On the number of circuits in random graphs

Enzo Marinari\textsuperscript{1} and Guilhem Semerjian\textsuperscript{2}

\textsuperscript{1} Dipartimento di Fisica and INFN, Università di Roma La Sapienza, Piazzale Aldo Moro 2, 00185 Roma, Italy
\textsuperscript{2} Dipartimento di Fisica and CNR, Università di Roma La Sapienza, Piazzale Aldo Moro 2, 00185 Roma, Italy
E-mail: enzo.marinari@roma1.infn.it and guilhem.semerjian@lpt.ens.fr

Received 27 March 2006
Accepted 6 June 2006
Published 29 June 2006

Abstract. We apply in this paper (non-rigorous) statistical mechanics methods to the problem of counting long circuits in graphs. The outcomes of this approach have two complementary flavours. On the algorithmic side, we propose an approximate counting procedure, valid in principle for a large class of graphs. On a more theoretical side, we study the typical number of long circuits in random graph ensembles, reproducing rigorously known results and stating new conjectures.

Keywords: cavity and replica method, message-passing algorithms, random graphs, networks

ArXiv ePrint: cond-mat/0603657
On the number of circuits in random graphs

Contents

1. Introduction ........................................... 3

2. A statistical mechanics model and its Bethe approximation .......................... 4
   2.1. Derivation of the BP equations .......................... 4
   2.2. An approximate counting algorithm .......................... 9

3. The typical number of circuits in random graphs ensembles .......................... 10
   3.1. Definitions .................................... 10
   3.2. The quenched computation ............................ 11

4. The limit of small circuits ................................ 14

5. The limit of longest circuits .............................. 16
   5.1. In the absence of degree two sites in the two-core .......................... 17
   5.2. In the presence of degree two sites in the two-core .......................... 18
       5.2.1. Bounds on $\ell_{\text{max}}$ ................................ 18
       5.2.2. The large $u$ limit in presence of sites of degree two ............. 20

6. Stability of the replica symmetric ansatz ........................................... 22

7. Exhaustive enumerations .................................. 25

8. Conclusions and perspectives ................................ 27

Acknowledgments ........................................... 28

Appendix A. Combinatorial approach .......................... 28
   A.1. Notation and definitions ............................. 28
   A.2. First moment computations ........................... 29
       A.2.1. Generalities .................................. 29
       A.2.2. Erdős–Rényi ensembles .......................... 30
       A.2.3. Arbitrary connectivity distribution and regular ................ 31
   A.3. Second moment computations ........................... 32
       A.3.1. Generalities .................................. 32
       A.3.2. Erdős–Rényi ensembles .......................... 34
       A.3.3. Arbitrary connectivity distribution and regular ................ 34
   A.4. On the union of vertex disjoint circuits .......................... 36

Appendix B. Analysis of the leaf removal algorithm on random graphs with arbitrary connectivity distribution .......................... 36

Appendix C. An alternative derivation of $\ell_{\text{max}} = 1$ when the minimal connectivity is three .......................... 39

References ........................................... 40

doi:10.1088/1742-5468/2006/06/P06019
1. Introduction

Random graphs [1, 2] appeared in the mathematical literature as a convenient tool to prove the existence of graphs with a certain property: instead of a direct constructive proof exhibiting such a graph, one can construct a random ensemble of graphs and show that this property is true with a positive probability. Soon afterwards, the study of random graphs acquired interest on its own and led to many beautiful mathematical results. A large class of problems in this field can be formulated in the following generic way: a graph $H$ being given, what is the probability that a graph $G$ extracted from the random ensemble under consideration contains $H$ as a subgraph? With a more quantitative ambition, one can define $N_H(G)$ as the random variable counting the number of occurrences of distinct copies of $H$ in $G$, and study its distribution. These problems are relatively simple when the pattern $H$ remains of a finite size in the thermodynamic limit, i.e. when the size of the random graph $G$ diverges. The situation can become much more involved when $H$ and $G$ have large sizes of the same order, as $N_H(G)$ can grow exponentially with the system size.

In this paper we shall consider these questions when the looked for subgraph $H$ is a long circuit (also called loop or cycle), i.e. a closed self-avoiding path visiting a finite fraction of the vertices of the graph. The level of accuracy of the rigorous results on this problem depends strongly on the random graph ensemble [1]–[3]. The regular case (when all vertices of the graph have the same degree $c$) is the best understood one. It has for instance been shown that $c$-regular random graphs with $c \geq 3$ have with high probability Hamiltonian circuits (circuits which visit all vertices of the graph) and the distribution of their numbers is known [4, 5]. This study has been generalized to circuits of all length in [6]. Less is known for the classical Erdős–Rényi ensembles, where the degree distribution of the vertices converges to a Poisson law of mean $c$. Most results concern either the neighbourhood of the percolation transition at $c = 1$ [7]–[9], or the opposite limit of very large mean connectivity, either finite with respect to the size of the graph [10] or diverging like its logarithm [11] (it is in this latter regime that the graphs become Hamiltonian). We shall repeatedly come back in the following on this discrepancy between regular random graphs where probabilistic methods have been proved so successful and the other ensembles for which they do not seem powerful enough and might be profitably complemented by approaches inspired by statistical mechanics. We will discuss in particular a conjecture formulated by Wormald [3], according to which random graph ensembles with a minimal degree of three (and bounded maximal degree) should be Hamiltonian with high probability.

Besides this probabilistic point of view (what are the characteristics of the random variable associated to the number of circuits?), the problem has also an algorithmic side: how to count the number of circuits in a given graph. Exhaustive enumeration, even using smart algorithms [12], is restricted to small graphs as the number of circuits grows exponentially with the size. More formally, the decision problem of knowing if a graph is Hamiltonian (i.e. that it contains a circuit visiting all vertices) is NP-complete [13]. A probabilistic algorithm for the approximate counting of Hamiltonian cycles is known [14], but is restricted to graphs with large minimal connectivity.

Random graphs have also been largely considered in the physics literature, mainly in the real-world networks perspective [15], i.e. in order to compare the characteristics of observed networks, of the Internet for instance, with those of proposed random models.
Empirical measures for short loops in real world graphs were for instance presented in [16]. Long circuits visiting a finite fraction of the vertices were also studied in [17]. The behaviour of cycles in the neighbourhood of the percolation transition was considered in [18], and the average number of circuits for arbitrary connectivity distribution was computed in [19].

In this paper we shall turn the counting problem into a statistical mechanics model, which we treat within the Bethe approximation. This will lead us to an approximate counting algorithm, cf section 2. We will then concentrate on random graph ensembles and compute the typical number of circuits with the cavity replica-symmetric method [20] in section 3. The next two sections will be devoted to the study of the limits of short and longest circuits, then we shall investigate the validity of the replica-symmetry assumption in section 6. We perform a comparison with exhaustive enumerations on small graphs in section 7 and draw our conclusions in section 8. Three appendices collect more technical computations. A short account of our results has been published in [21].

2. A statistical mechanics model and its Bethe approximation

2.1. Derivation of the BP equations

Let us consider a graph $G$ on $N$ vertices (also called sites in the following) $i = 1, \ldots, N$, with $M$ edges (or links) $l = 1, \ldots, M$. The notation $l = \langle ij \rangle$ shall mean that the edge $l$ joins the vertices $i$ and $j$. The degree, or connectivity, of a site is the number of links it belongs to. The graphs are assumed in the main part of the text to be simple, i.e. without edges from one vertex to itself or multiple edges between two vertices. We denote by $\partial_i$ the set of neighbours of the vertex $i$, and use the symbol $\setminus$ to subtract an element of a set: if $j$ is a neighbour of $i$, $\partial_i \setminus j$ will be the set of all neighbours of $i$ distinct of $j$. The same symbol $\partial_i$ will be used for the set of edges incident to the vertex $i$; the context will always clarify which of the two meanings is understood.

A circuit of length $L$ is an ordered set of $L$ different vertices, $(i_1, \ldots, i_L)$, such that $\langle i_n i_{n+1} \rangle$ is an edge of the graph for all $n \in [1, L - 1]$, as well as $\langle i_L i_1 \rangle$. Two circuits are distinct if they do not share the same set of edges (i.e. the starting point and the orientation of a tour along the vertices is not relevant), and we denote by $N_L(G)$ the number of distinct circuits of length $L$ in a graph $G$.

The degrees of freedom of our model are $M$ variables $S_l \in \{0, 1\}$ placed on the edges of the graph, with their global configuration called $\underline{S} = \{S_1, \ldots, S_M\}$. We shall also use $\underline{S}_i = \{S_l | l \in \partial_i\}$ for the configuration of the variables on the links around the vertex $i$. We introduce the following probability law on the space of configurations:

$$p(\underline{S}) = \frac{1}{Z(u)} w(\underline{S}), \quad w(\underline{S}) = \left( \prod_{l=1}^{M} \hat{w}_l(S_l) \right) \left( \prod_{i=1}^{N} w_i(\underline{S}_i) \right),$$

where $Z(u)$ is the normalization constant, and the weights $\hat{w}_l$, $w_i$ are given by

$$\hat{w}_l(S_l) = u^{S_l}, \quad w_i(\underline{S}_i) = \begin{cases} 1 & \text{if } \sum_{l \in \partial_i} S_l \in \{0, 2\} \\ 0 & \text{otherwise.} \end{cases}$$

By convention $w_i = 1$ if $\partial i = \emptyset$, that is to say if the vertex $i$ is isolated. The relevance of this model for the counting of circuits is unveiled by the following reasoning. Each
configuration \( \Sigma \) can be associated with a subgraph of \( G \), retaining only the edges \( l \) such that \( S_l = 1 \). The probability (with respect to law (1)) of such a subgraph is non-zero only if the retained edges form closed circuits (any site \( i \) is constrained by \( w_i \) to be surrounded by either zero or two edges of the subgraph), and in this case it is proportional to \( u^L \) with \( L \) the number of its edges. This implies that the normalization factor \( Z(u) \) is the generating function of the numbers \( \mathcal{N}_L(G) \),

\[
Z(u) = \sum_L u^L \mathcal{N}_L(G).
\]

A precision should be made at this point: we defined above a circuit as a self-avoiding closed path. From the weights on the configurations defined by equations (1), (2), \( \mathcal{N}_L(G) \) counts in fact the number of configurations made of possibly several vertex disjoint circuits, of total length \( L \). In the following we shall concentrate on the limit of large graph and of long circuits, and we expect the leading order behaviour of \( \mathcal{N}_L \) not to be affected by this subtlety (see appendix A.4 for a combinatorial argument in favour of this thesis), that will be kept understood from now on.\(^3\) Note also that [22] proposed a Monte Carlo Markov chain algorithm for the evaluation of such a partition function.

We are thus performing a canonical computation where the length of the circuits is allowed to fluctuate around a mean value fixed by the conjugate external parameter \( u \). In the thermodynamic limit \( N \to \infty \) the saddle-point method can be used to evaluate the sum (3). Defining \( f(u) = (1/N) \ln Z(u) \) and \( \sigma(\ell) = (1/N) \ln \mathcal{N}_{L=\ell N} \), where \( \ell = L/N \) is a reduced intensive length, one obtains

\[
f(u) = \max_{\ell} [\ell \ln u + \sigma(\ell)].
\]

In this limit the fluctuations of the intensive circuit length in the canonical ensemble vanish; \( \ell \) is concentrated around its mean value \( \ell(u) = uf'(u) \). The (concave hull of the) microcanonical entropy can thus be obtained from the canonical free energy (with a slight abuse of terminology we shall use this denomination for \( f(u) \)) by an inverse Legendre transform,

\[
\sigma(\ell) = \min_u [f(u) - \ell \ln u].
\]

We shall now use the Bethe approximation to obtain an estimation of the generating function \( Z(u) \). We sketch first the general strategy to derive Bethe approximations of statistical models (see [23] for a comprehensive discussion), before applying it to the present case.

Consider a (non-negative) weight function \( w(\Sigma) \) defined on a space of configurations \( \{ \Sigma \} \). The computation of the partition function, \( Z = \sum_{\Sigma} w(\Sigma) \), can be reformulated as an extremization problem. Indeed, the Gibbs functional free energy,

\[
F_{\text{Gibbs}}[p_\nu] = \sum_{\Sigma} p_\nu(\Sigma) \ln \left( \frac{p_\nu(\Sigma)}{w(\Sigma)} \right),
\]

\(^3\) The reader might think this problem would be solved by enlarging the space of the configurations \( S_i \) to a Potts-like spin, \( S_i \in \{0,1,\ldots,q\} \), with the weight \( w_i \) enforcing that either all variables around \( i \) are vanishing, or two are non-zero and of the same colour \( 1,\ldots,q \). In the bivariate generating function \( Z(u,q) \) is then conjugated to the number of disconnected circuits, and the limit \( q \to 0 \) should allow to eliminate configurations made of several disconnected circuits. However, the Bethe approximation of this model is pathological, and we shall not pursue this road here.
is minimal (in the space of normalized variational distributions) for \( p_v(\mathcal{S}) = p(\mathcal{S}) = w(\mathcal{S})/Z \), where it takes the value \((-\ln Z)\). In general, finding the minimum of this functional is not simpler than a direct computation of \( Z \); however, this formulation opens the way to variational approaches: the minimum of \( F_{\text{Gibbs}} \) in a restricted set of trial distributions \( p_v \), more easily parametrized than generic ones, yields an upper bound on \((-\ln Z)\). The simplest implementation of this idea is the mean-field approximation, in which the trial distributions are factorized, \( p_v(\mathcal{S}) = \prod_i p_i(S_i) \). A natural refinement consists in introducing correlations between neighbouring variables in the trial distributions. Consider for instance a weight function of the form (1), for arbitrary \( w_l \) and \( w_i \). One can easily show that if the underlying graph \( G \) were a tree the true probability distribution would be given by

\[
p(\mathcal{S}) = \left( \prod_{l=1}^{M} p_l(S_l) \right)^{-1} \left( \prod_{i=1}^{N} p_i(S_i) \right),
\]

where \( p_l \) and \( p_i \) are the exact marginals (for instance, \( p_l(S_l) = \sum_{\mathcal{S} \setminus S_l} p(\mathcal{S}) \)) of the law \( p \). When the graph is not a tree, this expression is not valid any more. The Bethe approximation consists however in assuming that trial probability distributions can be approximately written under this form even if the graph contains cycles. This yields the so-called Bethe free energy,

\[
F_{\text{Bethe}}[\{p_l\}, \{p_i\}] = \sum_{i=1}^{N} \sum_{\mathcal{S}_i} p_i(S_i) \ln \left( \frac{p_l(S_l)}{w_l(\mathcal{S}_l)} \right) - \sum_{i=1}^{M} \sum_{S_i} p_i(S_i) \ln (p_i(S_i) \hat{w}_i(S_i)).
\]

This free energy is to be minimized with respect to the approximate marginals \( p_l, p_i \), which have to respect two types of constraints.

- \( p_l \) and \( p_i \) are normalized.
- They are consistent, i.e. for each link \( l = (ij) \), one has

\[
p_l(S_l) = \sum_{\mathcal{S}_i \setminus S_l} p_i(S_i) = \sum_{\mathcal{S}_j \setminus S_l} p_j(S_j).
\]

This constrained minimization can be performed considering the \( \{p_l\}, \{p_i\} \) as independent, at the price of the introduction of Lagrange multipliers to enforce the conditions (9). It is well known that such a procedure amounts to looking for a fixed point of the corresponding belief propagation (BP) equations [23]. In this setting the Lagrange multipliers are interpreted as messages sent by variables to neighbouring constraints, and vice versa.

