

# On Certain Perturbations of the Erdős–Renyi Random Graph

Stéphane Coulomb<sup>1</sup> and Michel Bauer<sup>1</sup>

Received October 8, 2003; accepted January 21, 2004

---

We study perturbations of the Erdős–Renyi model for which the statistical weight of a graph depends on the abundance of certain geometrical patterns. Using the formal correspondance with an exactly solvable effective model, we show the existence of a percolation transition in the thermodynamical limit and derive perturbatively the expression of the threshold. The free energy and the moments of the degree distribution are also computed perturbatively in that limit and the percolation criterion is compared with the Molloy–Reed criterion.

---

**KEY WORDS:** Random graphs; degree correlations; percolation.

## 1. INTRODUCTION

Random graphs were introduced more than forty years ago by mathematicians and have proved since then to be a very useful and versatile concept. The most studied example is the Erdős–Renyi model,<sup>(5)</sup> where the edges are independent. Balanced with the simplicity of its definition, the richness and deepness of mathematical results are really fascinating.

On the other hand, there is evidence that the Erdős–Renyi model is a poor idealization of real networks, those which pop out naturally in sociology, biology, communication sciences,... . For instance, the degree distribution (i.e., the statistics of the number of edges incident at a vertex) of most of the real life examples exhibits statistical, scale-free, properties very far from the poissonian behavior predicted by the Erdős–Renyi model.<sup>(1,4)</sup>

Many random graph models are now on the market, some constructed *ad hoc* to reproduce certain desired features needed to fit real data, some

---

<sup>1</sup> Service de Physique Théorique de Saclay, Laboratoire de la Direction des Sciences de la Matière du Commissariat à l’Energie Atomique, URA2306 du CNRS, CE Saclay, 91191 Gif sur Yvette, France; e-mail: {coulomb, bauer}@spht.saclay.cea.fr

constructed according to general principles. Belonging to the second category, the Molloy–Reed model<sup>(6)</sup> concentrates, inside the space of all labeled graphs with uniform probability, on the subspace of graphs with an arbitrarily given degree distribution. Within this model, many relevant quantities can be computed analytically, and there is a general percolation criterion given in terms of cumulants of the edge degree distribution. Another general class of random graphs is the so-called Markov graph model (see, e.g., refs. 10 and 11), in which short-range correlations between edges are introduced.

Our aim is to study another family of random graphs, with longer range correlations, for which explicit computations are also possible. The idea is roughly as follows. Suppose that to each graph  $G$  one assigns a weight  $u(G)$ . From the weight  $u$  one can construct another weight  $w(G) = \sum_{G' \subset G} u(G')$  where the sum is over graphs  $G'$  with the same vertex set as  $G$  and edge set included in that of  $G$ . Conversely, from any weight function  $w$  one can extract a unique weight function  $u$ , but the expression of  $u$  in terms of  $w$  involves minus signs.

We shall introduce a model for which the weight  $u$  is a counting function for certain structures on graphs. See, for instance, ref. 9 for a slightly different approach of a similar problem, and ref. 12 for a study of nonlocal interactions. This weight  $u$  has two further properties: first it is permutation invariant, i.e., the weight of a graph does not depend on the labelling of its vertices, and second it factors over connected components, i.e., the weight for a graph with several connected components is the product of the weight of each component. Note that by standard combinatorial arguments, these two properties are inherited by the weight  $w$ .

Then we study the thermodynamic finite connectivity limit, when the size of the system (i.e., the number of vertices of the graph) becomes large but the average number of neighbors of any given vertex has a fixed finite value. In this regime, the relevant features of the weights  $u$  and  $w$  are encoded in two tree generating functions,  $u$  and  $w$  respectively.

The idea is that because  $u$  counts less objects than  $w$  (which is not true for arbitrary  $w$  because then  $u$  does not have a simple combinatorial interpretation in general), the generating function  $u$  has better convergence properties than the corresponding generating function  $w$ . We shall make the (crucial) assumption that the first singularity in the generating function of  $w$  can be obtained from the functional relation that ties it to the generating function of  $u$ , without having to know the singularities of the generating function for  $u$  itself. This is certainly true, as we shall recall later, for the Erdős–Renyi model. It is also true order by order in perturbation theory around the Erdős–Renyi model for the models we introduce. We shall have little to say analytically on non-perturbative properties, but the numerical simulations are encouraging.

Under this assumption we are able to give expressions for the free energy, the size distribution of connected components and for the percolation criterion and size of the giant component when it exists. The expressions are not very explicit, because they involve the function  $u$ , which is very complicated in terms of the original parameters of the model. But we show how to perform explicit perturbative computations of the physical quantities. We also introduce an effective model for which the relationship between  $u$  and  $u$  is directly computable, and we illustrate these computations on three easy examples.

Our motivations are the following. First the models we study form a natural and reasonably manageable family of random graph models. Our point is to emphasize the connection with quantum field theory. We do not claim that the relation is very deep, but many random graph phenomena have quantum field theory counterparts, and quantum field theory gives a very convenient language and insight. Second, one of the interests of studying models with nontrivial degree correlations is that attacks (see, e.g., refs. 7 and 8) automatically induce such features, even if they were absent to begin with. Third, at a more basic level, we can contrast with the Molloy–Reed model. This is useful for the purpose of general comparison, but especially because heuristic arguments, always based on non-explicit assumptions, allow to recover the Molloy–Reed percolation criterion without using the particular hypothesis of the Molloy–Reed model, thereby suggesting that the Molloy–Reed percolation criterion has a much wider range of validity. This is probably wrong, and the model solved in this paper is definitely not in this range.

## 2. THE MODEL

After recalling the elementary graph theoretic definitions, we present our basic assumptions. We use the framework of statistical mechanics, i.e., we assign to each labelled graph of size  $N$  a weight (real positive number), which we use as an unnormalized probability distribution. For the Erdős–Renyi model, the weight is simply  $p^{E(G)}(1-p)^{\frac{N(N-1)}{2}-E(G)}$  where  $E(G)$  is the number of edges of  $G$ . We shall choose a weight function that depends on more detailed local features of the graph, namely the abundance of certain geometric motifs.

### 2.1. A Few Definitions and Notations

**Simple Unoriented Graphs, Connected Graphs, Trees.** A (simple unoriented) graph  $G$  is a couple  $(V, E)$  where  $V \neq \emptyset$  is the vertex set and  $E \subset \{\{i, j\}; i, j \in V, i \neq j\}$  is the edge set. If  $V = \{1, \dots, N\}$  for some integer  $N$ ,

then  $G$  is called a labelled graph. The set of labelled graphs of size  $N$  is denoted  $\mathcal{G}_N$ .

If  $G$  is a graph, we denote by  $V(G)$  the vertex set of  $G$  or the cardinal of this vertex set, depending on the context, i.e., whether a set or a number is expected at that place.<sup>2</sup> Similarly,  $E(G)$  will denote either the edge set of  $G$  or the cardinal of this edge set.

A connected component of  $G$  is a minimal graph  $(V', E')$  with  $V' \subset V$  such that if  $(i, j) \in V' \times V$  and  $\{i, j\} \in E$  then  $j \in V'$  and  $\{i, j\} \in E'$ .

A connected graph is a graph which has only one connected component.

A circuit of  $G$  of size  $s$  is a sequence  $(i_1, \dots, i_s)$  of distinct vertices with  $s \geq 3$  such that  $\{i_1, i_2\}, \dots, \{i_{s-1}, i_s\}$  and  $\{i_s, i_1\}$  are edges.

A tree is a connected graph without circuits, and the set of labelled trees of size  $N$  is denoted by  $\mathcal{T}_N$ .

If  $\{i, j\}$  is an edge of  $G$ , we say that  $i$  and  $j \in V$  are neighbours in  $G$ . The number of neighbours of a given vertex  $i \in V$  in a graph  $G$ , also called the degree of  $G$  at vertex  $i$ , is denoted  $l_i(G)$ , or  $l_i$  when there is no ambiguity. It is the number of elements of  $E$  in which  $i$  appears.

**Adjacency Matrix of a Graph, Operations on Matrices.** The adjacency matrix  $A(G)$  (or simply  $A$ ) of a labelled graph  $G \in \mathcal{G}_N$  is the  $N$  by  $N$  matrix defined by  $A_{i,j} = 1$  if  $\{i, j\}$  is an edge of  $G$ ,  $A_{i,j} = 0$  else. Note that the set of adjacency matrices is the set of symmetric 0, 1 matrices with vanishing diagonal elements.

The sum of all elements of any matrix  $M$  will be written  $\|M\| \equiv \sum_{i,j} M_{i,j}$ . If  $M$  is the adjacency matrix  $A(G)$  of a graph  $G$ ,  $\frac{1}{2}\|M\|$  is  $E(G)$ , the number of edges of  $G$ .

The sum of all diagonal elements of a square matrix  $M$  is the trace of  $M$ , written  $\text{Tr}(M) \equiv \sum_i M_{ii}$ . Note that  $\frac{1}{2}\text{Tr}(A(G)^2)$  is again equal to  $E(G)$ .

## 2.2. Probability Distribution, Partition Function

To emphasize the similarities between the random graph model studied in this paper and quantum field theory, we split the weight of graphs in a product of a free part and an interacting part. The free weight is  $w_0(G) \equiv q^{E(G)}$  where  $q \in ]0, +\infty[$ . For later convenience, we also introduce  $p = q/(1+q) \in ]0, 1[$ . The interacting part is  $w_I(G) \equiv e^{S_I(G)}$ , where

$$S_I(G) \equiv \sum_k \frac{t_k}{2k} \text{Tr} A(G)^k + \sum_k \frac{S_k}{2} \|A(G)^k\|.$$

<sup>2</sup> This should cause no confusion, though from a fundamental point of view, a number is a set as well.

