a saddle-point approximation, we obtain that in the thermodynamic limit equation (41) is

$$f_1 = \frac{1}{Z_1} \left\langle \sum_{s(G)} c(s) e^{-hs} \right\rangle = \sum_s c(s) e^{-hs}, \quad (44)$$

where  $s = S/N$  and  $c(s) = C(S)/N$ . Differentiating the previous equation with respect to the external field we obtain that the node probability to be in the percolating cluster is linked to the free energy function of the Potts model in the limit  $q \rightarrow 1$

$$\lim_{h \rightarrow 0^+} 1 + \frac{\partial f_1}{\partial h} = 1 - \sum_s s c(s) = \mathcal{P}(\{p_{ij}\}). \quad (45)$$

While the second derivative gives the mean clusters per nodes. Using the Potts model, we are also able to compute the probability of the two given nodes that belong to the percolating component. Let us introduce the node–node correlation in the limit  $h \rightarrow 0$

$$D_{ij}(q) = \sum_{\sigma} e^{-\beta E(\sigma)} \delta_{\sigma_i, \sigma_j} = \langle \delta_{\sigma_i, \sigma_j} \rangle \quad (46)$$

that measures the probability of the two nodes that have the same color. We could easily compute this quantity and we obtain

$$\lim_{q \rightarrow 1} \frac{q}{q-1} D_{ij}(q) = 1 - \langle \mathcal{C}_{ij} \rangle, \quad (47)$$

where  $\mathcal{C}_{ij}$  is the indicator function: if nodes  $i$  and  $j$  are in the same cluster it has the value 1, otherwise it vanishes. We want to underline the fact that the probability that the two nodes are in the same non-percolating component is defined through the following relation:

$$\mathcal{H}_{ij} = \langle \mathcal{C}_{ij} \rangle - \mathcal{P}^2(\{p_{kl}\}). \quad (48)$$

This shows how solving the Potts model in the  $q \rightarrow 1$  limit gives us information on the percolating transition in a generalized network ensemble.

### 5. Free energy of the Potts model and the percolation phase transition

In order to solve the mean-field Potts model we introduce the order parameters

$$c_{\theta\alpha}(\sigma) = \frac{1}{N_{\theta\alpha}} \sum_{i=1}^N \delta(\sigma, \sigma_i) \delta(\theta, \theta_i) \delta(\alpha, \alpha_i), \quad (49)$$

where

$$N_{\theta\alpha} = \sum_i \delta(\theta, \theta_i) \delta(\alpha, \alpha_i) \quad (50)$$

are the number of nodes with given *hidden variables*  $\theta$  and  $\alpha$ . The order parameters  $c_{\theta\alpha}$  satisfy their proper normalization

$$\sum_{\sigma} c_{\theta\alpha}(\sigma) = 1. \quad (51)$$

The energy of the Potts model in the absence of the external field  $h = 0$ , expressed in terms of the order parameters  $c_{\theta\alpha}(\sigma)$ , takes the form

$$E[\{c_{\theta\alpha}(\sigma)\}] = -\frac{N^2}{2} \sum_{\sigma, \theta, \theta', \alpha, \alpha'} p_{\theta\alpha} p_{\theta'\alpha'} c_{\theta\alpha}(\sigma) c_{\theta'\alpha'}(\sigma) J(\theta, \theta', \alpha, \alpha') + O(N), \quad (52)$$

where we have explicitly shown the dependence of the coupling from external parameters  $\theta$  and  $\alpha$ . In order to express the partition function as a sum over the collective variables  $c_{\theta\alpha}(\sigma)$ , we need to take into account the entropic contribution, counting the number of microscopic configurations with a given value of  $c_{\theta\alpha}(\sigma)$ . To the leading order in  $N$  we get

$$Z = \sum_{\{c_{\theta\alpha}(\sigma)\}} e^{-\beta E[c_{\theta\alpha}(\sigma)]} \prod_{\theta, \alpha} \left( \frac{N_{\theta\alpha}!}{\prod_{\sigma} [N c_{\theta\alpha}(\sigma)]!} \right) = \sum_{c_{\theta\alpha}(\sigma)} e^{-\beta N f[\{c_{\theta\alpha}(\sigma)\}]}, \quad (53)$$

where the free energy density functional reads

$$\begin{aligned} \beta f[\{c_{\theta\alpha}(\sigma)\}] = & -\frac{N}{2} \sum_{\sigma \theta \theta' \alpha \alpha'} p_{\theta\alpha} p_{\theta'\alpha'} c_{\theta\alpha}(\sigma) c_{\theta'\alpha'}(\sigma) \beta J(\theta, \theta', \alpha, \alpha') \\ & + \sum_{\sigma \theta\alpha} p_{\theta\alpha} c_{\theta\alpha}(\sigma) \ln c_{\theta\alpha}(\sigma). \end{aligned} \quad (54)$$

In the large- $N$  limit one can evaluate the sum in (53) by the saddle-point method. As a function of  $q$ , the Potts model undergoes a phase transition. For  $q \leq q_c$  the order parameter is invariant under the permutation of the spin values  $\sigma = 0, \dots, q-1$ . Nevertheless above the percolation transition, for  $q > q_c$  the ground state breaks the symmetry of the Hamiltonian.

### 5.1. Symmetric saddle point

The free energy of the Potts model is invariant under the permutation of the  $q$  colors. When this symmetry is also shared by the ground state, the fraction of nodes of a given color could be written as

$$c_{\theta\alpha}(\sigma) = \frac{1}{q}, \quad (55)$$

which ensures that different colors are identical. Inserting this ansatz in equation (54) we get

$$\beta f^{\text{sym}}(q) = -\frac{N}{2q} \beta \sum_{\theta \theta' \alpha \alpha'} p_{\theta\alpha} p_{\theta'\alpha'} J(\theta, \theta', \alpha, \alpha') - \ln q. \quad (56)$$

Computing the second-order derivative of the free energy density functional, we can study the stability of the symmetric solution. When the eigenvalue of the Hessian matrix of the free energy (54) changes sign and becomes negative the ansatz (55) is no longer correct. The Hessian matrix reads

$$\begin{aligned} H_{\theta\theta'\alpha\alpha'}(\sigma, \tau) = & \frac{\partial^2 \beta f[c_{\ell\beta}(\rho)]}{\partial c_{\theta\alpha}(\sigma) \partial c_{\theta'\alpha'}(\tau)} = \delta(\sigma, \tau) \\ & \times \left[ \delta(\theta, \theta') \delta(\alpha, \alpha') \frac{p_{\theta\alpha}}{c_{\theta\alpha}(\sigma)} - p_{\theta\alpha} p_{\theta'\alpha'} N \beta J(\theta, \theta', \alpha, \alpha') \right] \end{aligned} \quad (57)$$

and the related eigenvalue problem is

$$(\lambda_{\theta\alpha} - p_{\theta\alpha} q) e_{\theta\alpha} = -p_{\theta\alpha} M_{\theta\alpha}, \quad (58)$$

where the quantity  $M_{\theta\alpha}$  is defined as

$$M_{\theta\alpha} = \sum_{\theta'\alpha'} p_{\theta'\alpha'} N \beta J(\theta, \theta', \alpha, \alpha') e_{\theta'\alpha'}. \quad (59)$$