We would like to emphasize that the minimization of the Bethe free energy does not yield a rigorous bound on \( \ln Z \) when \( G \) is not a tree: in such a case a trial probability distribution of the form (7) should contain a normalizing prefactor, hence the Gibbs and Bethe free energies differ by a correcting term neglected in the Bethe approximation.

Let us now apply this formalism to the specific weights defined in equation (2). A peculiarity of \( w_l \) has to be kept in mind: it can be strictly vanishing when the geometrical constraint of having zero or two present edges around each vertex is not fulfilled. As a consequence, the approximate variational marginals \( p_i(S_i) \) have to respect this constraint,
On the number of circuits in random graphs

Figure 1. The messages involved in equation (13).

and vanish when \( w_i(S_i) = 0 \). The Bethe free energy now reads

\[
F_{\text{Bethe}}[\{p_i\}, \{p_l\}] = \sum_{i=1}^{N} \sum_{S_i} p_i(S_i) \ln(p_i(S_i)) - \sum_{l=1}^{M} \sum_{S_l} p_l(S_l) \ln(p_l(S_l)u^{S_l}), \tag{10}
\]

where the convention \( 0 \ln 0 = 0 \) has been used, for the strictly forbidden configurations with \( w_i(S_i) = 0 \) not to contribute to \( F_{\text{Bethe}} \). The constraints of normalization and of consistency of the marginals can be enforced by adding to this expression the adequate Lagrange multipliers,

\[
\sum_{l} \lambda_l \sum_{S_l} p_l(S_l) + \sum_{i} \lambda_i \sum_{S_i} p_i(S_i) + \sum_{l=(ij)} \left\{ \sum_{S_l} \lambda_{l \rightarrow j}(S_l) \left[ p_l(S_l) - \sum_{S_j \backslash S_l} p_j(S_j) \right] \right\} = 0,
\]

(11)

After a reparametrization of the Lagrange multipliers one obtains for the solution of this constrained extremization of the Bethe free energy:

\[
p_l(S_l) = \frac{1}{C_l} (u_{l \rightarrow j} y_{j \rightarrow i})^{S_l}, \quad p_i(S_i) = \frac{1}{C_i} w_i(S_i) \prod_{j \in \partial i} (u_{j \rightarrow i})^{S(ij)}, \tag{12}
\]

where the \( C_l \) and \( C_i \) are normalization constants, and for each link \( \langle ij \rangle \) of the graph a pair of directed (real positive) messages has been introduced, \( y_{i \rightarrow j} \) and \( y_{j \rightarrow i} \). These messages obey the following BP equations:

\[
y_{i \rightarrow j} = \frac{u \sum_{k \in \partial i \backslash j} y_{k \rightarrow i}}{1 + (1/2)u^2 \sum_{k,k' \in \partial i \backslash j} y_{k \rightarrow i} y_{k' \rightarrow i}}, \tag{13}
\]

cf figure 1 for a graphical representation. Roughly speaking, \( y_{i \rightarrow j} \) is proportional to the probability that the edge \( \langle ij \rangle \) would be present if the constraint \( w_j \) and the weight \( u^{S_i} \) were to be discarded. Hence the form of equation (13): the numerator corresponds to the situation where \( \langle ij \rangle \) is present; the constraint \( w_i \) then imposes that exactly one of the edges of \( \partial i \backslash j \) is also present. The denominator states on the contrary that if \( \langle ij \rangle \) is absent either none or two of the edges of \( \partial i \backslash j \) are present.
On the number of circuits in random graphs

The normalization constants of the marginals are easily computed,

\[ C_l = 1 + u y_{i \rightarrow j} y_{j \rightarrow i}, \quad C_i = 1 + \frac{1}{2} u^2 \sum_{k,k' \in \partial_i} y_{k \rightarrow i} y_{k' \rightarrow i}, \]  

(14)

and one can check, using the BP equations, that the consistency conditions are indeed respected by these expressions of the marginals. Moreover the value of \( F_{\text{Bethe}} \) at its minimum can be expressed in terms of the normalization constants \( C_i \) and \( C_l \). Using this value as an approximation for \(-\ln Z(u)\), the free energy in the Bethe approximation can be written as

\[ N f(u) = - \sum_{i=1}^M \ln(C_i) + \sum_{i=1}^N \ln(C_i). \]  

(15)

One should also compute the length of the circuits in the configurations selected at a given value of \( u \), \( \ell(u) = uf'(u) \). It is rather unwise to use equation (15) to compute the derivative \( f'(u) \), as this expression involves the messages which are solutions of the BP equations and hence have a non-trivial dependence on \( u \). On the contrary, the expression (10) being variational, it is enough to compute its explicit derivative with respects to \( u \) to obtain

\[ \ell(u) = \frac{1}{N} \sum_{l=1}^M p_l(1) = \frac{1}{N} \sum_{\langle ij \rangle} \frac{u y_{i \rightarrow j} y_{j \rightarrow i}}{1 + u y_{i \rightarrow j} y_{j \rightarrow i}}. \]  

(16)

The first equality is natural, the average length of the circuits being equal to the sum of the probabilities of presence of all the edges of the graph. Note also that the marginal probabilities contain individually some local information: for instance, \( p_l(1) \) is the fraction of circuits of length \( \ell(u) \) which go through the particular link \( l \).

Let us now come back for an instant to the BP equations and underline two simple properties they possess. In equation (13) we used the natural convention that sums on empty sets are null. The first consequence is that \( y_{i \rightarrow j} = 0 \) if \( j \) is the only neighbour of \( i \), as \( \partial i \setminus j = \emptyset \). In such a situation \( i \) is indeed a leaf of the graph, and no circuits can go through the edge \( \langle ij \rangle \). Even if in general the directed message in the reverse direction \( y_{j \rightarrow i} \) is non-zero, one can easily check that the edge \( \langle ij \rangle \) does not contribute to the free energy. In other words the physical observables are unaffected by the leaf removal process, in which the graph \( G \) is deprived of the dangling edge \( \langle ij \rangle \). Moreover this simplification can be iteratively repeated, until no leaves are present in the remaining graph. An illustration of this process in terms of the null directed messages is given in the left part of figure 2. This property of the BP equations reflects the fact that the circuits of a graph are necessarily part of its two-core, that is to say the largest of its subgraphs in which all sites have connectivity at least two.

Consider now a site \( i \) with two neighbours \( j \) and \( k \), for which the BP equations read \( y_{i \rightarrow j} = uy_{k \rightarrow i} \) and \( y_{i \rightarrow k} = uy_{j \rightarrow i} \). This implies that along a chain of degree two sites the directed messages follow a geometric progression, of the right part of figure 2, and in consequence one easily shows that the marginal probabilities of all edges in a chain are equal: if a circuit goes through one of the edges of the chain, it must go through all of it.
2.2. An approximate counting algorithm

The presentation of the Bethe approximation in terms of messages [23] followed in the previous section suggests in a very natural way the following algorithm for the approximate counting of long circuits in a given graph.

- Initialize messages $y_{i \rightarrow j}$ for each directed edge of the graph to some random positive values.
- Iterate the BP equations (13) at a given value of $u$ until convergence has been reached.
- Using the message solution of the BP equations, compute $f(u), \ell(u)$ from equations (15), (16), and $\sigma(\ell(u)) = f(u) - \ell(u) \ln u$ (cf equation (5)).
- Repeat this procedure for different values of $u$ to obtain a plot of $\sigma(\ell)$ parametrized by $u$.

This algorithm is of course far from being exact. A first limitation is that the BP equations are not a priori convergent; on the contrary, it is easy to construct counterexamples of small graphs on which they do not reach any fixed point. It would thus be interesting to determine under which conditions the convergence towards an unique (non-trivial) fixed point is ensured. This kind of question has been the subject of recent interest; see for instance [24,25]. Another possible criticism is that even in the case of convergence of the BP equations the prediction for the number of loops relies on the Bethe approximation, which is an uncontrolled one. This being said, one should however keep in mind that for large graphs with numerous circuits, an exact enumeration [12] is computationally very expensive and reaches very soon the limitations of present day computers. The approximate algorithm we introduced here can then serve as an efficient alternative, even if its predictions should be treated with caution.

We presented in [21] the results of such a procedure when applied to a real-world network of the Autonomous System Level description of the Internet [26], allowing us to estimate the total number of circuits, the length of the most numerous circuits and the maximal length circuits, obtaining numbers which are far beyond the possibilities of exhaustive counting. We also checked the compatibility of our results with the direct enumeration of very short circuits.
3. The typical number of circuits in random graphs ensembles

3.1. Definitions

The rest of the paper shall be devoted to the study of the number of circuits in graphs $G$ belonging to random ensembles. In the regime we are interested in (long circuits of large graphs with finite mean degree), the common wisdom about the statistical mechanics of disordered systems is that the random variable $\log(N_L = N \ell)/N$ should be concentrated around its average, the quenched entropy $\sigma_q(\ell)$. More formally, one expects the existence of a constant $\ell_{\text{max}} \in [0, 1]$ and a function $\sigma_q(\ell) > 0$ defined on $]0, \ell_{\text{max}}]$ such that for any sequence $L_N$ with $L_N/N \to \ell$ if $\ell > \ell_{\text{max}}$, \[ \text{Prob}[N_{L_N} > 0] \to 0, \quad (17) \]
if $\ell \in ]0, \ell_{\text{max}}]$, $\forall \epsilon > 0$ \[ \text{Prob} \left[ \frac{1}{N} \log N_{L_N} - \sigma_q(\ell) \geq \epsilon \right] \to 0. \quad (18) \]

In the second line $\log(0)$ should be interpreted as $-\infty$, i.e. outside any finite interval.

The standard probabilistic methods for proving this kind of results rely essentially on the combinatorial computation of the average and variance of $N_{L_N}$, which are then used in the Markov and Chebyshev inequalities (first and second methods). The rigorous results on the number of circuits in regular random graphs [3]–[6] have indeed been obtained through a refined version of the second moment method (see theorem 4.1 in [3]). In this context this approach is limited to cases where the second moment of $N_{L_N}$ is not exponentially larger than the square of its first moment\(^4\). The quenched entropy is then shown to be equal to the annealed one, $\sigma_a(\ell) = \lim \log(N_{L_N})/N$, where the overline denotes the average over the random graph ensemble.

We believe that in all ensembles of graphs which are not strictly regular and have a fast decaying connectivity distribution the second moment of the number of long circuits is exponentially larger that the square of its first moment\(^4\). The quenched entropy is then shown to be equal to the annealed one, $\sigma_a(\ell) = \lim \log(N_{L_N})/N$, where the overline denotes the average over the random graph ensemble.

We shall now follow the cavity method [20], which is particularly well suited to tackle this problem, ubiquitous in the field of disordered statistical mechanics models [28]. According to this view, the quenched entropy controlling the leading behaviour of the number of circuits in the typical graphs depends on the graph ensemble only through its limiting degree distribution $q_k$.\(^5\) We can for instance assume that the graphs are drawn uniformly among all graphs on $N$ vertices which have this degree distribution. Let us recall the existence in this case of a percolation transition [29, 30] between a low connectivity regime, where the connected components of the graph are essentially trees of finite size, and a percolated phase, where one giant component contains a finite fraction of the vertices.

\(^4\) In a different problem, namely the random ensemble of $k$-satisfiability formulae, this limitation has been overcome by a weighted second moment method [27].

\(^5\) This is not true for the annealed entropy which depends on the ‘microscopic details’ of the ensemble. For instance, the two classical ensembles $G(N, p = c/N)$ and $G(N, M = N c/2)$ have the same Poisson degree distribution but distinct annealed entropies; see section 7 and appendix A.
Before proceeding with the computations, we introduce some notations used in the following. \( c = \sum_k kq_k \) denotes the mean connectivity of the graph, hence the number of edges is in the thermodynamic limit \( M = Nc/2 \). \( \tilde{q}_k \) will be the offspring probability, that is to say the probability of finding a site of degree \( k+1 \) when selecting at random an edge of the graph and then one of its two vertices. As a site is encountered in such a selection with a probability proportional to its degree, \( \tilde{q}_k \) is proportional to \((k+1)q_{k+1}\). By normalization,

\[
\tilde{q}_k = \frac{(k+1)q_{k+1}}{c},
\]

To simplify notations we shall also define the factorial moments of \( q_k \) and \( \tilde{q}_k \) as

\[
\mu_n = \sum_{k=n}^{\infty} q_k (k-1) \ldots (k-n+1), \quad \tilde{\mu}_n = \sum_{k=n}^{\infty} \tilde{q}_k (k-1) \ldots (k-n+1), \quad \mu_n = c\tilde{\mu}_{n-1},
\]

where the last relation is a simple consequence of equation (19).

The condition for percolation \([29,30]\) reads with these notations \( \tilde{\mu}_1 > 1 \). We shall assume in the following that this condition is met: the long circuits we are studying cannot be present if the graph has no giant component.

We restrict ourselves to fast (i.e. faster than any power law) decaying distributions of connectivities, such that all their moments are finite. After stating the results for arbitrary \( q_k \) we shall often specialize to Poissonian graphs of mean connectivity \( c \), i.e. such that \( q_k = e^{-c}c^k/k! \).

3.2. The quenched computation

In essence the computation of the quenched entropy we undertake now amounts to performing the Bethe approximation of the statistical model defined by equations (1), (2) for graphs generated according to the connectivity distribution \( q_k \). The solution of the BP equations (13), which depends on the particular graph on which they are applied, leads then to a random set of messages \( y \). Taking at random a graph of the ensemble, and a directed edge of this graph, one finds a message \( y \) with probability distribution \( P(y;u) \). In the so-called cavity method at the replica-symmetric level \([20]\) one assumes that the incoming messages on this directed edge are independent random variables with the same probability law \( P(y;u) \). Using equation (13), this is turned into a self-consistent equation,

\[
P(y;u) = \tilde{q}_0 \delta(y) + \sum_{k=1}^{\infty} \tilde{q}_k \int_0^{\infty} dy_1 P(y_1;u) \ldots dy_k P(y_k;u) \delta(y-g_k(y_1,\ldots,y_k)),
\]

where we have defined

\[
g_1(y_1) = uy_1, \quad g_k(y_1,\ldots,y_k) = \frac{u \sum_{i=1}^{k} y_i}{1 + u^2 \sum_{1 \leq i < j \leq k} y_i y_j} \quad \text{for } k \geq 2.
\]
On the number of circuits in random graphs

The quenched free energy is then expressed in terms of this $P(y; u)$ as (cf equation (15))

$$f_q(u) = \sum_{k=2}^{\infty} q_k \int_0^{\infty} dy_1 P(y_1; u) \ldots dy_k P(y_k; u) \ln \left( 1 + u^2 \sum_{1 \leq i < j \leq k} y_i y_j \right)$$

$$- \frac{c}{2} \int_0^{\infty} dy_1 P(y_1; u) dy_2 P(y_2; u) \ln (1 + u y_1 y_2).$$

(23)

In a similar way the length of the circuits in the configurations selected by a given value of $u$, and the corresponding quenched entropy, read

$$\ell(u) = \frac{c}{2} \int_0^{\infty} dy_1 P(y_1; u) dy_2 P(y_2; u) \frac{u y_1 y_2}{1 + u y_1 y_2},$$

(24)

$$\sigma_q(\ell(u)) = f_q(u) - \ell(u) \ln u.$$

(25)

As appears clearly when considering equation (21), the distribution $P(y; u)$ contains a Dirac’s delta in $y = 0$; that is to say, a finite fraction of the messages are strictly vanishing. Let us call $\eta$ the fraction of non-trivial messages, and $\hat{P}(y; u)$ their (normalized) distribution, i.e. $P(y; u) = (1 - \eta) \delta(y) + \eta \hat{P}(y; u)$ where $\hat{P}$ does not contain a Dirac’s delta in $y = 0$.

