On large deviation properties of Erdös-Rényi random graphs

Andreas Engel[1,2], Rémi Monasson[2,3], and Alexander K. Hartmann[4]
[1] Institut für Physik, Universität Oldenburg,
26111 Oldenburg, Germany.
[2] CNRS-Laboratoire de Physique Théorique,
3 rue de l'Université, 67000 Strasbourg, France.
[3] CNRS-Laboratoire de Physique Théorique de l'ENS,
24 rue Lhomond, 75005 Paris, France.
[4] Institut für Theoretische Physik, Tammannstraße 1,
37077 Göttingen, Germany

April 30, 2004

Abstract

We show that large deviation properties of Erdös-Rényi random graphs can be derived from the free energy of the q-state Potts model of statistical mechanics. More precisely the Legendre transform of the Potts free energy with respect to $\ln q$ is related to the component generating function of the graph ensemble. This generalizes the well-known mapping between typical properties of random graphs and the $q \to 1$ limit of the Potts free energy. For exponentially rare graphs we explicitly calculate the number of components, the size of the giant component, the degree distributions inside and outside the giant component, and the distribution of small component sizes. We also perform numerical simulations which are in very good agreement with our analytical work. Finally we demonstrate how the same results can be derived by studying the evolution of random graphs under the insertion of new vertices and edges, without recourse to the thermodynamics of the Potts model.

PACS: 02.50.-r, 05.50.+q, 75.10.Nr

1 Introduction

Random graphs have kept being an issue of tremendous interest in probability and graph theory ever since the seminal work by Erdös and Rényi [1] more than four decades ago. In addition to fixed edge number and fixed edge probability distributions also random graphs with constant vertex degree [2] or power law degree distribution [3, 4] have been investigated. Most of the efforts devoted to the study of the properties of random graphs have taken advantage of the fact that these properties undergo some concentration process in the infinite size (number of vertices) limit. For instance, the numbers of vertices in the largest component or the number of connected components, which are stochastic in nature, become highly concentrated in this limit, and with high probability do not differ from their average values.

For large but finite sizes, properties as the one evoked above obviously fluctuate from graph to graph. The understanding of their statistical deviations are important for several problems in statistical physics, *e.g.* for the life-time of metastable states and the extremal properties of models defined on random graphs [5], as well as in computer science, *e.g.* for information-packet transmission in random networks [6, 7], resolution of random decision problems with search procedures [8, 9, 10] and others. Up to now apparently little attention has been paid to a quantitative characterization of large deviations in random graph ensembles [11, 12].

The present work is intended to contribute to an improved understanding of rare fluctuations in random graphs. Our main objective was to devise a microscopic "mean-field" approach permitting to handle such rare deviations in much the same way as for average properties of various similar problems, as *e.g.* bootstrap and rigidity percolation [13] and spin-glasses [14]. The mean-field approach relies on a statistical stability argument: a large graph is not strongly modified when adding an edge and/or a vertex. This statement can be translated into some self-consistent equations for the average value of physical properties of interest, as *e.g.* the magnetization for a spin system, or the probability of belonging to the k-core for bootstrap percolation. We will show in the present work that a similar self-consistent approach can also be successfully used to access large deviations in random graphs.

The main property we focus on throughout this paper is the number of connected components of a random graph. As established by Fortuin and Kasteleyn [15], several properties of random graphs with a *typical* number of components can be inferred from the knowledge of the thermodynamics of the q-state Potts model on a complete graph for values of q around 1. We will show that the thermodynamic properties for *general* values of q can be used to additionally characterize the properties of random graphs with an *atypical* number of components. This allows us to verify the validity of our microscopic mean-field approach.

This paper is organized as follows. In Section 2, we introduce the basic definitions and notations for the quantities studied. Section 3 is devoted to the derivation of rare graphs properties through the study of the Potts model. We show in Section 4 how these results can be rederived through the requirement of the statistical stability of very large atypical graphs against the addition of a vertex and its attached edges, or an edge. In Section 5 we describe our numerical procedure to simulate large deviation properties of random graphs ensembles. Some conclusion is finally proposed in Section 6.

2 Basic notions

We begin by fixing some vocabulary. For a detailed and precise account on random graphs we refer the reader to the textbook [16]. A graph G is a collection of *vertices* numbered by $i = 1, \ldots, N$ with *edges* (i, j), $i \neq j$, $i, j = 1, \ldots, N$ connecting them. The number of edges is between 0 (for the empty graph) and $N(N-1)/2$ (for the complete graph). A *component* of a graph is a subset of connected vertices which are disconnect from the rest of the graph. The *size* S of a component is the number of vertices it contains. Hence the empty graph consists of N components of size 1 whereas the complete graph is made from a single component of size N. The *number* of components of a graph G is denoted by $C(G)$. We are generally interested in properties of *large* graphs, $N \to \infty$.

We will consider *random* graphs in the sense that an edge between two vertices may be present or absent with a certain probability. The various joint probabilities to be discussed below will be denoted in the form $P(x_1, x_2, ...; a_1, a_2, ...)$ with the x_i representing the random variables and the a_i denoting the parameters of the distribution. In particular we consider random graphs in which each pair of vertices is connected by an edge with probability γ/N independently of all other pairs of vertices. The parameter γ characterizes the connectivity of the graph. Since each vertex establishes edges with probability γ/N with all the other $N-1$ vertices γ is in the limit $N \to \infty$ just the typical *degree* of a vertex denoted by d^*, giving the average number of edges emanating from it.

More precisely, in this limit the degree d of a vertex is a random variable obeying a Poisson law with parameter γ,

$$P(d; \gamma) = e^{-\gamma} \, \frac{\gamma^d}{d!} . \tag{1}$$

In particular, $P(d = 0; \gamma) = e^{-\gamma}$ is the fraction of isolated vertices. Hence the average number of components of a random graph of the described type is bounded from below by $Ne^{-\gamma}$. Note also that the typical degree d^* of a vertex remains finite for $N \to \infty$. Typical realization of such random graphs are therefore *sparse*.

The probability $P(G; \gamma, N)$ of one particular random graph G with N

vertices and parameter γ derives from the binomial law,

$$
\begin{aligned}
P(G;\gamma,N) &= \left(\frac{\gamma}{N}\right)^{L(G)} \left(1-\frac{\gamma}{N}\right)^{\binom{N}{2}-L(G)} \\
&= e^{-\frac{\gamma N}{2}+O(1)} \left(\frac{\gamma}{N}\right)^{L(G)}, \qquad (2)
\end{aligned}
$$

where $L(G) = O(N)$ denotes the number of edges of graph G. To describe the decomposition of a large random graph into its components, it is convenient to introduce the probability $P(C;\gamma,N)$ of a random graph with N vertices to have C components

$$
P(C;\gamma,N) = \sum_G P(G;\gamma,N)\,\delta(C,C(G)), \qquad (3)
$$

where $\delta(a,b)$ denotes the Kronecker delta.

A general observation is that for given γ and large N the probability $P(C;\gamma,N)$ gets sharply peaked at some *typical* value C^* of C and the probabilities for values of C significantly different from C^* being exponentially small in N. To describe this fact more quantitatively we introduce the number of components per vertex $c = C/N$ together with the quantity

$$
\omega(c,\gamma) = \lim_{N\to\infty} \frac{1}{N} \ln P(C;\gamma,N), \qquad (4)
$$

where the existence of the limit is assumed. Clearly $\omega(c,\gamma) \leq 0$ and the typical value c^* of c has $\omega(c^*,\gamma) = 0$. Averages with $P(G;\gamma,N)$ are therefore dominated by graphs with a typical number of components.

The focus of the present paper is on properties of random graphs which are *atypical* with respect to their number of components C. In order to get access to the properties of these graphs we introduce the *biased* probability distributions

$$
P(G;\gamma,q,N) = \frac{1}{Z(\gamma,q,N)}\, P(G;\gamma,N)\, q^{C(G)}, \qquad (5)
$$

with $Z(\gamma,q,N)$ defined by

$$
Z(\gamma,q,N) = \sum_G P(G;\gamma,N)\, q^{C(G)} = \sum_C P(C;\gamma,N)\, q^C. \qquad (6)
$$

The normalization constant $Z(\gamma,q,N)$ in (5) has hence the meaning of a component generating function of $P(G;\gamma,N)$. Contrary to averages with $P(G;\gamma,N)$ those with $P(G;\gamma,q,N)$ are dominated by graphs with an atypical number of components which is fixed implicitly with the parameter q. Values of q smaller than 1 shift weight to graphs with few components whereas for $q > 1$ graphs with many components dominate the distribution. The typical case is obviously recovered for $q = 1$.

Similar to $\omega(c, \gamma)$ it is convenient to introduce the function

$$\varphi(\gamma, q) = \lim_{N\to\infty} \frac{1}{N} \ln Z(\gamma, q, N). \tag{7}$$

From (6) and (4) it follows to leading order in N that

$$Z(\gamma, q, N) = \int_0^1 dc\ \exp(N[\omega(c, \gamma) + c \ln q]) \tag{8}$$

and performing the integral by the Laplace method for large N we find that $\varphi(\gamma, q)$ and $\omega(c, \gamma)$ are Legendre transforms of each other:

$$\varphi(\gamma, q) = \max_c [\omega(c, \gamma) + c \ln q] \qquad \omega(c, \gamma) = \min_q [\varphi(\gamma, q) - c \ln q] \tag{9}$$

$$q = \exp(-\frac{\partial \omega}{\partial c}) \qquad c = q\frac{\partial \varphi}{\partial q} \tag{10}$$

The large deviation properties of the ensemble of random graphs as characterized by $\omega(c, \gamma)$ can hence be inferred from $\varphi(\gamma, q)$. In the next section we show how $\varphi(\gamma, q)$ can be obtained from the statistical mechanics of the Potts model.

For later use we also note that from differentiating (7) with respect to γ we find using (6) and (2) to leading order in N

$$\bar{\ell}(\gamma, q) = \frac{\gamma}{2} + \gamma \frac{\partial \varphi}{\partial \gamma}. \tag{11}$$

Here $\bar{\ell}(\gamma, q)$ denotes the average number of edges per vertex in the graph where the average is performed with the distribution (5).

3 Thermodynamics of atypical graphs

3.1 The mean-field Potts model

It has long been known [15] that certain characteristics of random graphs are related to the thermodynamic properties of the Potts model [17, 18]. The Potts model is defined in terms of an energy function $E(\{\sigma_i\})$ depending on N spin variables $\sigma_i, i = 1, \ldots, N$, which may take on q distinct values $\sigma = 0, 1, ..., q-1$. In the mean-field variant the energy function reads

$$E(\{\sigma_i\}) = -\frac{1}{N} \sum_{i<j} \delta(\sigma_i, \sigma_j) - h \sum_{\sigma=0}^{q-1} u_\sigma \sum_i \delta(\sigma, \sigma_i), \tag{12}$$

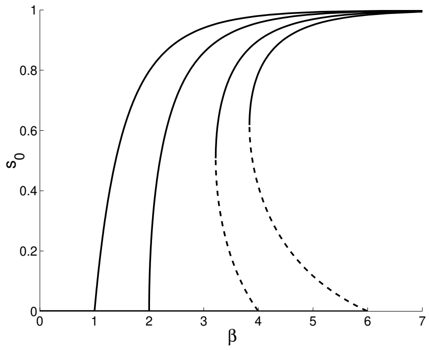

Figure 1: Solution $\overline{s}_0(\beta, q)$ of the saddle-point Eq. (16) as function of β for $q = 1, 2, 4, 6$ (from left to right). For $q \leq 2$ the non-trivial solution $\overline{s}_0 > 0$ branches off continuously from the high-temperature solution $\overline{s}_0 = 0$ at $\beta = q$. For $q > 2$ the new solution appears discontinuously at the spinodal point $\beta^{\mathrm{s}} < q$ by a subcritical bifurcation.

where hu_σ is an auxiliary field parallel to the direction σ. The thermodynamic properties of the system at inverse temperature β can be derived from the partition function

$$\mathcal{Z}(\beta, h, q, \{u_\sigma\}, N) = \sum_{\{\sigma_i\}} \exp(-\beta E(\{\sigma_i\})) \tag{13}$$

where the sum runs over all q^N spin configurations $\{\sigma_i\}$. A standard analysis (cf. the appendix) gives for the free energy

$$f(\beta, h, q, \{u_\sigma\}) = -\lim_{N\to\infty} \frac{1}{\beta N} \ln \mathcal{Z}(\beta, h, q, \{u_\sigma\}, N) \tag{14}$$

at $h = 0$ the result

$$\begin{aligned} f(\beta, q) = \operatorname*{extr}_{s_0} \Bigg[& -\frac{1}{2q} - \frac{q-1}{2q} s_0^2 - \frac{1}{\beta} \ln q \\ & + \frac{1 + (q-1)s_0}{\beta q} \ln(1 + (q-1)s_0) + \frac{q-1}{\beta q} (1 - s_0) \ln(1 - s_0) \Bigg]. \end{aligned} \tag{15}$$

The saddle-point value $\overline{s}_0(\beta, q)$ extremizing the expression in the brackets is the stable solution of the equation

$$e^{\beta \overline{s}_0} = \frac{1 + (q-1)\overline{s}_0}{1 - \overline{s}_0}. \tag{16}$$

Clearly $\overline{s}_0 = 0$ is always a solution of this equation. It is, however, unstable for large β and another, non-trivial solution becomes stable which describes the spontaneous appearance of order in the low temperature phase. Fig. 1 displays the solutions of (16) as function of β for different values of q. Note the subcriticial bifurcation in $\overline{s}_0(\beta, q)$ for $q > 2$.