The full weight is  $w(G) \equiv w_0(G) w_I(G)$ . The normalization factor included in the definition of the partition function

$$Z_N \equiv (1-p)^{\frac{N(N-1)}{2}} \sum_{G \in \mathcal{G}_N} w(G)$$

is chosen in such a way that  $Z_N$  can be expressed as an average over Erdős–Renyi weights:

$$Z_N = \langle e^{S_I(G)} \rangle_{ER(p)},$$

where  $ER(p)$  assigns probability  $p^E(1-p)^{\frac{N(N-1)}{2}-E}$  to any graph on  $N$  vertices with  $E$  edges. We view  $w_0$  as describing a gas of independent edges (the Erdős–Renyi model), and  $S_I$  as describing the interactions between edges, the  $t_k$ 's and  $s_k$ 's being arbitrary real parameters that regulate the abundance of certain local geometric features of  $G$ . Note that if  $G$  is made of several connected components,  $G_1, \dots, G_l$ ,  $S_I(G) = S_I(G_1) + \dots + S_I(G_l)$ , so that  $w_I$  factors as a product over connected components. This is also true of  $w_0$ . This multiplicativity of the weight plays a crucial role to simplify the analysis below. Another striking feature is that  $w(G)$  is invariant under permutations of the vertex set. Many other interactions with this property could be incorporated, for instance by including products of traces and norms into the interaction, however these would break the multiplicativity property. The “simple” multiplicative model does not seem to be exactly solvable, and we have to rely on perturbation theory to make explicit computations. Even the “trivial” case when all parameters except, say,  $s_3$  vanish is already complicated enough (see also ref. 9 for a study of the case where all parameters but  $t_3$  vanish). This is why we insist on keeping the two properties: multiplicativity and permutation invariance.

The above model is perfectly well defined for all parameter values as long as  $N$  is finite. However, we shall be interested in taking a large  $N$  limit such that the average degree is a fixed number (so that the number of edges is proportional to  $N$ ). This will impose some constraints, see below.

Our model is a cousin of the matrix models. For instance in hermitian matrix models, the averages are taken not over the discrete set  $\mathcal{G}_N$  of adjacency matrices but over the continuous set  $\mathcal{H}_N$  of hermitian matrices. The group of linear transformations preserving hermiticity is the unitary group, which plays a role analogous to the permutation group for our random graph models. A case abundantly studied in the literature is when the partition function reads

$$Z_N = \int_{\mathcal{H}_N} dH e^{-NS(H)},$$

with  $S(H) = S_0(H) + S_I(H)$ ,  $S_0(H) = \frac{1}{2} \text{Tr}(H^2)$ ,  $S_I(H) = \sum_k \frac{g_k}{2k} \text{Tr}(H^k)$ , and  $dH = \prod dH_{ii} \prod_{i < j} d \text{Re}(H_{ij}) d \text{Im}(H_{ij})$ . These choices ensure that  $dH e^{-NS(H)}$  is invariant under the transformation  $H \rightarrow UHU^{-1}$  when  $U$  is a unitary matrix. Note that the trace and norm of a matrix are permutation invariant, but only the trace is unitary invariant. Many quantities have been computed analytically for these matrix models. The introduction of products of traces in the matrix model action does not break unitary symmetry, and an analytic solution would have important applications. However the resulting models are much more difficult. For our random graph models, the presence of products of traces and norms in the action is incompatible with factorization over connected components and this explains why they are much more complicated, but the meaning of factorization is not clear in the context of matrix models.

### 2.3. Examples

Before elaborating further on this model we give a few examples in order to show what kind of objects appear in  $S_I(G)$  and  $e^{S_I(G)}$ .

For example,  $\|A(G)^k\| = \sum_{1 \leq i_1, \dots, i_{k+1}} a_{i_1 i_2} \cdots a_{i_k i_{k+1}}$  is the number of walks with  $k$  steps on  $G$ , and  $\text{Tr} A(G)^k$  is the number of closed walks on  $G$ , that is walks whose first and last vertices are the same. Note that these walks are not necessarily paths (one may have  $i_j = i_{j'}$  as long as these two vertices are not consecutive on the walk). Figure 1 summarizes the topologies corresponding to these walks for a few small values of  $k$ .

When expanding  $e^{S_I(G)}$  in powers of the  $t_k$ 's and  $s_k$ 's, one is led to deal with products of norms and traces, each of these having the interpretation just depicted above. In terms of geometrical motifs, these products are the superpositions of elementary, first order, motifs. So even if all but a finite number of the  $t_k$ 's and  $s_k$ 's vanish, the sizes of the corresponding motifs

	Trace	Norm
k=2		
k=3		
k=4		

Fig. 1. Illustration of the walks on  $k$  steps appearing in  $S_I(G)$ .

become arbitrarily large as the order of the expansion increases. Yet, a simplification occurs from the fact that, unlike the first order motifs which are connected by principle, these walks may contain several, uncorrelated, connected components. In the following section, we show how to deal with this simplification. In the next one, we give more details on the graphical interpretation of the expansion of  $e^{S_I(G)}$  and present a procedure to reorganize it in the purpose of explicit computations.

### 3. TWO COMBINATORIAL FORMULAE FOR THE PARTITION FUNCTION

In this section, we shall derive two formulæ related to exponentiation in combinatorics.

#### 3.1. Connected Components

Suppose we consider the grand canonical partition function

$$\mathcal{E} = \sum_G \frac{w(G)}{V(G)!} z^{V(G)}$$

as a power series in  $z$ : this is a formal sum over all graphs of any size, from which  $Z_N$  can be recovered as

$$Z_N = (1-p)^{\frac{N(N-1)}{2}} N! \oint \frac{dz}{z^{N+1}} \mathcal{E}(z)$$

where for the time being, the symbol  $\oint \frac{dz}{z^{N+1}}$  is not viewed as a real contour integral, but simply as the operation of taking the term of order  $N$  in the  $z$  expansion of a formal power series. If  $G$  has  $l$  connected components  $G_1, \dots, G_l$ , we observe that

$$\frac{w(G)}{V(G)!} z^{V(G)} = \frac{1}{(\sum_{k=1}^l V(G_k))!} \prod_{k=1}^l q^{E(G_k)} w_I(G_k) z^{V(G_k)}.$$

Summing over  $G$  is the same as summing over the connected components. Up to now the vertices of the  $G_k$  were labelled as subgraphs of  $G$ . Thanks to permutation invariance, one can instead sum over abstract finite sequences of  $l = 1, 2, \dots$  labelled connected graphs but weight the terms in the sum by a combinatorial factor

$$\frac{1}{l!} \frac{(\sum_{k=1}^l V(G_k))!}{\prod_k V(G_k)!}$$

to take into account all possibilities to order them and to label the union of their vertex sets from 1 to  $\sum_k V(G_k)$ . Then  $\Xi = e^W$  where  $W = \sum_G^c \frac{w(G)}{V(G)!} z^{V(G)} \equiv \sum_n \frac{W_n}{n!} z^n$  is defined exactly as  $\Xi$  except that the sum is only over connected graphs (this is what is meant by the symbol  $\sum_G^c$ ). The main formula of this section, the first exponential formula, relates  $Z_N$  and  $W(z)$  as

$$Z_N = (1-p)^{\frac{N(N-1)}{2}} N! \oint \frac{dz}{z^{N+1}} e^{W(z)}. \quad (1)$$

### 3.2. Reorganization of the Perturbative Expansion

The case when  $t_k = s_k = 0$  for all  $k$  corresponds to the Erdős–Renyi model. The probability of a graph only depends on the number of its edges, and many quantities such as degree distributions, component distributions and percolation threshold take a simple form. When one or more of the  $t_k$ 's and  $s_k$ 's are not vanishing, the finer structure of the graph becomes relevant, and this will be the case of interest in this paper.

Let us fix  $G$  and start from the expansion of  $e^{S_I(G)}$  in powers of  $t_k$ 's and  $s_k$ 's. This gives a linear combination of terms of the form:

$$\prod_k ((\text{Tr } A(G)^k)^{m_k} \|A(G)^k\|^{n_k}).$$

If we expand each matrix product, such a term becomes a sum of products of matrix elements of  $A(G)$  of generic form  $A(G)_{i_1 j_1} A(G)_{i_2 j_2} \cdots A(G)_{i_n j_n}$  and we may assume that  $i_1 \neq j_1, \dots, i_n \neq j_n$  because otherwise the product is 0 for any (simple graph) adjacency matrix. On the other hand, to any sequence  $i_1 j_1 \cdots i_n j_n$  with  $i_1 \neq j_1, \dots, i_n \neq j_n$ , we may associate a graph  $H$  with vertex set  $[1, N]$  and edge set  $\{\{i_1, j_1\}, \dots, \{i_n, j_n\}\}$ . The product  $A(G)_{i_1 j_1} A(G)_{i_2 j_2} \cdots A(G)_{i_n j_n}$  vanishes unless all edges of  $H$  are edges of  $G$ , in which case it has value 1.

With this observation in mind, we define  $\overline{e^{S_I(G)}}$  by keeping, in the expansion of  $e^{S_I(G)}$ , only those terms  $A(G)_{i_1 j_1} A(G)_{i_2 j_2} \cdots A(G)_{i_n j_n}$  such that  $\{\{i_1, j_1\}, \dots, \{i_n, j_n\}\}$  exhausts the edge set of  $G$  (maybe with repetitions). Then by definition,

$$e^{S_I(G)} = \sum_{H, E(H) \subset E(G)} \overline{e^{S_I(H)}},$$

where the sum is over all graphs on the same vertex set as  $G$  whose edge set is a subset of that of  $G$ . The reciprocal formula is given by  $\overline{e^{S_I(H)}} = \sum_{G, E(G) \subset E(H)} (-)^{E(H)-E(G)} e^{S_I(G)}$  and the multiplicative property of  $e^{S_I(G)}$



ensures that  $e^{\overline{S_I(H)}}$  is also multiplicative: if  $H$  has  $l$  connected components  $H_1, \dots, H_l$ ,

$$\overline{e^{S_I(H)}} = \overline{e^{S_I(H_1)}} \dots \overline{e^{S_I(H_l)}}.$$

Now

$$\begin{aligned} Z_N &= \sum_{G \in \mathcal{G}_N} (1-p)^{\frac{N(N-1)}{2} - E(G)} p^{E(G)} e^{S_I(G)} \\ &= \sum_{G \in \mathcal{G}_N} \sum_{H, E(H) \subset E(G)} (1-p)^{\frac{N(N-1)}{2} - E(G)} p^{E(G)} \overline{e^{S_I(H)}} \\ &= \sum_{H \in \mathcal{G}_N} \sum_{G, E(G) \supset E(H)} (1-p)^{\frac{N(N-1)}{2} - E(G)} p^{E(G)} \overline{e^{S_I(H)}}. \end{aligned}$$