Inserting this definition in equation (21), one obtains the equation satisfied by $\eta$:

$$1 - \eta = \sum_{k=0}^{\infty} \hat{q}_k (1 - \eta)^k.$$

(26)

Besides the trivial solution $\eta = 0$, this equation has another positive solution as soon as $\tilde{\mu}_1 > 1$, i.e. when the graph is in the percolating regime. One also realizes that $\hat{P}$ satisfies the equation obtained from equation (21) by replacing the offspring distribution $\tilde{q}$ by $\tilde{r}$, defined as

$$\tilde{r}_0 = 0, \quad \tilde{r}_k = \sum_{n=k}^{\infty} \tilde{q}_n \binom{n}{k} \eta^{k-1} (1 - \eta)^{n-k} \quad \text{for } k \geq 1.$$

(27)

Finally the free energy and the typical length of circuits (cf equations (23), (24)) can also be expressed in terms of the simplified distribution $\hat{P}$, if one replaces $q$ by the following distribution $r$:

$$r_0 = 1 - \sum_{k=2}^{\infty} r_k, \quad r_1 = 0, \quad r_k = \sum_{n=k}^{\infty} q_n \binom{n}{k} \eta^{k-1} (1 - \eta)^{n-k} \quad \text{for } k \geq 2.$$

(28)

It is easily verified that this modified distribution has mean $c \eta^2$, and that a relation similar to equation (19) holds between $r$ and $\tilde{r}$:

$$\tilde{r}_k = \frac{(k + 1) r_{k+1}}{c \eta^2}.$$

(29)

---

$^6 \eta$ was denoted $1 - \zeta$ in [21].

doi:10.1088/1742-5468/2006/06/P06019
Let us now give the interpretation of this simplification process. We have shown the equality of the quenched entropy of the circuits in the two ensembles defined one by \( q_k \), the other by \( r_k \). As we explained at the end of section 2, the circuits of a graph \( G \) necessarily belong to its two-core, that is to say the largest subgraph of \( G \) in which all vertices have a degree at least equal to two. On the dangling ends, i.e. the edges that do not belong to the two-core, at least one of the two directed messages \( y \) is equal to zero. It is thus very natural to interpret the elimination of null messages in terms of the typical properties of the two-core of graphs drawn from the ensemble defined by the distribution \( q_k \). The fraction of edges in the two-core should be \( \eta^2 \), as both directed messages have to be non-zero for the edge to belong to the two-core; \( r_k \) (respectively \( \tilde{r}_k \)) should be the connectivity (respectively offspring) distribution of the two-core. This interpretation is indeed confirmed by a direct study of a leaf-removal algorithm which iteratively removes the dangling ends of a graph, that we present in appendix B. In the following we shall use the distribution \( q \) or \( r \), depending on which is simpler in the encountered context.

For future use we give the explicit expressions in the case of Poissonian random graphs,

\[
\eta = 1 - e^{-cn}, \quad r_k = \frac{e^{-cn}(cn)^k}{k!} \quad \text{for } k \geq 2, \quad \tilde{r}_k = \frac{1}{\eta} \frac{e^{-cn}(cn)^k}{k!} \quad \text{for } k \geq 1. \tag{30}
\]

We now come back to the predictions of the quenched entropy and consider as a first example the case of regular graphs of connectivity \( c \), for which \( \tilde{q}_k = \delta_{k,c-1} \). Equation (21) on the distribution of messages has then a very simple solution, \( P(y; u) = \delta(y - y_t(u, c)) \), with

\[
y_t(u, c) = \sqrt{\frac{2u(c - 1) - 2}{u^2(c - 1)(c - 2)}}. \tag{31}
\]

It is then straightforward to express \( \ell(u) \) and \( \sigma_q(\ell(u)) \) from this solution. One can also eliminate the parametrization by \( u \) to obtain the entropy,

\[
\sigma_q(\ell, c) = -(1 - \ell) \ln(1 - \ell) + \left( \frac{c}{2} - \ell \right) \ln \left( 1 - \frac{2\ell}{c} \right) + \ell \ln(c - 1), \tag{32}
\]

which corresponds to the known results mentioned above [6,31].

The peculiarity of the regular case for which annealed and quenched averages coincide is hinted at by the simplicity of this solution \( P(y; u) \) with a single Dirac peak. As soon as \( \tilde{q}_k \) is positive for more than one connectivity, the distribution \( P(y; u) \) acquires a non-vanishing support, which we expect to show up as larger fluctuations in the numbers of circuits, and hence a difference between quenched and annealed computations.

The equation on \( P(y; u) \) is not solvable analytically for an arbitrary connectivity distribution. Two complementary roads can then be followed: this distributional equation can be easily solved with a population dynamics algorithm [20]. One represents \( P(y; u) \) by a sample of a large number of \( y \); at each time step one draws a number \( k \) following the law \( \tilde{r}_k \), extracts \( k \) values \( y_1, \ldots, y_k \) randomly from the population, computes the new value \( g_k(y_1, \ldots, y_k) \), and replaces one of the representants of the population by this new value. Starting from a random sample of messages, the population converges to

\[ \text{doi:10.1088/1742-5468/2006/06/P06019} \]
Figure 3. Quenched entropy for a Poissonian graph of mean connectivity three.

4. The limit of small circuits

We shall investigate in this section the behaviour of the quenched entropy in the limit of small circuits, computing analytically its first two derivatives at the origin, $\sigma_q'(0)$ and $\sigma_q''(0)$.

Let us first show the existence of a critical value $u_m$ below which the typical configurations are deprived of any circuit. This transition is signalled by an instability of the trivial solution of equation (21), $P(y) = \delta(y)$. Perturbing this distribution infinitesimally, one can expand $g_k$ as $g_k(y_1, \ldots, y_k) = u \sum_{i=1}^k y_i + O(y_i^3)$. In this limit, if one inserts in the rhs of equation (21) some $P(y)$ with an infinitesimal mean $P_1$, one obtains another distribution with a mean $(u/u_m)P_1$, with

$$u_m^{-1} = \sum_{k=1}^\infty k \tilde{q}_k = \tilde{\mu}_1. \quad (33)$$

If $u < u_m$, the mean of the perturbed distributions decreases upon iteration, so $P(y) = \delta(y)$ is a stable solution. In contrast, if $u > u_m$ this solution is unstable and must flow to a non-trivial stable fixed point.

We shall now set up an expansion around the stability limit, $u = u_m + \epsilon$ with $\epsilon \to 0^+$. In this limit the messages $y$ are supported on a scale which vanishes with $\epsilon$; we shall consequently define $y = x \epsilon^a + o(\epsilon^a)$ with $x$ finite, and $a$ a positive exponent to be determined in a few lines. Let us denote by $Q(x)$ the distribution of the rescaled messages and relate the moments of $P$ and $Q$ as

$$P_n(\epsilon) = \int_0^\infty dy \, y^n P(y; u_m + \epsilon) \sim \epsilon^{na} \int_0^\infty dx \, x^n Q(x) = \epsilon^{na} Q_n. \quad (34)$$
Expanding equation (22) as
\[ x \sim u_m \sum_{i=1}^k x_i + \epsilon \sum_{i=1}^k x_i - u_m^2 \epsilon^2 \left( \sum_{i=1}^k x_i \right) \left( \sum_{1 \leq i < j \leq k} x_i x_j \right) , \]
we obtain at the lowest order
\[ \epsilon = u_m^4 \tilde{\mu}_2 Q_2 \epsilon^{2a} + \frac{1}{2} u_m^4 \tilde{\mu}_3 Q_1^2 \epsilon^{2a} , \]
\[ Q_2 \epsilon^{2a} = u_m Q_2 \epsilon^{2a} + u_m^2 \tilde{\mu}_2 Q_1^2 \epsilon^{2a} . \]
This fixes the scale \( a = 1/2 \) and the values of \( Q_1 \) and \( Q_2 \), the first two moments of the distribution solution of
\[ Q(x) = \tilde{q}_0 \delta(x) + \sum_{k=1}^\infty \tilde{q}_k \int_0^\infty dx_1 Q(x_1) \ldots dx_k Q(x_k) \delta \left( x - u_m \sum_{i=1}^k x_i \right) . \]

Let us now consider the consequences of this scaling on the observables \( f_q, \ell, \sigma_q \) in the limit \( u \to u_m \). Taking into account both their explicit dependence on \( u \) and the implicit one through the distribution \( P(y; u) \), one finds after a short computation the following.

- The expansion of \( f_q(u_m + \epsilon) \) starts at the second order,
\[ f_q(u_m + \epsilon) = f_q^{(2)} \epsilon^2 + O(\epsilon^3) , \]
\[ f_q^{(2)} = \frac{c}{2} Q_1^2 + \frac{c}{4} u_m^2 (1 - u_m) Q_2^2 - \frac{1}{2} u_m^4 \tilde{\mu}_3 Q_1^2 Q_2 - \frac{1}{8} u_m^4 \tilde{\mu}_4 Q_1^4 . \]

- The intensive length \( \ell \) is, at its first non-trivial order,
\[ \ell(u_m + \epsilon) = \ell^{(1)} \epsilon + O(\epsilon^2) , \]
\[ \ell^{(1)} = \frac{c}{2} u_m Q_1^2 . \]

- The first two derivatives of \( \sigma_q(\ell) \) in \( \ell = 0 \) can be obtained from the previous expressions:
\[ \sigma_q'(0) = -\ln u_m , \]
\[ \sigma_q''(0) = \frac{2 f_q^{(2)} }{(\ell^{(1)})^2} - \frac{2}{u_m \ell^{(1)}} . \]

Solving for \( Q_{1,2} \) and plugging their values in the expression (42) of the derivatives of the entropy one finally obtains
\[ \sigma_q'(0) = \ln \tilde{\mu}_1 , \]
\[ \sigma_q''(0) = -\frac{1}{c} \left( \frac{\tilde{\mu}_3}{\tilde{\mu}_1} + \frac{2 \tilde{\mu}_2}{\tilde{\mu}_1 (\tilde{\mu}_1 - 1)} \right) . \]

We can now turn to the discussion of these results, and in particular to the comparison with the annealed computation of Bianconi and Marsili [19]. Expanding their result (reproduced in equation (A.16)) in powers of \( \ell \), one obtains
\[ \sigma_a'(0) = \ln \tilde{\mu}_1 , \]
\[ \sigma_a''(0) = -\frac{1}{c} \frac{\tilde{\mu}_3 + 4 \tilde{\mu}_2 - 2 \tilde{\mu}_1 (\tilde{\mu}_1 - 1)}{\tilde{\mu}_1^2} . \]
Let us first consider a large but non-extensive circuit length, $1 \ll L \ll \ln N$. The number $\mathcal{N}_L$ of such circuits is, in the thermodynamic limit, a Poisson distributed random variable with a mean equal to

$$\frac{1}{2L} \left( \frac{\sum_k k(k-1)q_k}{\sum_k kq_k} \right)^L = \frac{1}{2L} (\bar{\mu}_1)^L. \quad (45)$$

When $L \gg 1$ the most probable value of this random variable is equal to its mean, in which one can neglect the polynomial prefactor $1/(2L)$ (we recall that we assume $\bar{\mu}_1 > 1$ to be in the percolated regime). Consequently, the quenched and annealed computation of the first derivative of the entropy at $\ell = 0$ coincide and match the result for $1 \ll L \ll \ln N$:

$$\mathcal{N}_L \sim e^{N\sigma(L/N)} \sim e^{L\sigma^\prime(0)} = (\bar{\mu}_1)^L. \quad (46)$$

In contrast the second derivatives differ, in general, in the two computations:

$$\sigma''_a(0) - \sigma''_q(0) = \frac{2}{c\bar{\mu}_3^2(\bar{\mu}_1 - 1)} (\bar{\mu}_2 - \bar{\mu}_1(\bar{\mu}_1 - 1))^2. \quad (47)$$

As expected the annealed entropy is always greater than the quenched one at this order of the expansion. Moreover, it is straightforward to show from the above expression that $\sigma''_a(0) - \sigma''_q(0)$ vanishes only if the distribution $\bar{q}_k$ is supported by a single integer, in other words in the random regular graph case.

We performed this computation using the degree distribution $q_k$ of the graph; however, the reader will easily verify that equation (43) remains unchanged if one replaces $q_k$ by the connectivity distribution $r_k$ of its two-core (the factorial moments $\bar{\mu}_n$ are multiplied by $\eta_n^{-1}$, the mean connectivity $c$ by $\eta^2$).

For completeness we state here the results for Poissonian graphs of average degree $c$,

$$\sigma'_q(0) = \ln c, \quad \sigma''_q(0) = -\frac{c + 1}{c - 1}. \quad (48)$$

5. The limit of longest circuits

A more interesting limit case is the one of maximal length circuits. Some questions arise naturally in this context: what is the maximal length, $\ell_{\text{max}}$, for which circuits of $N\ell_{\text{max}}$ edges are present with high probability in a given random graph ensemble? In particular, under which conditions are these graphs Hamiltonian, that is to say $\ell_{\text{max}} = 1$? Finally, what is the number of such longest circuits, measured by the corresponding quenched entropy $\sigma_q(\ell_{\text{max}})$? From the properties of the Legendre transform (cf equation (4)), these quantities can be determined by investigating the limit $u \to \infty$ of the free energy:

$$f_q(u) \sim \ell_{\text{max}} \ln u + \sigma_q(\ell_{\text{max}}). \quad (49)$$

This corresponds, in the jargon of the statistical mechanics approach to combinatorial optimization problems, to a zero temperature limit, where $-\ell_{\text{max}}$ (respectively $\sigma_q(\ell_{\text{max}})$) is the ground-state energy (respectively entropy) density.

It turns out that the answers to the above questions crucially depend on the presence or not of degree two sites in the two-core of the random graphs under study; we shall
thus divide the rest of this section according to this distinction. Before this we state an equivalent expression of the free energy which will prove more convenient in this limit,

$$f_q(u) = \sum_{k=3}^{\infty} q_k \left( \frac{k}{2} I_{k-1}(u) - \frac{k-2}{2} I_k(u) \right),$$

(50)

where we have defined some logarithmic moments of the distribution $P$,

$$I_k(u) = \int_0^\infty dy_1 P(y_1; u) \ldots dy_k P(y_k; u) \ln \left( 1 + u^2 \sum_{1 \leq i < j \leq k} y_i y_j \right) \quad \text{for } k \geq 2.$$  

(51)

This form of $f_q(u)$ is obtained from equation (23) by using the equation (21) on $P(y; u)$ and the identity

$$1 + u y_0 g_k(y_1, \ldots, y_k) = \frac{1 + u^2 \sum_{0 \leq i < j \leq k} y_i y_j}{1 + u^2 \sum_{1 \leq i < j \leq k} y_i y_j}.$$  

(52)

### 5.1. In the absence of degree two sites in the two-core

One can gain some intuition on the limit $u \to \infty$ by inspecting the behaviour of the messages in the regular case. Indeed, the expansion of equation (31) shows a scaling of the form $y \sim x u^{-1/2}$, with $x$ (an evanescent field in the jargon of optimization problems) finite. Consider now the more general case of random graph ensembles with a minimal connectivity of three, i.e. $q_0 = q_1 = q_2 = 0$, which consequently implies $\hat{q}_0 = \hat{q}_1 = 0$. Thanks to the vanishing of $\hat{q}_1$, the equations (21), (22) have a solution with the above scaling of $y$ with $u$. One can also check numerically in particular cases that the distributions $P(y)$ concentrate according to this behaviour for large but finite values of $u$.