3.2 Diagrammatic expansion of the Potts model

The relation between the Potts model and the random graph ensemble introduced in section 2 becomes apparent when considering the high-temperature expansion of the free energy (14) of the Potts model. Since the Kronecker delta can take only the values zero or unity, the partition function (13) can be recast into the form [15]

$$\mathcal{Z}(\beta, h, q, \{u_\sigma\}, N) = \sum_{\{\sigma_i\}} \prod_{i<j} [1 + w\, \delta(\sigma_i, \sigma_j)] \; e^{\beta\, h \sum_\sigma u_\sigma \sum_i \delta(\sigma_i, \sigma)}, \quad (17)$$

where

$$w = \exp(\frac{\beta}{N}) - 1 = \frac{\beta}{N} + O(\frac{1}{N^2}). \quad (18)$$

When expanding the product appearing in (17) we obtain a sum of $2^{N(N-1)/2}$ terms each of which is in one–to–one correspondence with a graph. The N vertices of this graph represent the Potts variables σ_i, whereas an edge (i, j) stands for a factor $w\, \delta(\sigma_i, \sigma_j)$. Performing the trace over the configurations $\{\sigma_i\}$ for each term in the sum, *i.e.* for each graph, separately, the Kronecker deltas constrain the Potts variables belonging to one component of the graph to the same value. As a result we find the Potts partition function as a sum over graphs in the form

$$\mathcal{Z}(\beta, h, q, \{u_\sigma\}, N) = \sum_G w^{L(G)} \prod_{n=0}^{C(G)-1} \left(\sum_\sigma e^{\beta h u_\sigma S_n} \right) \quad (19)$$

where the product is over all components of the graph and S_n denotes the size of the n-th component. We will assume that $n = 0$ refers to the largest component.

From (19) and (18) we find

$$\mathcal{Z}(\beta, h = 0, q, N) = \sum_G \left(\frac{\beta}{N} \right)^{L(G)} q^{C(G)}. \quad (20)$$

and comparison with (6) and (2) yields to leading order in N

$$\mathcal{Z}(\gamma, h = 0, q, N) = e^{\frac{\gamma N}{2}} Z(\gamma, q, N). \quad (21)$$

Correspondingly from (7) and (14) it follows that

$$f(\gamma, q) = -\frac{1}{2} - \frac{1}{\gamma}\varphi(\gamma, q). \tag{22}$$

Eqs. (21) and (22) establish the relation between the random graph ensemble defined in section 2 and the statistical mechanics of the Potts model sketched in section 3.1. In particular we obtain from (22) and (15)

$$\begin{aligned}\varphi(\gamma, q) = \operatorname*{extr}_{s_0} \Bigg[& \frac{\gamma}{2}\frac{q-1}{q}(s_0^2 - 1) + \ln q \\ & -\frac{1+(q-1)s_0}{q}\ln(1+(q-1)s_0) - \frac{q-1}{q}(1-s_0)\ln(1-s_0)\Bigg]\end{aligned} \tag{23}$$

from which $\omega(c, \gamma)$ follows with the help of the Legendre transform (9), (10). The equation for the saddle-point value $\overline{s}_0(\gamma, q)$ in (23) is from (16)

$$e^{\gamma \overline{s}_0} = \frac{1+(q-1)\overline{s}_0}{1-\overline{s}_0}. \tag{24}$$

We also note for later convenience that using (24) in (23) the expression for $\varphi(\gamma, q)$ can be rewritten as

$$\varphi(\gamma, q) = -\frac{\gamma}{2}\,\frac{q-1}{q}\,(1+\overline{s}_0^2) - \frac{\gamma \overline{s}_0}{q} + \ln(q - 1 + e^{\gamma \overline{s}_0}). \tag{25}$$

To elucidate the meaning of $\overline{s}_0$ for the random graph ensemble we consider the case $u_\sigma = \delta(\sigma, 0)$. Replacing β by γ in (19) we then find

$$\mathcal{Z}(\gamma, h, q, N) = \sum_G \left(\frac{\gamma}{N}\right)^{L(G)} \prod_{n=0}^{C(G)-1} \left(q - 1 + e^{\gamma h S_n}\right). \tag{26}$$

For the derivative of the free energy (14) with respect to h this implies

$$\begin{aligned}&\frac{\partial f}{\partial h}(\gamma, h, q) = \\ &- \lim_{N\to\infty} \frac{1}{N}\frac{1}{\mathcal{Z}} \sum_G \left(\frac{\gamma}{N}\right)^{L(G)} \prod_{n=0}^{C(G)-1} \left(q-1+e^{\gamma h S_n}\right) \sum_m S_m \,\frac{e^{\gamma h S_m}}{q-1+e^{\gamma h S_m}}.\end{aligned} \tag{27}$$

Assuming the existence of exactly one extensive component, i.e. $S_0 = O(N)$ and $S_m = o(N)$ for $m > 0$, we have

$$\lim_{h\to 0}\lim_{N\to\infty} \frac{e^{\gamma h S_m}}{q-1+e^{\gamma h S_m}} = \begin{cases} 1 & \text{if} \quad m = 0 \\ \frac{1}{q} & \text{else} \end{cases}. \tag{28}$$

Observing finally $\sum_{m>0} S_m = N - S_0$ we find

$$\lim_{h\to 0} \frac{\partial f}{\partial h}(\gamma, h, q) = \frac{1}{q}\big(1 + (q-1)\frac{\langle S_0\rangle}{N}\big), \tag{29}$$

where the average $\langle \ldots \rangle$ is with respect to the biased probability distribution (5). Comparing this expression with the result obtained for the same quantity from (126) we realize that the stable solution $\overline{s}_0(\gamma, q)$ of (24) is nothing but the average fraction of vertices in the largest component, $\overline{s}_0 = \langle S_0\rangle/N$, in an ensemble of random graphs with biased probability (5). Hence the phase transition in the Potts model describing the appearance of a spontaneous magnetization at sufficiently low temperature $1/\beta$ corresponds to a percolation transition in the random graph ensemble giving birth to a *giant component* with extensively many vertices at sufficiently large connectivity parameter γ.

It is finally useful to write (19) with β replaced by γ in the form

$$\begin{aligned}\mathcal{Z}(\gamma, h, q, \{u_\sigma\}, N) &= \mathcal{Z}(\gamma, h=0, q, N) \\ &\Big\langle \exp\Big(\sum_{n=0}^{C(G)-1} \ln(\sum_\sigma e^{\gamma h u_\sigma S_n(G)}) - C(G)\ln q\Big)\Big\rangle , \end{aligned}\tag{30}$$

where the average is again taken with (5). We now concentrate on cases with $u_0 = 1$ and $|u_\sigma| < 1$ for all $\sigma \neq 0$. Singling out a possible giant component of size $S_0 = N s_0$ and grouping together all small components of the same size we obtain for the free energy (14)

$$\begin{aligned}&f(\gamma, h, q, \{u_\sigma\}) = f(\gamma, h=0, q) - \\ &\lim_{N\to\infty} \frac{1}{\gamma N} \ln\Big\langle \exp(N\big[\gamma h s_0(G) + \sum_S \psi(S,G) \ln \sum_\sigma e^{\gamma h u_\sigma S} - c\ln q\big])\Big\rangle .\end{aligned}\tag{31}$$

Here we have introduced the number of components of size S of graph G divided by N

$$\psi(S,G) = \frac{1}{N} \sum_{n=1}^{C(G)-1} \delta(S, S_n(G)) . \tag{32}$$

Eq. (31) forms a suitable starting point for the characterization of the distribution of small components from the Potts free energy.

3.3 Properties of atypical graphs

The connection between the Potts free energy and the component generating function of Erdös-Rényi graphs allows to elucidate several large deviation properties of the random graph ensemble. First we get for the average number of edges per vertex from (11) and (23)

$$\overline{\ell}(\gamma,q) = \frac{\gamma}{2q}\big(1+(q-1)\overline{s}_0^2(\gamma,q)\big)\ . \tag{33}$$

The dependence of $\overline{\ell}$ on the relative size of the giant component $\overline{s}_0(\gamma,q)$ for $q \neq 1$ indicates a non-trivial internal organization of edges in rare graphs.

For the number of components of graphs dominating the distribution (5) we find from (23) and (10)

$$\overline{c}(\gamma,q) = \big(1-\overline{s}_0(\gamma,q)\big)\Big(1-\frac{\gamma}{2q}\big(1-\overline{s}_0(\gamma,q)\big)\Big)\ . \tag{34}$$

The above equations already give access to some microscopic information on edges and vertices belonging to the giant component or to the small components. Call $\overline{\ell}_{\rm in}$ and $\overline{\ell}_{\rm out}$ the average numbers of edges inside and outside the giant component divided by N respectively. Obviously, $\overline{\ell}_{\rm in} + \overline{\ell}_{\rm out} = \overline{\ell}$. In addition, since almost all small components are trees (cf. section 4.1), the number of these components is related to the number of edges they contain through $\overline{c} = 1-\overline{s}_0-\overline{\ell}_{\rm out}$. From these two relations, we obtain

$$\begin{aligned}\overline{\ell}_{\rm in}(\gamma,q) &= \frac{\gamma}{2q}\big(2\,\overline{s}_0+(q-2)\,s_0^2\big)\\ \overline{\ell}_{\rm out}(\gamma,q) &= \frac{\gamma}{2q}\big(1-\overline{s}_0\big)^2\end{aligned} \tag{35}$$

from which we deduce the average degrees

$$\begin{aligned}\overline{d}_{\rm in}(\gamma,q) &= \frac{\gamma}{q}\big(2+(q-2)\,\overline{s}_0\big)\\ \overline{d}_{\rm out}(\gamma,q) &= \frac{\gamma}{q}\big(1-\overline{s}_0\big)\end{aligned} \tag{36}$$

of vertices inside and outside the giant component respectively. The dependence of these degrees on q for one particular value of γ is shown in Fig. 2 together with results from numerical simulations described in section 5.

In order to calculate the complete spectrum $\omega(c,q)$ using the Legendre transform (9) we need to know $\varphi(\gamma,q)$ for general real $q>0$. We have hence to study the extremization over s in (23) for fixed γ and variable q. This is somewhat complementary to what is done in the statistical mechanics of the Potts model where the free energy (15) is minimized for integer $q \geq 2$ and different values of β. Here we have to keep in mind that the extremum in (23) is a *minimum* if $q>1$ but a *maximum* if $0<q<1$[1].