Inserting equation (58) into (59), we find

$$M_{\theta\alpha} = - \sum_{\theta'\alpha'} \frac{p_{\theta'\alpha'}^2 N \beta J(\theta, \theta', \alpha, \alpha')}{\lambda_{\theta'\alpha'} - p_{\theta'\alpha'} q} M_{\theta'\alpha'}, \quad (60)$$

defining the eigenvalues of the Hessian matrix in (57). In order to obtain the critical values for the external parameters that cause instability in the free energy density, we have to find when eigenvalues change sign. Upon imposing  $\lambda_{\theta\alpha} = 0$  we find this condition is

$$M_{\theta\alpha} = \sum_{\theta'\alpha'} \frac{1}{q} p_{\theta'\alpha'} N \beta J(\theta, \theta', \alpha, \alpha') M_{\theta'\alpha'}. \quad (61)$$

In the general case  $\beta J(\theta, \theta', \alpha, \alpha') = \theta\theta' W(\alpha, \alpha')$ , the stability condition can be expressed as

$$q \leq q_c = \Lambda \quad (62)$$

with  $\Lambda$  indicating the maximal eigenvalue of the matrix

$$K_{\alpha,\alpha'} = N \sum_{\theta} p_{\theta\alpha'} \theta^2 W(\alpha, \alpha'). \quad (63)$$

In the following, we study in detail the critical point  $q_c$  defined by (62) and (63) in a few relevant cases of the generalized network ensembles.

- *The  $G(N, p)$  ensemble.* In the special case of the networks in the  $G(N, p)$  ensemble networks with a delta-like distribution  $p_{\theta} = \delta(\theta, \sqrt{\langle k \rangle / N})$ , the critical point for percolation  $q = 1$  provided by expressions (62) and (63) is the well-known percolation condition for a random network  $\bar{k} = \langle k \rangle = 1$ .
- *The configuration model.* In the case of the configuration model the couplings factorize,  $\beta J(\theta, \theta', \alpha, \alpha') \sim \theta\theta'$ . The stability conditions (62) and (63) become

$$q_c = N \langle \theta^2 \rangle. \quad (64)$$

In the case in which the network is uncorrelated we have  $\theta_i = \bar{k}_i / \sqrt{\langle k \rangle N}$  and the degree  $k_i$  of a node  $i$  is a Poisson variable with average  $\bar{k}_i$ . The critical point (62) can be then expressed in terms of the actual degree of the canonical configuration ensemble as

$$q_c = \frac{\langle k(k-1) \rangle}{\langle k \rangle}. \quad (65)$$

In the typical case limit, i.e.  $q = 1$ , the previous equation corresponds to the condition for the percolation transition in configuration networks [4, 11, 15].

- *Structured networks.* In the general case of the structured networks the complete eigenvalue problem in equations (62) and (63) has to be solved on a case-by-case basis in order to find the percolation critical point.

Nevertheless, in the following, we present two simple cases in which the problem can be simplified.

- *First case.* We present a case in which a perturbative analysis can give a good approximation to the critical point. The case under study is the case in which the network has a detailed structure made of  $A$  different communities labeled with an index  $\alpha = 1, \dots, A$  and  $A \simeq \mathcal{O}(1)$ . Each community has well-defined features such as the average degree and the number of links shared with other communities. This naturally leads to an interaction between nodes which depends on the community they belong to, encoded in the following matrix:

$$W(\alpha, \alpha') = \begin{cases} \psi & \text{if } \alpha = \alpha' \\ \xi & \text{if } \alpha \neq \alpha'. \end{cases} \quad (66)$$

In this hypothesis, the matrix  $K$  (63) takes the form

$$K_{\alpha\alpha'} = N W(\alpha, \alpha') \sum_{\theta} p_{\theta\alpha'} \theta^2 = W(\alpha, \alpha') \langle \theta^2 \rangle_{\alpha'}, \quad (67)$$

where we indicated with  $\langle x \rangle_\beta = \sum_\theta p_{\theta\beta} x$  the average over one single component  $\beta$ . The eigenvalue problem (62) that we have to solve to find the critical point of the Potts model can be solved perturbatively in the limit  $\Delta = \xi - \psi \ll 1$ . In this case, the matrix  $K$  is

$$K = \psi (D + \Delta H), \tag{68}$$

where  $D$  is a diagonal matrix and  $H$  has vanishing diagonal elements

$$\begin{aligned} D_{\alpha\alpha'} &= N \langle \theta^2 \rangle_\alpha \delta(\alpha, \alpha'), \\ H_{\alpha\alpha'} &= \begin{cases} 0 & \text{if } \alpha = \alpha' \\ N \langle \theta^2 \rangle_{\alpha'} & \text{if } \alpha \neq \alpha'. \end{cases} \end{aligned} \tag{69}$$

It is well known in the perturbation theory for non-degenerate states that the eigenvalues of this problem show second-order corrections to the diagonal entries  $D_{\alpha\alpha'}$  in the parameter  $\Delta$ . Finally, we obtain that the onset of instability occurs when the following relation is satisfied:

$$q_c = N \max_{\alpha=1, \dots, A} \psi \left[ \langle \theta^2 \rangle_\alpha + \Delta^2 \sum_{\alpha' \neq \alpha} \frac{\langle \theta^2 \rangle_\alpha \langle \theta^2 \rangle_{\alpha'}}{|\langle \theta^2 \rangle_\alpha - \langle \theta^2 \rangle_{\alpha'}|} + o(\Delta^3) \right]. \tag{70}$$

This set of coupled equations reduces to the value found in the configuration model, i.e.  $q_c = N \langle \theta^2 \rangle$  when there is only one single community. Here we report the condition for the leading term in  $O(\Delta^0)$  that has the following form:

$$q_c = \psi N \max_{\alpha=1, \dots, A} \langle \theta^2 \rangle_\alpha. \tag{71}$$

We want to underline the new percolation condition which becomes

$$N \max_{\alpha=1, \dots, A} \langle \theta^2 \rangle_\alpha = \frac{1}{\psi} \tag{72}$$

meaning that the percolation transition depends strongly on the number of links of the most connected community.