Denoting $V_0(x)$ the distribution of the evanescent fields, one easily obtains the integral equation it obeys:

$$V_0(x) = \sum_{k=2}^{\infty} \hat{q}_k \int_0^\infty dx_1 V_0(x_1) \ldots dx_k V_0(x_k) \delta(x - h_k(x_1, \ldots, x_k)),$$  

(53)

with

$$h_k(x_1, \ldots, x_k) = \frac{\sum_{i=1}^k x_i}{\sum_{1 \leq i < j \leq k} x_i x_j}.$$  

Moreover, the logarithmic moments defined in equation (51) have the following scaling in this limit:

$$I_k(u) \sim \ln u + J_k, \quad \text{with } J_k = \int_0^\infty dx_1 V_0(x_1) \ldots dx_k V_0(x_k) \ln \left( \sum_{1 \leq i < j \leq k} x_i x_j \right).$$  

(54)

Plugging this equivalent in the expression (50) for the quenched free energy, and identifying the maximal length of the circuits with the coefficient of order $\ln u$, and their entropy with the constant term, we obtain

$$\ell_{\max} = 1, \quad \sigma_q(1) = \sum_{k=3}^{\infty} q_k \left( \frac{k}{2} J_{k-1} - \frac{k-2}{2} J_k \right).$$  

(55)
The identity $\ell_{\text{max}} = 1$ (of which we present an alternative derivation in appendix C) reproduces the conjecture of Wormald (conjecture 2.27 in [3]) that random graphs with a minimum connectivity of three are, with high probability, Hamiltonian. Obviously the methods we used are far from rigorous and do not provide a valid proof of the conjecture. However, they give it a quantitative flavour with the prediction of $\sigma_q(1)$, the typical entropy of such Hamiltonian circuits. We performed a numeric resolution of equation (53), again by a population dynamics algorithm, to compute the moments $J_k$ and from them the quenched entropy $\sigma_q(1)$. As an illustrative example, figure 4 presents the results of such a computation in the case of random graphs with a fraction $\epsilon$ of degree three vertices, and $1 - \epsilon$ of degree four. As a function of $\epsilon$ the entropy interpolates between the rigorously known values at $\epsilon = 0$, $\epsilon = 1$, for which the graphs are regular. Note that the quenched and annealed entropies, even if strictly different when $0 < \epsilon < 1$, are found to be numerically close. For instance when $\epsilon = 0.5$, one has $\sigma_q(1) \approx 0.2489$ and $\sigma_a(1) \approx 0.2501$.

5.2. In the presence of degree two sites in the two-core

Let us now consider the question of the longest cycles in random graph ensembles with connectivity distribution $q_k$ which do not fulfil the condition $q_0 = q_1 = q_2 = 0$ we assumed in the previous subsection. We first present simple combinatorial arguments which lead to bounds on $\ell_{\text{max}}$ and to an asymptotic expansion when there are very few degree two sites in the two-core, before coming back to the limit $u \to \infty$ of the cavity approach.

5.2.1. Bounds on $\ell_{\text{max}}$

Let us call a graph drawn at random from such an ensemble $G_1$, and its two-core $G_2$, determined for instance by the leaf removal algorithm detailed in appendix B. $G_2$ has the connectivity distribution $r_k$ defined and computed in section 3.2 and appendix B. The number of sites in the two-core is $N\ell_{\text{core}}$ (cf equation (B.9)), with $\ell_{\text{core}} < 1$ unless the original graph was deprived of any isolated sites and of leaves (i.e. $q_0 = q_1 = 0$). This $\ell_{\text{core}}$ is clearly an upper bound on $\ell_{\text{max}}$, as circuits cannot be longer than the number of available sites in the two-core.

One can derive a lower bound on $\ell_{\text{max}}$ with the following reasoning. From $G_2$, the two-core of $G_1$, eliminate recursively sites of degree two, identifying the two edges which were previously incident to it (see figure 5). When all sites of degree two have been
removed, one ends up with a graph, call it $G_3$, on $N(\ell_{\text{core}} - r_2)$ sites, where the minimal connectivity is three. Using the result of the previous section, this reduced graph typically contains circuits of length $N(\ell_{\text{core}} - r_2)$. Each of the circuits of $G_3$ can be unambiguously associated with a circuit of $G_2$, reinserting the edges which were simplified during the construction of $G_3$. Obviously the reconstructed circuits of $G_2$ are longer than the ones of $G_3$, hence $l_{\text{lb}} = (\ell_{\text{core}} - r_2)$ should be a lower bound for $\ell_{\text{max}}$. These bounds have been used under a stronger form in the case of Erdős–Rényi random graphs very close to the percolation threshold (for $c = 1 + \delta$ with $N^{-1/3} \ll \delta \ll 1$) in [8].

One can then wonder if whether the upper bound is saturated, in other words whether the two-core is Hamiltonian. In general the answer is no, as explained by the following remark. Consider a site of degree strictly greater than two, surrounded by at least three neighbours of degree two: obviously, no circuit can go through more than two of these neighbours. As soon as $r_2 > 0$ there will be an extensive number of such forbidden vertices, hence in such a case $\ell_{\text{max}} < \ell_{\text{core}}$. The equality is possible only if $r_2 = 0$, which was the case investigated in the previous section.

Note that the gap between the lower and upper bounds closes when $r_2$ vanishes, as the two-core becomes Hamiltonian in this limit. A conjecture on the behaviour of $\ell_{\text{max}}$ in the limit $r_2 \to 0$ can be formulated as

$$\ell_{\text{max}} = \ell_{\text{core}} - \sum_{k=3}^{\infty} r_k \binom{k}{3} \tilde{r}_1^3 + O(\tilde{r}_1^4).$$

(56)

This expression has a very simple interpretation: a forbidden site in the above argument appears if a vertex of degree greater than three is surrounded by at least three vertices of degree two. At the lowest order these forbidden sites are far apart from each other; $N\ell_{\text{max}}$ is thus reduced by one unit each time this appears. This conjecture will come out of the cavity analysis of next subsection, we preferred to anticipate it here because of its simple combinatorial interpretation.

Let us exemplify the bounds and the conjecture in two particular cases. Consider an ensemble of graphs where a fraction $1 - \epsilon$ of vertices have degree $c_0 \geq 3$, the others being of degree two, with $0 \leq \epsilon < 1$. The above bounds and estimation read

$$1 - \epsilon \leq \ell_{\text{max}} \leq 1, \quad \ell_{\text{max}} = 1 - \frac{4(c_0 - 1)(c_0 - 2)}{3c_0^2} \epsilon^2 + O(\epsilon^4).$$

(57)
As a second example we consider Poissonian random graphs of mean degree \( c \), for which the bounds read (cf equation (30))
\[
\ell_{fb} = 1 - (1 - \eta) \left( 1 + c \eta + \frac{1}{2} (c \eta)^2 \right), \quad \ell_{core} = 1 - (1 - \eta) (1 + c \eta),
\]
where \( \eta \) is the solution of \( \eta = 1 - \exp(-c \eta) \). In the limit \( c \to \infty \) the fraction of degree two vertices in the two-core vanishes; the above conjecture then reads
\[
\ell_{max} = 1 - (c + 1) e^{-c} - c^2 e^{-2c} - \frac{c^2}{2} \left( \frac{c^4}{3} + 3c - 1 \right) e^{-3c} + O(c^3 e^{-4c}),
\]
where \( n \) is some positive integer. Most of these terms come from the expansion of \( \ell_{core} \), the only non-trivial one being \(-c^2 e^{-3c}/6\). This is in agreement with a rigorous result of Frieze [10],
\[
1 - (1 + \delta(c)) c e^{-c} \leq \ell_{max} \leq 1 - (c + 1) e^{-c} \quad \text{with } \delta(c) \to 0.
\]

5.2.2. The large \( u \) limit in presence of sites of degree two. We come back to the cavity approach and investigate the large \( u \) limit in this case. The simple ansatz \( y \sim x u^{-1/2} \) is not compatible with equations (21), (22) any longer, because of the non-vanishing value of \( \hat{r}_1 \). The most natural generalization which allows us to close this set of equations is then \( y \sim x u^{b-1/2} \), where \( p \) is a relative integer. In the jargon of optimization problems, this is a hard field, i.e. non-vanishing in the zero temperature limit. We denote by \( V_p(x) \) the probability distribution of the evanescent fields \( x \) associated with hard fields \( p \). For notational simplicity we take the \( V_p \) to be unnormalized, with \( \int_0^\infty dx \, V_p(x) = v_p \), and impose the condition \( \sum_{p \in \mathbb{Z}} v_p = 1 \), in formula
\[
P(y; u) \sim \sum_{p \in \mathbb{Z}} V_p(y u^{1/2-p}) u^{1/2-p}.
\]
Expanding equations (21), (22) with this ansatz, one finds
\[
V_p(x) = \sum_{k=1}^\infty \hat{r}_k \sum_{p_1, \ldots, p_k \in \mathbb{Z}^k} \delta_{p, \varepsilon_k(p_1, \ldots, p_k)} \int_0^\infty dx_1 V_{p_1}(x_1) \ldots dx_k V_{p_k}(x_k) \times \delta(x - h_k(p_1, x_1, \ldots, p_k, x_k)).
\]
In order to simplify the expression of \( e_k \) and \( h_k \), we shall denote by \([n]\) a permutation of the indices which orders the hard fields in decreasing order, \( p_{[1]} \geq p_{[2]} \geq \cdots p_{[k]} \). Then \( e_1(p_1) = 1 + p_1, \quad e_k(p_1, \ldots, p_k) = \min(1 + p_{[1]}, -p_{[2]}) \quad \text{for } k \geq 2 \).

We also define
\[
d_k(p_1, x_1, \ldots, p_k, x_k) = \begin{cases} 
1 & \text{if } p_{[1]} + p_{[2]} < -1 \\
1 + \hat{d}_k(p_1, x_1, \ldots, p_k, x_k) & \text{if } p_{[1]} + p_{[2]} = -1 \\
\hat{d}_k(p_1, x_1, \ldots, p_k, x_k) & \text{if } p_{[1]} + p_{[2]} > -1,
\end{cases}
\]
\[
\hat{d}_k(p_1, x_1, \ldots, p_k, x_k) = \begin{cases} 
\sum_{i<j,|p_i=p_j|=p_{[1]}} x_i x_j & \text{if } p_{[1]} = p_{[2]} \\
-x_{[1]} & \text{if } p_{[1]} > p_{[2]}.
\end{cases}
\]
in terms of which the evanescent contribution reads
\[ h_1(p_1, x_1) = x_1, \]
\[ h_k(p_1, x_1, \ldots, p_k, x_k) = \frac{\sum_{i|p_i=p_1} x_i}{d_k(p_1, x_1, \ldots, p_k, x_k)} \quad \text{for } k \geq 2. \] (66)

Within this ansatz the logarithmic moments (cf equation (51)) behave as
\[ I_k(u) \sim J_k^{(h)} \ln u + J_k , \] (67)
\[ J_k^{(h)} = \sum_{p_1, \ldots, p_k \in \mathbb{Z}^k} v_{p_1} \cdots v_{p_k} \max(1 + p_{[1]} + p_{[2]}, 0), \] (68)
\[ J_k = \sum_{p_1, \ldots, p_k \in \mathbb{Z}^k} \int_0^\infty dx_1 V_{p_1}(x_1) \cdots dx_k V_{p_k}(x_k) \ln d_k(p_1, x_1, \ldots, p_k, x_k). \] (69)

From the behaviour of \( f_\ell \) we thus obtain
\[ \ell_{\max} = \sum_{k=3}^\infty r_k \left( \frac{k}{2} J_{k-1}^{(h)} - \frac{k-2}{2} J_k^{(h)} \right) , \quad \sigma_\ell(\ell_{\max}) = \sum_{k=3}^\infty r_k \left( \frac{k}{2} J_{k-1}^{(h)} - \frac{k-2}{2} J_k^{(h)} \right). \] (70)

Obviously this set of expressions reduces to the ones of section 5.1 when all the hard fields are null, which is a solution if and only if \( \tilde{r}_1 = 0 \).

Let us first discuss the computation of \( \ell_{\max} \). This quantity can be obtained from the distribution \( v_p \) of the hard fields, independently of the evanescent ones. Integrating away the evanescent fields in equation (62), one obtains
\[ v_p = \sum_{k=1}^\infty \tilde{r}_k \sum_{p_1, \ldots, p_k \in \mathbb{Z}^k} v_{p_1} \cdots v_{p_k} \delta_{p,e(p_1, \ldots, p_k)}. \] (71)

Besides the population dynamics method, a faster and more precise method can be devised to solve this equation. Let us for this purpose define \( w_p \), the integrated form of the distribution, and \( \varphi(x) \), the generating function of the \( \tilde{r}_k \):
\[ w_p = \sum_{p'=\infty}^{p-1} v_{p'}, \quad \varphi(x) = \sum_{k=1}^\infty \tilde{r}_k x^k. \] (72)

One can then rewrite equation (71), after a few lines of computation based on the expression of \( e_k(p_1, \ldots, p_k) \), under the form
\[ w_{p+1} = 1 - (1 - w_p) \varphi'(w_p) \quad w_{-p} = 1 - (1 - w_{p+1}) \varphi'(w_{p+1}) + \varphi(w_{p+1}) - \varphi(w_{p+1}) \quad \text{for } p \geq 0. \] (73)

It is easy to solve them numerically by iteration (both \( w_{-p} \) and \( 1 - w_p \) vanish exponentially fast when \( p \to +\infty \); a cut-off on \( p \) can thus be safely introduced), and to deduce \( \ell_{\max} \) from the solution \( v_p \) (see equations (68), (70)). We present the results of this procedure for Poissonian graphs in the left panel of figure 6, along with the bounds discussed above.

We now sketch the way to compute the expansion of \( \ell_{\max} \) stated in equation (56). In the limit \( \tilde{r}_1 \to 0 \), the distribution \( v_p \) tends to \( \delta_{p,0} \). A more precise inspection reveals that \( v_p = O(\tilde{r}_1^p) \), \( v_{-p} = O(\tilde{r}_1^{2p}) \) for \( p > 0 \). In order to obtain equation (56), it is thus enough...
On the number of circuits in random graphs

Figure 6. The length of longest cycles in Poissonian graphs of mean connectivity $c$ (left) and the associated quenched entropy (right). Dotted lines in the left panel are the bounds of equation (58). The dashed part of the curves corresponds to the regime $c \leq c^{(+)_{\ast}}$, where we expect the replica symmetry to be broken. The inset of the right figure shows a blow-up for small connectivities; the replica symmetric entropy presents a local maximum (respectively minimum) at $c \approx 2.48$ (respectively $c \approx 2.71$).