[1]The reason for this is that due to the constraint (124) the free energy depends on $(q-1)$ variables, a number which becomes negative for $q<1$.

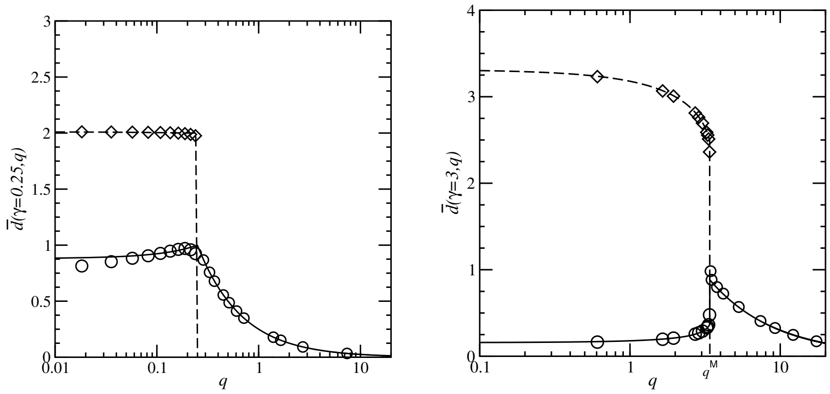

Figure 2: Average degrees $\overline{d}_{\rm in}(\gamma, q)$ (dashed top) and $\overline{d}_{\rm out}(\gamma, q)$ (full) as functions of q according to (36) for $\gamma = 0.25$ (left) and $\gamma = 3$ (right). The symbols indicate numerical results (diamond=inside, circle=outside the giant component). The statistical error bars are much smaller than the symbol size.

The dependence of $\varphi(\gamma, q)$ on q is qualitatively different for $\gamma \leq 2$ and $\gamma > 2$ as shown in Fig. 3. For $\gamma \leq 2$ the stable solution of (24) is positive for $q < \gamma$, goes to zero for $q \to \gamma$, and is identically zero for $q > \gamma$ (cf. left inset in Fig. 3). Accordingly $\varphi(\gamma, q)$ shows a second order phase transition at $q = \gamma$ as displayed in the left part of Fig. 3. For $\gamma > 2$ the small-q solution $\overline{s}_0(\gamma, q) > 0$ remains stable up to $q = q^{\rm s} > \gamma$ and coexists for $\gamma < q < q^{\rm s}$ with the solution $\overline{s}_0 = 0$ which is stable for $q > \gamma$ as before, see right inset in Fig. 3. Accordingly the phase transition is now first order and takes place at the Maxwell point $q = q^{\rm M}$ where the two values of $\varphi(\gamma, q)$ coincide as shown in the right panel of Fig. 3. At the transition, the value of $\overline{s}_0$ jumps *discontinuously*, and so does the derivative of $\varphi(\gamma, q)$ with respect to q.

Let us now turn to the discussion of $\omega(c, \gamma)$. The bifurcation point $q = \gamma$ of $\varphi(\gamma, q)$ maps according to (9) and (10) onto $c = 1/2$ for *all* values of γ. On the other hand the different behaviour of $\varphi(\gamma, q)$ for $\gamma \leq 2$ and $\gamma > 2$ implies qualitative differences of $\omega(c, \gamma)$ in the two cases as well.

For $\gamma \leq 2$ we find from the Legendre transform (9) that for all values of q there is exactly *one* corresponding value of c. Accordingly the transition from the percolating phase $\overline{s}_0 > 0$ at $c < 1/2$ to the small component phase at $c > 1/2$ is smooth as shown by the curves for $\gamma = 0.25, 1, 2$ in Fig. 4. From the expansion of (23) and (24) for values of q near γ we find that together with $\omega(c)$ also its first and second derivatives with respect to c

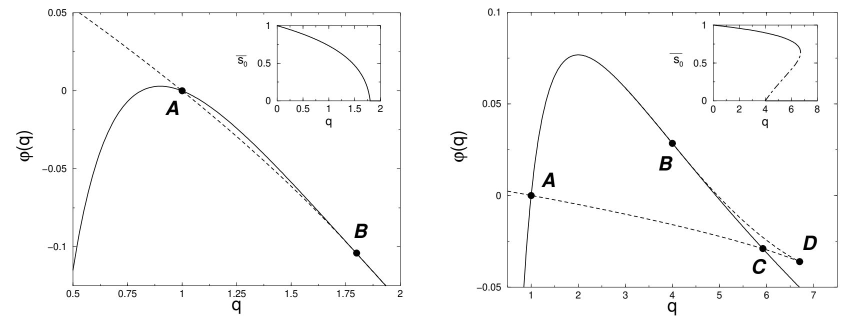

Figure 3: Free energy (23) of the random graph ensemble as a function of q for connectivities $\gamma = 1.8$ (left) and $\gamma = 4$ (right). The full and dashed curve correspond to the small clusters phase ($\overline{s}_0 = 0$) and the giant component phase ($\overline{s}_0 > 0$) respectively. Both free energies coincide in $q = 1$ (point A). For $\gamma \leq 2$ (left), a second order phase transition arises when q crosses $q = \gamma$ (point B) with the size of the giant component approaching zero continuously (left inset). When $\gamma > 2$ (right), the transition takes place at point C with abscissa $q^{\mathrm{M}} > \gamma$ and is first order. Both the slope of the free energy and the size of the giant component (right inset) are discontinuous at the transition. Branches BD and CD correspond to unstable (local maximum) and metastable (secondary local minimum) solutions respectively.

are continuous at $c = 1/2$ whereas the third derivative exhibits a jump if $\gamma \neq 1$. The transition is therefore third order in the thermodynamical sense if $\gamma \neq 1$. Except for $q = 1$ the appearance of the giant component takes place in graphs with exponentially small probabilities, $\omega(c = 1/2, \gamma) < 0$. For $\gamma < 1$ this happens in the increasing part of $\omega(c, \gamma)$, for $\gamma > 1$ in the decreasing one in accordance with the fact that the slope of $\omega(c, \gamma)$ is given by $-\ln q$, cf. (10), and that $\gamma = q$ at the transition.

For $\gamma > 2$ the first order transition in $\varphi(\gamma, q)$ implies via (9) that for one particular value of q, namely $q = q^M$, there are *two* corresponding values, c_1^{M} and c_2^{M}, of c. Hence the biased probability distribution $P(C; \gamma, q^M, N)$ is bimodal and $\omega(c, \gamma)$ is non-convex in the interval $c_1^{\mathrm{M}} < c < c_2^{\mathrm{M}}$. At the same time the Legendre transform (9) does only yield the convex hull of $\omega(c, \gamma)$ and therefore includes a linear part with slope $-\ln q^{\mathrm{M}}$ interpolating between $\omega(c_1^{\mathrm{M}}, \gamma)$ and $\omega(c_2^{\mathrm{M}}, \gamma)$ as shown exemplarily for $\gamma = 3$ with the dotted line in Fig. 4. A random graph ensemble generated according to $P(C; \gamma, q^M, N)$ is hence inhomogeneous in the sense that it contains realizations with $c = c_1^{\mathrm{M}}$ (and with giant component) and with $c = c_2^{\mathrm{M}}$ (and

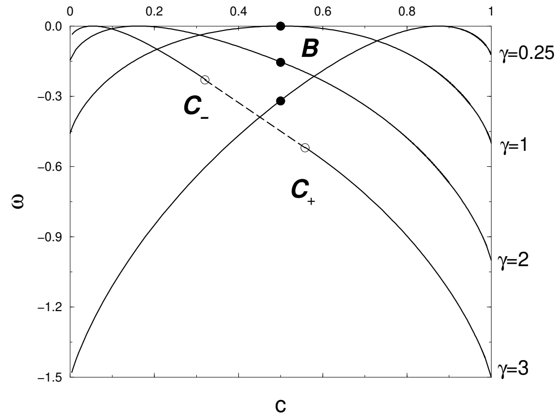

Figure 4: Logarithmic probability distribution $\omega(c,\gamma)$ of the number of components per vertex, c, for different values of the connectivity parameter $\gamma = 0.25, 1, 2, 3$ (left bottom to top). ω is maximal and zero for the most probable fraction of components c^* given by (51). For $\gamma \leq 2$, there is a second order percolation transition at $c = 1/2$ (points B) marking the appearance of a giant component for $c < 1/2$. When $\gamma > 2$, a first order transition separates the giant component phase (left to point C_-) from the phase without giant component (right to C_+). In between, both phases coexist and the convex hull of ω is linear in c (dashed line).

without giant component). The value of $\overline{c}$ in such an ensemble depends on the relative fraction of these two realizations and is determined by pre-exponential factors in $P(C;\gamma,q^M,N)$. The fraction of realizations without giant component is zero for $\overline{c} = c_1^{\mathrm{M}}$, increases linearly with $\overline{c}$, and reaches one at $\overline{c} = c_2^{\mathrm{M}}$.

The above analytical results for $\omega(c,\gamma)$ including the bimodal distribution $P(C;\gamma,q,N)$ for $q = q^M$ are in very good agreement with extensive numerical simulations described in section 5. This is exemplified for $\gamma = 0.25$ and $\gamma = 3$ in Fig. 5.

For $c > \max(1/2, c_2^{\mathrm{M}})$, *i.e.* in the region where $\overline{s}_0(\gamma,q) = 0$, it is possible to perform the Legendre transform (9) analytically to find

$$\omega(c,\gamma) = -\frac{\gamma}{2} + (1-c)(1 + \ln\frac{\gamma}{2} - \ln(1-c)). \tag{37}$$

Hence we have $\omega(c=1,\gamma) = -\gamma/2$ for all values of γ which is, of course, consistent with Fig. 4. This result holds as long as γ is finite. Another interesting large-q limit is obtained if q and γ tend to infinity simultaneously with the ratio $r = \ln q/\gamma$ being kept constant [20]. The tendency

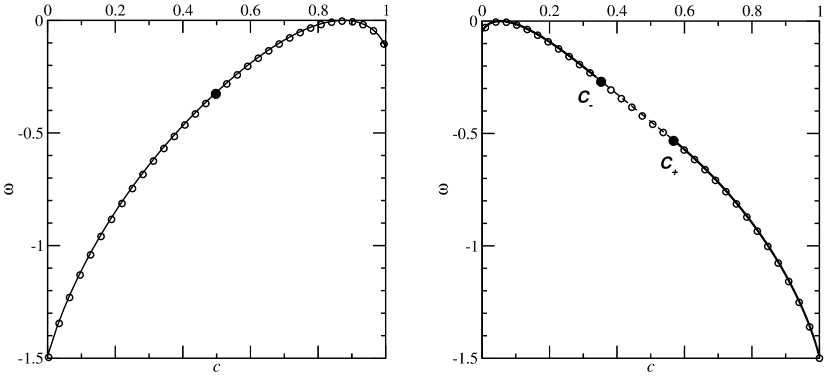

Figure 5: Comparison between analytical (full lines) and numerical (symbols) results for the logarithmic probability $\omega(c,\gamma)$ of Erdös-Renyi graphs with atypical number of components for $\gamma = 0.25$ (left) and $\gamma = 3$ (right). The simulations were done for $N = 1000$ and are described in section 5, the statistical error bars are much smaller than the symbol size. The big black dots have the same meaning as in Fig. 4.

to prefer graphs with many components implied by $q \to \infty$ may then be counterbalanced by the large connectivity parameter γ. In fact for $r > 1/2$ we have $q > q^M(\gamma)$ and hence $\overline{s}_0 = 0$ which brings us back to (37). On the other hand for $r \leq 1/2$ we find from (24) to leading order $\overline{s}_0 = r$ and hence from (34) $\overline{c} = 1 - r$. Therefore in this case γ is large enough to set up a giant component while the other vertices are essentially isolated in order to make the number of components as large as possible.