Whenever different communities have the same distribution, i.e. the same second moment  $\langle \theta^2 \rangle_\alpha = \langle \theta^2 \rangle$ , we are able to perform the calculation exactly and the critical value  $q_c$  reads

$$q_c = (\psi + (A - 1)\phi) N \langle \theta^2 \rangle. \tag{73}$$

– *Second case.* The second case that we consider is formed by sparse structured networks with the couplings  $\beta J(\theta, \theta', \alpha, \alpha')$  taking expression (40) that we write here for convenience

$$\beta J(\theta_i, \theta_j, \alpha_i, \alpha_j) = \bar{k}_i \bar{k}_j \frac{W(\alpha_i, \alpha_j)}{\mathcal{N}_{\alpha_i} \mathcal{N}_{\alpha_j}}. \tag{74}$$

In the further approximation that the density of nodes with ‘hidden variables’  $\theta$  and  $\alpha$  is factorizable, i.e.  $p_{\theta\alpha} = p_\theta \hat{p}_\alpha$  we can simplify the eigenvalue problem (62), (63) to find the critical point of the Potts model as

$$q_c = \overline{\langle k(k-1) \rangle} \hat{\Lambda}, \tag{75}$$

where  $\hat{\Lambda}$  is the maximal eigenvalue of the matrix  $\hat{K}$  defined as

$$\hat{K}_{\alpha, \alpha'} = N \hat{p}_{\alpha'} \frac{W(\alpha, \alpha')}{\mathcal{N}_\alpha \mathcal{N}_{\alpha'}}. \tag{76}$$

### 5.2. Asymmetric saddle point

Below the phase transition the symmetric solution (55) is no longer stable, as shown in the previous section. In the stationary state of the Potts model a giant component appears, and a more complicated saddle point has to be found. Due to the fact that one single color becomes dominant, generalizing for similar ansatz made for the Potts model with homogeneous couplings [13], the following ansatz on the parameter  $c_{\theta\alpha}$  is proposed

$$\begin{aligned} c_{\theta\alpha}(\sigma, s_{\theta\alpha}) &= \frac{1}{q} (1 + (q - 1)s_{\theta\alpha}) && \text{if } (\sigma = 0) \\ c_{\theta\alpha}(\sigma, s_{\theta\alpha}) &= \frac{1}{q} (1 - s_{\theta\alpha}) && \text{else.} \end{aligned} \tag{77}$$

And thus the density functional free energy reads

$$\begin{aligned} \beta f[\{c_{\theta\alpha}(\sigma, s_{\theta\alpha})\}] &= \frac{N}{2q} \sum_{\theta\theta'\alpha\alpha'} p_{\theta\alpha} p_{\theta'\alpha'} \beta J(\theta, \theta', \alpha, \alpha') [(1 - q)s_{\theta\alpha}s_{\theta'\alpha'} - 1] \\ &\quad - \log q + \sum_{\theta\alpha} \frac{p_{\theta\alpha}}{q} [(q - 1)(1 - s_{\theta\alpha}) \log(1 - s_{\theta\alpha}) \\ &\quad + (1 + (q - 1)s_{\theta\alpha}) \log(1 + (q - 1)s_{\theta\alpha})], \end{aligned} \tag{78}$$

where we have to minimize over the variational parameters  $s_{\theta\alpha}$ . Solving the equation

$$\left. \frac{\partial \beta f[\{c_{\theta\alpha}(\sigma, s_{\theta\alpha})\}]}{\partial s_{\theta\alpha}} \right|_{s_{\theta\alpha}^*} = 0 \tag{79}$$

we finally obtain the self-consistent condition for the parameter we solved numerically

$$s_{\theta\alpha}^* = \frac{e^{\theta\rho_\alpha} - 1}{q - 1 + e^{\theta\rho_\alpha}} \tag{80}$$

with  $\rho_\alpha$  given by

$$\rho_{\alpha'} = N \sum_{\alpha\theta} \theta W(\alpha, \alpha') p_{\alpha\theta} s_{\theta\alpha}. \tag{81}$$

Therefore, equation (80) can be expressed as a close expression for  $\rho_\alpha$ , which is the order parameter for the Potts phase transition. In particular, we find

$$\rho_{\alpha'} = N \sum_{\theta, \alpha} p_{\theta\alpha} \theta W(\alpha, \alpha') \frac{e^{\theta\rho_\alpha} - 1}{q - 1 + e^{\theta\rho_\alpha}}. \tag{82}$$

The solution of this equation is  $\rho_\alpha = 0$  for  $q \leq q_c$  and develops a non-zero solution for  $q > q_c$ . The transition can be continuous or discontinuous. In all the ensembles studied in this paper,  $q_c = 2$  signals the crossover between a second-order phase transition and a first-order one. This could be understood in the general framework of Landau theory. As it is well known, the Hamiltonian of the Potts model is invariant under the permutation symmetry of the  $q$  colors, which in the case  $q \leq 2$  is accidentally equivalent to the  $Z_2$  symmetry. In equation (78), it is easy to show that the free energy is explicitly even under the transformation  $s_{\theta\alpha} \rightarrow -s_{\theta\alpha}$  when  $q \leq 2$ , while for higher  $q$ , the free energy density  $f$  contains all possible powers of the order parameter  $s_{\theta\alpha}$ . As a consequence, within the Landau theory, the property of the free energy for  $q \leq 2$  necessary reflects into a continuous phase transition at least in the absence of an external field. Thus, on general grounds, we expect the crossover from second- to first-order transition could occur only at the value  $q = 2$  independently on the network we choose. If we expand (82) to the first order in  $\rho_\alpha$  we get the equation

$$\rho_\alpha = \frac{1}{q} \sum_{\alpha'} K_{\alpha, \alpha'} \rho_{\alpha'}, \tag{83}$$

with the matrix  $K_{\alpha,\alpha'}$  given by (63). We recover therefore the same critical point (62),  $q_c = \Lambda$  where  $\Lambda$  is the maximal eigenvalue of the matrix  $K$  as was found by studying the stability of the Potts model above the phase transition.

For the case of the percolation transition, i.e.  $q \rightarrow 1$  [14] we have a continuous phase transition and we can study equation (82) for small values of  $\rho_\alpha$  to find the critical exponents of the percolation transition.

- *The  $G(N, p)$  ensemble.* In this case, the order parameter  $\rho_\alpha$  is independent of  $\alpha$ ,  $\rho_\alpha = \rho$  and the self-consistent equation (82) simplifies to

$$\rho = N\theta \frac{e^{\theta\rho} - 1}{q - 1 + e^{\theta\rho}}, \tag{84}$$

with  $\theta = \sqrt{\langle k \rangle} / N$ . The expansion for small value of  $x = \theta\rho$  and  $q = 1$  gives

$$x = \langle k \rangle \left( x - \frac{x^2}{2} \right), \tag{85}$$

therefore we can derive the known result that

$$\rho \propto (\langle k \rangle - 1)^{\hat{\beta}} \tag{86}$$

with the mean field critical exponent  $\hat{\beta}$  given by  $\hat{\beta} = 1$ .