To find $\{v_{-1}, v_0, v_1, v_2, v_3\}$ at order $\tilde{r}_1^3$, as a function of the connectivity distribution $\tilde{r}_k$. The result follows by collecting the terms of order $\tilde{r}_1^3$ in equations (68) and (70). This expansion could be in principle pursued at any higher order, at the price of more tedious computations.

If one is not only interested in the length of the longest circuits, but also in the associated entropy $\sigma_q(\ell_{\text{max}})$, one has to solve the complete equation (62) on the distribution of both hard and evanescent fields. This is easily done by a population dynamics algorithm, following population of couples $(p, x)$; see the right panel of figure 6 for the results in the Poissonian case. However, when the value of $\tilde{r}_1$ is too large, there appears an instability in the resolution of equation (62). For the sake of definiteness let us consider the Poissonian case and postpone a more general discussion to the next section. For values of $c$ larger than a critical value $c^{(+)_{\ast}} \approx 2.88$, the evanescent fields distributions converge, whereas below $c^{(+)_{\ast}}$ the iteration brings some of them towards diverging or vanishing values. The origin of this instability can be traced back to the behaviour of the original messages $y$ at large but finite $u$. A closer inspection of numerically obtained histograms of the $y$ reveals that in this limit they indeed obey a scaling of the form $y \sim xu^{p-1/2}$, but $p$ is a relative integer only for $c \geq c^{(+)_{\ast}}$. For lower connectivities, a continuously growing fraction of the hard fields are half-integers. This fraction reaches one at $c^{(-)_{\ast}} \approx 2.67$, below which all the $p$ are half-integers. If one allows the hard fields to be both integers and half-integers in equations (62), (68), and (69), this instability problem is cured, which allowed us to obtain the (dashed) low connectivity part of the curves in figure 6. We shall come back in the next section to the interpretation of this phenomenon.

6. Stability of the replica symmetric ansatz

The cavity computations we have presented so far were based on the assumption of replica symmetry (RS), valid if the space of configurations is smooth enough. In disordered
systems this assumption can be violated; we shall thus investigate its validity in the present model. More precisely, we consider the local stability of the RS ansatz in the enlarged space of one step replica symmetry breaking (1RSB) order parameters \([20]\) (we leave aside the possibility of a discontinuous transition). In the 1RSB setting, the messages \(y\) are replaced by probability distributions \(Q(y)\) over the states, and the recursion \(y \leftarrow g_k(y_1, \ldots, y_k)\) becomes

\[
Q(y) = \frac{1}{Z} \int dy_1 Q_1(y_1) \cdots dy_k Q_k(y_k) \delta(y - g_k(y_1, \ldots, y_k)) W(y_1, \ldots, y_k)^m, \tag{74}
\]

where \(Z\) is a normalization constant, \(m\) is the Parisi 1RSB parameter, and \(W(\{y_{k-i}\})\) is a reweighting factor whose explicit form is not needed here. The distributions \(Q\) are themselves drawn from a distribution over distributions, \(Q[Q]\).

The replica symmetric solution studied in the main part of the text is recovered by taking the distributions \(Q\) concentrated on a single value \(y\). To investigate its local stability, one gives them an infinitesimal variance \(v\). Expanding equation (74) in the limit of vanishing variances \(v\), one obtains the following relation:

\[
(y, v) \leftarrow \left( g_k(y_1, \ldots, y_k), \sum_{j=1}^k \left( \frac{\partial g_k(y_1, \ldots, y_k)}{\partial y_j} \right)^2 v_j \right). \tag{75}
\]

For the RS solution to be stable against this perturbation, the variances of the 1RSB order parameters should decrease upon iterations of the above relation. This can be studied numerically for any random graph ensemble, by iterating the above relation on a population of couples \((y, v)\), the value of \(k\) being drawn from \(\hat{\mu}_k\). The variances \(v\) can be initially all taken as unity (note that equation (75) is linear in the \(v\)); in the course of the dynamics the \(v\) are periodically divided by a number \(\lambda\), chosen each time to maintain the average value of \(v\) constant. After a thermalization phase \(\lambda\) converges (in order to gain numerical precision one computes the average over the iterations of \(\ln \lambda\)), its limit being \(>1\) (respectively \(<1\)) if the RS solution is unstable (respectively stable). This method, pioneered in the context of the instability of the 1RSB solution in \([33]\), can be replaced by the computation of the associated non-linear susceptibility; see for instance \([34]\).

For regular random graphs of connectivity \(c\), where all RS messages take the same value \(y_r(u, c)\) given in equation (31), one can readily compute the value of the stability parameter,

\[
\lambda_r(u, c) = \frac{(2 - u(c - 1))^2}{u^2(c - 1)^3}. \tag{76}
\]

It is easy to check that \(0 \leq \lambda_r(u, c) \leq (c - 1)^{-1} < 1\) when \(u\) lies in its allowed range \([u_m = (c - 1)^{-1}, \infty[\), confirming the validity of the RS ansatz. It would have been surprising anyhow to discover an instability in this case where the annealed computation is exact.

Another case which is analytically solvable is the limit \(\ell \to 0\) (i.e. \(u \to u_m\)). Indeed, we have seen that the messages scale then as \(x(u - u_m)^2\), and it turns out that \(\partial g_k/\partial y_i \to u_m\) independently of the rescaled messages \(x\). Recalling that \(\sum_k \tilde{r}_kk = \tilde{\mu}_1 = u_m^{-1}\), one finds \(\lambda = \tilde{\mu}_1^{-1} < 1\) in this limit: for any connectivity distribution, the RS ansatz is always stable in the small \(\ell\) regime.

All the numerical investigations of \(\lambda\) we conducted for ensembles with minimal connectivity three suggest that in this case the replica symmetric solution is stable for
On the number of circuits in random graphs

Figure 7. Left: the stability parameter $\lambda$ for Poissonian random graphs; from left to right $c = 1.2, 2, 3$. Right: sketched behaviour of the quenched entropy for generic families of random graphs. From top to bottom a control parameter drives the graphs towards a continuous percolation transition; the maximal length of the circuits is reduced. In the neighbourhood of the percolation transition replica symmetry breaking takes place for large enough circuits, and should be taken into account to compute the dashed part of the curve.

The situation is less fortunate for Poissonian graphs. The reader may have anticipated the appearance of non-integer hard fields in the zero temperature limit for mean connectivities lower than $c^{(+)}_s$ as a hint of RSB. The data presented in the left panel of figure 7 show indeed that for small $c$, the stability parameter $\lambda$ crosses unity when $u$ is increased above some finite value $u_s(c)$. This critical value of $u$ increases with the mean connectivity, and an educated guess makes us conjecture that it diverges at $c^{(+)}_s$. The rightmost curve for $c = 3$ shows indeed $\lambda < 1$ for all the values of $u$ we could numerically study. A precise extrapolation of $u_s(c)$ turned out however to be rather difficult. Note also that the study directly at $u = \infty$ is greatly complicated here by the fact that the hard fields do not take a finite number of distinct values as is often the case in usual optimization problems [35], but extend on the contrary on all relative integers. In summary, the conjectured scenario is that at high enough connectivities the whole curve $\sigma_q(\ell)$, and in particular its zero temperature limit, is correctly described by the RS computation. For lower connectivities there will be a critical length above which replica symmetry breaks down. We also believe that this scenario, sketched in the right part of figure 7, is valid not only in the Poissonian case, but for all families of random graph ensembles (with a fast decaying connectivity distribution) with a control parameter which drives the graphs towards a continuous percolation transition, the fraction of degree two sites in the two-core growing as the transition is approached.

Let us finally propose an interpretation for the occurrence of replica symmetry breaking for the largest circuits in the presence of a large fraction of degree two sites in the two-core, by relating it to an underlying extreme value problem [36]. In the discussion of section 5.2.1, one could indeed tag the edges $l$ of the reduced graph $G_3$ with a strictly
on the number of circuits in random graphs

positive integer, by counting the number of edges of $G_2$ which were collapsed onto $l$. The
length of a circuit of $G_2$ is thus the weighted length of the corresponding circuit of $G_3$,
i.e. the sum of the labels on the edges it visits. These weighted lengths are correlated
random variables, because of the structural constraint defining a circuit: for a given graph
$G_3$, not all the sums of $L$ tags correspond to circuits of length $L$. When the fraction of
degree two sites is small enough, these correlations are sufficiently weak for the RS ansatz
to treat them correctly; when long chains of degree two vertices become too numerous
they somehow pin the longest circuits, which cluster in the space of configurations and
cause the replica symmetry breaking.

7. Exhaustive enumerations

We present in this section the results of the numerical experiments we have conducted in
order to check our analytical predictions. These experiments are based on the exhaustive
enumeration algorithm of [12] which allows us to generate all the circuits of a given graph
$G$, and in particular to compute the numbers $N_L(G)$ of circuits of a given length. This
algorithm runs in a time proportional to the total number of circuits, hence exponential
in the size of the graphs for the cases we are interested in, which obviously puts a strong
limitation on the sizes we have been able to study.

Let us begin with the investigation of the Erdős–Rényi ensembles $G(N,p)$ and
$G(N,M)$. In the former, each of the $N(N-1)/2$ potential edges between the $N$ vertices
of the graph is present with probability $p$, independently of each other; in the latter a
set of $M$ among the $N(N-1)/2$ edges is chosen uniformly at random. With $p = c/N$
and $M = cN/2$, these two ensembles are expected to be equivalent in the large-size
limit. In particular, the vertex degree distribution converges in both cases to a Poisson
law of mean $c$; the cavity computation thus predicts that their typical properties should
be the same in the thermodynamic limit. This is not true for the annealed entropies
$\sigma_a(\ell;N) = \log(N_{\ell N})/N$, which are easily computed exactly even at finite sizes (see
appendix A) and which remain distinct in the thermodynamic limit. In the left part
of figure 8 we present the annealed and quenched entropies for both ensembles, computed
from 10 000 graphs of size $N = 36$ and mean connectivity $c = 3$. The finite size
quenched entropy has been estimated using the median of the random variables $N_L$.
The annealed entropies are very different in both ensembles (and in perfect agreement
with the computation of appendix A), and clearly different from the quenched ones. The
striking feature of this plot is the almost perfect coincidence of the median in the two
ensembles; this was expected in the thermodynamic limit, but is already very clear at
this moderate size. On the right panel of figure 8 the quenched entropy is plotted for two
graph sizes, along with its extrapolated values in the thermodynamic limit, which agrees
with the cavity computation.

As argued above, the difference between annealed and quenched entropies can be
also seen in the exponentially larger value of the second moment of $N_L$ with respect to
the square of the first moment. This fact is illustrated in figure 9, where the analytic
computation of the ratio $\log(\langle N_L^2 \rangle/\langle N_L \rangle^2)/N$ presented in appendix A is confronted with its
numerical determination.

We also considered the largest circuits in each graph, of length $L_{\text{max}}$ and degeneracy
$N_{L_{\text{max}}}$, and computed the averages $L_{\text{max}}/N$ and $\ln N_{L_{\text{max}}}/N$ for various connectivities.

doi:10.1088/1742-5468/2006/06/P06019
Figure 8. Left: annealed and quenched entropies for the Erdős–Rényi ensembles $G(N,p)$ and $G(N,M)$ of mean connectivity $c = 3$, for graphs of size $N = 36$, computed from the mean and the median of $N_L$ on samples of 10000 graphs. Right: the quenched entropy for $G(N,M)$ at $N = 36$ and $N = 54$; symbols are the extrapolation in the limit $N \to \infty$ from several values of $N$; the solid line is the replica symmetric cavity computation.

Figure 9. Left: the ratio of the first two moments of $N_L$ for $G(N,M)$ at $c = 3$, the symbols are numerically determined values which converge in the large size limit to the solid line, analytically computed in appendix A. Right: finite size analysis for $\ell = 1/4$; the solid line is a best fit of the form $a + b/N + c/N^2$, where $a$ is constrained to its analytic value (dashed line), and the form of the fit is justified in appendix A.

Their extrapolated values in the thermodynamic limit are compatible with the predictions $\ell_{\text{max}}$ and $\sigma_q(\ell_{\text{max}})$ of the cavity method, within the numerical accuracy we could reach. This is also true for connectivities smaller than $c_{3}^{(+)}$, where we argued above in favour of a violation of the replica symmetry hypothesis: the corrections due to RSB should be smaller than the numerical precision we reached.

Another set of experiments concerned uniformly generated graphs with an equal number of degree three and four vertices. We checked that the probability for such graphs to be Hamiltonian converges to unity when increasing their size. The values for the annealed and quenched entropies for the Hamiltonian circuits are too close to be distinguished numerically. However, the study of the ratio of the first two moments of $N_N$ (see figure 10) indicates that they should be strictly distinct in the thermodynamic limit.
8. Conclusions and perspectives

Let us summarize the main results presented in this paper. We have proposed an approximative counting algorithm that runs in a linear time with respect to the size of the graph. We also presented a heuristic method to compute the typical number of circuits in random graph ensembles, which yields a quantitative refinement of Wormald’s conjecture on the typical number of Hamiltonian cycles in ensembles with minimal degree three (equation (55)) and a new conjecture on the maximal length of circuits in ensembles with a small fraction of degree two vertices in their two-cores (equation (56)).

Several directions are opened for future work. First of all we believe that a rigorous proof of Wormald’s conjecture, which seems difficult to reach by variations around the second moment method, could be obtained by statistical mechanics inspired techniques. In recent years there has been indeed a series of mathematical achievements in the formalization of the kind of method used in this paper. One line of research is based on Guerra’s interpolation method [37], and culminated in Talagrand’s proof of the correctness of the Parisi free-energy formula for the Sherrington–Kirkpatrick model [38]. These ideas have also been applied to sparse random graphs in [39, 40]. Alternatively the local weak convergence method of Aldous [41] has been successfully applied to similar counting problems in random graphs [42].

There has also been a recent interest [43]–[45] in the corrections to the Bethe approximation for general graphical models. It would be of great interest to implement these refined approximations for the counting problem considered in this paper. This should lead on one hand to a more precise counting algorithm, and on the other hand give access to the finite-size corrections of the quenched entropy. We expect in particular that the difference between circuits and unions of vertex disjoint circuits will become relevant for these corrections.

The convergence in probability of \( \log \mathcal{N}_L/N \) expressed by equation (18) can a priori be promoted to a stronger large deviation principle: according to the common wisdom, the finite deviations of this quantity are exponentially small. A general method for computing these rate functions has been presented in [46] and could be of use in the present context. An interesting question could be to compute the exponentially small probability that a random graph is not Hamiltonian in ensembles where typical graphs are so.
In the algorithmic perspective, one could try to take advantage of the local information provided by the messages. In particular, they could be useful to explicitly construct long cycles, in a ‘belief inspired decimation’ fashion [47]: most probable edges in the current probability law would be recursively forced to be present, and the BP equations re-run in the new simplified model.

The neighbourhood of the percolation transition should also be investigated more carefully, in particular the effects of replica-symmetry breaking on the structure of the configuration space.