For fixed  $H$ ,  $\sum_{G, E(G) \supset E(H)} (1-p)^{\frac{N(N-1)}{2} - E(G)} p^{E(G)} = p^{E(H)}$ , and we find that

$$Z_N = \sum_{H \in \mathcal{G}_N} p^{E(H)} \overline{e^{S_I(H)}}.$$

Defining

$$u(H) = p^{E(H)} \overline{e^{S_I(H)}} \quad U = \sum_H^c \frac{u(H)}{V(H)!} z^{V(H)} \equiv \sum_n \frac{U_n}{n!} z^n,$$

we can repeat the steps leading from multiplicativity and permutation invariance to Eq. (1) to obtain the second exponential formula

$$Z_N = N! \oint \frac{dz}{z^{N+1}} e^{U(z)}. \tag{2}$$

### 3.3. Consequences of the Exponential Formulæ

The two expressions obtained for the partition function, one in terms of  $W(z)$  and the other in terms of  $U(z)$ , show that for every nonnegative integer  $N$

$$(1-p)^{-\frac{N(N-1)}{2}} \oint \frac{dz}{z^{N+1}} e^{U(z)} = \oint \frac{dz}{z^{N+1}} e^{W(z)}.$$

It is convenient to eliminate  $N$  by the following trick: putting  $1-p = e^{-\beta}$ ,  $(1-p)^{-\frac{N(N-1)}{2}}$  takes the form of a gaussian integral  $\frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\beta} + yN - \beta \frac{N}{2}}$ . From the change of variable  $x = ze^{y - \frac{\beta}{2}}$ , it follows that

$$e^{W(z)} = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\beta} + U(ze^{y - \frac{\beta}{2}})} dy. \tag{3}$$

There is no good reason why  $e^{-\frac{y^2}{2\beta} + U(ze^{y-\frac{\beta}{2}})}$  should be integrable in  $y$  along the full real axis. However, if one expands this function in powers of  $z$ , term by term integration is ok, and for the time being, Eq. (3) is a shorthand notation for the fact that this term by term integration leads to the formal power series of  $e^{W(z)}$ .

#### 4. PRACTICAL PERTURBATIVE EXPANSION

Our aim is to organize the perturbative expansion to make explicit computations. We would like to make a systematic enumeration of the terms that appear in perturbation theory.

A typical term in the perturbative expansion is of the form  $A_{i_1 j_1} \cdots A_{i_n j_n}$  to which we associate the sequence  $i_1 j_1 \cdots i_n j_n$ , i.e., a word written using the alphabet  $[1, N]$ . For a graph  $G$  with adjacency matrix  $A_{ij}$ , the product  $A_{i_1 j_1} A_{i_2 j_2} \cdots A_{i_n j_n}$  is 1 if  $\{i_1, j_1\}, \dots, \{i_n, j_n\}$  are amongst the edges of  $G$  and 0 else. If  $l$  is the number of distinct edges among these  $n$  2-sets, the average is simply the sum of the Erdős–Renyi weights of all graphs containing these  $l$  edges. This is known to yield  $\langle A_{i_1 j_1} A_{i_2 j_2} \cdots A_{i_n j_n} \rangle_{ER(p)} = p^l$ . This average is invariant under permutations of  $[1, N]$ , all vertices play the same role in  $\text{Tr } A(G)^k$  and  $\|A(G)^k\|$ . So we regroup the words  $i_1 j_1 \cdots i_n j_n$  in classes under the action of the permutation group, compute the size of each class and find a representative in each class. Then we enumerate the representatives and take multiplicities into account.

The idea is the following: suppose that you have a finite word written using any alphabet (i.e., any set of symbols) on, say,  $N$  letters. To each letter that appears in the word, associate an integer as follows: assign 1 to the first letter of the word, then assign 2 to the next new (i.e., distinct from the first) letter appearing in the word, then 3 to the next new (i.e., distinct from the first and the second) and so on until all letters appearing in the word have been assigned a number, the highest one being, say,  $v$  ( $v$  is the number of distinct letters used to compose the word, which may well be strictly smaller than the length of the word, because the same letter can appear more than once). Replacing each letter of the word by its number leads to a new word, the alphabet being  $[1, v] \subset [1, N]$  this time. The words obtained by this procedure are characterized by the fact that 1 appears before 2 which appears before 3 and so on. Say that two words in the original alphabet are equivalent if they yield the same numerical word by the above procedure. Then each class contains  $\frac{N!}{(N-v)!}$  words.

In our case, the original alphabet is already  $[1, N]$ , and we are led to the concept of normalized sequences, an elaboration of a procedure introduced in a slightly simpler context in ref. 3.

### 4.1. Normalized Sequences

For an arbitrary sequence  $i_1 j_1 i_2 j_2 \cdots i_n j_n$  (with alphabet  $[1, N]$ ) such that  $i_1 \neq j_1, \dots, i_n \neq j_n$ , we define  $v = \#\{i_1, j_1, \dots, i_n, j_n\}$ , the number of distinct vertices in the sequence, and  $l = \#\{\{i_1, j_1\}, \{i_2, j_2\}, \dots, \{i_n, j_n\}\}$ , the number of distinct edges in the sequence.

We shall say that a sequence  $i_1 j_1 i_2 j_2 \cdots i_n j_n$  is normalized with respect to  $\prod_k (\text{Tr } A(G)^k)^{m_k} \prod_k \|A(G)^k\|^{n_k}$  or more simply with respect to  $(m_k, n_k)$  if

- $n = \sum_k k(m_k + n_k)$ .
- In this sequence, 1 comes before 2, which comes before 3,... which comes before  $v$ .
- $i_1 \neq j_1, \dots, i_n \neq j_n$ .

• The sequence has a correct structure as regards  $\text{Tr}$  and  $\| \cdot \|$ . That is, to  $\text{Tr } A(G)^p$  ( $m_k = \delta_{p,k}, n_k = 0$ ) and  $\|A(G)^p\|$  ( $m_k = 0, n_k = \delta_{p,k}$ ) correspond the constraints  $j_1 = i_2, \dots, j_{p-1} = i_p$ , with the additional constraint  $j_p = i_1$  for  $\text{Tr } A(G)^p$ . If more than one term of  $(m_k, n_k)$  is nonzero, then we choose an arbitrary ordering: increasing  $k$ 's, all traces coming before norms. This allows to decompose the sequence in subsequences, which correspond either to a trace or a norm, and are accordingly constrained. For instance, when  $m_3 = 1$  and  $n_1 = 1, n_3 = 1$  are the only nonvanishing elements of  $(m_k, n_k)$ , the sequence has a correct structure if it is of the form  $i_1 j_1 i_2 j_2 i_3 j_3 i_4 j_4 i_5 j_5 i_6 j_6 i_7 j_7$  where  $j_1 = i_2, j_2 = i_3, j_3 = i_1$  and  $j_5 = i_6, j_6 = i_7$  (to  $\|A(G)\|$  correspond no constraint of structure).

We write  $\mathcal{M}_{v,l,(m_k,n_k)}$  for the number of normalized sequences with  $v$  vertices and  $l$  edges. By our previous remarks, the class containing a normalized sequence has  $\frac{N!}{(N-v)!}$  members, each of which leads to the same average. Hence

$$\left\langle \prod_k ((\text{Tr } A(G)^k)^{m_k} \|A(G)^k\|^{n_k}) \right\rangle_{ER(p)} = \sum_{v,l} \frac{N!}{(N-v)!} p^l \mathcal{M}_{v,l,(m_k,n_k)}. \tag{4}$$

In doing explicit computations, which can be painful, there is a useful check of the formula, namely a sum rule corresponding to  $p = 1$ , in which case only the complete graph contributes, and there is no average to compute. It is straightforward to check that if  $G$  is the complete graph on  $N$  vertices,

$$\text{Tr } A(G)^k = (N-1)^k + (-1)^k (N-1) \quad \|A(G)^k\| = N(N-1)^k.$$

Hence

$$\sum_{v,l} \frac{N!}{(N-v)!} \mathcal{M}_{v,l,(m_k,n_k)} = \prod_k ((N-1)^k + (-)^k (N-1))^{m_k} (N(N-1)^k)^{n_k}.$$

## 4.2. Graphical Expansion

Although the interpretation in terms of normalized sequences is adequate for the purpose of numerical computations, there is another useful graphical representation of the perturbation series which we present briefly now.

Expanding  $\text{Tr } A^k = \sum_{i_1, \dots, i_k} A_{i_1, i_2} A_{i_2, i_3} \cdots A_{i_k, i_1}$  we represent each term as a colouring of a labelled cycle on  $k$  vertices with  $N$  colours, vertex  $j$  carrying color  $i_j$  for  $j = 1, \dots, k$ . In the same way, we represent each term in  $\|A^k\| = \sum_{i_1, \dots, i_{k+1}} A_{i_1, i_2} A_{i_2, i_3} \cdots A_{i_k, i_{k+1}}$  as a colouring of a labelled segment on  $k+1$  vertices with  $N$  colours, vertex  $j$  carrying color  $i_j$  for  $j = 1, \dots, k+1$ . The expansion of  $e^{S_l}$  in powers of  $t_k$ 's and  $s_k$ 's, is then represented as a sum over colorings, with  $N$  colours, of labelled graphs whose connected components are cycles and segments. Pick one term, call it  $\Gamma$ , in this sum. Each cycle of length  $k$  yields a factor  $t_k/(2k)$ , each segment on  $k+1$  vertices yields a factor  $s_k/2$ , there is a factor  $1/m_k!$  if there are  $m_k$  cycles of length  $k$ , and a factor  $1/n_k!$  if there are  $n_k$  segments on  $k+1$  vertices.

The probabilistic average of  $\Gamma$  (over the set of incidence matrices) is zero if some edge of  $\Gamma$  has its two extremities of the same colour. If not, let  $e(\Gamma)$  be the number of distinct pairs of colors that appear as extremities of edges of  $\Gamma$ . The probabilistic average over the set of incidence matrices multiplies the former weight of  $\Gamma$  by  $p^{e(\Gamma)}$ .

Let  $v(\Gamma)$  be the number of distinct colours in the colouring of  $\Gamma$ . Say that terms  $\Gamma$  and  $\Gamma'$  are equivalent if there is a permutation of  $[1, N]$  (the set of colours) that maps  $\Gamma$  to  $\Gamma'$ . The equivalence class of  $\Gamma$  is made of  $N!/(N-v(\Gamma))!$  graphs with the same weight. The equivalence class of  $\Gamma$  has a graphical representation: starting from  $\Gamma$ , draw a dashed line between two vertices if and only if they carry the same color. Then remove the colors. In this way, obtain a graph with two kind of edges, solid and dashed. The graphs that appear in this operation have two properties. First, the solid components are cycles and segments, and the dashed components are complete graphs. Second, two vertices cannot be adjacent for solid and dashed edges at the same time. In lack of a better denomination, we call graphs satisfying these two conditions (labelled) admissible graphs.