The opposite limit $c \to 0$ corresponds to $q \to 0$. The random graph ensemble is for very small q dominated by graphs with very few components and for $q \to 0$ only fully connected graphs (i.e. those with $C = 1$) survive. From (24) and (25) we find in this limit

$$\overline{s}_0(\gamma, q) = 1 - \frac{q}{e^\gamma - 1} + O(q^2) \tag{38}$$

$$\varphi(\gamma, q) = \ln(1 - e^{-\gamma}) + q\,\frac{\gamma(e^\gamma + 1)}{2(e^\gamma - 1)^2} + O(q^2). \tag{39}$$

This results via (10) and (9) in

$$\overline{c}(\gamma, q) = q\,\frac{\gamma(e^\gamma + 1)}{2(e^\gamma - 1)^2} + O(q^2) \tag{40}$$

consistent with $C \to 1$ and

$$\omega(c, \gamma) = \ln(1 - e^{-\gamma}) - c \ln c + O(c) \tag{41}$$

The first term in this small c expansion, $\omega(c=0,\gamma) = \ln(1-e^{-\gamma})$, again agrees with Fig. 4. It characterizes graphs with just a single component and reproduces the known rigorous result that the probability of an Erdös-Rényi random graph to be connected is asymptotically given by $(1-e^{-\gamma})^N$ [11]. The next term in the expansion, $c \ln c$, is of entropic origin and describes the number of ways in which $C \ll N$ components may be formed from N vertices.

We may finally extract useful information on the size distribution of small components from the Potts free energy. Let us denote by

$$\overline{\psi}(S,\gamma,q) = \sum_G P(G;\gamma,q,N)\ \psi(S,G) \tag{42}$$

with $\psi(S,G)$ defined by (32) the average distribution of small components in a graph ensemble characterized by the biased distribution $P(G;\gamma,q,N)$. Consistent with the meaning of $\psi(S,G)$ we then find to leading order in N

$$\sum_S \overline{\psi}(S,\gamma,q) = \overline{c}(\gamma,q) \tag{43}$$

$$\sum_S \overline{\psi}(S,\gamma,q)\ S = 1 - \overline{s}_0(\gamma,q)\ . \tag{44}$$

To get in addition an expression for the second moment of $\overline{\psi}(S,\gamma,q)$ it is useful to consider the second derivative of $f(\gamma,h,q,\{u_\sigma\})$ with respect to h at $h=0$ for field configurations with

$$u_0 = 1 \qquad \text{and} \qquad |u_\sigma| < 1 \quad \text{for} \quad \sigma = 1,...,q-1\ . \tag{45}$$

Denoting by $\langle s_0^2\rangle_c = \langle s_0^2\rangle - \langle s_0\rangle^2$ the second cumulant of the relative size of the giant component and using the abbreviations

$$\widehat{u} = \frac{1}{q}\sum_\sigma u_\sigma \qquad \text{and} \qquad \widehat{u^2} = \frac{1}{q}\sum_\sigma u_\sigma^2 \tag{46}$$

on can show from (31) that

$$-\frac{1}{\gamma}\,\frac{\partial^2 f}{\partial h^2}(\gamma,h=0,q) = (\widehat{u^2} - \widehat{u}^2)\sum_S \overline{\psi}(S,\gamma,q)\ S^2 + (1-\widehat{u})^2\ N\langle s_0^2\rangle_c\ . \tag{47}$$

On the other hand one finds from (126) for the same quantity after some algebra

$$\begin{aligned} -\frac{1}{\gamma}\,\frac{\partial^2 f}{\partial h^2}(\gamma,h=0,q) &= (\widehat{u^2} - \widehat{u}^2)\ \frac{q(1-\overline{s}_0)}{q-\gamma(1-\overline{s}_0)} \\ &+ (1-\widehat{u})^2\ \frac{q^2\ \overline{s}_0\ (1-\overline{s}_0)}{(q-\gamma(1-\overline{s}_0))(q-\gamma(1-\overline{s}_0)(1+(q-1)\overline{s}_0))}\ . \end{aligned} \tag{48}$$

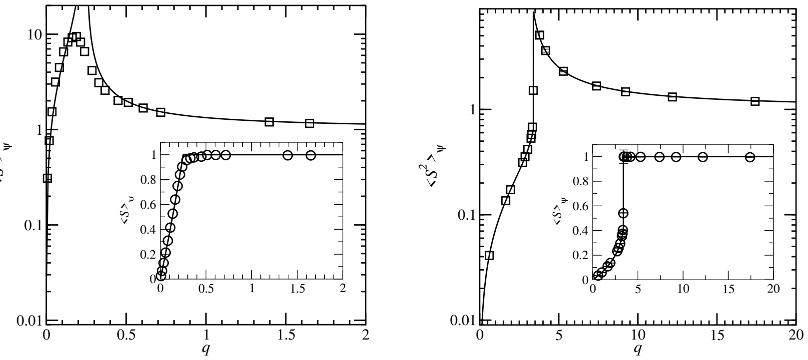

Figure 6: First and second moment of the distribution of non-extensive component sizes $\overline{\psi}(S,\gamma,q)$ as function of q for $\gamma = 0.25$ (left) and $\gamma = 3$ (right). Full lines are the analytical expressions describing the thermodynamic limit $N \to \infty$, symbols give results of numerical simulations for $N = 1000$ described in section 5.

Since (47) and (48) must be identical for any choice of the fields u_σ consistent with (45) we are left with

$$\sum_S \overline{\psi}(S,\gamma,q)\; S^2 = \frac{q\;(1-\overline{s}_0)}{q-\gamma(1-\overline{s}_0)} \tag{49}$$

$$\langle s_0^2\rangle_c = \frac{1}{N}\;\frac{q^2\;\overline{s}_0\;(1-\overline{s}_0)}{(q-\gamma(1-\overline{s}_0))(q-\gamma(1-\overline{s}_0)(1+(q-1)\overline{s}_0))}\;. \tag{50}$$

Fig. 6 shows the first and second moment of $\overline{\psi}(S,\gamma,q)$ for two values of γ as function of q together with results from the numerical simulations. Note that for $\gamma \leq 2$ the soft transition at $\gamma = q$ gives rise to a diverging second moment of $\overline{\psi}(S,\gamma,q)$ (left panel) whereas for $\gamma > 2$ it remains finite at the transition (right panel). Accordingly the finite size corrections at the transition are much larger in the first case.

It appears to be possible to extend the above procedure to obtain also higher moments of $\overline{\psi}(S,\gamma,q)$, however the calculations become increasingly tedious. We have not been able to derive a closed expression for the complete distribution $\overline{\psi}(S,\gamma,q)$ from the Potts free energy except for the case $q = 1$ which is discussed in the next section. A general expression for $\overline{\psi}(S,\gamma,q)$ will, however, be derived in section 4.6 using our microscopic approach.

3.4 Properties of typical random graphs

In this subsection we rederive some of the central results for typical graphs as special cases of our more general framework. As discussed in section 2 the random graph ensemble is for large values of N dominated by graphs contributing to the maximum of $\omega(c,\gamma)$. Since at this maximum $\partial\omega/\partial c = 0$ we find from (10) that typical properties of random graphs can be extracted from the Potts free energy in the vicinity of $q = 1$. This is well known [15] and is, of course, also clear from the definition (5) implying $P(G;\gamma,q=1,N) = P(G;\gamma,N)$.

Explicitly we find for the typical number of components c^* from (10) and (23)

$$c^* = (1-s_0^*)(1-\frac{\gamma}{2}(1-s_0^*)) \tag{51}$$

where the relative size of the giant component s_0^* is the stable solution of the equation

$$1 - s_0^* = e^{-\gamma s_0^*} \tag{52}$$

which follows from (24) for $q = 1$. Eqs. (51) and (52) are classical results of Erdös and Rényi [1]. For small values of γ almost all components of the graph are small *trees*. Hence $s_0^* = 0$ and each edge reduces the number of components by one. With the typical number of edges per vertex given by (cf. (11) and (23))

$$\ell^* = \overline{\ell}(\gamma, q=1) = \frac{\gamma}{2} \tag{53}$$

this implies $c^* = 1-\gamma/2$ which coincides with (51) for $s_0^* = 0$. For $\gamma > 1$ there is on average more than one edge attached to each vertex and hence the connectivity may spread out through the whole system resulting in the emergence of a giant component. Its size s_0^* is an increasing function of γ. At the same time the giant component has a denser connectivity than a tree involving loops which slows down the decrease of the number of components c^* with γ as described by (51). The dependence of s_0^* and c^* on γ is shown in Fig. 7. The reason for the remarkable similarity between the results (51) and (34) for the number of components in typical and atypical graphs respectively will become clear in section 4.5.

For the case of typical graphs considered in this subsection it is possible to obtain some more detailed results. A simple application of Bayes' equation for conditional probabilities [19] yields the complete degree distribution inside and outside the giant component. The probability of a vertex to have d edges is given by (1). The probability not to belong to the giant component conditioned to having d edges is clearly $P(\text{out}; d) = (1-s_0^*)^d$. The complementary probability to belong to the giant component is hence $P(\text{in}; d) = 1-(1-s_0^*)^d$. Then from Bayes' theorem we get for the probability to have degree d conditioned to being not part and being part of the

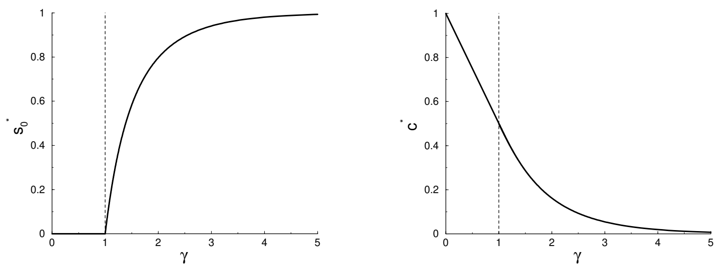

Figure 7: Properties of typical random graphs. Shown are the fraction s_0^* of vertices in the largest component (left), and the number c^* of components per vertex (right) as function of the connectivity parameter γ. The vertical dashed line, $\gamma = 1$, indicates the location of the percolation transition.

giant component respectively

$$P^*_{\rm out}(d) = \frac{P({\rm out};d)\,P(d)}{P({\rm out})} = \frac{(1-s_0^*)^d}{1-s_0^*}\,e^{-\gamma}\frac{\gamma^d}{d!} = e^{-\gamma(1-s_0^*)}\frac{(\gamma(1-s_0^*))^d}{d!} \tag{54}$$

$$P^*_{\rm in}(d) = \frac{P({\rm in};d)\,P(d)}{P({\rm in})} = \frac{1-(1-s_0^*)^d}{s_0^*}\,e^{-\gamma}\frac{\gamma^d}{d!}. \tag{55}$$

The last equality in (54) in which we have used (52) shows that the degree distribution outside the giant component is still Poissonian. On the other hand the distribution inside the giant component clearly deviates from a Poissonian law. Calculating the averages of the distributions (54) and (55) we find

$$\begin{aligned} d^*_{\rm in}(\gamma) &= \gamma(2-s_0^*) \\ d^*_{\rm out}(\gamma) &= \gamma(1-s_0^*), \end{aligned} \tag{56}$$

consistent with (36) for $q = 1$.