- *The configuration model.* In the case of the configuration model the order parameter  $\rho_\alpha$  is independent of  $\alpha$ , i.e.  $\rho_\alpha = \rho$  and the self-consistent equation (82) reduces to

$$\rho = \sum_{\theta} \frac{p_{\theta}\theta}{\langle \theta \rangle} \frac{e^{\theta\rho} - 1}{q - 1 + e^{\theta\rho}}, \tag{87}$$

where  $\rho = \sum_{\theta} \frac{\theta}{\langle \theta \rangle} p_{\theta} s_{\theta}$ . The expansion of this equation for small value of  $\rho$  provides the critical exponents for networks in the configuration model and generalizes the results of uncorrelated networks to a network with the correlations imposed by the configuration model. In the case in which  $\langle \theta^3 \rangle$  is finite, the expansion of (82) for  $q = 1$  gives

$$\rho = N \sum_{\theta} p_{\theta} \theta^2 \left( \rho - \frac{1}{2} \theta \rho^2 \right). \tag{88}$$

The order parameter close to the percolation phase transition goes like

$$\rho \propto (N \langle \theta^2 \rangle - 1)^{\hat{\beta}} \tag{89}$$

with  $\hat{\beta} = 1$  as in the  $G(N, p)$  ensemble. In contrast, in the case in which  $p_{\theta} \propto \theta^{-\gamma}$  with  $\gamma \in (3, 4]$  the expansion of equation (87) gives

$$\rho = N \langle \theta^2 \rangle \rho - \rho^{\gamma-2} I \tag{90}$$

giving the critical exponent  $\hat{\beta} = \frac{1}{\gamma-3}$ . Finally, we study the case in which  $p_{\theta} = \theta^{-\gamma}$  and  $\gamma < 3$ . In this case the self-consistent equation (87) can be written as

$$\rho = \rho^{\gamma-2} \hat{I} \tag{91}$$

giving  $\rho = \hat{I}^{\hat{\beta}}$  with the critical exponent  $\hat{\beta} = \frac{1}{3-\gamma}$ .

- *Structured networks.* In general, the study of the percolation transition for structured networks might be considered on case-by-case basis depending on the  $p_{\theta\alpha}$  distribution and on the function  $W(\alpha, \alpha')$  under consideration. Here we consider the case in which

the distribution  $p_{\theta\alpha}$  is factorizable, i.e.  $p_{\theta\alpha} = p_\theta \hat{p}_\alpha$ . In this case, if the moment  $\langle \theta^3 \rangle$  is finite, expanding equation (82) for  $q = 1$  we get

$$\begin{aligned} \rho_\alpha &= N \sum_{\theta, \alpha'} p_{\theta\alpha} \theta^2 W(\alpha, \alpha') \left( \rho_{\alpha'} - \frac{1}{2} \theta \rho_{\alpha'}^2 \right) \\ &= \sum_{\alpha'} K_{\alpha, \alpha'} \rho_{\alpha'} - \frac{1}{2} \sum_{\theta, \alpha'} p_{\theta\alpha} W(\alpha, \alpha') \theta^3 \rho_{\alpha'}^2. \end{aligned} \tag{92}$$

If we write  $\rho_\alpha$  in terms of the eigenvectors  $\vec{u}^\lambda$  of the matrix  $K$ , i.e.

$$\rho_\alpha = \sum_{\lambda} c_\lambda u_\alpha^\lambda \tag{93}$$

equation (92) can be written as an equation for the constants  $c_\lambda$ . Solving perturbatively assuming that  $c_\Lambda \gg c_\lambda$  for  $\lambda \neq \Lambda$  with  $\Lambda$  given by the maximal eigenvalue of the matrix  $K$ , we get

$$\begin{aligned} c_\Lambda &\propto (\Lambda - 1) \\ c_\lambda &\propto (\Lambda - 1)^2 \quad \text{for } \lambda \neq \Lambda. \end{aligned} \tag{94}$$

Therefore in this case there are two critical exponents  $\hat{\beta} = 1$  for the maximal eigenvector and  $\hat{\beta}' = 2$  for all the other eigenvectors. In the case  $p_\theta \propto \theta^{-\gamma}$  with  $\gamma \in (3, 4]$  we have the expansion

$$\rho_\alpha = \sum_{\alpha'} K_{\alpha, \alpha'} \rho_{\alpha'} - \sum_{\theta, \alpha'} p_\theta \hat{p}_\alpha W(\alpha, \alpha') H(\alpha) \rho_{\alpha'}^{\gamma-2}. \tag{95}$$

In this case we have

$$\begin{aligned} c_\Lambda &\propto (\Lambda - 1)^{1/(\gamma-3)} \\ c_\lambda &\propto (\Lambda - 1)^{2/(\gamma-3)} \quad \text{for } \lambda \neq \Lambda \end{aligned} \tag{96}$$

and the critical exponent  $\hat{\beta} = 1/(\gamma - 3)$  and  $\hat{\beta}' = 2/(\gamma - 3)$  which generalizes the results for scale-free networks. In the case in which  $p_\theta \propto \theta^{-\gamma}$  and  $\gamma < 3$  the expansion of equation (87) gives

$$\rho_\alpha = \sum_{\alpha'} \rho_{\alpha'}^{\gamma-2} W(\alpha, \alpha') \hat{H}_{\alpha'} \tag{97}$$

getting the critical exponents  $\hat{\beta} = 1/(3 - \gamma)$  and  $\hat{\beta}' = (4 - \gamma)/(3 - \gamma)$ .

Using equations (80) and (81) we find an explicit expression of the free energy of the Potts model in the asymmetric phase (78) as a function of the order parameter vector  $\rho_\alpha$

$$\begin{aligned} \beta f(q) &= -\frac{1}{2q} \sum_{\theta, \theta', \alpha, \alpha'} p_{\theta\alpha} p_{\theta'\alpha'} N \theta \theta' W(\alpha, \alpha') - \frac{q-1}{q} \sum_{\theta, \alpha} p_{\theta\alpha} \theta \rho_\alpha \frac{e^{\theta \rho_\alpha - 1}}{q-1 + e^{\theta \rho_\alpha}} \\ &\quad + \sum_{\theta, \alpha} p_{\theta\alpha} \left[ \frac{\theta \rho_\alpha e^{\theta \rho_\alpha}}{q-1 + e^{\theta \rho_\alpha}} - \log(q-1 + e^{\theta \rho_\alpha}) \right]. \end{aligned} \tag{98}$$

Using this expression we find for the function  $\phi(q)$  the explicit expression as a function of  $\rho_\alpha$ ,

$$\begin{aligned} \phi(q) &= \frac{1-q}{2q} \left[ \sum_{\theta, \theta', \alpha, \alpha'} p_{\theta\alpha} p_{\theta'\alpha'} N \theta \theta' W(\alpha, \alpha') - \sum_{\theta, \alpha} p_{\theta\alpha} \theta \rho_\alpha \frac{e^{\theta \rho_\alpha - 1}}{q-1 + e^{\theta \rho_\alpha}} \right] \\ &\quad - \sum_{\theta, \alpha} p_{\theta\alpha} \left[ \frac{\theta \rho_\alpha e^{\theta \rho_\alpha}}{q-1 + e^{\theta \rho_\alpha}} - \log(q-1 + e^{\theta \rho_\alpha}) \right]. \end{aligned} \tag{99}$$