### 4.3. Exponentiation

Computations on a sheet of paper are more economical using admissible graphs, but systematic machine enumeration is best carried out using normalized sequences.

It is not difficult to convince oneself that the notion of connectedness of normalized sequences or of admissible graphs is the same, and coincides with the notion of connectedness used to establish Eq. (2). We infer that

$$U(z) = z + \sum_H^c \frac{1}{s(H)} p^{e(H)} z^{v(H)} \prod_k t_k^{m_k(H)} s_k^{n_k(H)},$$

where  $\sum_H^c$  is the sum over unlabelled connected admissible graphs or equivalently that

$$U(z) = z + \sum_{v,l,\{m_\cdot, n_\cdot\}} z^v p^l \frac{\tilde{t}_\cdot^m \tilde{s}_\cdot^n}{m_\cdot! n_\cdot!} \mathcal{M}_{v,l,\{m_\cdot, n_\cdot\}}^c,$$

where  $\tilde{t}_\cdot^m \tilde{s}_\cdot^n \equiv \prod_k \binom{t_k}{2k}^{m_k} \binom{s_k}{2}^{n_k}$  and  $\mathcal{M}_{v,l,\{m_k, n_k\}}^c$  is the number of normalized connected sequences.

## 5. FINITE CONNECTIVITY LARGE $N$ BEHAVIOUR

Our aim is to use the identities Eqs. (1) and (2) and their consequence Eq. (3) to derive mean field type identities valid in the limit  $N \rightarrow \infty$ ,  $pN, t_k, s_k$  being  $N$  independent, or more generally having finite limits for large  $N$ . It is customary to define  $\alpha \equiv \lim_{N \rightarrow \infty} pN$ .

### 5.1. Caveat

In this regime, when  $t_k = s_k = 0$  for all  $k$ 's (the Erdős–Renyi model), the event that a graph contains a component with much more edges than vertices has a vanishingly small probability, and the connected components look locally like trees. This is called the dilute regime. In particular, the complete graph (a caricature of a non-dilute graph) has a negligible weight in the dilute regime.

We are looking for an analogous regime for the perturbed Erdős–Renyi model. However, in that case, the complete graph has weight

$$w = q^{N(N-1)/2} e^{\sum_k \frac{t_k}{2k} ((N-1)^k + (-)^k (N-1)) + \sum_k \frac{s_k}{2} N(N-1)^k}.$$

Compare this to a union of isolated vertices (a caricature of a dilute graph), which has weight  $w = 1$ . Consider for instance, the case when there is only

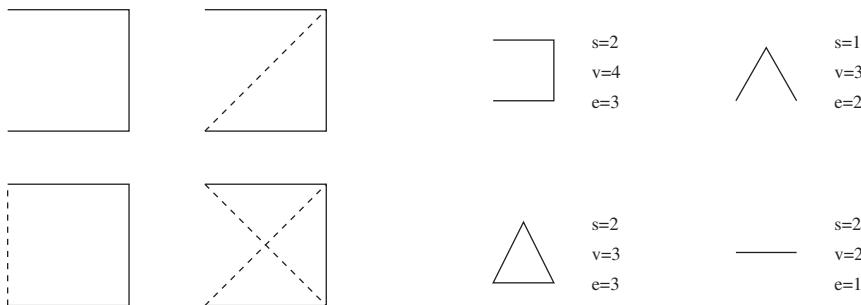


Fig. 3. Graphical representation for the coefficients of  $s_3$ .

one nonvanishing perturbation parameter, say  $s_3$ . The two weights are equal if  $qe^{s_3(N-1)} = 1$ . That this crude balance gives the correct qualitative frontier between a dilute regime and a dense regime is confirmed by numerical simulations.

If  $p \equiv \alpha/N$  and  $s_3 < 0$ , the complete graph is indeed strongly suppressed (in fact much more drastically than for the pure Erdős–Renyi model). However, if  $s_3 > 0$ , the weight of the complete graph submerges the weight of dilute configurations.

To summarize, the following discussion makes sense only if  $S_l$  does not become positive and large ( $\gg N^2$ ) for non-dilute configurations. An easy way to ensure that is to take the sign of all perturbations negative. Another possibility would be to take only a finite number of nonzero perturbations, and then impose that the dominant one be negative.

## 5.2. Main Equations

With these observations in mind, we start from

$$\begin{aligned} Z_N &= N! \oint \frac{dy}{y^{N+1}} e^{U(y)} \\ &= (1-p)^{\frac{N(N-1)}{2}} N! \oint \frac{dx}{x^{N+1}} e^{W(x)} \end{aligned}$$

and

$$e^{W(x)} = \frac{1}{\sqrt{2\pi\beta}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2\beta} + U(xe^{z-\frac{\beta}{2}})} dz.$$

Recall that  $p = q/(1+q) = 1 - e^{-\beta}$ . We fix  $p = p_N \equiv \alpha/N$  where  $\alpha$  is a constant, and make changes of variables  $y \rightarrow y/p$ ,  $x \rightarrow x/q$  in the above integrals. In the expansions

$$U(y/p) = \sum_H^c \left( \frac{\alpha}{N} \right)^{E(H)-V(H)} \frac{y^{V(H)}}{e^{S_I(H)} V(H)!}$$

and

$$W(x/q) = \sum_H^c \left( \frac{\alpha}{N-\alpha} \right)^{E(H)-V(H)} e^{S_I(H)} \frac{x^{V(H)}}{V(H)!}$$

the sum is over connected graphs, and by Euler formula,  $E - V = L - 1$  where  $L \geq 0$  is the number of loops. Hence we may write formally

$$U(y/p) = \frac{N}{\alpha} u(y) + o(N) \quad W(x/q) = \frac{N}{\alpha} w(x) + o(N),$$

where  $u(y) = \sum_T \frac{e^{S_I(T)} y^{V(T)}}{V(T)!}$  and  $w(x) = \sum_T e^{S_I(T)} \frac{x^{V(T)}}{V(T)!}$  are sums over trees (connected graphs with  $L = 0$ ). If we use a naïve version of the saddle point approximation and write  $Z_N = e^{NF + o(N)}$ , we find

$$F = -1 - \log \frac{y^*}{\alpha} + \frac{1}{\alpha} u(y^*) \quad (5)$$

$$F = -\frac{\alpha}{2} - 1 - \log \frac{x^*}{\alpha} + \frac{1}{\alpha} w(x^*) \quad (6)$$

$$w(x) = -\frac{\hat{z}^2}{2} + u(xe^{\hat{z}}), \quad (7)$$

where the  $x^*$ ,  $y^*$ , and  $\hat{z}$  are appropriate saddle point values:

$$\alpha = x^* w'(x^*) = y^* u'(y^*) \quad \hat{z} = x e^{\hat{z}} u'(x e^{\hat{z}}).$$

We end this section with the following remarks. The average number of edges is more or less the variable conjugate to  $p$ . More precisely, the average number of edges is

$$q \frac{\partial}{\partial q} \log((1-p)^{-N(N-1)/2} Z_N).$$



We infer that in the thermodynamic regime with  $N \rightarrow \infty$  and  $p_N = \alpha/N$ , the average number of neighbors of a given point (i.e.,  $2/N$  times the average number of edges) is

$$c = \alpha + 2\alpha \frac{\partial F}{\partial \alpha}. \quad (8)$$

For the pure Erdős–Renyi model, the weights form a probability distribution,  $Z_N = 1$ ,  $F = 0$ , and  $c = \alpha$ . In the perturbed models,  $\alpha$  is not so easily measured on the graph, and only the parameter  $c$  has direct physical meaning. From the point of view of quantum field theory, it is natural to view  $\alpha$  as the bare connectivity and  $c$  as the physical connectivity. For each coupling constant  $t_k$  or  $s_k$ , it would be desirable to find analogous physical quantities that first, one can compute directly on a random graph without knowing *a priori* the sampling measure and that second one can reduce  $t_k$  or  $s_k$  to first order in perturbation theory. This is very ambiguous and we have not found an elegant way to select such physical observables systematically.

### 5.3. Discussion

We have seen before that a dilute regime for the perturbed Erdős–Renyi model with fixed values of the  $t_k$ 's and  $s_k$ 's cannot exist if  $S_I$  becomes large positive for graphs with many loops. Here we discuss a related limitation even if one considers only loopless graphs.

Instead of considering the complete graph, look at the star shaped tree on  $n$  vertices, whose adjacency matrix we denote by  $S$ , with a center connected to the  $n-1$  other vertices. From  $\text{Tr } S = 0$ ,  $\|S\| = \text{Tr } S^2 = 2(n-1)$ ,  $\|S^2\| = n(n-1)$ , and  $S^3 = (n-1)S$ , it is easy to compute recursively that  $\text{Tr } S^{2k+1} = 0$  and  $\|S^{2k+1}\| = 2(n-1)^{k+1}$  for  $k \geq 0$ , and that  $\text{Tr } S^{2k} = 2(n-1)^k$  and  $\|S^{2k}\| = n(n-1)^k$  for  $k \geq 1$ . As an example, consider again the case when there is only one nonvanishing perturbation parameter, say  $s_3$ . The contribution of star shaped trees to  $w$  is  $\sum_n \frac{1}{(n-1)!} e^{s_3(n-1)^2} x^n$ . As all trees give a positive contribution to  $w$ , no compensation is possible and we conclude that if  $s_3 > 0$ , the series for  $w$  has a vanishing radius of convergence. So it is meaningless to deform contours, and Eq. (6) is meaningless as well. Then so is Eq. (5) because analyticity of  $u(y)$  at small  $y$  implies analyticity of  $w(x)$  at small  $x$  via Eq. (7). On the other hand, if  $s_3 < 0$ , the star-shaped trees of large size are very strongly suppressed. Let us note however as shown in the next section that, in the realm of formal power series, Eq. (7) describes the correct combinatorial relationship between  $u$  and  $w$  even if both series have a vanishing radius of convergence.

More generally, if  $S_T(T)/V(T)$  is bounded above (an easy way to ensure that is to take the sign of all perturbations negative, another possibility would be to take only a finite number of nonzero perturbations, and then impose that the dominant one be negative),  $w$  is analytic near the origin. Indeed, if  $S_T(T)/V(T) \leq \tau$  for all trees, using the fact that there are  $n^{n-2}$  labelled trees on  $n$  vertices, we see that  $0 < \frac{w_n}{n!} \leq \frac{n^{n-2}}{n!} e^{\tau n}$ , leading to a nonzero radius of convergence.