Also the complete distribution of component sizes can be determined from the Potts free energy. In fact for the special choice $u_\sigma = \delta(\sigma, 0)$ we find from (31)

$$\frac{\partial^2 f}{\partial q\,\partial h}(\gamma, h, q=1) = \sum_S \psi^*(S,\gamma)\;S\;e^{-\gamma h S}\;, \tag{57}$$

with $\psi^*(S,\gamma) = \overline{\psi}(S,\gamma,q=1)$. On the other hand from (126) we get for the same configuration of the fields u_σ the result

$$\frac{\partial^2 f}{\partial q\,\partial h}(\gamma, h, q=1) = 1 - \tilde{s}_0(\gamma, h)\;, \tag{58}$$

where $\tilde{s}_0(\gamma, h)$ is the solution of

$$1 - \tilde{s}_0 = e^{-\gamma(h+\tilde{s}_0)}. \tag{59}$$

Comparing (57) and (58) we hence find

$$\sum_{S=1}^{\infty} \psi^*(S,\gamma)\, S\, e^{-\gamma h S} = 1 - \tilde{s}_0(\gamma, h)\ . \tag{60}$$

From (59) and (60) it is straightforward to produce equations for all the moments of $\psi^*(S,\gamma)$ through successive differentiation in $h = 0$. A more direct way to obtain $\psi^*(S,\gamma)$ is to get from (59) the explicit dependence of $\tilde{s}_0(\gamma, h)$ on γ and h with the help of the Lagrange inversion theorem [21]

$$1 - \tilde{s}_0 = \frac{1}{\gamma} \sum_{S=1}^{\infty} \frac{S^{S-1} e^{-\gamma h S}}{S!} \left(\gamma\, e^{-\gamma}\right)^S . \tag{61}$$

From (60) and (61) we then infer

$$\sum_{S=1}^{\infty} \psi^*(S,\gamma)\, S\, e^{-\gamma h S} = \frac{1}{\gamma} \sum_{S=1}^{\infty} \frac{S^{S-1} e^{-\gamma h S}}{S!} \left(\gamma\, e^{-\gamma}\right)^S . \tag{62}$$

Matching powers of $e^{-\gamma h}$ we finally obtain

$$\psi^*(S,\gamma) = \frac{1}{\gamma} \frac{S^{S-2}}{S!} \left(\gamma e^{-\gamma}\right)^S , \tag{63}$$

another classical result of Erdös and Rényi [1]. For the second moment of this distribution we easily find

$$\sum_{S=1}^{\infty} \psi^*(S,\gamma)\, S^2 = \frac{\gamma(1-s_0^*)}{1-\gamma(1-s_0^*)} \tag{64}$$

which reproduces (49) for $q = 1$. We also note that from the complementary equation (50) we get for $q = 1$

$$\langle s_0^2 \rangle_c = \frac{1}{N}\, \frac{s_0^*\, (1-s_0^*)}{(1-\gamma(1-s_0^*))^2}\ , \tag{65}$$

a result consistent with rigorous findings about the fluctuations of the relative size of the giant component of typical Erdös-Rényi graphs [22].

4 Evolution of atypical graphs

In the present section we will look at rare graphs from a more microscopic point of view focusing on individual vertices and edges. Our aim will be to rederive several of the thermodynamic results presented above without reference to the Potts model. To this end we will study the evolution of rare random graphs under the addition of a new vertex or a new edge. This is similar in spirit to the so-called *cavity method* in the statistical mechanics of disordered systems [14]. The main motivation of what follows is to find an alternative way to quantitatively characterize rare graphs. It may be helpful in the analysis of graphs which are atypical with respect to other properties than the number of components, as *e.g.* the size of the giant component or the number of loops. In these cases the relation to the Potts model is no longer helpful and no thermodynamic approach seems to be known. Finally we will also derive some new results including the complete degree distributions inside and outside the giant component and the size distribution of non-extensive components. These results were obtained above for typical graphs only.

4.1 Energetic versus entropic costs

From a microscopic point of view we may identify two qualitatively different reasons for the exponentially small probability of a graph G. On the one hand the *number* of edges in the graph may deviate by an extensive amount from the typical number. On the other hand the *distribution* of edges among the vertices of the graph may differ from the typical one. We will refer to these two different sources for an exponentially small probability as *energetic* and *entropic* contribution respectively.

The energetic cost is completely fixed by the probability distribution of edges. The probability for a random graph to have $L = \ell N$ edges is given by (cf. (2))

$$\begin{aligned} P(L;\gamma,N) &= \binom{N^2/2}{L} \left(\frac{\gamma}{N}\right)^L \left(1-\frac{\gamma}{N}\right)^{N^2/2-L} \\ &= \exp\left(N\left[\ell \ln\frac{\gamma}{2\ell} + \ell - \frac{\gamma}{2}\right] + O(1)\right). \end{aligned} \tag{66}$$

The expression in the brackets is zero for $\ell = \ell^* = \gamma/2$ reproducing (53). It is negative for $\ell \neq \ell^*$ and hence all other values of ℓ have probabilities exponentially small in N.

To leading order in N we find from (66)

$$P(L+1;\gamma,N) = \frac{\gamma}{2\,\ell} P(L;\gamma,N), \tag{67}$$

which gives the change in the energetic contribution to the probability when one edge is added. For $\ell < \gamma/2$ the probability increases by the insertion of an edge, for $\ell > \gamma/2$ it decreases in accordance with the fact that $\ell = \gamma/2$ is the typical case.

Let us then consider a graph with N vertices, no giant component, and an atypically large number C of non-extensive components. A possible realization of such a graph has all components as trees and an atypically small number, $L = N - C$, of edges. However, this may be not the optimal way to build the graph. In fact from (67) we see that the probability of the graph increases by a factor of order 1 if we add another edge. On the other hand, in order not to decrease at the same time the number of components we have to put the new edge between two vertices of one of the *already existing* components. For a component with $S = O(1)$ vertices (and hence $(S-1)$ edges) the chance to put the new edge between two of its vertices is roughly $(S-1)(S-2)/N^2 = O(N^{-2})$. Multiplying by the number of components we find that the probability not to reduce this number by putting the new edge is of order $O(1/N)$. For large N this decrease in probability cannot be compensated by the $O(1)$ energetic gain. Hence, also in the case of rare graphs the non-extensive components are predominantly *trees*.

The situation for graphs G without giant component is hence rather clear. Since all components are trees their number is given by $C(G) = N - L(G)$. This implies a simple relation between the generating functions for rare graphs ($q \neq 1$) and typical graphs ($q = 1$) which follows from (6) and (2):

$$Z_{s_0=0}(\gamma, q, N) = q^N Z_{s_0=0}(\frac{\gamma}{q}, 1, N). \tag{68}$$

Therefore these graphs are characterized by an effective connectivity parameter γ/q and the number of edges per vertex is given by $\overline{\ell} = \gamma/(2q)$ consistent with (33) for $\overline{s}_0 = 0$. The probability of such a graph is solely determined by the energetic cost (66) yielding

$$P(G; \gamma, q, N, \overline{s}_0 = 0) = \exp\left(N\left[\frac{\gamma}{2q}\left(1 - q + \ln q\right)\right]\right). \tag{69}$$

Replacing q in this expression by c according to (10) and (23) we find back the logarithmic probability (37).

The situation changes in the presence of a giant component of size $S_0 = O(N)$. From the same kind of reasoning as used above it is clear, that the entropic cost for putting an additional edge inside the giant component is $O(1)$ and may hence well be over-compensated by the energetic gain in probability [2]. The precise balance between energetic and entropic contributions in this case will be investigated in the next subsection.

[2] We expect the probability to have more than one extensive component to be negli-

4.2 Adding an edge

To quantify the entropic contribution to the probability let us consider the probability $P(C;L)$ for a graph with L edges to have C components. Upon adding one more edge the number of components may change and we have quite generally

$$P(C;L+1) = \sum_{\Delta C} K(\Delta C)\, P(C+\Delta C;L). \tag{70}$$

The kernel $K(\Delta C)$ is easily determined. If the new edge lies with both ends inside the giant component of the graph the number of components does not change at all, otherwise it is reduced by one,

$$K(\Delta C) = s_0^2\, \delta(\Delta C, 0) + (1-s_0^2)\, \delta(\Delta C, 1). \tag{71}$$

Combining (67) and (70) we hence find for the probability $P(C,L;\gamma,N) = P(C;L)P(L;\gamma,N)$ the evolution equation

$$P(C,L+1;\gamma,N) = \frac{\gamma}{2\,\ell} \left(s_0^2\, P(C,L;\gamma,N) + (1-s_0^2)\, P(C+1,L;\gamma,N) \right). \tag{72}$$

For the biased probability

$$P(L;\gamma,q,N) = \frac{1}{Z(\gamma,q,N)} \sum_C P(C,L;\gamma,N)\, q^C, \tag{73}$$

where $Z(\gamma,q,N)$ is defined by (6) this implies

$$P(L+1;\gamma,q,N) = \frac{\gamma}{2\,\ell\, q} \,(1+(q-1)s_0^2)\, P(L;\gamma,q,N) \tag{74}$$

as follows from (72) by multiplying with q^C and summing over C. Summing (74) over L and using the fact that this sum is dominated by graphs with $\ell = \overline{\ell}(\gamma,q)$ and $s_0 = \overline{s}_0(\gamma,q)$ we find for the average number of edges per vertex

$$\overline{\ell} = \frac{\gamma}{2\,q}(1+(q-1)\overline{s}_0^2). \tag{75}$$

reproducing (33).

In a similar way we may rederive the results (35) for the average number of edges inside and outside the giant component. To this end we decompose the kernel (71) into contributions corresponding to the cases of the new edge being connected to the giant component or not

$$K(\Delta C) = K_{\text{in}}(\Delta C) + K_{\text{out}}(\Delta C). \tag{76}$$

gible for the graphs atypical with respect to the number of components considered here, similarly to what happens for typical random graphs [23]. See also section 4.6.

Clearly

$$K_{\rm in}(\Delta C) = s_0^2\,\delta(\Delta C,0) + 2s_0(1-s_0)\,\delta(\Delta C,1) \tag{77}$$
$$K_{\rm out}(\Delta C) = (1-s_0)^2\,\delta(\Delta C,1). \tag{78}$$

Proceeding as above we get for the biased probability to have a graph with $L+1$ edges with the last edge added being connected to the giant component

$$P_{\rm in}(L+1;\gamma,q,N) = \frac{\gamma}{2\,\ell\,q}\,(2+(q-2)s_0^2)\,P(L;\gamma,q,N). \tag{79}$$

Summing over L this gives

$$P_{\rm in}(\gamma,q,N) = \frac{\gamma}{2\,\overline{\ell}\,q}\,(2+(q-2)\overline{s}_0^2) \tag{80}$$

and hence

$$\overline{\ell}_{\rm in} = \overline{\ell}\,P_{\rm in}(\gamma,q,N) = \frac{\gamma}{2\,q}\,(2+(q-2)\overline{s}_0^2), \tag{81}$$

which is identical with (35). Similarly one may rederive the result for $\overline{\ell}_{\rm out}$. The results for the average total number of edges and the average number of edges inside and outside the giant component respectively of atypical graphs are hence directly linked with the balance between energetic and entropic contributions to the probability of these graphs.

4.3 Adding a vertex

Several interesting results may be obtained by investigating the evolution of atypical graphs under addition of a new vertex. Compared with the same procedure for typical graphs some special care is needed in the present case of atypical graphs. The reason is the following. In order to keep the statistical properties of the new vertex as simple as possible we will assume that it is characterized by the simple Poissonian degree distribution (1). In this sense we add a *typical* vertex to an *atypical* graph. This in turn implies a change in the "degree of unlikeliness" of the graph which needs to be monitored.

To make this argument more quantitative consider the following basic step of adding one vertex. The probability of a graph G with N vertices and parameter γ is from (2) given by

$$P(G;\gamma,N) = e^{-\frac{\gamma N}{2}+\gamma(\frac{1}{2}-\frac{\gamma}{4}+\frac{L(G)}{N})+o(1)}\,\left(\frac{\gamma}{N}\right)^{L(G)}, \tag{82}$$

and depends on its number of edges, $L(G)$, only. The new vertex is assumed to have d incident edges with probability $P(d;\gamma) = e^{-\gamma}\,\gamma^d/d!$, cf.