The case of heavy tailed (scale-free) degree distributions also deserves further work. The assumption of fast decay we made here is indeed crucial for some of our results: Bianconi and Marsili showed in [19] that scale-free graphs, even with a minimal connectivity of three, can fail to have Hamiltonian cycles. Other random graph models (generated by a growing process [48], or incorporating correlations between vertex degrees [49]), could also be investigated.

Let us finally mention two closely related problems which are currently studied by very similar means. Circuits can be defined as a particular case of $k$-regular graphs, with $k = 2$. Replacing the number of allowed edges around any site from 2 to $k$ in equation (2), one can similarly study the number of $k$-regular subgraphs in a random graph ensemble. The case $k = 1$ corresponds to matchings, which was largely studied in the mathematical literature [50,51] and has been reconsidered by statistical mechanics methods in [52]. The appearance of $k \geq 3$-regular subgraphs in random graphs was first considered in [53]; see [54] for a statistical mechanics treatment.

Acknowledgments

We warmly thank Rémi Monasson with whom the first steps of this work were taken. We also acknowledge very useful discussions with Ginestra Bianconi, Andrea Montanari, Andrea Pagnani, Federico Ricci-Tersenghi, Olivier Rivoire, Martin Weigt and Lenka Zdeborová.

The work was supported by EVERGROW, integrated project No 1935 in the Complex Systems Initiative of the Future and Emerging Technologies Directorate of the IST Priority, EU Sixth Framework.

Appendix A. Combinatorial approach

A.1. Notation and definitions

We collect in this appendix the combinatorial arguments for the computation of the first and second moments of the number of circuits in various random graph ensembles. Let us denote by $N_L(G)$ the number of circuits of length $L$ in a graph $G$, and $C_L$ the set of circuits of length $L$ in the complete graph of $N$ vertices, its cardinality being

$$|C_L| = \mathcal{M}_L = \frac{N!}{2L(N-L)!}. \tag{A.1}$$

Indeed, choosing such a circuit amounts to selecting an ordered list of the $L$ vertices it will visit, modulo the orientation and the starting point of the tour. Introducing $\mathbb{I}(H;G)$,
On the number of circuits in random graphs

the indicator function equal to unity if $H$ is a subgraph of $G$, zero otherwise, we can write

$$N_L(G) = \sum_{C \in C_L} I(C; G).$$  \hfill (A.2)

Let us now describe the random graph ensembles we shall consider in the following. The first two are the classical Erdős–Rényi random graph ensembles. In $G(N, p)$, each of the $N(N - 1)/2$ edges is present with probability $p$, independently of the others. In $G(N, M)$, a set of $M$ distinct edges is chosen uniformly at random among the $N(N - 1)/2$ possible ones. We shall concentrate on the thermodynamic limit $N \to \infty$, $p = c/N$ and $M = cN/2$ with the mean connectivity $c$ kept finite. In this regime $G(N, p)$ and $G(N, M)$ are essentially equivalent: drawing at random from $G(N, p)$ amounts to drawing $M$ from a binomial distribution of parameters $(p, N(N - 1)/2)$, and then drawing at random a graph from $G(N, M)$. In the limit described above, the number of edges in $G(N, p)$ is weakly fluctuating around $M = cN/2$. Moreover, the degree of a given vertex in the graph converges in both cases to a Poisson random variable of parameter $c$.

For an arbitrary degree distribution $q_k$ of mean $c$, one can define the uniform ensemble of graphs obeying this constraint of degree distribution. A practical way of drawing a graph from this ensemble is the so-called configuration model [62], defined as follows. Each of the vertices is randomly attributed a degree, in such a way that $Nq_k$ vertices have degree $k$ (we obviously skip some technical details [29]: $q_k$ should be a function of $N$, such that $Nq_k$ is an integer). $2k$ half-links go out of each vertex of degree $k$. Then one generates a random matching of the $cN = 2M$ half-links and puts an edge between sites which are matched. In general one obtains in this way a multigraph, i.e. there appear edges linking one vertex with itself, or multiple edges between the same pair of vertices. However, discarding the non-simple graphs leads to a uniform distribution over the simple ones [3]. To compute averages over the graph ensemble, one can thus use the configuration model and condition on the multigraph to be simple. For clarity we shall denote by $N_L^*(G)$ the number of circuits in the unconditioned multigraph ensemble. Note also that regular random graphs are a particular case of this ensemble, with $q_k = \delta_{k,c}$.

A.2. First moment computations

A.2.1. Generalities. Taking the average over the graphs of equation (A.2) leads to

$$\mathcal{N}_L(G) = \sum_{C \in C_L} \overline{I}(C; G) = \mathcal{M}_L \mathcal{P}_L$$  \hfill (A.3)

for the ensembles we are considering, where the probability $\mathcal{P}_L = \overline{I}(C; G)$ for a circuit $C \in C_L$ to be present is independent of $C$. Before inspecting the various cases, let us state the asymptotic behaviour of $\mathcal{M}_L$ in the limit $N, L \to \infty$, $\ell = L/N$ finite, obtained with the Stirling formula:

$$\mathcal{M}_{\ell N} = \frac{1}{N} \frac{1}{2\ell \sqrt{1 - \ell}} N^{L \ell} e^{N(-h(1 - \ell) - \ell)} (1 + O(N^{-1})) \quad \text{for } 0 < \ell < 1,$$  \hfill (A.4)

$$\mathcal{M}_N = \frac{1}{\sqrt{N}} \sqrt{\frac{\pi}{2}} N^N e^{-N} (1 + O(N^{-1})).$$  \hfill (A.5)

In the first formula we have introduced the function $h(x) = x \ln x$. 

doi:10.1088/1742-5468/2006/06/P06019
A.2.2. Erdős–Rényi ensembles. In $G(N, p)$ the probability $P_L$ has a very simple expression, $P_L = (c/N)^L$. The mean number of circuits thus reads

$$\bar{N}_L(G) = \frac{1}{2L} \frac{N!}{(N-L)!} \left( \frac{c}{N} \right)^L = \frac{1}{N2\ell\sqrt{1-\ell}} e^{N\sigma(\ell)}(1 + O(N^{-1})), \quad (A.6)$$

where the first expression is valid for any $N, L$, and the second one has been obtained in the thermodynamic limit with $0 < \ell < 1$. The annealed entropy for this first ensemble is

$$\sigma(\ell) = -(1 - \ell) \ln(1 - \ell) + \ell((\ln c) - 1). \quad (A.7)$$

Note that if $\ell = 1$ the algebraic prefactor in (A.6) is slightly different,

$$\bar{N}_N(G) = \frac{\pi}{\sqrt{2N}} e^{N((\ln c) - 1)}(1 + O(N^{-1})). \quad (A.8)$$

In $G(N, M)$ the probability $P_L$ reads

$$P_L = \frac{\left( \binom{N}{2} - L \right)!}{\left( \binom{N}{2} \right)!} \frac{M!}{(M-L)!} \quad (A.9)$$

Obviously this expression has a meaning only for $L \leq M$, as there cannot be circuits longer than the total number of edges. This gives an exact expression for $\bar{N}_L(G) = M_L P_L$ for any $N$ and $L \leq \min(N, M)$. The expansion in the thermodynamic limit with $0 < \ell < \min(1, c/2)$ leads to

$$\bar{N}_L(G) = \frac{1}{N2\ell\sqrt{1-\ell}} e^{\ell(\ell+1)} e^{N\sigma(\ell)}(1 + O(N^{-1})) \quad (A.10)$$

with the annealed entropy

$$\sigma(\ell) = -(1 - \ell) \ln(1 - \ell) + \left( \ell - \frac{c}{2} \right) \ln \left( 1 - \frac{2\ell}{c} \right) + \ell((\ln c) - 2). \quad (A.11)$$

Again the different algebraic prefactor in (A.10) can be easily computed also for $\ell = \min(1, c/2)$.

Let us now make a few comments on these results. First, when $c < 1$, both annealed entropies are negative for all values of $\ell > 0$ where they are well defined. Consequently $\bar{N}_L(G)$ is exponentially small in the thermodynamic limit, and thanks to the so-called Markov inequality (or first moment method) valid for positive integer random variables,

$$\text{Prob}[\bar{N}_L(G) > 0] \leq \bar{N}_L(G), \quad (A.12)$$

with high probability there are no circuits of extensive lengths in these graphs. This could be expected: the percolation transition occurs at $c = 1$, in this non-percolated regime the size of the largest component is of order $\ln N$, and thus extensive circuits cannot be present.

As a second remark, let us note that for $c > 1$ the annealed entropy of the first ensemble is strictly positive for $\ell \in [0, \ell_c(c)]$, with $\ell_c(c)$ an increasing function which reaches unit value at $c = e$. The average number of such circuits is in consequence exponentially large. However one can easily convince oneself that this cannot be the typical behaviour. Indeed, it turns out that $\ell_c(c)$ is larger than the typical number of vertices in the two-core of the graphs $\ell_{\text{core}}(c)$. When $\ell$ belongs to the interval

$$\text{Prob}[\bar{N}_L(G) > 0] \leq \bar{N}_L(G),$$

with high probability there are no circuits of extensive lengths in these graphs. This could be expected: the percolation transition occurs at $c = 1$, in this non-percolated regime the size of the largest component is of order $\ln N$, and thus extensive circuits cannot be present.

As a second remark, let us note that for $c > 1$ the annealed entropy of the first ensemble is strictly positive for $\ell \in [0, \ell_c(c)]$, with $\ell_c(c)$ an increasing function which reaches unit value at $c = e$. The average number of such circuits is in consequence exponentially large. However one can easily convince oneself that this cannot be the typical behaviour. Indeed, it turns out that $\ell_c(c)$ is larger than the typical number of vertices in the two-core of the graphs $\ell_{\text{core}}(c)$. When $\ell$ belongs to the interval

$$\text{Prob}[\bar{N}_L(G) > 0] \leq \bar{N}_L(G),$$

with high probability there are no circuits of extensive lengths in these graphs. This could be expected: the percolation transition occurs at $c = 1$, in this non-percolated regime the size of the largest component is of order $\ln N$, and thus extensive circuits cannot be present.
On the number of circuits in random graphs

\[ \ell_{\text{core}}(c), \ell_a(c) \], typically the graphs cannot contain circuits of \( L = \ell N \) edges; however, an exponentially small fraction of the graphs have two-core larger than their typical sizes. These untypical graphs contribute with an exponential number of circuits to the annealed mean \( \mathcal{N}_L(G) \), which is in consequence not representative of the typical behaviour of the ensemble.

Finally, let us underline that the annealed entropies (A.7), (A.7) for the two ensembles are definitely different. For instance, in the second ensemble, the entropy is defined only for \( \ell \leq c/2 \): the number of edges in the graph being fixed at \( M = Nc/2 \), no circuits can be longer than the number of edges. In contrast, in the first ensemble, arbitrary large deviations of the number of edges from its typical value are possible, even if with an exponential small probability.

A.2.3. Arbitrary connectivity distribution and regular. The expectation of the number of circuits of length \( L \) in the multigraph ensemble extracted with the configuration model was presented in [19]. For the sake of completeness and to make the study of the second moment simpler we reproduce the argument here. In this case one has

\[
P_L = \frac{1}{(\ell)} \left( \sum \prod \frac{Nq_k}{L_k} (k(k-1))^{L_k} \right) \frac{(cN - 2L - 1)!!}{(cN - 1)!!},
\]  

where the sum is over \( L_2, L_3, \ldots \) positive integers constrained by \( \sum_{k=2}^{\infty} L_k = L \), and we used the classical notation \((2p-1)!! = (2p-1)(2p-3)\ldots1\). The \( L_k \) are the number of sites of degree \( k \) in the circuit, which are to be distributed among the \( Nq_k \) sites of degree \( k \). The term \((k(k-1))^{L_k}\) accounts for the choice of the half-links around each site, and finally the ratio of the double factorials is the probability that the matching of half-links contains the desired configuration. Introducing the integral representation of the Kronecker symbol, \( \delta_n = \oint (d\theta/2i\pi) \theta^{-n-1} \), where \( \theta \) is a complex variable integrated along a closed path around the origin, this expression can be simplified as

\[
P_L = \frac{1}{(\ell)} \left( \sum \prod \frac{Nq_k}{L_k} (k(k-1))^{L_k} \right) \oint \frac{d\theta}{2\pi i} \theta^{-L-1} \prod_{k=2}^{\infty} (1 + \theta k(k-1))^{Nq_k}. \]

(A.14)

In the thermodynamic limit the integral can be evaluated by the saddle-point method, combining the expansion with the one of \( \mathcal{M}_L \) yields

\[
\mathcal{N}_L^*(G) \approx e^{N\sigma(\ell)},
\]  

(A.15)

\[
\sigma(\ell) = \frac{1}{2} h(c - 2\ell) - \frac{1}{2} h(c) + h(\ell) + \text{ext} \left[ \sum_{k=2}^{\infty} q_k \ln(1 + k(k-1)\theta) - \ell \ln \theta \right],
\]  

(A.16)

where here and in the following \( \approx \) stands for equivalence up to subexponential terms, i.e. \( x_N \approx y_N \) means \((1/N) \log(x_N/y_N) \to 0 \) as \( N \to \infty \).

In the regular case one has

\[
P_L = (c(c - 1))^L \frac{(cN - 2L - 1)!!}{(cN - 1)!!},
\]  

(A.17)

\[\text{doi:10.1088/1742-5468/2006/06/P06019}\]
from which the prefactors are more easily computed
\begin{align}
\overline{\mathcal{N}}_L^2(G) &= \frac{1}{N} \frac{1}{2\sqrt{1 - \ell}} e^{N\sigma(\ell)} (1 + O(N^{-1})) \quad \text{for } 0 < \ell < 1, \\
\overline{\mathcal{N}}_N(G) &= \frac{\pi}{\sqrt{2N}} e^{N\sigma(1)} (1 + O(N^{-1})), \\
\sigma(\ell) &= -h(1 - \ell) - \frac{1}{2}h(c) + \frac{1}{2}h(c - 2\ell) + \ell \ln(c(c - 1)).
\end{align}
Moreover, the conditioning on the multigraph being simple can be explicitly done in the regular case [6], thanks to the relative concentration of \( \mathcal{N}_L^2(G) \). This yields
\begin{align}
\frac{\overline{\mathcal{N}}_L(G)}{\overline{\mathcal{N}}_L^2(G)} &\to \exp \left[ \frac{(c - 2)\ell}{c - 2\ell} \left( 1 + \frac{c(1 - \ell)}{c - 2\ell} \right) \right] \quad \text{for } 0 < \ell \leq 1.
\end{align}
We checked that the numerical findings of [31] were in perfect agreement with these exact results.

Note that in the regular case this conditioning modifies the value of \( \overline{\mathcal{N}}_L^2(G) \) only by a constant factor, thus the annealed entropy is the same in the graph and in the multigraph ensemble. It is not clear to us whether this fact should remain true for arbitrary connectivity distributions.

