4

RANDOM GRAPHS, AND THE PERCOLATION TRANSITION

Rigorous results obtained by Erdös and Rényi in the sixties on the percolation and giant component properties of random graphs may be found back by means of the mean-field Potts model. This calculation allows the non physicist reader to get familiar with various tools e.g. analytic continuations, Legendre transforms, ... , and to see how statistical physics may be practically applied to combinatorial problems.

4.1 Random graphs, models and giant component

4.1.1 *Models of random graphs*

Two popular models of random graphs over N vertices are (Erdös and Rényi 1960, Bollobàs 1985, 1998),

- *Model I (fixed number of edges):*
  Consider the complete graph K_N over N vertices. We define $\mathcal{G}_{N,N_E}$ as the set of graphs obtained by taking only $N_E = cN/2$ among the $\binom{N}{2}$ edges of K_N in all possible different ways. Within Distribution I, a random graph G is a randomly chosen element of $\mathcal{G}_{N,N_E}$ with the flat measure,

$$\mathcal{P}_{\mathcal{I}}(G) = \frac{1}{\binom{\binom{N}{2}}{N_E}} \qquad . \tag{4.1}$$

- *Model II (fixed probability of edge deletion):*
  Another way of generating random graphs from the complete graph is through edge deletion. Start from K_N, and delete every edge with probability $1 - c/N$. This choice ensures that the average number of edges left at the end of the deletion process equals N_E defined in Distribution I. The probability $\mathcal{P}(G)$ of drawing a random graph G depends on the number $N_E(G)$ of its edges, and follows the Binomial law,

$$\mathcal{P}_{\mathcal{II}}(G) = \left(\frac{c}{N}\right)^{N_E(G)} \left(1 - \frac{c}{N}\right)^{\binom{N}{2} - N_E(G)} \qquad . \tag{4.2}$$

In the large N limit, both distributions of random graphs share common properties which depend on the control parameter c. We shall make explicit the distribution considered when necessary.[69]

[69]Distribution I is easier to implement on a computer. Drawing a graph amounts to chose N_E distinct pairs of vertices among the $N(N-1)/2$ possible ones, a task which can be carried

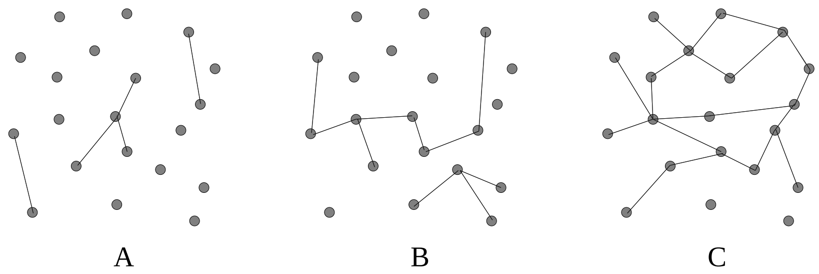

FIG. 4.1. Examples of random graphs generated at fixed number N_E of edges (Distribution I). All graph include $N = 20$ vertices (grey dots). The average degrees of valency, $c = 2N_E/N$, equal $c = 0.5$ (**A**), $c = 1$ (**B**), and $c = 2$ (**C**). The labels of the vertices have been permuted to obtain planar graphs, *i.e.* avoid crossing of edges.

Fig. 4.1 shows examples of graphs obtained from Distribution I and for various values of c. Notice the qualitative change of structure of graphs when the connectivity c varies from low values (graphs are mostly made of small isolated trees) to higher ones (a large part of vertices are now connected together). This change is known as the percolation transition in physics, or the appearance of a giant component in mathematics literature.

Before reviewing some of the aspects of the percolation transition rigorously established by mathematicians, let us mention an important fact on the valency of vertices. As a result of the randomness of the graph generation process, each node share edges with a variable number of neighboring vertices. For both distributions, in the large N limit, the degree v of a vertex, *i.e.* the number of its neighbors, is indeed a random variable obeying a Poisson law with parameter c,

$$\rho(v) = e^{-c}\,\frac{c^v}{v!} \qquad . \tag{4.3}$$

In particular, $\rho(0) = e^{-c}$ is the fraction of isolated vertices. Control parameter c may be thus seen as the average degree of nodes. This is what is meant in the following when referring to c as the connectivity of the graph.