In the specific case of the configuration model the preceding equations (98) and (99) can be simplified to the following equation for the free energy density  $f(q)$ :

$$\beta f(q) = \frac{\langle \theta \rangle}{2q} ((1-q)\rho^2 - 1) + \sum_{\theta} p_{\theta} \left[ \frac{\theta \rho e^{\theta \rho}}{q - 1 + e^{\theta \rho}} - \log(q - 1 + e^{\theta \rho}) \right] \quad (100)$$

and the following equation for the function  $\phi(q)$ :

$$\phi(q) = \frac{\langle \theta \rangle (q - 1)}{2q} (\rho^2 - 1) - \sum_{\theta} p_{\theta} \left[ \frac{\theta \rho e^{\theta \rho}}{q - 1 + e^{\theta \rho}} - \log(q - 1 + e^{\theta \rho}) \right]. \quad (101)$$

## 6. Cluster distribution

The large deviations  $\omega(c)$  of the cluster distribution can be calculated using (24), by performing a Legendre transformation of the function  $\phi(q)$  calculated by evaluating expression (99) at the self-consistent solution of equation (82),

$$\omega(c) = \min_q (\phi(q) - c \log q). \quad (102)$$

Therefore we have shown that by solving the heterogeneous Potts model with couplings  $\beta J(\theta, \theta', \alpha, \alpha') = \theta \theta' W(\alpha, \alpha')$  we can directly characterize the critical point of the percolation phase transition, the critical exponent  $\hat{\beta}$  of this transition and the large deviation function of the number of clusters  $c = C/N$  present in the networks of the ensemble.

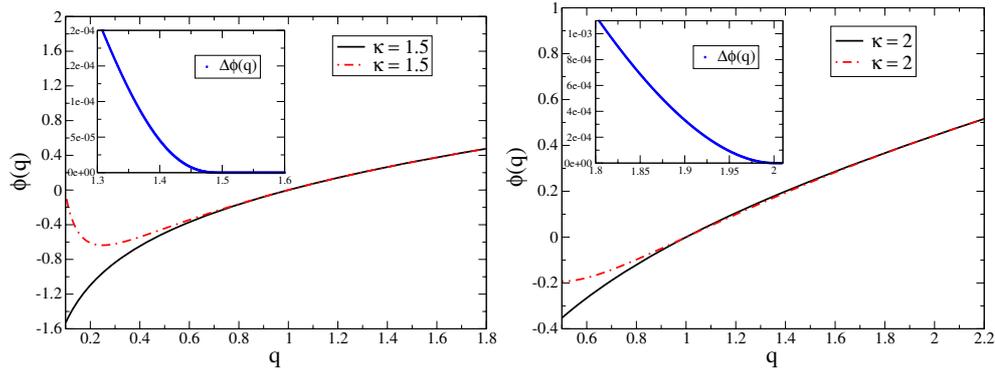
## 7. Numerical results

In this section, we present the study of the large deviation of the cluster distribution for different examples of generalized network ensembles.

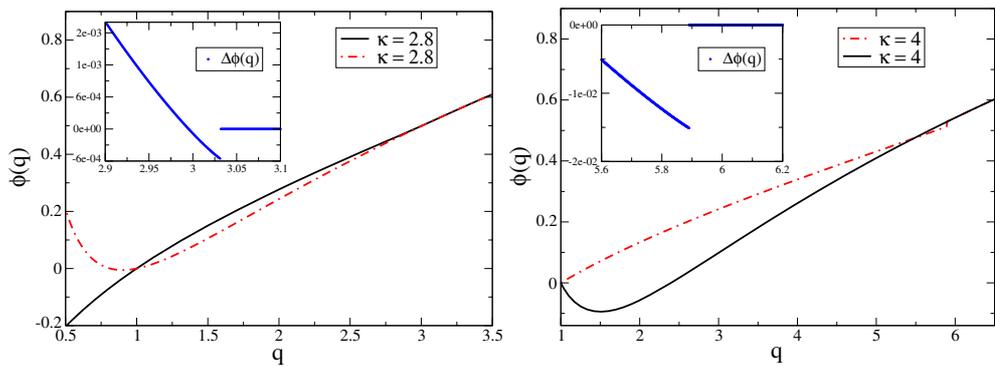
- *The  $G(N, p)$  ensemble.* We consider the simple case of  $G(N, p)$  network ensembles where the average degree of each node  $\theta_i$  is independent of  $i$ , i.e.  $\theta_i = \sqrt{\langle \bar{k} \rangle} / N$ , i.e.  $p = \langle \bar{k} \rangle / N$ . Equation (84) is the self-consistent equation for the Potts model phase transition. This equation has been already studied [13] where it was found that the Potts model has a phase transition in  $q_c = \langle \bar{k} \rangle$  of second order for a value of  $q_c \leq 2$  and of first order for  $q_c > 2$ . We suggest the reader to refer to [13, 14] for a full account of this case.
- *The configuration model.* In particular, we study the limit of weak heterogeneity when we assume that the average degree of the nodes  $\{\bar{k}\}$  is Poisson distributed and the case of strong heterogeneity of the degree of the nodes when the hidden variables  $\{\theta_i\}$  are distributed according to a power law.

– *Poisson hidden variable distribution.* We consider the case in which the distribution of the mean values for the connectivity's  $\bar{k}_i$  is Poisson and  $\theta_i = \sqrt{\bar{k}_i} / N$ . This ensemble introduces a small heterogeneity with respect to pure Erdős and Rényi networks. In this case the critical point is equal to  $q_c = N \langle \theta^2 \rangle = \langle \bar{k} \rangle + 1$ . Therefore the percolation transition happens at  $\langle \bar{k} \rangle + 1 = 1$  revealing that the percolated phase is already present when the mean connectivity is  $\langle \bar{k} \rangle \rightarrow 0$ .

As predicted by the theoretical results we found a phase transition for  $q_c = N \langle \theta^2 \rangle = \langle \bar{k} \rangle + 1$  depending on the mean value of connectivity  $\langle \bar{k} \rangle$ . In figures 1 and 2 we show the function  $\phi(q)$  as a function of  $q$  for different values of the parameter  $\kappa = \langle \bar{k} \rangle + 1$  that has the same role of the inverse temperature in the associated Potts model. The phase transition is of the second order when  $q_c = \langle \bar{k} \rangle + 1 \leq 2$  (see figure 1). In contrast, when  $\langle \bar{k} \rangle + 1 > 2$ , see figure 2.

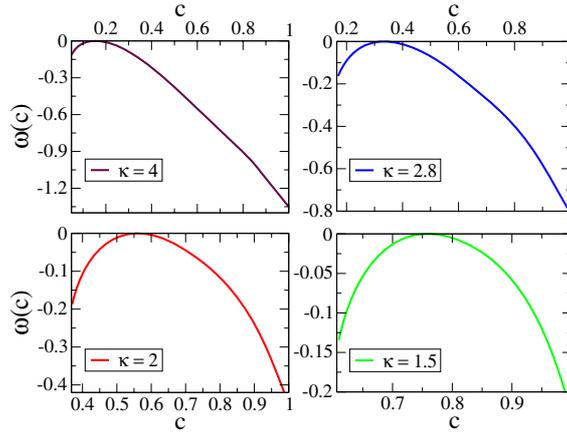


**Figure 1.** The two branches of the  $\phi(q)$  function across the Potts model phase transition for the hidden variable ensembles with average degree distributed according to a Poissonian. The solid line indicates the function  $\phi(q)$  calculated at the asymmetric solution of the Potts model and the dot-dashed line indicates the function  $\phi(q)$  calculated for the symmetric solution. For  $\kappa \leq 2$  the free energy at the phase transition varies continuously, from the  $\rho = 0$  solution to the asymmetric solution. The inset shows the difference of the free energy calculated on the two branches. In the left figures  $\kappa = 1.5$ , while in the right  $\kappa = 2$ .