(1). There are $\binom{N}{d}$ different ways to connect these d edges with existing vertices. The new graph is one of these possible "wirings", and has therefore probability

$$P(d;\gamma)\,\frac{1}{\binom{N}{d}}\,P(G;\gamma,N)$$
$$= e^{-\frac{\gamma}{2}N+\gamma(\frac{1}{2}-\frac{\gamma}{4}+\frac{L(G)}{N})-\gamma+o(1)}\;\left(\frac{\gamma}{N}\right)^{L(G)+d}\left(1+O(\frac{1}{N})\right). \quad (83)$$

But the new graph, G', is one particular graph with $N+1$ vertices and $L(G') = L(G) + d$ edges. Its probability should therefore be

$$P(G';\gamma,N+1) = e^{-\frac{\gamma}{2}(N+1)+\gamma(\frac{1}{2}-\frac{\gamma}{4}+\frac{L(G)+d}{N+1})+o(1)}\;\left(\frac{\gamma}{N+1}\right)^{L(G)+d}. \quad (84)$$

Equality of expressions (83) and (84) imposes that

$$\frac{\gamma}{2} = \ln\left(\frac{N+1}{N}\right)^{L(G)+d} = \frac{L(G)}{N} + O(\frac{1}{N}). \quad (85)$$

This is fulfilled for typical graphs since for these the mean number of edges per vertex is indeed $\gamma/2$, cf. (53). However, if we add a typical vertex to an atypical graph with $\ell \neq \gamma/2$ we have to introduce an *extra multiplicative weight factor*

$$w(\gamma,q) = \exp\left(\frac{\gamma}{2} - \ell(\gamma,q)\right) \quad (86)$$

in order to make the new graph an unbiased representative of the new ensemble.

Let us now investigate how the probability $P(C;\gamma,N)$ for a graph to have C components as defined in (3) changes when we add a new vertex. The number of components will decrease by a stochastic amount ΔC, and we have similarly to (70)

$$P(C;\gamma,N+1) = \sum_{\Delta C} K(\Delta C)\; P(C+\Delta C;\gamma,N). \quad (87)$$

The new kernel K has now to comprise both the extra weight (86) of the new vertex and the probability for the change ΔC when adding the new vertex. The degree d of the new vertex, which is also the number of new edges added to the graph, is a stochastic variable with Poissonian distribution $e^{-\gamma}\gamma^d/d!$. Of these d edges, d_0 may be connected with the giant component whereas the remaining $d - d_0$ ones are connected with small components which (with probability 1 for $N \to \infty$) are all different from each other. The number of components is hence reduced by $d - d_0$, except for the case $d_0 = 0$ where it changes by $d - 1$. We therefore find

$$K(\Delta C) = \sum_{d\geq 0} e^{-\frac{\gamma}{2}-\ell}\,\frac{\gamma^d}{d!}\sum_{d_0=0}^{d}\binom{d}{d_0} s_0^{d_0}\,(1-s_0)^{d-d_0}\,\delta(\Delta C, d-d_0-\delta(d_0,0)) \tag{88}$$

where s_0 is the relative size of the giant component *prior* to the insertion of the new vertex.

In order to obtain results for atypical graphs from (87) we again multiply by q^C and sum over C to find

$$Z(\gamma, q, N+1) = \Sigma(\gamma, q)\; Z(\gamma, q, N). \tag{89}$$

where

$$\Sigma(\gamma, q) = \sum_{\Delta C} \overline{K}(\Delta C)\; q^{-\Delta C}. \tag{90}$$

and $\overline{K}(\Delta C)$ results from $K(\Delta C)$ by replacing ℓ and s_0 by $\overline{\ell}(\gamma, q)$ and $\overline{s}_0(\gamma, q)$ as given by (24) and (75) respectively. Using (88), performing the sum over ΔC, d_0 and finally over d we are left with

$$\Sigma(\gamma, q) = (q - 1 + e^{\gamma \overline{s}_0})\;\exp\big(-\frac{\gamma}{2} - \overline{\ell} + \frac{\gamma}{q}\,(1 - \overline{s}_0)\big). \tag{91}$$

Iterating (89) we find

$$\lim_{N\to\infty}\frac{1}{N}\ln Z(\gamma, q, N) = \ln \Sigma(\gamma, q). \tag{92}$$

Comparison with (7) and insertion of (75) shows that (91) reproduces the result for the free energy $\varphi(\gamma, q)$ in the form (25). To complete the rederivation of results of the thermodynamic approach we still have to produce the self-consistent equation (24) for $\overline{s}_0(\gamma, q)$.

4.4 The giant component

More detailed results can be obtained by again decomposing the kernel $K(\Delta C)$ in (87) into different contributions. *E.g.* in order to reproduce the self-consistent equation for $\overline{s}_0$ we decompose $K(\Delta C)$ into parts corresponding to the possible values of d_0:

$$K(\Delta C) = \sum_{d_0\geq 0} K_{d_0}(\Delta C), \tag{93}$$

where

$$K_{d_0}(\Delta C) = e^{\frac{\gamma}{2}-\ell} \sum_{d=d_0}^{\infty} e^{-\gamma} \frac{\gamma^d}{d!} \binom{d}{d_0} s_0^{d_0} (1-s_0)^{d-d_0} \times \\ \delta(\Delta C, d-d_0-\delta(d_0,0)). \quad (94)$$

For the probability of a graph with $N+1$ vertices to have C components and the last vertex added making d_0 connections with the giant component we then get

$$P(C, d_0; \gamma, N+1) = \sum_{\Delta C} K_{d_0}(\Delta C)\ P(C+\Delta C; \gamma, N). \quad (95)$$

Multiplying with $q^C/Z(\gamma, q, N+1)$, summing over C, and specifying to the case $d_0 = 0$ we get for the biased probability that the new vertex does *not* belong to the giant component

$$\begin{aligned} P(d_0=0; \gamma, q, N+1) &= \frac{1}{Z(\gamma, q, N+1)} \sum_{\Delta C} \overline{K}_{d_0=0}(\Delta C)\ q^{-\Delta C}\ \ Z(\gamma, q, N) \\ &= \frac{\Sigma_{d_0=0}(\gamma, q)}{\Sigma(\gamma, q)}, \end{aligned} \quad (96)$$

where we have used (6), (89), and introduced

$$\Sigma_{d_0=0}(\gamma, q) = \sum_{\Delta C} \overline{K}_{d_0=0}(\Delta C)\ q^{-\Delta C}. \quad (97)$$

Performing the sum over ΔC and d we find

$$\Sigma_{d_0=0}(\gamma, q) = q\ \exp\left(-\frac{\gamma}{2} - \overline{\ell} + \frac{\gamma}{q}(1-\overline{s}_0)\right). \quad (98)$$

On the other hand for large N the probability (96) has to be identified with $1-\overline{s}_0(\gamma, q)$. Using (91) and (98) this yields finally

$$1 - \overline{s}_0 = \frac{q}{q-1+e^{\gamma \overline{s}_0}} \quad (99)$$

which coincides with (24).

4.5 The degree distribution

It is finally possible to derive expressions for the complete degree distribution in rare graphs by decomposing the kernel $K(\Delta C)$ in (87) into different parts according to the value of d (rather than d_0 as done above)

$$K(\Delta C) = \sum_{d \geq 0} K_d(\Delta C), \quad (100)$$

where now

$$K_d(\Delta C) = e^{-\frac{\gamma}{2}-\ell}\ \frac{\gamma^d}{d!} \sum_{d_0=0}^{d} \binom{d}{d_0} s_0^{d_0}\ (1-s_0)^{d-d_0} \times$$
$$\delta(\Delta C, d-d_0-\delta(d_0,0)). \quad (101)$$

For the probability of a graph with $N+1$ vertices to have C components and the last vertex added having d edges this implies

$$P(C,d;\gamma,N+1) = \sum_{\Delta C} K_d(\Delta C)\ P(C+\Delta C;\gamma,N), \qquad (102)$$

Multiplying by $q^C/Z(\gamma,q,N+1)$ and summing over C we find for the biased probability that the added vertex has degree d,

$$P(d;\gamma,q,N+1) = \frac{\Sigma_d(\gamma,q)}{\Sigma(\gamma,q)}\ , \qquad (103)$$

where we have again used (89), and defined

$$\Sigma_d(\gamma,q) = \sum_{\Delta C} \overline{K}_d(\Delta C)\ q^{-\Delta C}. \qquad (104)$$

Calculating explicitly the above sum, and using (91) and (99), we obtain the degree distribution,

$$P(d;\gamma,q) = \frac{1-\overline{s}_0}{q} e^{-\frac{\gamma}{q}(1-\overline{s}_0)}\ \frac{\gamma^d}{d!}\ \left[(\overline{s}_0 + \ \frac{1-\overline{s}_0}{q})^d + (q-1)(\frac{1-\overline{s}_0}{q})^d\right]\ . \qquad (105)$$

This distribution reduces to the Poissonian law expected from (1) for $q=1$ only. For $q \neq 1$ we find deviations from a Poissonian degree distribution, where for large values of γ and q even a bimodal distribution may occur. For the average degree we obtain from (105)

$$\bar{d}(\gamma,q) = \frac{\gamma}{q}\big(1+(q-1)\overline{s}_0^2(\gamma,q)\big) \qquad (106)$$

where we have made use of the self-consistent equation (99). Since each edge is connected with two vertices this result is consistent with (75).

While (105) gives the distribution of degrees for a randomly chosen vertex in the graph, we may ask for more detailed information depending on whether the vertex belongs, or does not belong to the giant component. Let us call $P_{\rm in}(d;\gamma,q)$ and $P_{\rm out}(d;\gamma,q)$ the biased distributions of degrees for a vertex, respectively, inside and outside the giant component. The generalization of the above calculation is straightforward. $P_{\rm out}(d;\gamma,q)$ and

$P_{\rm in}(d;\gamma,q)$ are obtained from specializing the kernel K to $d, d_0 = 0$ and $d, 1 \leq d_0 \leq d$ respectively. The calculations are very similar to the one presented above, the results read

$$P_{\rm out}(d;\gamma,q) = e^{-\frac{\gamma}{q}(1-\overline{s}_0)}\ \frac{1}{d!}\ \left(\frac{\gamma(1-\overline{s}_0)}{q}\right)^d\ , \tag{107}$$

$$P_{\rm in}(d;\gamma,q) = \frac{1-\overline{s}_0}{q\overline{s}_0}\ e^{-\frac{\gamma}{q}(1-\overline{s}_0)}\ \frac{\gamma^d}{d!}\ \left[(\overline{s}_0 + \ \frac{1-\overline{s}_0}{q})^d - (\frac{1-\overline{s}_0}{q})^d\right]\ . \tag{108}$$

These equations give a rather detailed description of the connectivity in atypical graphs. For $q = 1$ they reproduce (54) and (55) respectively. The corresponding average degrees $\overline{d}_{\rm out}$ and $\overline{d}_{\rm in}$ are in agreement with (36). The remarkable fact that $P_{\rm out}(d;\gamma,q) = P^*_{\rm out}(d,\gamma/q)$ generalizes the mapping (68) to the case $\overline{s}_0 \neq 0$ and explains the similarity between the expressions (51) and (34) for the number of components in typical and atypical graphs respectively. For large γ and q the distribution $P_{\rm out}(d;\gamma,q)$ is peaked at small values of d whereas $P_{\rm in}(d;\gamma,q)$ is maximal for larger d which gives rise to the possible bimodal form of the total distribution $P(d;\gamma,q)$. We also note

$$P_{\rm out}(d=1;\gamma,q) = P_{\rm in}(d=1;\gamma,q) = P(d=1;\gamma,q) \tag{109}$$

for all values of γ and q showing the special role of leaves in the graphs. For all other values of d the distributions $P_{\rm out}(d;\gamma,q)$ and $P_{\rm in}(d;\gamma,q)$ differ from each other. Of course $P_{\rm in}(d=0;\gamma,q)$ since no isolated vertex may belong to the giant component.