4.1.2 *Overview of rigorous results*

We start with some vocabulary. A (connected) component of a graph G is called cluster. The size (order) of a cluster is the number of vertices it contains. An isolated vertex is a cluster of size unity. The number of components of G is denoted by $\Phi(G)$. We shall indicate its normalized fraction, $\Phi(G)/N$, by $\varphi(G)$, which is bounded from above by unity, and from below by e^{-c}, the fraction of

out in $O(N_E)$ ($= O(N)$ for finite c) steps. On the contrary, within Distribution II, all $O(N^2)$ edges of K_N have to been looked at in the course of the deletion process. However, Distribution II does not induce any correlation between different edges and is therefore more convenient from an analytical point of view.

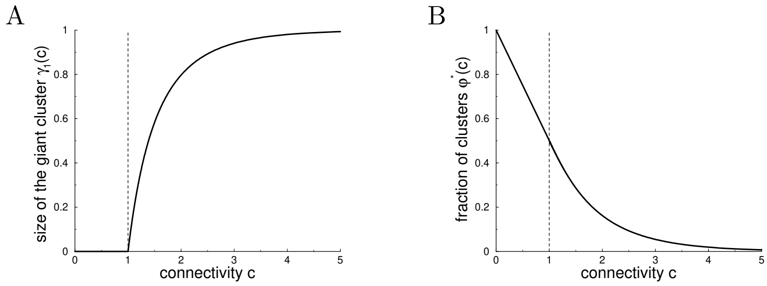

FIG. 4.2. The percolation transition in the random graph. Fraction $\gamma_1(c)$ of vertices in the largest component (**A**), and number $\varphi^*(c)$ of clusters per vertex (**B**) a function of the average connectivity c of the random graph. The vertical dot–dashed line $c = 1$ indicates the location of the percolation threshold.

isolated vertices. Components may be sorted by decreasing sizes $\Gamma_i(G)$, from the largest ($i = 1$) to the smallest one ($i = \Phi(G)$).

Erdös and Rényi (1960) were able to prove the following results on the sizes of the largest components (Bollobàs 1985),

- When $c < 1$, the largest clusters include with high probability a number of vertices scaling asymptotically as,

$$\Gamma_i(G) \sim \frac{\ln N - \frac{5}{2}\ln\ln N}{c - 1 - \ln c} \qquad (c < 1) \ , \tag{4.4}$$

  for all $1 \leq i \leq I$, with I finite as N gets large.
- At $c = 1$, the largest components contain $O(N^{2/3})$ vertices. More precisely, there exists two positive constants α_1, α_I such that

$$\alpha_I \, N^{2/3} < \Gamma_I(G) \leq \Gamma_{I-1}(G) \leq \ldots \leq \Gamma_1(G) \leq \alpha_1 \, N^{2/3} \qquad (c = 1), \tag{4.5}$$

  with I finite as N gets large.
- When $c > 1$, a drastic separation takes place between the largest cluster and all other smaller components. With high probability, $\Gamma_1(G)/N = \gamma_1(c)$ where $\gamma_1(c)$ is the unique positive solution of

$$1 - \gamma = e^{-c\,\gamma} \qquad . \tag{4.6}$$

  Again, the sizes of smaller components ($i \geq 2$) are given by eqn (4.4).

SAY SOMETHINNG HERE ABOUT THE DISTRIBUTION OF SIZES OF SMALL TREES

The phenomenon taking place at $c = 1$ is called percolation. While below the percolation threshold $c \leq 1$, most components are small with $O(\ln N)$ sizes, they percolate at the transition and give birth to a giant component with $O(N)$

vertices. Fig. 4.2A shows the fraction $\gamma_1(c)$ of vertices belonging to the giant cluster. We also report on Fig. 4.2B the average fraction of clusters,

$$\varphi^*(c) = -\frac{c}{2}\big(1-\gamma_1(c)^2\big) + \big[1-\gamma_1(c)\big]\big[1-\ln\big(1-\gamma_1(c)\big)\big] \qquad , \tag{4.7}$$

which will be useful in the following. Both are non analytic functions of their argument γ at the critical point $c=1$.