**Figure 2.** The two branches of the  $\phi(q)$  function across the Potts model phase transition for the configuration model with the average of the degrees at each node distributed according to a Poissonian. The solid line indicates the function  $\phi(q)$  calculated at the asymmetric solution of the Potts model and the dot-dashed line indicates the function  $\phi(q)$  calculated for the symmetric solution. The free energy at the transition varies discontinuously at  $q_c \geq \kappa = \langle \bar{k} \rangle + 1$  when the metastable configuration  $\rho \neq 0$  disappears. The inset shows explicitly the discontinuity in the difference  $\Delta\phi$  between the function  $\phi$  calculated on the two solutions.

In figure 3, we show the function  $\omega(c)$  for this ensemble for different values of  $\kappa = 1 + \langle \bar{k} \rangle$ . We obtain that for a value of  $c \leq \hat{c} = e^{-\langle \bar{k} \rangle}$  the function  $\omega(c)$  tends to infinity, i.e.  $\omega(c) \rightarrow -\infty$ . This value is relative to the minimum number of components that are equivalent to the number of isolated vertices. Therefore we plot the function  $\omega(c)$  only for  $c > \hat{c}$ . As a function of the parameter  $\langle \bar{k} \rangle$  we found that different numbers of connected components are dominant and in particular for  $\langle \bar{k} \rangle + 1 > 2$  we find that the typical number of clusters is less than 0.5 and that in the limit of high  $\langle \bar{k} \rangle$  this number vanishes (see figure 3).



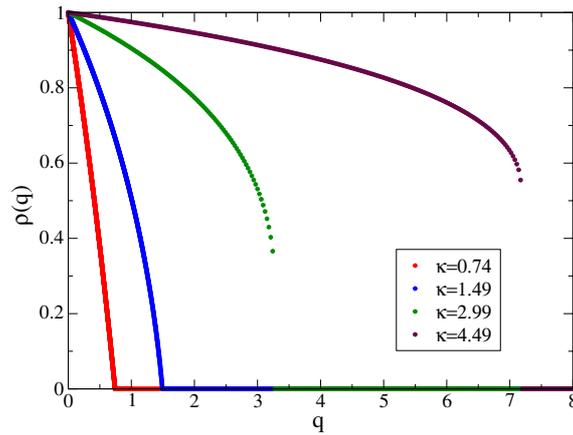
**Figure 3.** We show the numerically evaluated  $\omega(c)$  function for the hidden variable configuration model in which the average degrees are distributed according to a Poisson distribution with  $\kappa = 1 + \langle \bar{k} \rangle$ . We show the dependence on the typical number  $c^*$  as a function of the mean connectivity. The higher  $\kappa$  the smaller the number of expected typical components.

- *Scale-free degree distribution.* We use as degree distribution  $p_\theta \sim \theta^{-\gamma}$  with  $\gamma > 3$  so that the second moment does not vanish. We fixed the exponents  $\gamma$ , letting the value of infrared cutoff change in order to fix all the moments  $\langle \theta^m \rangle$ . In figure 4 we show the behavior of the solution  $\rho$  of the self-consistent equation (87) as a function of the parameter  $q$  for scale-free networks. When  $q < q_c$  defined by equation (62), i.e.  $q_c = \langle \bar{k}(k-1) \rangle / \langle \bar{k} \rangle$ , a non-zero solution  $\rho \neq 0$  is found while for  $q > q_c$  the  $\rho = 0$  solution becomes the stable one. The value of the stable solution  $\rho$  of equation (87) close to the phase transition  $q \simeq q_c$  varies continuously in the case  $q_c \leq 2$  and discontinuously for  $q > 2$ . There is a second-order phase transition for values of  $\frac{\langle \bar{k}(k-1) \rangle}{\langle \bar{k} \rangle} \leq 2$  and a first-order phase transition for  $\frac{\langle \bar{k}(k-1) \rangle}{\langle \bar{k} \rangle} > 2$  where the free energy discontinuously goes to zero as shown in figure 5.

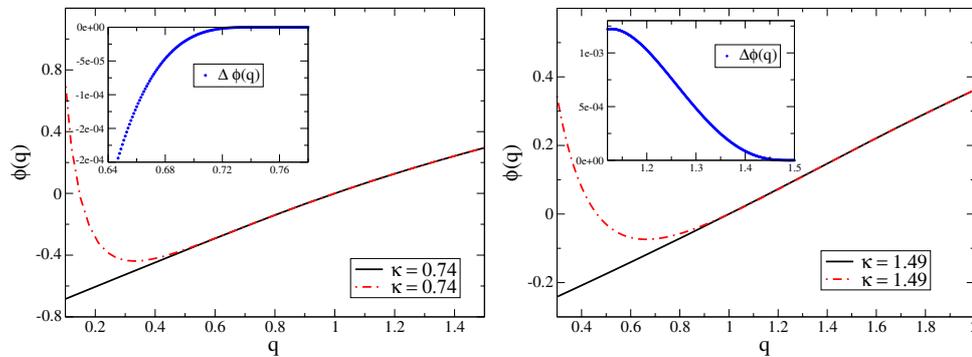
In figures 5 and 6, we plot the difference of the  $\phi(q)$  functions calculated on the two branches of the solution to (87) (the solution with  $\rho = 0$  and the other non-trivial solution stable for  $q < q_c$ ). From these figures we can see that  $\Delta\phi(q)$  has a discontinuity in the regime  $\frac{\langle \bar{k}(k-1) \rangle}{\langle \bar{k} \rangle} > 2$  and vanishes continuously in the regime  $\frac{\langle \bar{k}(k-1) \rangle}{\langle \bar{k} \rangle} \leq 2$ . In figure 7 the probability of large deviation is shown in the number of clusters for the configuration model with power-law distributed value of  $\{\theta\}$ 's. The typical number of clusters  $c^*$  is a decreasing function of  $\kappa = \langle \bar{k}(k-1) \rangle / \langle \bar{k} \rangle$ .

- *Comparison of the typical number of cluster for the configuration model with Poisson and power-law distributed  $\{\theta\}$ 's.* The typical value of the number of clusters  $c^*$  depends on the parameter  $N \langle \theta^2 \rangle$ . These dependences are shown in figure 8. In the case of power-law distributed  $\{\theta\}$ 's, the characteristic scale above which the number  $c^*$  vanishes is higher than in the Poisson case.

- *Simple case of structure network: network with four equal communities.* In general the case of structure networks with non-trivial  $W(\alpha, \alpha')$  functions has to be studied on a case-by-case approach.