For instance, when the sign of every perturbation is negative, the radius of convergence is a nonincreasing function of the  $t_k$ 's and  $s_k$ 's: it gets larger and larger as the  $t_k$ 's and  $s_k$ 's get more negative. To see that it remains finite, consider the linear graph on  $n$  vertices, whose adjacency matrix we denote by  $L$ . For this graph, for fixed  $k$  and large  $n$ ,  $\text{Tr } L^k$  and  $\|L^k\|$  grow at most linearly with  $n$ : they count  $k$  steps walks, and if the starting point is given, at each step there are at most two choices, so there is the obvious upper bound  $n2^k$ . There are  $n!/2$  ways to label the linear graph (the symmetry group is of order 2). So the contribution of the linear trees to  $w$  decreases at most geometrically with the size. As all trees give a nonnegative contribution,  $w$  has its first singularity on the real positive axis, and at a finite distance.

In the situation when  $w$  has a finite radius of convergence, we conclude that there is a forest-like regime for the perturbed Erdős–Renyi model that extends the forest-like regime of the pure Erdős–Renyi model, and that it is described by Eqs. (5)–(7), at least in the small  $\alpha$  phase. We shall elaborate on this point in the sequel.

## 5.4. Combinatorial Remarks

As we have seen before, the above formulæ for the free energy rely on crucial assumptions. What we would like to show in this subsection, before embarking on a detailed discussion of analytic features of these equations, is that the combinatorics embodied in Eq. (7) is correct. Suppose that we forget about the random graph model for a moment, and consider instead a random forest model,  $w$  being the generating function for random weighted trees.

Expand the function  $u(xe^z)$  in powers of  $z$ :  $u(xe^z) = \sum_l \frac{z^l}{l!} (x \frac{d}{dx})^l u(x)$ . It is well-known from quantum field theory that the formal expansion of

$$\int dz e^{\frac{1}{\hbar}(-\frac{z^2}{2} + u(xe^z))}$$

is a weighted sum of all connected Feynmann graphs. The weight of a Feynmann graph is computed as follows: each edge gives a factor  $\hbar$

(propagator), each vertex of degree  $l$  gives a factor  $\hbar^{-1}(x \frac{d}{dx})^l u(x)$  and finally one divides by the order of the symmetry group of the graph. The logarithm is given by the same sum, but restricted to connected graphs. For connected graphs, the power of  $\hbar$  is the number of loops minus 1, so the dominant contribution in the small  $\hbar$  limit restricts the sum to connected loopless graphs, i.e., trees. On the other hand, the small  $\hbar$  limit is given by the saddle point approximation, i.e., Eq. (7). So  $w$  is a sum over all trees, each vertex of degree  $l$  giving a factor  $(x \frac{d}{dx})^l u(x)$ .

But  $u(x)$  itself is a tree generating function, so  $(x \frac{d}{dx})^l u(x)$  is the generating function for trees with  $l$  marked vertices (a vertex can be marked more than once). So Eq. (7) means that to construct  $w$ , one takes arbitrary trees, (call them naked trees) and then blows up every vertex of degree  $l$  into a new tree with  $l$  marked vertices from which naked edges emerge. Note that a naked vertex can be blown up in a trivial tree, corresponding to the term  $x$  in  $u(x) = x + \dots$ .

As we have emphasized before, if  $T$  is a tree, each term  $t$  in the expansion of  $e^{S_I(T)}$  in terms of matrix elements of the adjacency matrix  $A$  of  $T$  defines a subgraph of  $T$ , i.e., a forest with the same vertex set as  $T$ , edge  $\{i, j\} \in E(T)$  being present in the forest if and only if the term  $t$  contains the factor  $A_{ij}$  or  $A_{ji}$ . But the connected components of the forest being given, one reconstructs  $w$  by connecting the different components with appropriate edges. This is exactly the procedure described by Eq. (7) if  $w$  is the generating function for  $e^{S_I(T)}$  and  $u$  the one for  $e^{\overline{S_I(T)}}$ .

## 5.5. Effective Model

If we have the original model in mind, each  $u_k$  is itself a highly non-trivial kind of partition function. However, if we take each  $u_k$  as an independent parameter, we can make a rather general analysis. In fact, there is a simple model for which the  $u_k$ 's are the fundamental microscopic parameters in the sense that they appear directly in the definition of the weights. We call this model an effective model for the following reasons.

In quantum field theory, the term “effective” often means that one renounces to deal with all observables of a system and only concentrates on certain degrees of freedom, so that the other ones can be averaged. For instance, to compute the long distance behaviour, one first averages over the short distance fluctuations. We are going to do something analogous here: we renounce to observe the local structure of connected components and are only interested in the distribution of their size. So instead of keeping track of the weight of each detailed connected component, we can as well give all components of a given size the same weight,

namely the average weight given by the original model for components of that size.

Now, to the precise definition. Choose parameters  $c_1, c_2, \dots$  and define an effective weight  $u^{(\text{eff})}(H) = p^{E(H)} c_k$  for any connected graph  $H$  of size  $k$ , and assume multiplicativity, so that for an arbitrary graph  $u^{(\text{eff})}(H) = p^{E(H)} \prod_k c_k^{n_k(H)}$  where  $n_k(H)$  is the number of components of size  $k$  of  $H$ . If we trade  $c_k$  for  $\lambda^k c_k$ , we multiply the weight  $u^{(\text{eff})}(H)$  by a trivial factor  $\lambda^{V(H)}$  so we shall assume that the  $c_k$ 's are normalized by  $c_1 = 1$  (the special case  $c_1 = 0$  would need a separate treatment). Define the corresponding effective weight  $w^{(\text{eff})}(G) = q^{E(G)} \sum_{H, E(H) \subset E(G)} u^{(\text{eff})}(H) p^{-E(H)}$ , where the sum is over all graphs on the same vertex set as  $G$  whose edge set is a subset of that of  $G$ . Note that contrary to the weight  $u^{(\text{eff})}$ , the weight  $w^{(\text{eff})}(G)$  does in general depend on the detailed structure of the graph, and not only on the sizes of connected components. Our interest however is in the distribution of sizes of connected component of graphs of large size  $N \rightarrow \infty$  sampled using the weight  $w^{(\text{eff})}$ . Following the same steps as for the original model, we find that this distribution can be obtained in the thermodynamic limit from tree generating functions  $u^{(\text{eff})}$  and  $w^{(\text{eff})}$  satisfying the very same coupled equations (5)–(7) as the original  $u$  and  $w$ . The coefficients of  $u^{(\text{eff})}$  are very simple in terms of  $c_1, c_2, \dots$  because all components of the same size have the same weight, and by Cayley's theorem there are  $k^{k-2}$  trees on  $k$  vertices. Hence  $u^{(\text{eff})} = \sum_{k \geq 1} \frac{k^{k-2}}{k!} c_k y^k$ . Hence if one sets  $c_k = k^{2-k} \sum_{T \in \mathcal{T}_k} e^{S_T(T)}$  where the sum is over trees of size  $k$ , the effective model has the same component size distribution as the original one.

For all these reasons, we shall remove in the sequel the superscript  $(\text{eff})$  from  $u^{(\text{eff})}$  and  $w^{(\text{eff})}$ , even if we sometimes keep the distinction between the weights  $u$  and  $w$  and the effective weights  $u^{(\text{eff})}$  and  $w^{(\text{eff})}$ . Accordingly, we shall analyse Eqs. (5)–(7), which involve only the component size distribution, without making explicitly the distinction between the original model and the effective model.

## 5.6. Connected Components and Percolation

We return to the finite  $N$  arbitrary  $p$  case to start the argument. As  $W_k$  is, modulo an overall multiplicative factor, the total weight of connected graphs of size  $k$ , we infer from Eq. (1) that the mean number of connected components on  $k$  vertices is

$$N_k = \frac{W_k}{Z_N} \frac{\partial Z_N}{\partial W_k} = \frac{N!}{k! (N-k)!} W_k \frac{Z_{N-k}}{Z_N} (1-p)^{\frac{N(N-1) - (N-k)(N-k-1)}{2}}.$$

Taking into account that when  $pN = \alpha$ ,  $p(N - k) = \alpha(1 - k/N)$ , we find that in the dilute regime, for fixed  $k$  and  $N \rightarrow \infty$ ,  $\frac{Z_{N-k}}{Z_N} \sim e^{-k(F + \alpha \frac{\partial F}{\partial \alpha})}$  and  $(1 - p)^{\frac{N(N-1) - (N-k)(N-k-1)}{2}} \sim e^{-k\alpha}$  so that

$$N_k/N \sim \frac{w_k}{k!} \alpha^{k-1} e^{-k(\alpha + F + \alpha \frac{\partial F}{\partial \alpha})}.$$

As expected, in this regime only trees contribute thermodynamically to the finite components.

From these equations for the abundance of connected component of each size, we can easily derive a percolation criterion. Indeed, by construction,  $\sum_k kN_k/N = 1$ , but what about the approximate sum

$$\sum_k k \frac{w_k}{k!} \alpha^{k-1} e^{-k(\alpha + F + \alpha \frac{\partial F}{\partial \alpha})}?$$

For each fixed  $k$  and  $N \rightarrow \infty$ , the  $k$ th term is a good approximation to  $kN_k/N$ , but there is problem of inversion of limits. Physically, the approximate sum counts the fraction of points in components of finite size, so it is  $\leq 1$ .