### A.3. Second moment computations

#### A.3.1. Generalities
We now turn to the computation of the second moment of the number of circuits, which has been inspired by [6]. Taking the square of equation (A.2) and averaging over the ensemble leads to
\begin{align}
\overline{\mathcal{N}}_L^2(G) &= \sum_{C_1, C_2 \in \mathcal{C}_L} \mathbb{I}(C_1; G) \mathbb{I}(C_2; G) = \mathcal{N}_L(G) + \sum_{C_1 \neq C_2 \in \mathcal{C}_L} \mathbb{I}(C_1; G) \mathbb{I}(C_2; G) \\
&= \mathcal{N}_L(G) + \sum_{X, Y, Z} \mathcal{M}_{LXYZ} \mathcal{P}_{LXYZ}.
\end{align}
We have indeed isolated the term \( C_1 = C_2 \) in the sum, which is readily computed, from the off-diagonal terms. The last expression is more easily understood after having a look at figure A.1, where we sketched the shape of the union of two distinct circuits. This pattern is characterized by \( X \), the number of common paths shared by \( C_1 \) and \( C_2 \), \( Y \), the number of edges in these paths, and \( Z \), the number of vertices which belongs to both circuits but are not neighboured by any common edge. One finds \( 2X \) vertices at the extremities of the common paths, \( Y - X \) vertices in the interior of the common paths, hence \( X + Y + Z \) vertices belong to both circuits, \( N - 2L + X + Y + Z \) to none of them. In consequence the sum is over \( X, Y, Z \) non-negative integers subject to the constraints
\begin{align}
2L - N \leq X + Y + Z \leq L, \quad X \leq Y, \quad X = 0 \Rightarrow Y = 0.
\end{align}
\( \mathcal{M}_{LXYZ} \) is the number of pairs of distinct circuits of the complete graph whose union has the characteristics \( (X, Y, Z) \), and \( \mathcal{P}_{LXYZ} \) is the (ensemble-dependent) probability that such a pattern appears in a random graph. Let us show that
\begin{align}
\mathcal{M}_{LXYZ} &= \frac{1}{4} 2^X \frac{N! (L - Y - 1)!^2}{X! Z! ((L - X - Y - Z)!)^2 (N - 2L + X + Y + Z)!} \binom{Y - 1}{X - 1},
\end{align}
doi:10.1088/1742-5468/2006/06/P06019 32
On the number of circuits in random graphs

Figure A.1. The union of two circuits. The vertices of $C_1$ are on and above the horizontal central line, those of $C_2$ on and below. In this drawing $L = 13$, $X = 3$, $Y = 6$, $Z = 1$, $n_1 = n_2 = n_3 = 1$.

where the combinatorial factor $(-1)$ is by convention set to unity. To construct such a pattern, one has to choose among the $N$ vertices those which are in $C_1$ but not in $C_2$, in $C_2$ but not in $C_1$ (both of these categories contain $L - X - Y - Z$ vertices), those in the common paths $(X + Y)$ and those shared by the circuits but with no adjacent common edges $(Z)$. This can be done in

$$\frac{N!}{Z!(X+Y)!((L-X-Y-Z)!)^2(N-2L+X+Y+Z)!}$$

(A.26)
distinct ways.

Let us call $n_i$ the number of common paths of $i$ edges, for $1 \leq i \leq L-1$, which obey the constraints $\sum n_i = X$ and $\sum n_i = Y$. The $X + Y$ sites can be distributed into such an unordered set of unorientated paths in

$$\frac{(X+Y)!}{2^X \prod_{i=1}^{L-1} n_i!}$$

(A.27)
distinct ways. We have to sum this expression on the values of $n_i$ satisfying the above constraints. By picking up the coefficient of $t^Y$ in

$$(t + t^2 + \cdots + t^{L-1})^X = \sum_{n_1,\ldots,n_{L-1}} \frac{X!}{\prod_{i=1}^{L-1} n_i!} t^{\sum n_i} \delta_{X,\sum n_i} = t^X \left(1 - t^{L-1} \right)^X, \quad (A.28)$$

one finds that

$$\sum_{n_1,\ldots,n_{L-1}} \frac{1}{\prod_{i=1}^{L-1} n_i!} \delta_{X,\sum n_i} \delta_{Y,\sum n_i} = \frac{1}{X!} \binom{Y-1}{X-1}. \quad (A.29)$$

When $X = Y = 0$ this factor should be one, in agreement with the above convention.

Finally, $C_1$ is formed by choosing an ordered list of the $L - X - Y - Z$ vertices which belong only to it, the $Z$ isolated common vertices, and of the $X$ orientated common paths, modulo the starting point and the global orientation of this tour, hence a factor

$$\frac{(L-Y-1)!}{2} 2^X,$$

(A.30)
and the same arises when constructing $C_2$. Equation (A.25) is obtained by multiplying the various contributions.
On the number of circuits in random graphs

In the thermodynamic limit with \((x, y, z) = (X/N, Y/N, Z/N)\) kept finite, the Stirling formula yields

\[
M_{LXYZ} \doteq N^{N(2\ell - y)} \exp[Nm(l, x, y, z)],
\]

(A.31)

\[
m(l, x, y, z) = y - 2\ell + x \ln 2 - 2h(x) + h(y) - h(z) - h(y - x)
+ 2h(\ell - y) - 2h(l - x - y - z) - h(1 - 2\ell + x + y + z).
\]

(A.32)

A.3.2. Erdős–Rényi ensembles. For both \(G(N, p)\) and \(G(N, M)\) ensembles, the probability \(P_{LXYZ}\) depends only on the number of edges present in the union of the two circuits, \(\mathcal{P}_{LXYZ} = \mathcal{P}_{2L-Y}\). For the non-trivial range of parameters \(\ell, c\) where the first moment is exponentially large, the first term in equation (A.23) can be neglected. The sum over \((X, Y, Z)\) can be evaluated with the saddle-point method, yielding

\[
\mathcal{N}^2_L(G) \doteq \exp[N\tau(\ell)], \quad \tau(\ell) = \max_y[p(2\ell - y) + \hat{m}(\ell, y)],
\]

(A.33)

where

\[
p(\ell) = \begin{cases} 
\ell \ln c & \text{for } G(N, p) \\
\frac{1}{2}h(c) - \frac{1}{2}h(c - 2\ell) - \ell & \text{for } G(N, M),
\end{cases}
\]

(A.34)

and we introduced

\[
\hat{m}(\ell, y) = \max_{x,z} m(\ell, x, y, z)
\]

(A.35)

\[
- 2h(1 - \ell) + y - 2\ell + h(y)
+ \max_x [x \ln 2 - 2h(x) + h(1 - x - y) - h(y - x)
+ 2h(\ell - y) - 2h(l - x - y)].
\]

(A.36)

The range of parameters in the various optimizations are such that \(2\ell - 1 \leq x + y + z \leq \ell\). The step between equations (A.35) and (A.36) amounts to maximizing \(m\) over \(z\), which can be done analytically. It is then very easy to determine the function \(\hat{m}(\ell, y)\) numerically. Finally, defining \(S(\ell) = \tau(\ell) - 2\sigma(\ell)\), we determined this function numerically (see figure 9) and found that \(S > 0\) for all parameters such that \(\sigma > 0\): the second moment of \(\mathcal{N}^2_L(G)\) is then exponentially larger than the square of the first moment, which forbids the use of the second moment method to determine the typical value of \(\mathcal{N}^2_L\).

A.3.3. Arbitrary connectivity distribution and regular. The computation of \(\mathcal{P}_{LXYZ}\) in the configuration model can be done similarly to the one of \(\mathcal{P}_L\) (cf equation (A.13)). To simplify notations let us define \(U = 2L - 3X - Y - 2Z, V = 2X\), and the multinomial coefficient

\[
\binom{N}{U, V, Z} = \frac{N!}{U!V!Z!(N - U - V - Z)!},
\]

(A.37)
for $U + V + Z \leq N$. We also use $(k)_n = k(k - 1) \cdots (k - n + 1)$. With these conventions one finds

$$
\mathcal{P}_{\text{LXYZ}} = \frac{1}{c N^{U,V,Z}} \left( \sum_{U_k, V_k, Z_k} \prod_{k=2}^{N} \left( \binom{N q_k}{U_k, V_k, Z_k} (k)_2^{U_k} (k)_3^{V_k} (k)_4^{Z_k} \right) \right) \frac{(cN - 2(2L - Y) - 1)!!}{(cN - 1)!!}.
$$

(A.38)

Indeed, $U$ (respectively $V$, $Z$) is the number of vertices with two (respectively three, four) half-edges involved in the pattern, and $(U_k, V_k, Z_k)$ the number of such vertices among the ones of degree $k$. In consequence the sum is over non-negative integers with $V_2 = Z_2 = Z_3 = 0$, $U_k + V_k + Z_k \leq N_k$, and $\sum_k U_k = U$, $\sum_k V_k = V$, $\sum_k Z_k = Z$. These last three constraints can be implemented using the complex integral representation of Kronecker’s delta, themselves evaluated by the saddle point method in the thermodynamic limit:

$$
\mathcal{P}_{\text{LXYZ}} \doteq N^{Y-2L} \exp[N p(\ell, x, y, z)],
$$

$$
p(\ell, x, y, z) = \frac{1}{2} h(c + 4\ell + 2y) - \frac{1}{2} h(c) + 2\ell - y + h(2\ell - 3x - y - 2z) + h(2x) + h(z) + h(1 - 2\ell + x + y + z) + \text{ext}_{\theta_1, \theta_2, \theta_3} \left[ \sum_{k=2}^{\infty} q_k \ln(1 + (k)_2 \theta_1 + (k)_3 \theta_2) + (k)_4 \theta_3) - (2\ell - 3x - y - 2x) \ln \theta_1 - 2x \ln \theta_2 - z \ln \theta_3 \right].
$$

(A.39)

Once this function has been determined for a given degree distribution, the exponential order of $\mathcal{N}_L^2(G)$ can be computed as

$$
\mathcal{N}_L^2(G) \doteq \exp[N \tau(\ell)], \quad \tau(\ell) = \max_{x,y,z} [p(\ell, x, y, z) + m(\ell, x, y, z)],
$$

(A.40)

where $m$ is given in equation (A.32).

In the regular case, the maximization over the six parameters can be performed analytically, and yields $\tau(\ell) = 2\sigma(\ell) [6]$, proving the concentration (at the exponential order) of $\mathcal{N}_L(G)$ around its mean. We expect that for any (fast decaying) connectivity distribution not strictly concentrated on a single integer, $\tau(\ell) > 2\sigma(\ell)$ when $\sigma(\ell) > 0$. A proof of this conjecture would be a quite painful exercise in analysis that we did not undertake. We verified this statement numerically however for the Hamiltonian circuits of random graphs with an equal mixture of vertices of degrees three and four, yielding $\tau(1) - 2\sigma(1) \approx 0.002$.

We have been rather loose in treating the algebraic prefactor hidden in $\doteq$ for the various expressions of $\mathcal{N}_L^2(G)$. However, it is rather simple to determine the power of $N$ in this prefactor, collecting the contributions which arise from the Stirling expansions, the transformation of sums into integrals, and the evaluation of the latter with the saddle point method. This leads to

$$
\frac{\mathcal{N}_L^2(G)}{\mathcal{N}_L(G)} = \text{cst} \left( 1 + O(N^{-1}) \right) \exp[N(\tau(\ell) - 2\sigma(\ell))],
$$

(A.41)

as we observed numerically in section 7.

doi:10.1088/1742-5468/2006/06/P06019
Note also that some information on the structure of the space of configurations can be obtained from this kind of computations. The average number of pairs of circuits at a given ‘overlap’ (number of common edges) is indeed obtained from the second moment computations if the parameter \( y \) is kept fixed.

### A.4. On the union of vertex disjoint circuits

In the statistical mechanics treatment of the main part of the text we used a model which counts the number \( N_L'(G) \) of subgraphs of \( G \) made of the union of vertex disjoint circuits of total length \( L \). We want to show in this appendix that, at the leading exponential order, the average of \( N_L'(G) \) equals the one of \( N_L(G) \) in the various ensembles considered in this appendix. Let us denote by \( C'_L \) the set of subgraphs of the complete graph on \( N \) vertices made of unions of vertex disjoint circuits of total length \( L \), and \( M'_L \) its cardinality. As such subgraphs are still made of \( L \) edges connecting \( L \) vertices, \( N_L'(G) = M'_L \mathcal{P}_L \), where the probability \( \mathcal{P}_L \) is the one defined previously for the computation of \( N_L(G) \). Let us define \( M'_{L,A} \), the cardinality of the subset of \( C'_L \) where the subgraphs are made of \( A \) disjoint circuits. A short reasoning leads to

\[
M'_{L,A} = \frac{N!}{(N-L)!} \frac{1}{2^A} \sum_{A_1,\ldots,A_L} \frac{1}{\prod_{i=3}^{L} A_i! A_i!} \delta_{\sum_i A_i} \delta_{\sum_i A_i,L},
\]

where the integers \( A_i \) are the number of circuits of length \( i \) in the subgraph. From this expression it is easy to check that \( M'_{L,1} = M_L \), and that as long as \( A \) is finite in the thermodynamic limit, \( M'_{L,A} \approx M_L \). More precisely, one can show that the leading behaviour of \( M'_L = \sum_A M'_{L,A} \) is not modified by contributions with \( A \) growing with \( N \). Indeed,

\[
\frac{M'_L}{M_L} = 2L [t^L] \exp \left[ \frac{1}{2} \left( \frac{t^3}{3} + \cdots \frac{t^L}{L} \right) \right],
\]

where \([t^x]f(t)\) denotes the coefficient of order \( x \) in the series expansion of \( f(t) \). Evaluating the right hand side with the saddle point method when \( L \to \infty \), one can conclude that \( M'_L \approx M_L \) and from the above remark \( N'_L(G) \approx N_L(G) \). As far as annealed computations are concerned, the distinction between circuits of (extensive) length \( L \) and union of disjoint circuits of total length \( L \) does not modify the entropy. The hypothesis made in the main part of the text is that this remains true for the quenched computations.

### Appendix B. Analysis of the leaf removal algorithm on random graphs with arbitrary connectivity distribution

We want to justify in this appendix the geometric interpretation of the null messages elimination we gave in section 3.2. Consider a random graph drawn uniformly among the ones with the connectivity distribution \( q_k \). The two-core of a graph is the largest subgraph in which all vertices have connectivity at least two. It can be determined using the following leaf removal algorithm, which reduces iteratively the graph. At each time step, if there is at least one vertex of degree one, choose randomly one of them, and remove the unique edge to which it belongs. When there is no vertex of degree one, the algorithm stops. At this point either all the edges have been removed and one is left with \( N \) isolated
vertices, or there remain some isolated vertices and a subgraph in which all vertices have at least degree two, i.e. the two-core of the initial graph.

One can define more generally the $q$-core of a graph as the largest subgraph with minimal degree $q$. For Erdős–Rényi random graphs, the thresholds for the appearance of giant $q$-cores have been obtained in [55]. These results have been recently extended to random graphs with arbitrary connectivity distributions in [56]; this appendix can thus be viewed as an informal presentation of these mathematical works, with the emphasis put on the quantitative results instead of the mathematical rigour (see also [57,58] for a heuristic derivation in the arbitrary connectivity distribution case, and [59,60] for new mathematical treatments of the problem). In the following we shall study the behaviour of the leaf removal algorithm through differential equations for the evolution of the average connectivity distribution along the execution of the leaf removal. This method is widely used in mathematics and computer science; see in particular [61] for a general presentation and a detailed derivation of the equations (B.3).