**Figure 4.** The solution of the self-consistent equation (87) for the parameter  $\rho$  as a function of the number of colors  $q$  for scale-free random networks with different values of the critical point  $q_c = \kappa = \langle k(k-1) \rangle / \langle k \rangle$  and power-law exponent  $\gamma = 5$ .



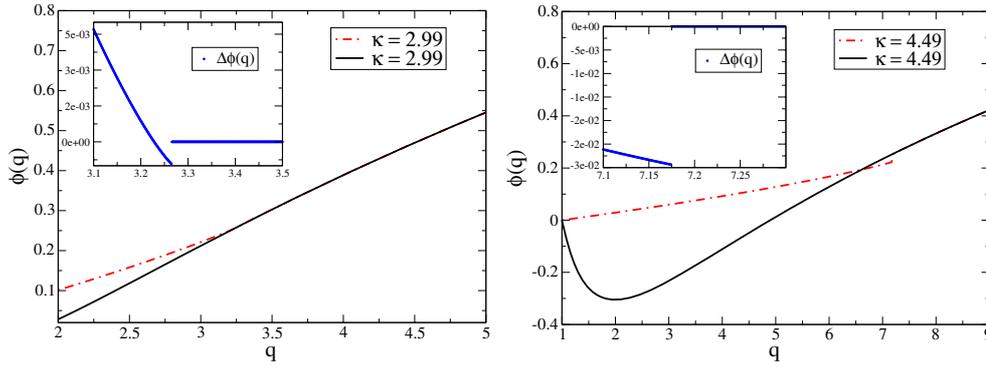
**Figure 5.** The two branches of the  $\phi(q)$  function for a configuration model within power-law distribution  $\{\theta\}$ 's with exponent  $\gamma = 5$ . The solid line is the function  $\phi(q)$  calculated at the asymmetric solution of the Potts model and the dot-dashed line is the function  $\phi(q)$  calculated for the symmetric solution. The transition point is at  $q_c = \kappa = N \langle \theta^2 \rangle \leq 2$ . In the inset, we show the difference in the free energy in the two branches which is continuous.

– *First case.* Here we consider the particular example in which the network is divided into four equivalent communities. These networks have been considered as benchmark networks with a community structure [27]. We consider in particular the case with  $N\theta_i = \text{const} \quad \forall i$  and the network divided in four communities  $\alpha = 1, 2, 3, 4$ , i.e.  $A = 4$  with

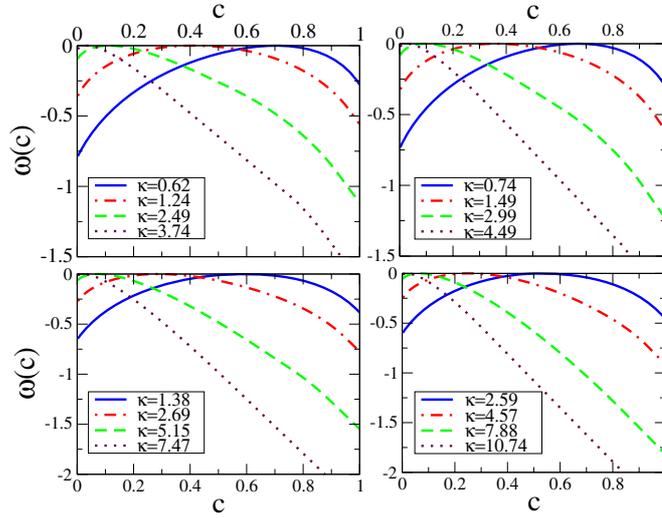
$$W(\alpha, \alpha') = \begin{cases} \frac{x}{64} & \text{for } \alpha = \alpha' \\ \frac{(1-x)}{192} & \text{for } \alpha \neq \alpha'. \end{cases} \quad (103)$$

In this case we find that for  $q > 1/64$  the only stable solution is the zero one while for lower value of the parameter  $q$  there is a non-zero solution that is stable everywhere independently of the value of  $x$ . We want to emphasize the fact that substituting equation (103) into (73), we obtain the critical relation

$$q_c = \frac{1}{64}, \quad (104)$$



**Figure 6.** The two branches of the  $\rho(q)$  function for the value of parameter  $\kappa = N(\theta^2) > 2$  for a scale free network  $\gamma = 5$ . In the inset we show the difference between the free energy associated with the symmetric solution  $\rho = 0$  and the free energy associated with the other asymmetric solution with  $\rho \neq 0$ . In the figure, we report the difference of the free energy calculated on the two branch solution showing evidence for the discontinuity in the free energy at the transition point.

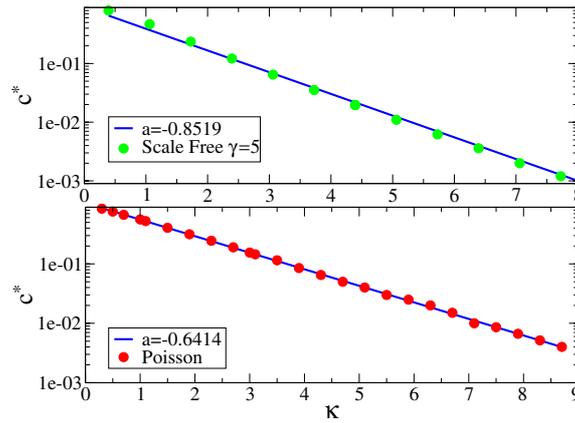


**Figure 7.** The logarithm of the probability density function  $\omega(c)$  versus the number of connected components  $c$ . It is easy to identify the typical value  $c^*$  when the function  $\omega$  touches the zero axis. This number depends strongly on  $\kappa = N(\theta^2)$ . Each network corresponds to different choices of the parameter  $\gamma = 7, 5, 3.5, 3.01$  starting from the top on the left.

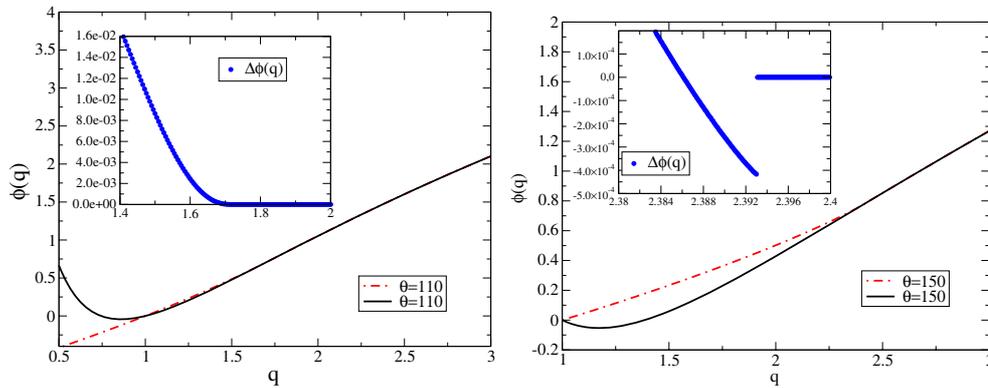
showing there is no phase transition for every value of the parameter  $x$  at fixed  $\theta$  and  $x$ .