At the phase transition, the largest components scale as $N^{2/3}$, showing that the fractions of vertices they contain scale as $N^{-1/\nu}$ with a finite size scaling exponent $\nu=3$, different from the Ising value $\nu=2$ (Chapter 3.1). The reader is referred to specialised literature for more elaborate results on the transition (Bollobàs 1984, Luczak 1990, Luczak, Pittel and Wierman 1994, Janson *et al.* 1993). To end with, let us mention that percolation is not at all limited to random graphs with no Euclidean underlying structure. There exists an enormous literature in physics devoted to percolation and the characterization of the phase transition in finite dimensions (Essam 1972, Lubensky 1979, Stauffer, Aharony 1992).

4.1.3 *Heuristic description of the giant component*

Some intuition about the above results may be gained from a simple argument. This reasoning is not mathematically correct but the results it leads to are. The starting point is an ubiquitous idea in probability and statistical physics, which could be phrased as follows: "if a system is very large, its statistical properties should be, in some sense, unaffected by a small increase in size". We now use this principle, which is merely a definition of what very large means, to determine the size γ of the giant component as a function of the connectivity c (4.6).

Consider a random graph G over N vertices, with connectivity c. Add a new vertex A to the graph to obtain G' (Fig. 4.3). If we want G' to be drawn from the same distribution as G, a number v of edges must be attached to A, where v an integer–valued random number following the Poisson distribution ρ (4.3). After addition of A, some connected components of G will merge in G'. In particular, with some probability p_v, A will be part of the giant component of G'. To estimate p_v, we note that this event will not happen if and only if none of the v neighbors of A in G' belongs to the giant component of G. Thus,

$$1 - p_v = (1-\gamma_1)^v \qquad , \tag{4.8}$$

where γ_1 is the size (fraction of vertices) of the latter. Summing both sides of (4.8) over the distribution (4.3) of v, and asserting that the change in size of the giant component between G and G' is $o(1)$ for large N, we obtain the desired equation (4.6) for γ_1.[70]

[70] We have ruled out here the coexistence of two distinct giant components above the threshold, which would be very likely to get connected after the addition to the graph of a large (but still small with respect to N) number of vertices and attached edges. The impossibility of having more than one giant component has been proven by ... (...).

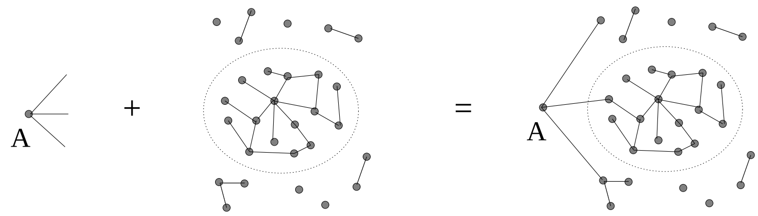

FIG. 4.3. Addition process of one vertex A and its $v = 3$ edges to a random graph G. Prior to merging, G is made of isolated nodes, small connected components, and a giant component (surrounded by the dotted line). After merging, the new vertex belongs to the giant component.

The above argument is plagued with unjustified assumptions and approximations. It is however at the heart of the cavity approach, a fructuous method we will use to tackle more complex problems (Onsager ..., Mézard, Parisi, Virasoro 1987).

4.2 Potts model and relationship with random graphs

The scope of this section is present a deep equivalence, first established by Kasteleyn and Fortuin (1969), between a statistical mechanics model, the Potts system (Potts 1952), and random graphs.