Fig. 8 shows the degree distributions inside and outside the giant component for $d = 2$ and $d = 3$ at $\gamma = 3$ as function of q together with results from numerical simulations. With increasing q the biased distribution (5) gets more and more dominated by graphs with many components. From Fig. 8 we infer that in this process both $P_{\rm out}(d=2;\gamma,q)$ and $P_{\rm in}(d=2;\gamma,q)$ increase. Nevertheless $P(d=2,\gamma,q)$ and therefore the total number of vertices carrying two edges decreases due to the shrinking of the giant component (cf. left inset in Fig. 3).

It is interesting to note that

$$\tilde{P}(d) = \frac{q\overline{s}_0}{1+(q-1)\overline{s}_0}\ P_{\rm in}(d;\gamma,q) + \frac{1-\overline{s}_0}{1+(q-1)\overline{s}_0}\ P_{\rm out}(d;\gamma,q) \tag{110}$$

$$= e^{-\frac{\gamma}{q}(1+(q-1)\overline{s}_0)}\ \frac{1}{d!}\ \left(\frac{\gamma}{q}(1+(q-1)\overline{s}_0)\right)^d , \tag{111}$$

and $\tilde{P}(d)$ is therefore Poissonian with parameter

$$\tilde{\gamma} = \frac{\gamma}{q}(1+(q-1)\overline{s}_0). \tag{112}$$

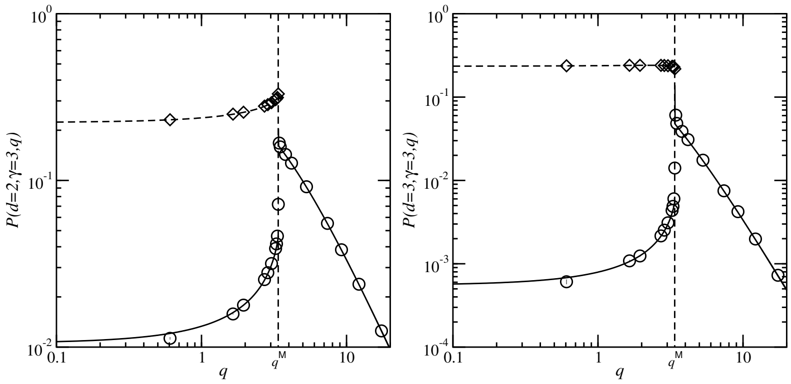

Figure 8: Degree distributions $P_{\rm in}(d;\gamma,q)$ (dashed top) and $P_{\rm out}(d;\gamma,q)$ (full) as given by (107) and (108) as functions of q for $\gamma = 3$, $d = 2$ (left) and $d = 3$ (right). The numerical results are shown by symbols (diamond=inside, circle=outside the giant component). The dashed vertical line indicates the critical value q^M, where the giant component ceases to exist. The statistical error bars are much smaller than the symbol size.

Unfortunately we have at the moment no clear idea about the relevance of this distribution for the properties of the random graph ensemble.

4.6 The distribution of small component sizes

The result (107) for the degree distribution outside the giant component allows to calculate the complete size distribution of non-extensive components $\overline{\psi}(S,\gamma,q)$. As noted in subsection 4.1, the small components are almost certainly trees, *i.e.* a component of size $S = O(1)$ involves $S-1$ edges. From the degree distribution (107) we get for the probability P to find among $N(1-\overline{s}_0)$ vertices a set of S vertices that make $S-1$ connections with each other and none with the remaining ones to leading order in N

$$P = \binom{N(1-\overline{s}_0)}{S} \left(\frac{\gamma}{qN}\right)^{S-1} \left(1-\frac{\gamma}{qN}\right)^{\left(N(1-\overline{s}_0)-S\right)S} \tag{113}$$

$$\sim \frac{N}{S!}\,\frac{q}{\gamma}\,\left(\frac{\gamma}{q}(1-\overline{s}_0)\right)^S e^{-\frac{\gamma}{q}(1-\overline{s}_0)S}. \tag{114}$$

Not all of these sets form trees however, since not all of them are connected. The number of (unlabeled) trees of S vertices is S^{S-2} [25]. For the number of small components of size S per vertex (of the complete graph) we hence

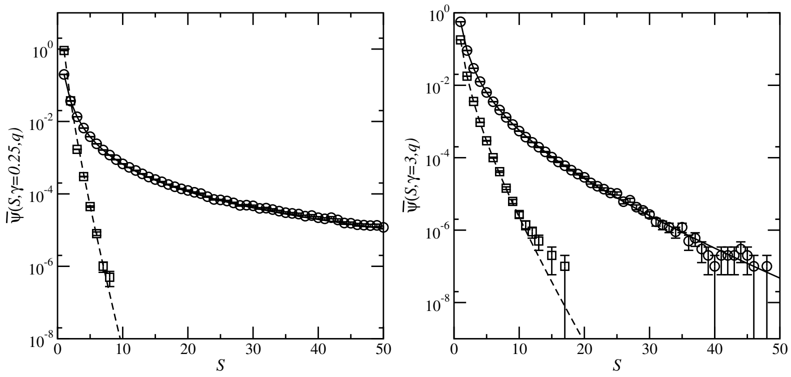

Figure 9: Distribution $\overline{\psi}(S,\gamma,q)$ of the size S of non-extensive components in graphs with an atypical number of components. Left panel is for $\gamma = 0.25$ showing the results for $q = 0.135$ (full line and circles) and $q = 2.72$ (dashed line and squares) respectively. Right panel is for $\gamma = 3.0$ and $q = 5.29$ (full line and circles) and $q = 2.72$ (dashed line and squares) respectively. Lines show the analytical result (115), symbols represent results from numerical simulations.

find

$$\overline{\psi}(S,\gamma,q) = \frac{1}{N}\, S^{S-2}\, P = \frac{q}{\gamma}\, \frac{S^{S-2}}{S!}\, \left(\frac{\gamma}{q}(1-\overline{s}_0)\, e^{-\frac{\gamma}{q}(1-\overline{s}_0)}\right)^{S}. \tag{115}$$

Fig. 9 compares this expression for $\overline{\psi}(S,\gamma,q)$ with results from numerical simulations described in section 5. The agreement is again very good except for relatively large components with correspondingly small probabilities, $\overline{\psi}(S,\gamma,q) \lesssim 10^{-6}$, where the statistical error in the simulation data prevents a meaningful comparison.

For $q = 1$ the result (115) for $\overline{\psi}(S,\gamma,q)$ reduces to (63) after using (52). Moreover comparison of (115) with (63) shows

$$\overline{\psi}(S,\gamma,q) = (1-\overline{s}_0(\gamma,q))\, \psi^*(S,\gamma') \tag{116}$$

with

$$\gamma' = \frac{\gamma}{q}(1-\overline{s}_0(\gamma,q)). \tag{117}$$

Hence in an atypical graph of the discussed type the vertices not belonging to the giant component can be considered to be a typical random graph of $N' = N(1-\overline{s}_0)$ vertices with effective connectivity parameter γ'. Multiplying (115) by S and summing over S we find

$$\gamma' = \sum_S \frac{S^{S-1}}{S!} (\gamma' e^- \gamma')^S \tag{118}$$

implying $\gamma' \leq 1$ [3]. We have hence always $s^*(\gamma') = 0$. On the one hand this implies that the outside vertices are not able to build up a giant component of their own and therefore shows the self-consistency of our assumption that there is only one giant component in rare random graphs of the considered type. On the other hand it allows to easily derive results for $q \neq 1$ from the corresponding ones for $q = 1, s_0^* = 0$. For the total number of components per vertex we find, *e.g.*, from (116) and (51)

$$\overline{c}(\gamma, q) = \sum_S \overline{\psi}(S, \gamma, q) = (1 - \overline{s}_0) \sum_S \psi^*(S, \gamma') \tag{119}$$

$$= (1 - \overline{s}_0)\, c^*(\gamma') = (1 - \overline{s}_0) \left(1 - \frac{\gamma'}{2}\right) \tag{120}$$

$$= (1 - \overline{s}_0) \left(1 - \frac{\gamma}{2q}(1 - \overline{s}_0)\right) \tag{121}$$

reproducing (34). Similarly one may rederive the expression (49) for the second moment of $\overline{\psi}(S, \gamma, q)$ which, however, follows more directly by differentiating (118) with respect to γ'.

We finally note that the evolution equations for the various probability distributions employed in this section are correct in the large N limit. For finite N, care must be paid to the fact that the addition of an edge or a vertex slightly changes the degrees of the vertices giving rise to $O(1/N)$ corrections. Similar corrections occur in the application of the cavity method to spin glasses as discussed in chapter V in [14]. These additional terms are, however, irrelevant in the calculations presented above.

5 Numerical simulations

In order to check the analytical results described above, we have performed Monte Carlo simulations to generate graphs with atypical numbers of components. We have performed simulations for graph sizes between $N = 50$ and $N = 10000$, the results shown are for $N = 1000$, where most simulations were performed. The rare-event algorithm [24] used works in the special case here as follows:

[3] This result may be obtained independently by using (24) in (117) to get $\gamma' = (1 - \overline{s}_0)/(q\overline{s}_0)\, \ln(1 + q\overline{s}_0/(1 - \overline{s}_0)) \leq 1$ for all $q > 0$ and all $\overline{s}_0$ with $0 \leq \overline{s}_0 \leq 1$.

One starts with an initial graph G, *e.g.* a typical random graph with connectivity γ and calculates the number of components $C(G)$. The simulation is performed for a given value of q [4]. Each Monte Carlo step consists of the following steps:

- A trial graph G' is generated:
  - Copy G to G'
  - Select one vertex i in G' randomly
  - Delete all edges adjacent to i
  - For all other vertices $j \neq i$ generate edge (i, j) with probability γ/N
- Calculate number of components $C(G')$
- Accept G' as new configuration G with the Metropolis probability $\min\{1, q^{C(G')-C(G)}\}$

In equilibrium, this procedure generates graphs distributed according to the probability distribution $P(G; \gamma, q, N)$ as given by (5). Equilibration was established in the following way. Two runs were started with two different initial configurations. One was a typical random graph the other one for $q < 1$ ($q > 1$) a graph having minimal (maximal) number of components, *i.e.* a fully connected graph (a graph without edges). In the simulation the number of components $C(t)$ was recorded as a function of Monte Carlo sweeps (MCS) t. The system was considered to be equilibrated after time t_0, if $C(t_0)$ agrees within the typical fluctuations for the two starting configurations. For $N = 1000$ this was the case for all values of q after $t_0 = 20$ MCS ($\gamma < 3$) respectively $t_0 = 50$ MCS ($\gamma = 3$). Hence the system equilibrates very quickly and does not show any sign of glassy behaviour. The average value of C found in the simulation depends on the value of q. For values $q < 1$ the average number of components is preferentially small, while for $q > 1$ it is high. After equilibration, we have taken every t_0 MCS graphs for analysis, 10^4 graphs for each value of q. This allows to obtain various quantities, as the degree distributions or size distributions of components, as a function of q.

In order to obtain numerical results for $\omega(c, \gamma)$ we need to determine $P(C; \gamma, N)$ for *all* values $C \in [1, N]$. Since simulations for one given value of q are dominated by graphs with number of components close to the typical number $N\overline{c}(\gamma, q)$ corresponding to this value of q, simulations at

[4] The parameter q corresponds to the temperature T used in Ref. [24] via $q = \exp(1/T)$. The number of components C corresponds to the energy H via $C = -\mathrm{sign}(T)H$.

various values of q have to be combined [24]. To this end one records during the simulation for each value of q the biased probability distribution $P(C;\gamma,q,N) = \sum_G P(G;\gamma,q,N)\delta(C,C(G))$. From (5) and (3) one finds for each q

$$P(C;\gamma,N) = q^{-C}\, Z(\gamma,q,N)\, P(C;\gamma,q,N). \tag{122}$$

In order to extract from this relation $P(C;\gamma,N)$ for values of C around $N\overline{c}(\gamma,q) \neq Nc^*$, we still have to determine $Z(\gamma,q,N)$. This in turn can be done by starting from $q=1$, where the value $Z(\gamma,1,N)=1$ is known. For values of q close to $q=1$, the measured ranges of the distributions $P(C;\gamma,1,N)$ and $P(C;\gamma,q,N)$ overlap with each other and $Z(\gamma,q,N)$ can be obtained from matching both distributions in this overlapping range. This procedure can be iterated to obtain $Z(\gamma,q,N)$ for values of q differing by an increasing amount from the starting point $q=1$. Using (122) the complete distribution $P(C;\gamma,N)$ can be determined. In our simulations using $N=1000$ and $\gamma=0.25,1,2,3$ between 22 and 27 different values of q where sufficient to obtain $P(C;\gamma,N)$ and therefore $\omega(c,\gamma)$ over the full range.