If we assume that  $u$  is analytic at small  $y$ , then  $F$  is analytic and small at small  $\alpha$  and  $w$  is analytic and small at small  $x$ . Moreover,  $x^*$  is an increasing function of  $\alpha$  at small  $\alpha$ . From Eq. (6), we infer that  $F + \alpha \frac{\partial F}{\partial \alpha} = -\alpha - \log \frac{x^*}{\alpha}$  or equivalently,  $\alpha e^{-\alpha - F - \alpha \frac{\partial F}{\partial \alpha}} = x^*$ . Then  $\sum_k k \frac{w_k}{k!} \alpha^{k-1} e^{-k(\alpha + F + \alpha \frac{\partial F}{\partial \alpha})} = \frac{x^*}{\alpha} w'(x^*) = 1$  for small enough  $\alpha$ . However, it may happen that as a function of  $\alpha$ ,  $x^* = \alpha e^{-\alpha - F - \alpha \frac{\partial F}{\partial \alpha}}$  is non-monotonic. There may be a value  $\alpha_c$  such that  $x^*$  increases in the interval  $[0, \alpha_c]$  but then starts to decrease, so that  $x^*(\alpha) \leq x^*(\alpha_c)$  in some interval strictly containing  $[0, \alpha_c]$ . One could build models where  $x^*(\alpha)$  has several oscillations, but in the sequel, we concentrate on the first. For a given  $\alpha$ , denote by  $\bar{\alpha} \leq \alpha_c$  the small solution to the equation  $x^*(\alpha) = x^*(\bar{\alpha})$ . Then we obtain the more general result that finite components occupy a fraction  $\frac{x^*}{\alpha} w'(x^*) = \frac{\bar{\alpha}}{\alpha} \leq 1$  of the sites in the system. If  $\alpha > \alpha_c$ , something else than finite components, in fact on general grounds one single giant component, occupies a fraction  $1 - \frac{\bar{\alpha}}{\alpha}$  vertices. Thus, the percolation criterion is that  $\alpha e^{-\alpha - F - \alpha \frac{\partial F}{\partial \alpha}}$  is maximum at  $\alpha = \alpha_c$ . So the transition point is when

$$\alpha + 2\alpha \frac{\partial F}{\partial \alpha} + \alpha^2 \frac{\partial^2 F}{\partial \alpha^2} = 1. \tag{9}$$

The first two terms yield simply the true average connectivity  $c = \alpha + 2\alpha \frac{\partial F}{\partial \alpha}$ , it would be nice to have a direct physical interpretation of the third term

$\alpha^2 \frac{\partial^2 F}{\partial \alpha^2}$ . This percolation criterion is expressed solely in terms of the free energy as a function of  $\alpha$ . But it can also be related to analytic properties of  $w$ . Indeed, the relevant saddle point equation is  $\alpha = x^* w'(x^*)$ . As  $\alpha$  approaches  $\alpha_c$ ,  $x^*$  reaches a maximum, so that the  $x^*$  derivative of  $x^* w'(x^*)$  has to get large, diverging at  $\alpha = \alpha_c$ . If  $\alpha_c$  is finite, this means that  $w$  and  $w'$  are finite at  $\alpha = \alpha_c$ , but  $w''$  is infinite. If the coefficients of  $w$  are nonnegative,<sup>3</sup> this means that  $x^*(\alpha_c)$  is the radius of convergence of  $w$ . From  $x^*(\alpha_c)$ , we recover  $\alpha_c$  itself by the general saddle point equation  $\alpha = x^* w'(x^*)$ .

Suppose now (we shall soon argue that this is true in many cases including interesting ones) that even if  $x^*(\alpha_c)$  is the radius of convergence of  $w$ , the function  $u$  is not singular at  $y^*(\alpha_c)$ . Hence  $u$  allows to compute the free energy  $F$  and show that it is analytic in some interval strictly containing  $\alpha_c$ . From that point of view, we observe that the saddle point equations imply that  $x^* = y^* e^{-\alpha} = y^* e^{-y^* u'(y^*)}$  from which the percolation criterion, i.e., the determination of the maximum of  $x^*$  becomes

$$y^* \frac{\partial \alpha}{\partial y^*} = y^* u'(y^*) + y^{*2} u''(y^*) = 1. \quad (10)$$

In the same spirit, the true average connectivity can be expressed as  $c = y^* u'(y^*) + 2 - 2 \frac{u(y^*)}{y^* u'(y^*)}$ .

In general, if  $u$  has nonnegative coefficients and Eq. (10) has a solution strictly within the disc of convergence, one can go through the above argument in the reverse order to prove the existence of a percolation transition with the announced characteristics. This is the case for instance if  $u$  is an entire function with nonnegative coefficients, or more generally if  $u$  is function with nonnegative coefficients such that  $u''$  is unbounded when the argument approaches the radius of convergence. It is worth to observe that if the  $y$  expansion of  $u$  has nonnegative coefficients, then the same is true of the  $x$  expansion of  $w$ . Indeed, from Eq. (7) and the corresponding saddle point equation we infer that  $xw'(x) = \hat{z}$ . Hence as functions of  $x$ ,  $w$ , and  $\hat{z}$  have the same singular points, and

$$w'(x) = e^{xw'(x)} u'(xe^{xw'(x)}).$$

Expand both sides of this identity to see that  $w_1 = 1$  and that  $w_{k+1} - u_{k+1}$  is a polynomial in  $u_1 = 1, u_2, \dots, u_k, w_1 = 1, w_2, \dots, w_k$  with nonnegative coefficients.

<sup>3</sup> This should be the case in statistical mechanics, and it is true by construction for our initial model as long as the parameters are real.

In the case of our original model, the situation is more tricky. We know by construction that the  $x$  expansion of  $w$  has nonnegative coefficients, but to ensure the existence of a dilute regime, the same cannot be true in general of  $u$ . In the sequel, we shall see that in perturbation theory at any finite order, we are in the following situation: the coefficients of  $u$  may be negative, but nevertheless  $u'(y)$  is analytic (in fact a polynomial) and positive in a interval strictly containing 0 and a solution of Eq. (10). Then our previous arguments can be made rigorous and there is a (perturbative) percolation transition with the announced characteristics. We do not know if this argument can be extended outside the realm of perturbation theory. The numerical simulations are encouraging, but the behaviour of some perturbative series is puzzling. Before discussing that, let us consider three simple but significant examples.

## 6. THREE EASY EXAMPLES

### 6.1. The Case of the Erdős–Renyi Model

Let us recover the Erdős–Renyi model in this framework. In that case, by construction,  $U(y) = u(y) = y$  and  $y^* = \alpha$ . Equation (5) leads to  $F = 0$  for all values of  $\alpha$  (no surprise, for the Erdős–Renyi model the weights are normalized as a probability distribution). Then Eq. (7) leads to  $\hat{z} = xe^{\hat{z}}$ , and from the Lagrange inversion formula,

$$\hat{z} = \sum_k \frac{k^{k-1}}{k!} x^k \quad w = \sum_k \frac{k^{k-2}}{k!} x^k$$

which are the classical (rooted and non-rooted) tree generating functions (in fact, this gives a proof of Cayley's formula for the number of trees). Note that if we use naively Eq. (6), we can deduce that  $F = 0$  only for  $\alpha \leq 1$ .

The number of connected components of size  $k$  is  $N_k \sim N \frac{k^{k-2}}{k!} \alpha^{k-1} e^{-k\alpha}$ , which is well-known to be true for fixed  $k$  and large  $N$ , for any value of  $\alpha$ . Notice again that the use of  $u$  plays a crucial role in our approach. Using only  $w$ , we would get the component distribution only for  $\alpha \leq 1$ . In fact, for the corresponding random forest model (which is thermodynamically equivalent to the random graph model for  $\alpha \leq 1$ )  $\lim_{N \rightarrow \infty} N_k/N$  is  $\frac{k^{k-2}}{k!} \alpha^{k-1} e^{-k\alpha}$  for  $\alpha \leq 1$  but is nonanalytic at  $\alpha = 1$ , which is the percolation transition.

The total number of points belonging to components of size  $k$  is  $\sim N \frac{k^{k-1}}{k!} \alpha^{k-1} e^{-k\alpha}$ .

For  $\alpha \leq 1$ ,  $\sum_k \frac{k^{k-1}}{k!} \alpha^k e^{-k\alpha} = \alpha$ , but for  $\alpha > 1$ ,  $\sum_k \frac{k^{k-1}}{k!} \alpha^k e^{-k\alpha} = \bar{\alpha}$ , where  $\bar{\alpha}$  is the smallest solution to  $\alpha e^{-\alpha} = \bar{\alpha} e^{-\bar{\alpha}}$ . The giant component occupies  $\sim N(1 - \bar{\alpha}/\alpha)$  sites.

## 6.2. The Nested Erdős–Renyi Model

As another example, suppose that  $u^{(\text{eff})}(H) = p^{E(H)}$  for all graphs, i.e., that  $c_k = 1$ ,  $k \geq 1$ . Then  $w^{(\text{eff})}(G) = q^{E(G)} \sum_{H, E(H) \subset E(G)} 1 = (2q)^{E(H)}$ . Both weights describe the Erdős–Renyi model, but with different values for the probability of an edge. Going to the large  $N$  finite connectivity limit, we find  $u = \sum_{k \geq 1} \frac{k^{k-2}}{k!} y^k$ , and from our previous analysis of the Erdős–Renyi model, we find that  $y u'(y) = \sum_{k \geq 1} \frac{k^{k-1}}{k!} y^k$  is the Lambert function  $L(y)$ , the solution of  $L(y) e^{-L(y)} = y$  analytic close to 0 and vanishing at 0. Hence  $\hat{z} = L(xe^{\hat{z}})$ , so that  $\hat{z} e^{-\hat{z}} = x e^{\hat{z}}$ . Hence  $2\hat{z} = L(2x)$ . Moreover, from  $u(y) = L(y) - \frac{L(y)^2}{2}$  we find  $2w^{(\text{eff})}(x) = L(2x) - \frac{L(2x)^2}{2}$ . So we recover the doubling of the edge probability when passing from the  $u^{(\text{eff})}$  weight to the  $w^{(\text{eff})}$ . The  $u^{(\text{eff})}$  percolation transition is at  $\alpha = 1$  but the  $w^{(\text{eff})}$  percolation transition occurs at  $\alpha = 1/2$ . Note that the equation  $y^* u'(y^*) = L(y) = \alpha$  cannot be solved for  $\alpha \geq 1$ , but that the free energy  $F = \alpha/2$  and the true connectivity  $c = 2\alpha$  have an analytic continuation for larger  $\alpha$ 's. That this analytic continuation is the true value of  $F$  cannot in principle be decided from our arguments (we would have to do one more step of the same construction to view the  $u^{(\text{eff})}$  weight itself as a composite weight). But this does not prevent us from finding and analysing correctly the  $w^{(\text{eff})}$  transition, because it occurs strictly before the  $u^{(\text{eff})}$  transition.