4.2.1 *Distribution of the number of components*

Consider the probability density $R(\varphi)$ of generating random graphs with $\Phi = \varphi N$ connected components (clusters),

$$R_N(\varphi, c) = \sum_{G \subset K_N} \mathcal{P}(G)\, \delta(\varphi - \varphi(G)) \qquad , \tag{4.9}$$

where δ indicates the Dirac distribution, and $\mathcal{P}(G)$ is the probability of drawing random graph G (which depends on the size N and the connectivity c). As the size N gets larger and larger, $R_N(\varphi, c)$ should become highly concentrated around $\varphi^*(c)$ (4.7), equal to the most probable number of clusters per vertex. Random graphs G with $\varphi(G)$ differing from $\varphi^*(c)$ by a finite amount are expected to be exponentially rare in N. The quantity

$$r(\varphi, c) = \lim_{N \to \infty} \frac{1}{N} \ln R_N(\varphi, c) \tag{4.10}$$

encodes information on the large deviations of the fraction of connected components with respect to its highly probable value. r should vanish for $\varphi = \varphi^*(c)$, and be strictly negative otherwise (Fig. 4.4A). In the following, we shall compute $r(\varphi, c)$ and thus derive through a unique calculation not only the typical number of clusters, but also its fluctuations with exponentially small probability (atypical graphs).

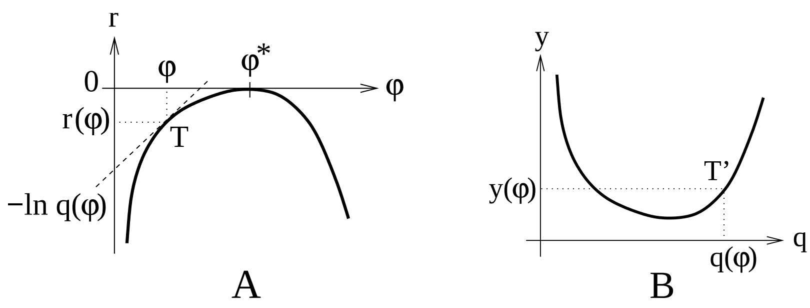

FIG. 4.4. Schematic representation of the large deviation distribution r of the numbers of clusters per vertex φ (**A**), and its Legendre transform $y(q)$ (**B**). r is maximal and zero at the most probable fraction of clusters φ^* (which depends upon connectivity c). Let T be a point on the representative curve of r with coordinates $\varphi, r(\varphi)$. We define $q(\varphi)$ from the slope $-\ln q$ of the tangent attached to T. To obtain the Legendre transform y of r, we define $y(\varphi) = r(\varphi) + \varphi \ln q(\varphi)$, and plot the point T' of coordinates $q(\varphi), y(\varphi)$. When T moves, so does T' and the whole representative curve of y is obtained in a parametric way. Notice that r may be found back from y following the same construction.

It results convenient to introduce the generating function of the cluster probability,

$$Y_N(q,c) = \sum_{G \subset K_N} \mathcal{P}(G) q^{\Phi(G)} = \int_0^1 d\varphi \, R_N(\varphi, c) \, q^{N\varphi} \qquad , \tag{4.11}$$

with q being a formal (eventually real) parameter. Insertion of eqn (4.10) into (4.11) and evaluation of the integral over φ by a Laplace calculation give the logarithm $y(q,c)$ of the cluster generating function,

$$y(q,c) \equiv \lim_{N \to \infty} \frac{1}{N} \ln Y_N(q,c) = \max_{0 \le \varphi \le 1} \left[r(\varphi, c) + \varphi \ln q \right] \qquad . \tag{4.12}$$

In other words, y and r are simply conjugated Legendre transforms (Fig. 4.4),

$$r(\varphi, c) = \min_{q \ge 0} \left[y(q,c) - \varphi \ln q \right] \qquad . \tag{4.13}$$

As an illustration, the number of connected components per vertex is, with high probability, given by the partial derivative of y with respect to q in $q = 1$.[71] We show below how statistical mechanics gives access to f, and how r can be derived subsequently from eqn (4.13).

[71] From Fig. 4.4, the slope $-\ln q(\varphi^*)$ at the maximum of the representative curve of r vanishes, giving $q(\varphi^*) = 1$.

Notice that the partial derivative of y with respect to c has a simple interpretation for model II (4.2). Differentiating definition (4.11) and sending N to infinity, we obtain that the average vertex degree, $d(q,c)$, equals

$$d(q,c) = c + 2\,c\,\frac{\partial y}{\partial c}(q,c) \quad . \tag{4.14}$$

This identity, which will turn out to be useful to interpret the structure of atypical graphs studied in Section 4.3.3, holds for Model II only. Indeed, in the case of Model I, the average vertex degree is equal to c, independently of q.