- *Second case.* We consider the case with four communities  $\alpha = 1, 2, 3, 4$ , i.e.  $A = 4$  with  $W(\alpha, \alpha')$  given by (103) at fixed  $x$  for the value of the *hidden variables*  $\theta_i = \theta \forall i \in [1, N]$ . The behavior in this case is exactly the same as in the configuration model. There is a crossover between first- and second-order phase transition governed by the parameter  $\theta$ . The threshold value is given by the relation

$$\theta_c = 2(\psi + (A - 1)\phi)^{-1} \tag{105}$$

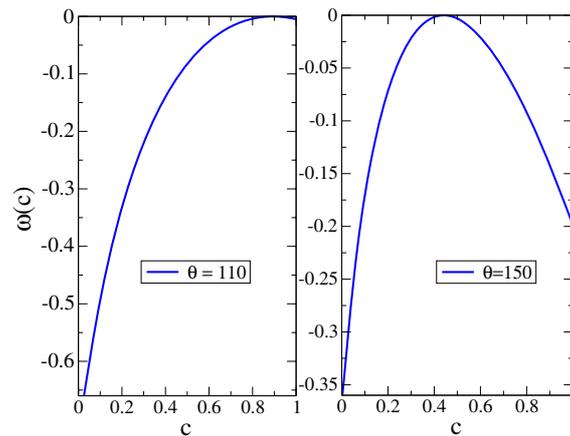


**Figure 8.** The typical number of clusters per vertex  $c^*$  vanishes exponentially fast with the increasing of  $\kappa = N(\theta^2)$ . Here we show in a logarithmic scale this relation for Poissonian degree distribution and power-law networks for a value of  $\gamma = 5$ . We report also the value from the best fitting calculation that has a good agreement with the data points.



**Figure 9.** In the figure we plot  $\phi(q)$  on the two branches of the solution for a network with four communities (second case under study). The network ensemble is characterized by  $W(\alpha, \alpha')$  given by (103) with  $x = 0.5$  and different values of  $\theta, \theta = 110$  and  $\theta = 150$ . This network shows the same crossover between a first-order and a second-order phase transition as soon as the parameter  $\theta > \theta_c$  calculated in equation (105),  $\theta_c = 120$ . The solid line indicates the function  $\phi(q)$  calculated at the asymmetric solution of the Potts model and the dot-dashed line indicates the function  $\phi(q)$  calculated for the symmetric solution. This is clearer by the discontinuity in the difference of  $\phi(q)$  calculated in two branches of the solution reported in the inset. The solid line corresponds to  $\rho = 0$  while the dashed one is  $\phi(q)$  calculated on the non-trivial solution  $\rho \neq 0$ .

and substituting the value of the parameter  $x = 0.5$  and  $A = 4$ , the critical value is of order  $\theta_c \sim 120$ . The free energy is shown in figure 9 where it is easy to notice the nature of the phase transition. For completeness we report also the logarithm of the cluster distribution in figure 10 that shows the same behavior of the configuration model.



**Figure 10.** We report the logarithm of the cluster probability  $\omega(c)$  for the study network ensemble with four communities (second case under study). The network ensemble is characterized by  $W(\alpha, \alpha')$  given by (103) with  $x = 0.5$  and the two different values of  $\theta$ ,  $\theta = 110$  and  $\theta = 150$ . The distribution is dominated by the typical number  $c^*$  that depends strongly on the average connectivity of the network ensembles.

## 8. Conclusions

In conclusion, we have studied the percolation transition and the large deviation of the cluster distribution in generalized canonical random network ensembles. The calculation has been performed by mapping the problem of finding the cluster distribution on a fully connected Potts model with heterogeneous couplings. The results generalize the known results for uncorrelated configuration models to correlated configuration models and are able to predict the behavior of the phase transition for generic structured networks with non-trivial community or spatial structure. Ongoing work will investigate what is the role of the percolation properties in generalized random network ensembles for the understanding of the critical behavior of dynamical models defined on them and for the characterization of their small loops distribution.

## Acknowledgments

This paper was supported by the project IST STREP GENNETEC contract No. 034952 and by MIUR grant 2007JHLPEZ.

## References

- [1] Dorogovtsev S N, Goltsev A V and Mendes J F F 2008 *Rev. Mod. Phys.* **80** 1275
- [2] Molloy M and Reed B A 1995 *Rand. Struct. Algorithms* **6** 161
- [3] Albert R, Jeong H and Barabási A-L 2000 *Nature* **406** 378
- [4] Cohen R, Erez K, ben Avraham D and Havlin S 2000 *Phys. Rev. Lett.* **85** 4626
- [5] Cohen R, Erez K, ben Avraham D and Havlin S 2001 *Phys. Rev. Lett.* **86** 3682
- [6] Dorogovtsev S N, Goltsev A V and Mendes J F F 2002 *Phys. Rev. E* **66** 016104
- [7] Leone M, Vazquez A, Vespignani A and Zecchina R 2002 *Eur. Phys. J. B* **28** 191
- [8] Bianconi G 2002 *Phys. Lett. A* **303** 166
- [9] Skantzos N S, Prez Castillo I and Hatchett J P L 2005 *Phys. Rev. E* **72** 066127
- [10] Coolen A C C *et al* 2005 *J. Phys. A: Math. Gen.* **38** 8289
- [11] Noh J D 2008 *Eur. Phys. J. B* **66** 251

- [12] Bollobas B B 2001 *Random Graphs* 2nd edn (Cambridge: Cambridge University Press) (1st edn in 1985)
- [13] Engel A, Monasson R and Hartmann A K 2004 *J. Stat. Phys.* **117** 387
- [14] Fortuin C M and Kasteleyn P W 1972 *Physica* **57** 536
- [15] Boguna M and Serrano M A 2005 *Phys. Rev. E* **72** 016106
- [16] Goltsev A V, Dorogovtsev S N and Mendes J F F 2008 *Phys. Rev. E* **78** 051105
- [17] Bianconi G 2008 *Europhys. Lett.* **81** 28005
- [18] Bianconi G 2008 arXiv:0802.2888
- [19] Park J and Newman M E J 2004 *Phys. Rev. E* **70** 066146
- [20] Sodeberg B 2002 *Phys. Rev. E* **66** 066121
- [21] Chung F and Lu L 2002 *Proc. Natl. Acad. Sci.* **100** 6313
- [22] Caldarelli G, Capocci A, De Los Rios P and Muñoz M A 2002 *Phys. Rev. Lett.* **85** 5468
- [23] Boguñá M and Pastor-Satorras R 2003 *Phys. Rev. E* **68** 036112
- [24] Lee D S, Goh K I, Kahng B and Kim D 2004 *Nucl. Phys. B* **696** 351–80
- [25] Wu F 1982 *Rev. Mod. Phys.* **54** 235
- [26] Lubensky T C 1979 Thermal and geometrical critical phenomena in random systems, in *La Matière Mal Condensée Les Houches 1978* ed R Balian, R Maynard and G Toulouse (New York: North-Holland)
- [27] Girvan M and Newman M E J 2002 *Proc. Natl. Acad. Sci.* **99** 7821