## 6.3. The Matching Model

When  $u_k = 0$  for  $k \geq 2$  we recover the Erdős–Renyi model, so let us try the next degree of difficulty, when  $u_k = 0$  for  $k \geq 3$  but  $u_2$  is a free parameter. Thus  $w^{(\text{eff})}(G)$  is the generating function for a gas of disjoint edges on  $G$ , that is, the generating function for (all, non-necessarily maximal) matchings on  $G$ . This is a rather natural weight from the point of view of combinatorics. It is plain that the detailed structure of  $G$  is relevant, and not simply the size of its connected components. On the other hand, the  $u^{(\text{eff})}$  weight is nonzero only for a finite number of connected graphs, so that the function  $U^{(\text{eff})}$  is simply  $U^{(\text{eff})} = z + qu_2 \frac{z^2}{2}$  and  $Z_N = N! \oint \frac{dz}{z^{N+1}} e^{z + qu_2 \frac{z^2}{2}}$ . In such a simple case, the saddle point approximation applies without subtleties, and we retrieve, in the large  $N$  finite connectivity limit, the expected equations. The function  $w(x)$  does not seem to be an elementary function. The small  $x$  and the perturbative small  $u_2$  expansions are



straightforward but become quickly ugly. However from  $u = y + u_2 \frac{y^2}{2}$ , we can easily find the percolation criterion. Parametrizing  $u_2 = \frac{1-y_c}{2y_c^2}$  (with  $y_c \in ]0, 1]$  for positive  $u_2$ ) and using Eq. (10), one finds that at the percolation threshold:

$$y^* = y_c \quad \alpha_c = \frac{1+y_c}{2} \quad c_{\text{perc}} = \frac{1+y_c}{2} + \frac{1-y_c}{1+y_c}.$$

So  $\alpha_c$  decreases from 1 to 1/2 when  $u_2$  grows, but the physical average connectivity  $c_{\text{perc}}$  increases from 1 to 3/2. The special case  $u_2 = 1$  is of special combinatorial significance, because the weight  $w^{(\text{eff})}(G)$  counts the number of configurations of nonadjacent edges on  $G$ . Then  $y_c = 1/2$ ,  $\alpha_c = 3/4$  and  $c_{\text{perc}} = 13/12$ . Consequently,  $x_c = \frac{1}{2} e^{-3/4}$ , from which we can derive a result of direct combinatorial significance:

$$\frac{1}{N!} \sum_{T \in \mathcal{T}_N} \# \text{ matchings of } T \sim C^{st} \frac{(2e^{3/4})^N}{N^{5/2}},$$

to be compared with  $\frac{1}{N!} \sum_{T \in \mathcal{T}_N} 1 = \frac{N^{N-2}}{N!} \sim \frac{1}{\sqrt{2\pi}} \frac{e^N}{N^{3/2}}$ . Hence, if we put the uniform probability law on labelled trees of size  $N$ , the average number of matchings on a random tree of size  $N$  behaves like  $C^{st} (\frac{16}{e})^{N/4}$ .

## 7. BACK TO THE ORIGINAL MODEL

### 7.1. Finite Orders in Perturbation Theory

Remember that we established in Section 4.3 that

$$U(z) = z + \sum_H^c \frac{1}{s(H)} p^{e(H)} z^{v(H)} \prod_k t_k^{m_k(H)} s_k^{n_k(H)},$$

where  $\sum_H^c$  is the sum over unlabelled connected admissible graphs (we could equivalently reason in terms of normalized connected sequences). Consider the coefficient of  $\prod_k t_k^{m_k} s_k^{n_k}$ : it is the sum over admissible graphs with  $m_k$  solid  $k$ -cycles and  $n_k$  solid segments on  $k+1$  vertices. There is only a finite number of ways to join these fixed solid components with any number of complete dashed graphs. So the coefficient of  $\prod_k t_k^{m_k} s_k^{n_k}$  is a polynomial in  $p$  and  $z$ . A fortiori, if we restrict to admissible graphs  $H$  such that  $v = l+1$ , which are the ones contributing to  $u$ , the sum is finite, and the coefficient of  $\prod_k t_k^{m_k} s_k^{n_k}$  in the perturbative expansion of  $u(y)$  is a polynomial in  $y$ .

Note that  $u(y) = y + O(y^2)$ , so  $yu'(y) = y + O(y^2)$  and  $yu'(y) + y^2u''(y) = y + O(y^2)$ , where the  $O(y^2)$  vanish to zeroth order in perturbation

theory. Hence to any finite order in perturbation theory,  $yu'(y)$  is analytic and increasing up in a large value of  $y$ , but  $yu'(y) + y^2u''(y) = 1$ , the signal of the percolation transition, occurs at a value of  $y$  of order 1. Hence generically to any finite order in perturbation theory our initial model exhibits a percolation transition described by our previous results. In the following we shall make explicit perturbative computations of the free energy, the percolation threshold, etc, for the special case  $t_k = 0$ ,  $s_k = 2\mu\delta_{k,3}$ . To compare with the prediction of the Molloy–Reed criterion, we need first to show how to compute it in perturbation theory for our model.

## 7.2. Moments of the Degree Distribution, Molloy–Reed’s Criterion

By degree distribution of a given labelled graph  $G$  on  $N$  vertices is meant the sequence  $(n_0(G), n_1(G), \dots, n_{N-1}(G))$  where  $n_i(G)$  is the number of vertices in  $G$  with exactly  $i$  neighbours. For fixed  $N$ , the Molloy–Reed model concentrates on the set of all those labelled graphs with a fixed degree distribution  $(n_0, n_1, \dots)$  and gives them uniform probability, see ref. 6. This represents a microcanonical point of view in the sense that the degree distribution is fixed and can not fluctuate. For a grand canonical presentation of the same idea, see ref. 2. If, for large  $N$ ,  $(n_0/N, n_1/N, \dots)$  converges (in a sense made precise by Molloy and Reed) to a probability distribution  $(f_0, f_1, \dots)$ , a limiting random graph model is obtained, which depends only on  $(f_0, f_1, \dots)$  and not on the details of the approximating sequence  $(n_0/N, n_1/N, \dots)$ . We now recall the percolation criterion for the Molloy–Reed model with arbitrary degree distribution.

For a given graph  $G$ , define  $k^{(q)}(G)$  as the following average over vertices of  $G$ :

$$k^{(q)}(G) \equiv \frac{1}{N} \sum_{i=1}^N l_i(G)^q.$$

For instance, when  $q = 1$ ,  $Nk(G) = 2E(G)$ .

The statistical average  $\langle k^q \rangle \equiv \frac{1}{\sum_G w(G)} \sum_G w(G) k^{(q)}(G)$  is called the  $q$ th moment of the degree distribution. Note that in the Molloy–Reed model, all graphs have the same degree distribution, so that  $\langle k^q \rangle = k^{(q)}(G)$  for all  $G$  in the relevant statistical ensemble.

The Molloy–Reed percolation criterion states that the Molloy–Reed random graph has a giant component if and only if the two first moments of the degree distribution verify  $\langle k^2 - 2k \rangle > 0$ . For the Erdős–Renyi model,  $\langle k \rangle = \alpha$ ,  $\langle k^2 \rangle = \alpha(\alpha + 1)$ , leading to the percolation threshold  $\alpha = 1$ .

Our present purpose is to compute in perturbation theory the first moments of the degree distribution for our model. In principle, it is possible to compute  $\langle k^q \rangle$  for any  $q \geq 1$ . In the definition

$$\langle k^q \rangle = \frac{1}{NZ_N} \sum_{G \in \mathcal{G}_N} e^{S_I(G)} w_0(G) \sum_{i, j_1, \dots, j_q} a_{ij_1} \cdots a_{ij_q}$$

of the  $q$ th moment,  $e^{S_I(G)} \sum_{i, j_1, \dots, j_q} a_{ij_1} \cdots a_{ij_q}$  may be viewed as the  $x$  derivative taken at  $x = 0$  of

$$\exp(S_I^{(q)}(G, x)) \equiv \exp\left(S_I(G) + x \sum_{i, j_1, \dots, j_q} a_{ij_1} \cdots a_{ij_q}\right).$$

Seen as a new term of interaction, this exponential is still multiplicative and permutation invariant. We thus follow the steps which led us to Eq. (2) (see Section 3.2) to prove that  $\langle k^q \rangle$  is the derivative taken at  $x = 0$  of

$$\frac{(N-1)!}{Z_N} \oint \frac{dz}{z^{N+1}} e^{U^{(q)}(z)} = \frac{1}{N} \frac{Z_N^{(q)}}{Z_N},$$

where  $Z_N^{(q)}$  is the partition function of the model obtained from the original model by replacing  $S_I$  by  $S_I^{(q)}$ . In the large  $N$  limit, we proceed just as in Section 5.2 to show that  $Z_N^{(q)} = e^{NF_q + o(N)}$ ,  $F_q$  being the new free energy:  $F_q = -1 - \log \frac{y_q^*}{\alpha} + \frac{1}{\alpha} u^{(q)}(y_q^*)$ . In this expression,  $u^{(q)}$  is the tree generating function for the new model and  $y_q^*$  is the corresponding saddle point.

We now take the derivative and put  $x = 0$  to yield

$$\langle k^q \rangle = \frac{1}{\alpha} \sum_T \sum_{i, j_1, \dots, j_q} \frac{e^{S_I(T)} a_{ij_1} \cdots a_{ij_q}}{V(T)!} y^{V(T)}.$$

Just as we did in the original model, we can use normalized sequences (or admissible graphs) to give a combinatorial interpretation of the overlined term. A sequence  $i_1 j_1 \cdots i_n j_n$  is said to be normalized with respect to  $\{m_k, n_k\}$ ,  $q$  if

- $n = q + \sum_k k(m_k + n_k)$ ,
- 1 comes before 2, which comes before 3, ... which comes before the number  $v$  of distinct elements among the sequence,
- $i_1 \neq j_1, \dots, i_n \neq j_n$ ,
- it has a correct structure. That is, the sequence of the  $2(n-q)$  first terms has a correct structure as regards  $\text{Tr}$  and  $\| \|$  and, moreover,  $i_n = i_{n-1} = \cdots = i_{n-q+1}$ .

We put  $\mathcal{M}_{v,l,(m_k,n_k),q}$  for the number of such sequences. Finally, the  $q$ th moment of the degree distribution is

$$\langle k^q \rangle = \frac{y^*}{\alpha} + \frac{1}{\alpha} \sum_{v, \{m_*, n_*\}} \frac{\tilde{t}^{m_*} \tilde{s}^{n_*}}{m_*! n_*!} \mathcal{M}_{v, (m_k, n_k), q}^t y^{*v}. \quad (11)$$

In particular, the Molloy–Reed's criterion can, in principle, be computed by means of this formula: it involves normalized sequences (of type  $(m_k, n_k)$ ) to which are concatenated subsequences of 2 elements for  $\langle k \rangle$  or 3 elements for  $\langle k^2 \rangle$ .