4.2.2 *Diagrammatic expansion of the Potts model*

The Potts model (Potts 1952) is defined in terms of an energy function which depends on N spin variables σ_i, one for each vertex of the complete graph K_N, which take q distinct values $\sigma_i = 0, 1, ..., q-1$. The energy function reads

$$E[\sigma_1, \sigma_2, \ldots, \sigma_N] = -\frac{1}{N}\sum_{i<j} \delta_{\sigma_i,\sigma_j} \quad , \tag{4.15}$$

where $\delta_{a,b}$ is the Kronecker delta function. Note the $1/N$ scaling of the coupling as in 3.1. The partition function of the Potts model is

$$Z_N(q,T) = \sum_{\{\sigma_i = 0,\ldots,q-1\}} \exp\left(\frac{1}{TN}\sum_{i<j}\delta_{\sigma_i,\sigma_j}\right) \tag{4.16}$$

where the summation runs over all q^N spin configurations. δ taking value zero or unity, the partition function may be recast in the form (Kasteleyn, Fortuin 1969, Fortuin, Kasteleyn 1972),

$$Z_N(q,T) = \sum_{\{\sigma_i\}}\prod_{i<j}\left[1 + w\,\delta_{\sigma_i,\sigma_j}\right] \quad , \tag{4.17}$$

with

$$w = \exp\left(\frac{1}{T\,N}\right) - 1 \quad . \tag{4.18}$$

When expanding the product appearing in (4.17), we obtain $2^{N(N-1)/2}$ terms t, each of which composed by two factors, the first one given by w raised to a power equal to the number of δs composing the second factor. We can represent graphically each term t in the expansion as a subgraph G_t of K_N. An edge connects the vertices i and j in G_t if and only if t includes the factor $\delta_{\sigma_i,\sigma_j}$. This one–to–one mapping between the terms of the expansion and and the subgraphs of K_N allows us to rewrite the partition function as

$$Z_N(q,T) = \sum_{\{\sigma_i\}}\sum_{G\subset K_N} w^{N_E(G)}\prod_{k=0}^{N_E(G)} \delta_{\sigma_{i_k},\sigma_{j_k}} \tag{4.19}$$

where $N_E(G)$ is the number of edges in the subgraph G and i_k, j_k are the vertices connected by the kth edge of the subgraph. We now exchange the order of the summations and perform the sum over the spin configurations first. Given a subgraph G with N_E edges and Φ clusters (isolated vertices included), the sum over spins configurations will give zero unless all the σs belonging to a connected component of G have the same value (as a consequence of the δ factors). In such a cluster, one can set the σs to any of the q different values. Hence,

$$Z_N(q,T) = \sum_{G \subset K_N} w^{N_E(G)}\, q^{\Phi(G)} \qquad . \tag{4.20}$$

Identity (4.20) is true for any temperature T. We set $T = T_p(c,N) \equiv 1/(N \ln(1+c/N))$ to obtain $w = c/N$ from (4.18). Eqns (4.2,4.11,4.20) then lead to the following identity between the Potts partition function and the cluster generating function,

$$Z_N\Big(q, T_p(c,N)\Big) = \Big(1 - \frac{c}{N}\Big)^{-\frac{N(N-1)}{2}} \; Y_N(q,c) \simeq e^{N\,c/2}\; Y_N(q,c) \quad , \tag{4.21}$$

for terms exponential in N. Taking the logarithm of (4.21), we see that the free energy $f(q,T)$ of the Potts model at temperature $T = T_p(c,N) \to 1/c$ is simply related to the logarithm $y(q,c)$ of the cluster generating function (4.12),

$$f\left(q, \frac{1}{c}\right) = -\frac{1}{2} - \frac{1}{c}\, y(q,c) \qquad . \tag{4.22}$$

As a consequence, the average fraction of clusters per node, $\varphi^*(c)$, can be obtained from the knowledge of the Potts free-energy $f(q,c)$ in the vicinity of $q=1$. Since the Potts model is originally defined for integer values of q only, an analytic continuation to real values of q is necessary. We now explain how to perform this continuation.