6 Conclusion

In the present paper we have investigated large deviation properties in ensembles of Erdös-Rényi random graphs. In particular we have studied graphs that are atypical with respect to their number of components. We have shown that several of their properties such as their probability, the relative size of their giant component, as well as the second moment of the distribution of component sizes can be obtained from the Legendre transform with respect to $\ln q$ of the mean-field free energy of the q-state Potts model. This generalizes the well-known connection between typical properties of random graphs and the $q \to 1$ limit of the Potts free energy. Therefore this free energy conveys also interesting information about the random graph ensemble for values of $q \neq 1$. In particular the well known first order phase transition in the mean-field Potts model for $q > 2$ gives rise to a non-convex part in the logarithmic probability of the graphs corresponding to a bimodal probability distribution $P(C;\gamma,q,N)$.

In a second part we have rederived these results without recourse to the Potts model by requiring the "statistical stability" of the random graph ensemble under the insertion of an additional vertex or edge. This approach is made possible by the mere existence of the thermodynamic limit in which the number of vertices N tends to infinity. Besides reproducing the results obtained previously we have also pointed out some subtleties of this method when applied to exponentially rare configurations. Moreover we obtained as additional results the complete degree distribution and the size distribution of non-extensive components in atypical graphs.

From the mathematical point of view both approaches rest on the assumptions that for the various quantities introduced the thermodynamic limit in which the number N of vertices of the graphs tends to infinity exist and that there is at most one extensive component comprising $O(N)$ vertices.

Our analytical findings describing the thermodynamic limit are in very good agreement with numerical simulations using a rare-event algorithm for $N = 1000$.

It is well known that the typical properties of Erdös-Rényi random graphs with fixed number of edges and fixed probability of edges are equivalent. This equivalence does, however, not carry over to the large deviation behaviour. In fact in an ensemble with fixed number of edges there is no energetic contribution to the probability of a rare graph and the large deviation characteristics will be rather different from those studied in the present paper.

Further work to improve our understanding of the relationship between the two processes (one more edge or one more vertex) would be useful. It would also be interesting to see how the microscopic approach may be generalized for the study of graphs which are atypical with respect to other properties than their number of components as, *e.g.* their number of vertex covers [10], their average degree or the size of their giant component, where the connection to the Potts model cannot be used. After completion of this work we became aware of another very recent application of the cavity method to characterize certain properties of rare random graphs [26].

A Appendix

In this appendix we give some intermediate results for the derivation of the Potts free energy (15), see also [18]. The explicit determination of the partition function is possible since the energy function (12) depends on the configuration of spins $\{\sigma_i\}$ solely through the fractions

$$x(\sigma, \{\sigma_i\}) = \frac{1}{N} \sum_{i=1}^{N} \delta(\sigma, \sigma_i) \tag{123}$$

of variables σ_i equal to σ. Clearly

$$\sum_{\sigma} x(\sigma, \{\sigma_i\}) = 1 \tag{124}$$

for all $\{\sigma_i\}$. The energy (12) may then be rewritten as

$$E(\{\sigma_i\}) = -\frac{N}{2}\sum_\sigma \left(x(\sigma,\{\sigma_i\})\right)^2 - Nh\sum_\sigma u_\sigma\, x(\sigma,\{\sigma_i\}) + O(1), \quad (125)$$

and the partition function (13) becomes to leading order in N

$$\begin{aligned}
\mathcal{Z}(\beta,h,q,\{u_\sigma\}) &= \sum_{\{\sigma_i\}} \exp\left(\frac{\beta N}{2}\sum_\sigma x(\sigma,\{\sigma_i\})^2 + Nh\sum_\sigma u_\sigma\, x(\sigma,\{\sigma_i\})\right) \\
&= \sum_{\{x(\sigma)\}} \frac{N!}{\prod_{\sigma=0}^{q-1}[N\,x(\sigma)]!}\ \exp\left(\frac{\beta N}{2}\sum_\sigma x(\sigma)^2 + \beta Nh\sum_\sigma u_\sigma\, x(\sigma)\right) \\
&= \int_0^1 \prod_{\sigma=0}^{q-1} dx(\sigma)\ \exp\left(N\left[\frac{\beta}{2}\sum_\sigma x(\sigma)^2 + \beta h\sum_\sigma u_\sigma\, x(\sigma) - \sum_\sigma x(\sigma)\ln x(\sigma)\right]\right).
\end{aligned}$$

The sum and the integral over $x(\sigma)$ are restricted to the normalized subspace defined by (124). In the limit of large N the integral may be evaluated by the Laplace method. The Potts free energy (14) then reads

$$f(\beta,h,q,\{u_\sigma\}) = \operatorname*{extr}_{\{x(\sigma)\}}\left[-\frac{1}{2}\sum_\sigma x(\sigma)^2 - h\sum_\sigma u_\sigma\, x(\sigma) + \frac{1}{\beta}\sum_\sigma x(\sigma)\ln x(\sigma)\right]. \quad (126)$$

We will need explicit results for the free energy and its derivatives for $h=0$ only. A suitable ansatz to perform the extremization in (126) for this case is

$$x(0) = \frac{1}{q}(1+(q-1)s_0) \quad (127)$$

$$x(\sigma) = \frac{1}{q}(1-s_0) \qquad \text{for} \qquad \sigma \neq 0 \quad (128)$$

with the yet undetermined parameter s_0. This ansatz allows for a possible spontaneous breaking of the Potts symmetry at low temperature and automatically fulfills the normalization (124). It gives rise to

$$\begin{aligned}
f(\beta,q) = \operatorname*{extr}_{s_0}\Bigg[&-\frac{1}{2q} - \frac{q-1}{2q}s_0^2 - \frac{1}{\beta}\ln q \\
&+\frac{1+(q-1)s_0}{\beta q}\ln(1+(q-1)s_0) + \frac{q-1}{\beta q}(1-s_0)\ln(1-s_0)\Bigg], \quad (129)
\end{aligned}$$

which is identical with (15). Differentiating the expression in the brackets in (129) with respect to s we find for the extremum value $\overline{s}_0(\beta,q)$ the equation

$$e^{\beta \overline{s}_0} = \frac{1 + (q-1)\overline{s}_0}{1 - \overline{s}_0}, \tag{130}$$

reproducing (16).

Acknowledgment: We are grateful to J.C. Angles d'Auriac, S. Cocco, O. Martin, A. Montanari, and M. Weigt for discussion and support during the completion of this work. Part the thermodynamical calculations on the Potts model were done in cooperation with A. M. Morgante during her Master in Theoretical Physics. A.E. acknowledges the kind hospitality at the Laboratoire de Physique Théorique at the Université Louis Pasteur in Strasbourg where this work was done as well as financial support from the *VolkswagenStiftung*. A.K.H. obtained financial support from the *VolkswagenStiftung* (Germany) within the program "Nachwuchsgruppen an Universitäten" and from the European Community via the Human Potential Program under contract number HPRN-CT-2002-00307 (DYGLAGE-MEM), via the High-Level Scientific Conferences (HLSC) program, and via the Complex Systems Network of Excellence "Exystence".

References

[1] Erdös, P. and Rényi, A. (1960) On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5**, 17-61.

[2] Wormald, N.C. (1999) Models of random regular graphs. Surveys in Combinatorics, 1999, J.D. Lamb and D.A. Preece, eds. London Mathematical Society, Lecture Note Series, vol 276, pp. 239–298. Cambridge University Press, Cambridge.

[3] Barabsi, A.L. and Albert, R. (2002) Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74**, 47-97.

[4] Bollobas, B., Riordan, O., Spencer, J. and Tusnady, G. (2001) The degree sequence of a scale free random graph process, *Random Structures and Algorithms* **18**, 279-290.

[5] Andreanov, A., Barbieri, F. and Martin, O.C. (2003) Large deviations in spin glass ground state energies. *preprint cond-mat/0307709.*

[6] Simonian, A. and Guibert, J. (1995) Large deviations approximation for fluid queues fed by a large number of on/off sources. *IEEE Journal of Selected Areas in Communications* **13**, 1017 - 1027.

[7] Mandjes, M. and Han Kim, J. (2001) Large deviations for small buffers: an insensitivity result. *Queueing Systems* **37**, 349-362.

[8] Semerjian, G. and Monasson, R. (2003) Relaxation and metastability in a local search procedure for the random satisfiability problem, *Phys. Rev. E* **67**, 066103; Barthel, W., Hartmann, A.K. and Weigt, M. (2003) Solving satisfiability problems by fluctuations: the dynamics of stochastic local search algorithms, *Phys. Rev. E* **67**, 066104.

[9] Cocco, S. and Monasson, R. (2002) Exponentially hard problems are sometimes polynomial, a large deviation analysis of search algorithms for the random Satisfiability problem, and its application to stop-and-restart resolutions. *Phys. Rev. E* **66**, 037101.

[10] Montanari, A. and Zecchina, R. (2002) Optimizing Searches via Rare Events. *Phys. Rev. Lett.* **88**, 178701.

[11] O'Connell, N. (1998) Some large deviation results for sparse random graphs. *Prob. Theor. Rel. Fields* **110**, 277

[12] Machado F.P. (1997) Large Deviations for the Number of Open Clusters per site in Long Range Bond Percolation, *Markov Processes and Related Fields* **3**, 367-376.

[13] Duxbury, P., Jacobs, D., Thorpe, M. and Moukarzel, C. (1999) Floppy modes and the free-energy: rigidity and connectivity percolation on Bethe lattices, *Phys. Rev. E* **59**, 2084; Moukarzel, C. (2003) Rigidity percolation in a field. *preprint cond-mat/0306380.*

[14] Mezard, M., Parisi, G. and Virasoro, M. (1987) Spin glasses and Beyond. World Scientific, Singapore.

[15] Fortuin, C.M. and Kasteleyn, P.W. (1972) On the random cluster model. I. Introduction and relation to other models. *Physica* **57**, 536.

[16] Bollobàs, B. (2001). *Random Graphs.* Cambridge University Press, 2nd edition, Cambridge.

[17] Potts, R. (1952), Some generalized order-disorder transformations. *Proc. Camb. Phil. Soc.* **48**, 106.

[18] Wu, F. (1982). The Potts model. *Rev. Mod. Phys.* **54**, 235.

[19] Feller, M. (1968) *An introduction to Probability Theory and Its Applications* Volume 1 Chapter V, Wiley, 3rd edition, New York.

[20] Angles d'Auriac, J-Ch., Igloi, F., Preissmann, M., Sebo, A. (2002), Optimal Cooperation and Submodularity for Computing Potts' Partition Functions with a Large Number of State. *J.Phys.* **35A**, 6973

[21] Segdewick, R. and Flajolet, P. (1995) *An introduction to the analysis of algorithms*, Chapter 3, Addison-Wesley, Boston.

[22] Barraez, D., Boucheron, S., Fernandez de la Vega, W. (2000), On the fluctuations of the giant component. *Combin. Probab. Comput.* **9** 287-304

[23] see Ref. [16], chapter 6.

[24] Hartmann, A. K. (2002), Sampling rare events: statistics of local sequence alignments. *Phys. Rev. E* **65**, 056102

[25] Wilf, H. S. (1994) *Generatingfunctionology*, Theorem 3.12.1, Chapter 3, Academic Press.

[26] Rivoire, O. (2004), Properties of atypical graphs from negative complexities. *cond-mat/0312501.*