We shall denote by $T$ the number of steps (elimination of one edge) already performed by the algorithm, and $t = T/N$ the reduced time variable. Let us call $R_k(t) = N\tilde{r}_k(t)$ the average (over the choice of the initial graph and the random decisions taken by the algorithm) number of sites of connectivity $k$ in the residual graph obtained after $Nt$ time steps of the algorithm. The initial condition reads obviously $\tilde{r}_k(t = 0) = q_k$. If one calls $\tilde{r}_k(t)$ the probability that the neighbour of the selected degree one vertex has connectivity $k + 1$, the average evolution of the $R$ during the time step $t \to t + 1/N$ reads

$$R_k(t + 1/N) - R_k(t) = \delta_{k,0} - \delta_{k,1} + \sum_{k'=0}^\infty \tilde{r}_{k'}(t)[-\delta_{k,k'+1} + \delta_{k,k'}]. \quad (B.1)$$

To close this set of equations we have to express the offspring probabilities $\tilde{r}_k(t)$ in terms of the connectivity distribution $r_k(t)$. As the graph is sequentially exposed by the algorithm, the residual graph at time $t$ is still uniformly distributed according to the connectivity distribution $r_k(t)$, hence

$$\tilde{r}_k(t) = \frac{(k + 1)r_{k+1}(t)}{\sum_k kr_k(t)} = \frac{(k + 1)r_{k+1}(t)}{c - 2t}. \quad (B.2)$$

In the last equality $c$ is the initial mean connectivity $\sum_k kr_k$, which is reduced by $2/N$ at each time step. In the thermodynamic limit the discrete time relations (B.1) become ordinary differential equations,

$$\dot{r}_0(t) = 1 + \frac{r_1(t)}{c\eta(t)^2},$$

$$\dot{r}_1(t) = -1 - \frac{r_1(t)}{c\eta(t)^2} + \frac{2r_2(t)}{c\eta(t)^2}, \quad (B.3)$$

$$\dot{r}_k(t) = -\frac{kr_k(t)}{c\eta(t)^2} + \frac{(k + 1)r_{k+1}(t)}{c\eta(t)^2} \quad \text{for } k \geq 2,$$

where dotted quantities are derivatives with respect to time, and we introduced the notation $\eta(t) = \sqrt{1 - (2t/c)}$.

For simplicity let us first assume the existence of a cut-off $k_m$ in the original distribution $q_k$, $q_k = 0$ for $k > k_m$. As the leaf removal procedure never increases

doi:10.1088/1742-5468/2006/06/P06019
the connectivity of one site, this cut-off remains present in \( r_k(t) \) for all times. The equations of rank \( k_m \) in the hierarchy (B.3) are then closed on \( r_{k_m} \). Using the fact that \( \dot{\eta}(t) = -1/(c\eta(t)) \), it can be written as

\[
\dot{r}_{k_m}(t) = -k_m \frac{\dot{\eta}(t)}{\eta(t)} r_{k_m}(t),
\]

and easily integrated with the initial condition \( r_{k_m}(t = 0) = q_{k_m} \) as

\[
r_{k_m}(t) = q_{k_m} \eta(t)^{k_m}.
\]

Now one can prove by a decreasing recurrence on \( k \) from \( k_m \) down to two that

\[
r_k(t) = \sum_{n=k}^{k_m} g_n \binom{n}{k} \eta(t)^k (1 - \eta(t))^{n-k}
\]

solves the hierarchy of equations (B.3). Note also that the initial conditions \( r_k(0) = q_k \) are enforced by equation (B.6) as \( \eta(0) = 1 \). Once \( r_k(t) \) has been computed for \( k \geq 2 \), the equation on \( r_1 \) yields

\[
r_1(t) = -c \eta(t)(1 - \eta(t)) + \sum_{n=1}^{k_m} n g_n \eta(t)(1 - \eta(t))^{n-1}.
\]

Finally, \( r_0(t) \) can be obtained from the normalization condition of the \( r_k \).

We introduced the cut-off \( k_m \) to have an explicit starting point of the downwards recurrence on \( k \). However, the expression (B.6) formally solves the hierarchy of equations (B.3) even for unbounded distributions \( q_k \); we shall therefore send the cut-off to infinity from now on, assuming that all the sums remain convergent.

The two-core is found when the leaf-removal algorithm stops, at the smallest time \( t_* \) for which the number of degree one vertices vanishes, \( r_1(t_*) = 0 \). This equation always admits \( t_* = c/2 \) as a solution (the graph has then been emptied); however, if there is a smaller solution the two-core is non-trivial and the algorithm stops before having removed all edges. Calling \( \eta = \eta(t_*) \), one obtains if \( \eta > 0 \)

\[
1 - \eta = \sum_{k=1}^{\infty} k q_k c (1 - \eta)^{k-1} = \sum_{k=0}^{\infty} \tilde{q}_k (1 - \eta)^k,
\]

which is nothing but equation (26) on the fraction of non-vanishing messages obtained in the main part of the text. The confirmation of the interpretation given in section 3.2 follows easily:

- the number of edges in the two-core is equal to the initial number of edges minus the number of steps performed by the algorithm before stopping, \( M_{\text{core}} = M - N t_* = M \eta^2 \);
- the distribution of the connectivities of the sites in the two-core is \( r_k(t_*) \), as expected from equation (28).

For completeness we also give the number of sites in the two-core:

\[
N_{\text{core}} = N \ell_{\text{core}}, \quad \ell_{\text{core}} = \sum_{k=2}^{\infty} r_k(t_*) = 1 - \sum_{k=0}^{\infty} q_k(1 - \eta)^k - cn(1 - \eta).
\]
On the number of circuits in random graphs

As it should, this number is smaller than the size of the giant component, which reads [29, 30]

\[ N_{\text{giant}} = N \left( 1 - \sum_{k=0}^{\infty} q_k (1-\eta)^k \right). \] (B.10)

Moreover the fraction of sites which are in the giant component but out of the two-core is proportional to \( \eta (1-\eta) \). Indeed, the corresponding edges bear exactly one non-null directed message, in the formalism of section 3.2: if both messages were non-null the edge would be in the two-core; if both vanished the edge would be out of the giant component.

Appendix C. An alternative derivation of \( \ell_{\text{max}} = 1 \) when the minimal connectivity is three

This appendix presents, as a consistency check, another derivation of the identity \( \ell_{\text{max}} = 1 \) for a random graph ensemble with a minimal connectivity of three. In the main part of the text (section 5.1), we obtained it by inspection on the behaviour of the free energy in the large \( u \) limit, because of the absence of hard fields. There exists however another expression of \( \ell_{\text{max}} \), in terms of the distribution of evanescent fields, obtained by taking the large \( u \) limit in equation (24):

\[ \ell_{\text{max}} = \frac{c}{2} \int_{0}^{\infty} dx_1 V_0(x_1) dx_2 V_0(x_2) \frac{x_1 x_2}{1 + x_1 x_2}. \] (C.1)

Let us introduce the following functional of any probability distribution law \( A \):

\[ F_k[A](x) = \int_{0}^{\infty} dx_1 A(x_1) \ldots dx_k A(x_k) \delta(x - h_k(x_1, \ldots, x_k)), \] (C.2)

such that equation (53) can be rewritten in a compact way as \( V_0 = \sum_k \tilde{q}_k F_k[V_0] \), and a bilinear form on the space of probability distribution functions,

\[ \langle A, B \rangle = \int_{0}^{\infty} dx \, dy A(x) B(y) \frac{xy}{1 + xy}. \] (C.3)

Consider now this form with its arguments being a distribution \( A \) and its image through the functional \( F_k \):

\[ \langle A, F_k[A] \rangle = \int_{0}^{\infty} dx_0 A(x_0) dx_1 A(x_1) \ldots dx_k A(x_k) \frac{x_0 h_k(x_1, \ldots, x_k)}{1 + x_0 h_k(x_1, \ldots, x_k)}. \] (C.4)

The rational fraction in the integral can be transformed in the following way:

\[ \frac{x_0 h_k(x_1, \ldots, x_k)}{1 + x_0 h_k(x_1, \ldots, x_k)} = \frac{x_0 \sum_{i=1}^{k} x_i}{x_0 \sum_{i=1}^{k} x_i + \sum_{1 \leq i < j \leq k} x_i x_j} = \frac{x_0 \sum_{i=1}^{k} x_i}{\sum_{0 \leq i < j \leq k} x_i x_j}. \] (C.5)

Both the denominator of this fraction and the integration measure \( \prod_{i=0}^{k} dx_i A(x_i) \) being invariant under the permutations of the \( k+1 \) \( x_i \), the integral can be computed by symmetrizing the numerator of the fraction. The normalization of \( A \) then gives

\[ \langle A, F_k[A] \rangle = \frac{2}{k+1}. \] (C.6)
The proof of $\ell_{\text{max}} = 1$ is now straightforward:
\[
\ell_{\text{max}} = \frac{c}{2} \langle V_0, V_0 \rangle = \sum_{k=2}^{\infty} \frac{c\tilde{q}_k}{2} \langle V_0, F_k[V_0] \rangle. \quad (C.7)
\]

Using the identity (C.6) and the relation between $\tilde{q}$ and $q$ (cf equation (19)), $\ell_{\text{max}}$ is found to be the sum of $q_k$ for $k \geq 3$, and hence is equal to unity by normalization.

We also verified numerically that in the presence of degree two vertices, and hence of non-trivial hard fields, the limit of equation (24), which involves both evanescent and hard fields, coincides with the expression equation (70) in terms of hard fields only. We believe this could be proved analytically, yet we have not found a simple way to do it.

References

[1] Bollobás B, 2001 Random Graphs (Cambridge: Cambridge University Press)
[2] Janson S, Luczak T and Rucinski A, 2000 Random Graphs (New York: Wiley)
[3] Wormald N C, Models of random regular graphs, 1999 Surveys in Combinatorics (London Mathematical Society Lecture Note Series vol 276) ed J D Lamb and D A Preece (Cambridge: Cambridge University Press) p 239
[4] Robinson R W and Wormald N C, 1994 Random Struct. Alg. 5 363
[5] Janson S, 1995 Comb. Probab. Comput. 4 369
[6] Garmo H, 1999 Random Struct. Alg. 15 43
[7] Janson S, 2003 Combin. Probab. Comput. 12 27
[8] Luczak T, 1991 Random. Struct. Alg. 2 421
[9] Flajolet P, Knuth D E and Pittel B, 1989 Discrete Math. 75 167
[10] Frieze A, 1986 Discrete Math. 59 243
[11] Pósa L, 1976 Discrete Math. 14 359
[12] Johnson D B, 1975 SIAM J. Comput. 4 77
[13] Garey M R and Johnson D S, 1983 Computers and Intractability: A Guide to the Theory of NP-Completeness (New York: Freeman)
[14] Dyer M E, Frieze A and Jerrum M R, 1998 SIAM J. Comput. 27 1262
[15] Albert R and Barabási A-L, 2002 Rev. Mod. Phys. 74 47
[16] Bianconi G, Caldarelli G and Capocci A, 2005 Phys. Rev. E 71 066116
[17] Rozenfeld H D, Kirk J E, Boltt E M and Ben-Avraham D, 2005 J. Phys. A: Math. Gen. 38 4589
[18] Ben-Naim E and Krapivsky P L, 2005 Phys. Rev. E 71 026129
[19] Bianconi G and Marsili M, 2005 J. Stat. Mech. P06005
[20] Mézard M and Parisi G, 2001 Eur. Phys. J. B 20 217
[21] Marinari E, Monasson R and Semerjian G, 2006 Europhys. Lett. 73 8
[22] Klemm K and Stadler P F, 2006 Phys. Rev. E 73 025101(R)
[23] Yedidia J S, Freeman W T and Weiss Y, 2001 Adv. Neural Inform. Processing Syst. 13 689
[24] Tatikonda S and Jordan M, 2002 Proc. UAI-2002 p 493
[25] Heskes T, 2004 Neural Comput. 16 2379
[26] http://www.netdimes.org/
[27] Achlioptas D and Peres Y, 2004 J. Am. Math. Soc. 17 947
[28] Mérzard M, Parisi G and Virasoro M A, 1987 Spin-glass Theory and Beyond (Singapore: World Scientific)
[29] Molloy M and Reed B, 1998 Combin. Probab. Comput. 7 295
[30] Newman M E J, Strogatz S H and Watts D J, 2001 Phys. Rev. E 64 026118
[31] Marinari E and Monasson R, 2004 J. Stat. Mech. P09004
[32] Müller M, Mézard M and Montanari A, 2004 J. Phys. A: Math. Gen. 37 2073
[33] Rivoire O, Biroli G, Martin O C and Mézard M, 2004 Eur. Phys. J. B 37 55
[34] Castellani T, Kraakala F and Ricci-Tersenghi F, 2005 Eur. Phys. J. B 47 99
[35] Pichon J P and Mézard M, 1997 J. Phys. A: Math. Gen. 30 7997
[36] Guerra F, 2003 Commun. Math. Phys. 233 1
[37] Talagrand M, 2006 Ann. Math. 163 221
[38] Franz S and Leone M, 2003 J. Stat. Phys. 111 535
[39] Franz S, Leone M and Toninelli F L, 2003 J. Phys. A: Math. Gen. 36 10967

doi:10.1088/1742-5468/2006/06/P06019
On the number of circuits in random graphs

[41] Aldous D and Steele J M, 2003 Discrete Combinatorial Probability ed H Kesten (Berlin: Springer)
[42] Bandyopadhyay A and Gamarnik D, 2005 Preprint math.PR/0510471
[43] Montanari A and Rizzo T, 2005 J. Stat. Mech. P10011
[44] Parisi G and Slanina F, 2005 Preprint cond-mat/0512529
[45] Chertkov M and Chernyak V Y, 2006 Preprint cond-mat/0601487
[46] Rivoire O, 2005 J. Stat. Mech. P07004
[47] Mézard M and Zecchina R, 2002 Phys. Rev. E 66 056126
[48] Barabási A-L and Albert R, 1999 Science 286 509
[49] Bianconi G and Marsili M, 2005 Preprint cond-mat/0511283
[50] Karp R M and Sipser M, 1981 Proc. FOCS 1981 364
[51] Aronson J, Frieze A and Pittel B, 1998 Random Struct. Alg. 12 111
[52] Zdeborová L and Mézard M, 2006 Preprint cond-mat/0603350
[53] Bollobás B, Kim J H and Verstraëte J, 2006 Random Struct. Alg. 29 1
[54] Pretti M and Weigt M, Sudden emergence of q-regular subgraphs in random graphs, 2006 Preprint cond-mat/0603819
[55] Pittel B, Spencer J and Wormald N C, 1996 J. Comb. Theory Ser. B 67 111
[56] Fernholz D and Ramachandran V, 2004 UTCS technical Report TR04-13 http://www.cs.utexas.edu/~vlr/pubs.html
Fernholz D and Ramachandran V, 2003 Preprint
[57] Dorogovtsev S N, Goltsev A V and Mendes J F F, 2006 Phys. Rev. Lett. 96 040601
[58] Goltsev A V, Dorogovtsev S N and Mendes J F F, 2006 Preprint cond-mat/0602611
[59] Janson S and Luczak M J, 2005 Preprint math.CO/0508453
[60] Riordan O, 2005 Preprint math.CO/0511093
[61] Wormald N C, The differential equation method for random graph processes and greedy algorithms, 1999 Lectures on Approximation and Randomized Algorithms ed M Karonski and H J Proemel (Warsaw: PWN) p 73
[62] Bollobás B, 1980 Eur. J. Combin. 1 311

doi:10.1088/1742-5468/2006/06/P06019