4.3 Statistical mechanics of the Potts model

4.3.1 *Free-energy and low connectivity regime*

As in the case of the Ising model of Chapter 3.1, a careful examination of the energy function (4.15) shows that the latter depends on the configuration of spins $\mathcal{C} = \{\sigma_i\}$ through the fractions $x(\sigma;\mathcal{C})$ of variables σ_i equal to σ $(= 0, 1, \cdots, q-1)$ (Wu 1982),

$$x(\sigma;\mathcal{C}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\sigma_i,\sigma}, \quad (\sigma = 0, 1, ..., q-1) \qquad . \tag{4.23}$$

Of course, the sum of the fractions (4.23) over σ must be equal to unity. Using these fractions, the energy (4.15) may be rewritten as

$$E[\mathcal{C}] = -\frac{N}{2} \sum_{\sigma=0}^{q-1} \big[x(\sigma;\mathcal{C})\big]^2 + \frac{1}{2} \qquad . \tag{4.24}$$

Note that the last term on the r.h.s. of (4.24) can be neglected with respect to the first term whose order of magnitude is $O(N)$. The partition function (4.16) at temperature $T = 1/c$ now becomes

$$
\begin{aligned}
Z_N\left(q, \frac{1}{c}\right) &= \sum_{\mathcal{C}\in\{0,1,\ldots,q\}^N} \exp\left(\frac{c}{2}\, N \sum_{\sigma=0}^{q-1} x(\sigma,\mathcal{C})^2\right) \\
&= \sum_{\{x(\sigma)=0,\frac{1}{N},\frac{2}{N},\ldots,1\}}^{(R)} \frac{N!}{\prod_{\sigma=0}^{q-1}[N\,x(\sigma)]!}\ \exp\left(\frac{c}{2}\, N \sum_{\sigma=0}^{q-1} x(\sigma)^2\right) \\
&= \int_0^1 {}^{(R)} \prod_{\sigma=1}^{q-1} dx(\sigma)\ \exp\Big(-N\,c\,\hat{f}[\{x(\sigma)\}]\Big)
\end{aligned}
\tag{4.25}
$$

to the leading order in N. The subscript (R) indicates that the sum and the integral must be restricted to the normalized subspace

$$\sum_{\sigma=0}^{q-1} x(\sigma) = 1 \qquad . \tag{4.26}$$

The free energy functional $\hat{f}$ appearing in (4.25) is the q-multivariate function,

$$\hat{f}\big[x(0), x(1), \ldots, x(q-1)\big] = \sum_{\sigma=0}^{q-1} \left\{ -\frac{1}{2}x(\sigma)^2 + \frac{1}{c}\, x(\sigma)\ln x(\sigma)\right\} \qquad . \tag{4.27}$$

In the limit of large N, the integral in (4.25) may be evaluated by the Laplace method. The Potts free energy then reads

$$f\left(q, \frac{1}{c}\right) = \min_{\{x(\sigma)\}} \hat{f}[\{x(\sigma)\}] \tag{4.28}$$

and the problem becomes that of analyzing the minima of f. Given the initial formulation of the problem, each possible value of σ among $0, \ldots, q-1$ plays the same role; indeed $\hat{f}$ is invariant under the permutation symmetry of the different q values. In other terms, if $\{x(\sigma)\}$ is a solution of the q coupled minimization equations,

$$\frac{\partial \hat{f}}{\partial x(\sigma)} = \lambda \qquad , \tag{4.29}$$

where λ is a Lagrange multiplier enforcing normalization (4.26), so is $\{x(\pi\sigma)\}$ for any permutation π of the symmetric group S_q. However, we should keep in mind that such a symmetry could be broken by the minimum (see Chapter 3.1). We shall see that depending on the value of the connectivity c, the permutation symmetry may or may not be broken, leading to a phase transition in the problem which coincides with the birth a giant component in the associated random graph.