We now study a simple example in which all quantities mentioned above can be explicitly (although perturbatively) computed.

### 7.3. Perturbation Theory: The Example $t_k = 0$ , $s_k = 2\mu\delta_{k,3}$

This is the simplest nontrivial case for which  $s_k \neq 0$ . The weight of a graph  $G$  is

$$w(G) = p^{E(G)} (1-p)^{-E(G)} e^{\mu \sum_{ijkl} a_{ij} a_{jk} a_{kl}}.$$

According to our previous discussion, we assume that  $\mu < 0$ . To get the percolation criterion, we use the general theory exposed in the preceding section. We set  $\mathcal{M}_{n,m}^t \equiv \mathcal{M}_{n,n-1,m}^c$ . The definition of  $y^*$  is  $y^* u'(y^*) = \alpha$ , i.e.,

$$y^* + \sum_{n,m} n y^{*n} \frac{\mu^m}{m!} \mathcal{M}_{n,m}^t = \alpha.$$

In fact, we were not able to find an expression of  $\mathcal{M}_{n,m}^t$  valid for all  $n, m$ , and we rely on a direct enumeration, up to order 6, of the normalized sequences, see Table I for the first five orders. At this moment, a fully automated enumeration algorithm starting from scratch and working in a reasonable time would need too much memory. To have some control over possible errors coming from human input, we have checked our results with two independent algorithms. On a 2.5 Ghz processor, the computation of the fifth order takes about 5 minutes, but the sixth order takes about 8 hours: the growth in complexity is extremely rapid, at least factorial.

Up to third order in  $\mu$ ,  $y^*$  takes the following form:

$$\begin{aligned} y^* = & \alpha - 2\alpha^2(\alpha+1)(2\alpha+1)\mu + 2\alpha^2(-1-17\alpha-56\alpha^2-57\alpha^3 \\ & - 15\alpha^4 + 4\alpha^5)\mu^2 - \frac{4}{3}\alpha^2(1+81\alpha+788\alpha^2+2485\alpha^3+3303\alpha^4 \\ & + 1808\alpha^5+159\alpha^6-126\alpha^7+8\alpha^8)\mu^3 + \dots \end{aligned}$$

Table I.  $\mathcal{M}_{n,m}^t$  for  $m = 1, \dots, 5$

$n \setminus m$	1	2	3	4	5
2	1	2	4	8	16
3	2	28	248	2032	16352
4	1	86	2236	44024	789616
5		108	7720	316784	10603040
6		66	14120	1152952	66713920
7		16	15424	2558624	248562304
8			10284	3781264	619455952
9			3888	3851664	1101864640
10			640	2698504	1444605680
11				1249712	1410932864
12				345600	1019814768
13				43264	531798240
14					189678720
15					41472000
16					4194304

The percolation criterion states that there exists a giant connected component if  $y^*u'(y^*) + y^{*2}u''(y^*) > 1$  and that, on the contrary, all connected components are of finite size if  $y^*u'(y^*) + y^{*2}u''(y^*) < 1$ . With  $u(y) = y + \sum_{n,m} \mathcal{M}_{n,m}^t \frac{\mu^m}{m!} y^n$ , the boundary between the percolating region and the non-percolating region is a curve in the  $(\alpha, \mu)$  plane, of equation:

$$\alpha - 1 + \sum_{n,m} n(n-1) y^{*n} \frac{\mu^m}{m!} \mathcal{M}_{n,m}^t = 0. \tag{12}$$

We can solve this equation for  $\alpha$  as a perturbative series in  $\mu$ . Up to order 5, this yields

$$\alpha_{\text{perc}} = 1 - 26\mu + 336\mu^2 - \frac{9500}{3}\mu^3 + \frac{49718}{3}\mu^4 - \frac{991328}{5}\mu^5 - \frac{41436164}{15}\mu^6 + \dots \tag{13}$$

Putting  $\alpha = \alpha_{\text{perc}}$  in formula (8) we find that, at the percolation threshold, the mean number of neighbours of a given vertex is

$$c_{\text{perc}} = 1 - 10\mu - 50\mu^2 - \frac{652}{3}\mu^3 - \frac{19786}{3}\mu^4 - \frac{3498268}{15}\mu^5 - \frac{67025012}{9}\mu^6 + \dots \tag{14}$$

In the preceding section, we saw how to infer the moments of the degree distribution from enumeration of the appropriate normalized sequences. Tables II and III show the result of these enumerations for  $\langle k \rangle$  and  $\langle k^2 \rangle$ .

Table II. Enumeration of the Sequences Appearing in  $\langle k \rangle$ 

$n \setminus m$	1	2	3	4	5
2	2	4	8	16	32
3	12	120	1008	8160	65472
4	18	692	14952	276560	4836768
5	8	1600	80800	2902784	91337088
6		1844	225648	14935280	779078400
7		1080	375408	45982304	3849121728
8		256	392360	93526304	12533947744
9			255312	131789760	28896796992
10			95040	130610064	49053023200
11			15488	89956640	62460050560
12				41178240	59854882464
13				11291904	42704264192
14				1404928	22060944640
15					7812720000
16					1698693120
17					171051008

Table III. Enumeration of the Sequences Appearing in  $\langle k^2 \rangle$ 

$n \setminus m$	1	2	3	4	5
2	2	4	8	16	32
3	20	184	1520	12256	98240
4	44	1336	27440	500320	8725184
5	38	3812	171208	5937552	184842528
6	12	5676	546752	33681040	1713610432
7		4804	1060024	113992144	9088370528
8		2212	1341416	257520720	31755109024
9		432	1127280	410985696	79109699392
10			611232	474725904	146874463968
11			194480	397440176	207952308800
12			27648	236315376	226475616384
13				94941392	189140564736
14				23156224	119320803648
15				2592000	55138687200
16					17635164160
17					3491452928
18					322486272

We compute  $\langle k \rangle$  (either by means of formula (8) or using the enumeration [2] together with Eq. (11)) and  $\langle k^2 \rangle$  as perturbative series in  $\mu$ , and then solve the equation  $\langle k^2 - 2k \rangle = 0$  in  $\alpha$  to find

$$\alpha_{\text{MR}} = 1 - 24\mu + 274\mu^2 - \frac{7324}{3}\mu^3 + \frac{28708}{3}\mu^4 - \frac{577988}{3}\mu^5 + \dots,$$

which does not coincide with  $\alpha_{\text{perc}}$ .

## 8. DISCUSSION AND PERSPECTIVES

In this paper, we have studied a class of perturbations of the Erdős–Renyi model which introduce correlations between the edges: the weight of a graph depends on the abundance of certain geometric features.

To solve this model, we have introduced an auxiliary model whose tree generating function  $u$  was expected to present better convergence properties than the original one  $w$ . The free energy  $F$  in the large  $N$  limit has been determined and a percolation transition has been established by means of an effective model: the percolation criterion is given by an equation, either on  $F$  or on  $u$ . We also have formulæ for the degree distributions.

On the basis of these general results we give explicit formulæ for the above quantities in the particular case where all parameters but one vanish. These perturbative results raise some crucial questions. Indeed, we hope that the thermodynamical model makes sense for  $\mu < 0$  but that  $\mu > 0$  has to be discarded because it gives too much weight to strongly connected configurations and cannot be treated like a diluted, tree-like, regime. In fact, up to sixth order, it is not so clear that the series for  $\alpha_{\text{perc}}$  is actually convergent for negative  $\mu$ , because its general term increases very fast. However, as suggested by the fifth and sixth terms, we hope that the following terms may all be negative, the series hence being possibly summable when  $\mu < 0$ . This interpretation is supported by the form of the perturbative expansion (14) of the physical connectivity parameter  $c_{\text{perc}}$ , which seems much better behaved, with negative coefficients for orders  $> 0$ .

We also have computed the Molloy–Reed criterion, which does not give an appropriate description of the percolation transition in this model. The  $\mu$  expansion of  $\alpha_{\text{MR}}$  seems to present the same pathology as  $\alpha_{\text{perc}}$ . A possibility is that this series is indeed divergent for negative  $\mu$ : the equation  $\langle k^2 - 2k \rangle$  may not admit any solution in  $\alpha$  as soon as  $\mu < 0$ . Another possibility is that, just as for  $\alpha_{\text{perc}}$  the series may stop to alternate at higher orders. Anyway, it would be desirable to determine a class of models for which the Molloy–Reed criterion is valid, and we believe that a minimal requirement may be a kind of locality. Indeed, the Molloy–Reed criterion

concentrates on the first two moments of the degree distribution, which are local quantities in the sense that  $k^{(q)}(G)$  can be computed as soon as the immediate environment of each vertex is known, independently of how the vertices are connected to each other. Even in the simple model that we used to illustrate perturbation theory, this information is not sufficient to compute the weight of a graph: one must also know the immediate environment of the first neighbours of each vertex.

Finally, we also believe that a more thorough understanding of degree correlations induced by attacks deserves a systematic treatment.

## REFERENCES

1. R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Modern Phys.* **74**:47 (2002).
2. M. Bauer and D. Bernard, Maximal entropy random networks with given degree distribution, ArXiv: cond-mat/0206150.
3. M. Bauer and O. Golinelli, Random incidence matrices: Moments of the spectral density, ArXiv: cond-mat/0007127.
4. S. N. Dorogovtsev and J. F. F. Mendes, Evolution of networks, *Adv. Phys.* **51**:1079 (2002).
5. P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5**:17–61 (1960).
6. M. Molloy and B. Reed, Statistical mechanics of complex networks, *Random Struct. Algorithms* **6**:161 (1995); *Comb. Proba. Comput.* **7**:295 (1998).
7. D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Network robustness and fragility: Percolation on random graphs, *Phys. Rev. Lett.* **85**:5468–5471 (2000).
8. R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, Breakdown of the internet under intentional attack, *Phys. Rev. Lett.* **86**:3682 (2001).
9. Z. Burda, J. Jurkiewicz, and A. Krzywicki, Network transitivity and matrix models, ArXiv: cond-mat/0310234.
10. O. Frank and D. Strauss, Markov graphs, *J. Amer. Stat. Assoc.* **81**:832–842 (1986).
11. J. Besag, Spatial interaction and the statistical analysis of lattice systems, *J. Roy. Stat. Soc. Ser. B* **36**:192–236 (1974).
12. J. Berg and M. Lässig, Correlated random networks, *Phys. Rev. Lett.* **89**:228701 (2002).