An obvious solution to eqn (4.29) is the symmetric (s) minimum, where all xs are identical,

$$x^s(\sigma) = \frac{1}{q}, \qquad \forall\, \sigma = 0, \ldots, q-1. \tag{4.30}$$

Inserting eqn (4.30) into the Potts free energy functional (4.27), we obtain the following expression for the logarithm of the cluster generating function,

$$y^s(q,c) = \ln q - \frac{c}{2}\left(1 - \frac{1}{q}\right) \quad . \tag{4.31}$$

Though the above expression has been derived for positive integer q, we now extend it to real values of q through an implicit analytic continuation. Taking the Legendre transform of y with respect to q, we get from eqn (4.13), we get for the logarithm r of the cluster distribution density,

$$r^s(\varphi, c) = -\frac{c}{2} + (1-\varphi)\left[1 + \ln\left(\frac{c}{2}\right) - \ln(1-\varphi)\right] \quad . \tag{4.32}$$

$r^s(\varphi, c)$ is maximal and null at $\varphi^{*,s}(c) = 1 - c/2$, a result obviously meaningless for connectivities larger than two. Comparison with the rigorous result (4.7) from random graph theory indicates that the symmetric result is exact as long as $c \leq 1$, and is false above the percolation threshold. The failure of the symmetric minimum in the presence of a giant component proves the onset of symmetry breaking.

The same conclusion can be reached from an analysis of the curvature of the free energy functional (4.27) around the symmetric solution (4.30). The $q \times q$ Hessian matrix

$$\mathcal{M}_{\sigma,\tau} = \left.\frac{\partial^2 \hat{f}}{\partial x(\sigma)\partial x(\tau)}\right|_{s,(R)} \quad , \tag{4.33}$$

can be exactly diagonalized for generic integer q. The normalised subspace (4.26) is spanned by $q-1$ eigenvectors, whose associated eigenvalues are: $\Lambda_0 = q(q-c)$ (non degenerate) and $\Lambda_1 = q-c$ with multiplicity $q-2$. The analytic continuation of the eigenvalues to real $q \to 1$ lead to the single value $\Lambda = 1-c$, which changes sign at the percolation threshold. Therefore, the symmetric saddle-point is not a local minimum of $\hat{f}$ above $c = 1$, showing that another solution of eqn (4.29) has to be sought, and that the percolation transition is of the second order.

4.3.2 *Symmetry breaking and percolation*

The simplest way to break the symmetry of eqn (4.30) is to look for solutions in which one among the q values appears more frequently than the others. Therefore we look for a minimum of the form

$$x(0) = \frac{1}{q}\left[1 + (q-1)\, s\right] \,,$$

$$x(\sigma) = \frac{1}{q}[1-s] \ , \qquad (\sigma = 1, ..., q-1), \tag{4.34}$$

which fulfills normalization condition (4.26). Parameter s may *a priori* take any real value, which enlarges the space of possible minima with respect to the symmetric case (4.30) corresponding to $s = 0$. The free energy $f(q, 1/c)$ of the Potts model is obtained by inserting the fractions (4.34) into (4.27). To obtain the average number of clusters $\varphi^*(c)$, it is sufficient to expand f around $q = 1$,

$$f\left(q, \frac{1}{c}\right) = -\frac{1}{2} + (q-1)\, f_1(s,c) + O\big((q-1)^2\big) \tag{4.35}$$

with

$$f_1(s,c) = \frac{1}{2}(1-s^2) - \frac{1}{c}\,(1-s)\big(1-\ln(1-s)\big) \qquad . \tag{4.36}$$

Minimization of $f_1(s,c)$ with respect to s shows that, for $c \leq 1$, the symmetric solution $s = 0$ is the only one, whereas, for $c > 1$, there exists a non vanishing optimal value $s^*(c)$ of s that is solution of the implicit equation

$$1 - s^* = \exp(-c\, s^*) \qquad . \tag{4.37}$$

Comparing eqns (4.37) and (4.6) allows to unveil the meaning of $s^*(c)$: it is simply the fraction of vertices $\gamma_1(c)$ belonging to the giant component. Note that the average fraction of connected components $\varphi^*(c)$ equals $-c\, f_1(s^*(c), c)$, in agreement with (4.7).

To understand the structure of the symmetry broken minimum (4.34), and why s^* coincides with the size of the giant component, we need a multivariate generating function accounting for the size of the components, and not only their number. This can be done through the introduction of a magnetic field h into the Potts Hamiltonian (4.15),

$$E[\sigma_1, \sigma_2, \ldots, \sigma_N] = -\frac{1}{N}\sum_{i<j} \delta_{\sigma_i,\sigma_j} - h \sum_i \delta_{\sigma_i,0}. \tag{4.38}$$

This field has a tendency to magnetize spins in the 0 state if $h > 0$, or in any of the remaining $q-1$ values if $h < 0$. All the calculations exposed in Section 4.3 can be repeated, with the free energy functional (4.27) added a $-h\, x(0)$ contribution. In the infinite size limit, the average fractions x may be derived from the density of free energy $f(q, T, h)$,

$$x(0) = \frac{\partial f}{\partial h}(q, T, h) \ , \qquad\qquad x(\sigma) = \frac{1 - x(0)}{q-1} \qquad (\sigma \geq 1) \qquad . \tag{4.39}$$

Notice that, in presence of a non zero field, the permutation symmetry of the free energy functional $\hat{f}$ is explicitly broken. It is therefore justified, and even necessary, to look for a minimum of the form (4.34).

The one–to–one correspondence between the expansion of partition function associated to the energy function (4.38) and the subgraphs of K_N exposed in

Section 4.2.2 now demands some multiplicative factor $\exp(h\,\delta_{\sigma_i,0}/T)$ to be associated to each vertex i of the subgraphs. As a consequence, the partition function reads

$$Z_N(q,T,h) = \sum_{G\subset K_N} w^{N_E(G)} \prod_{j=1}^{\Phi(G)} \left[q - 1 + \left(e^{h/T}\right)^{\Gamma(G_j)} \right] \quad , \tag{4.40}$$

where the G_js denote the components of G. To calculate $x(0)$ from (4.39), we must differentiate partition function (4.40) at a finite field, multiply by $-T/N/Z_N(q,T,h)$, send $N \to \infty$, and later make $h \to 0^+$. It can be easily checked that the output of this procedure is

$$x(0) = \frac{1-\gamma_1}{q} + \gamma_1 \quad , \tag{4.41}$$

where $\gamma_1 = \Gamma_1/N$ is the average fraction of sites in the giant component (if any). Comparing eqns (4.41) and (4.34), we see that s parametrising the broken minimum indeed coincides with γ_1.

(CHANGE WHAT FOLLOWS ACCORDING TO PAPER WITH ANDREAS)

Let us mention that the whole distribution of component sizes may be derived from this approach (Lubensky, McKane, 1981). First we differentiate eqn (4.40) with respect to h, then send $N \to \infty$, and finally send q to one to obtain

$$\sum_{C=1}^{\infty} J(\Gamma,c)\,\Gamma\, e^{-h\Gamma/T} = 1 - s(h,c) \quad , \tag{4.42}$$

where $J(\Gamma,c)$ is the typical number per vertex of components of size C, and $s(h,c)$ is the location of the maximum of $-c\,f_1(s,c) - h(1-s)$ where the expression of f_1 is given in eqn (4.36). The self-consistent equation for $s(h,\gamma)$ reads

$$1 - s = e^{-c\,s-h} \quad . \tag{4.43}$$

Using Lagrange inversion theorem[72], we obtain s as a function of h and identify the power of $e^{-h/T}$ on both sides of eqn (4.42) to obtain

$$J(\Gamma,c) = \frac{1}{c}\frac{\Gamma^{\Gamma-2}}{\Gamma!}\left(c\,e^{-c}\right)^{\Gamma} \quad , \tag{4.44}$$

for all $\Gamma \geq 1$, in agreement with eqn (**??**).

[72] We use Lagrange inversion theorem (...); let us define $t = (1-s)/c$, $u = c\exp(-c-h)$, then $u = t\varphi(t)$ where $\varphi(t) = \exp(-t)$. Inversion of $u(t)$ gives $t(u) = \sum_{n\geq 1} t_n u^n/n!$ where t_n is the $n-1$th derivative of $\varphi(t)^n$ in $t=0$. Going back to variables s, c, h gives eqn (**??**).