Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

The number of matchings in random graphs

Lenka Zdeborová and Marc Mézard

CNRS, Université Paris-Sud, UMR 8626, LPTMS, Orsay Cedex, F-91405, France
E-mail: zdeborov@lptms.u-psud.fr and mezard@lptms.u-psud.fr

Received 14 March 2006
Accepted 12 April 2006
Published 11 May 2006

Online at stacks.iop.org/JSTAT/2006/P05003
doi:10.1088/1742-5468/2006/05/P05003

Abstract. We study matchings on sparse random graphs by means of the cavity method. We first show how the method reproduces several known results about maximum and perfect matchings in regular and Erdös–Rényi random graphs. Our main new result is the computation of the entropy, i.e. the leading order of the logarithm of the number of solutions, of matchings with a given size. We derive both an algorithm to compute this entropy for an arbitrary graph with a girth that diverges in the large size limit, and an analytic result for the entropy in regular and Erdös–Rényi random graph ensembles.

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

ArXiv ePrint: cond-mat/0603350

 1742-5468/06/P05003+24$30.00

Contents

1. Introduction

We study a classical problem of graph theory, namely the size and number of matchings on various types of random graphs. This problem has been intensively studied for a long time by mathematicians and computer scientists [1]. Here we address it using some techniques developed in the statistical mechanics of spin glasses [2]. Such approaches have been used in recent years to describe successfully the typical cases of random combinatorial problems, e.g. weighted matching (or assignment) [3], the travelling salesman problem [4], the vertex cover on random graphs [5], K-satisfiability [6, 7], or the colouring of random graphs [8].

Here we apply the cavity method [9] to describe the matchings on ensembles of sparse random graphs with a given degree distribution. We work within the replica symmetric (RS) version of the cavity method, and we argue that it gives exact results for these problems. In fact we show how the method reproduces several known results about the size of the maximum matching (which is also the maximum number of self-avoiding dimers) and the existence of the perfect matchings (the possibility of covering the graph with $N/2$ dimers). This also confirms the previous result by Zhou and Ou-Yang [10], who also used the cavity method, but in a different way (we discuss below the differences of our approaches).

Our main new result is the computation of the entropy, i.e. the leading order of the logarithm of the number of solutions, of matchings with a given size in large sparse random graphs. We derive both an algorithm to compute this entropy for arbitrary graphs with a girth (the length of the shortest graph cycle) that diverges in the large size limit, and an analytic result for the entropy in regular and Erdös–Rényi random graph ensembles.

The cavity method is not yet proved to be a rigorous tool; however, it is widely believed—at least by physicists—to be exact, and in some cases its predictions have been confirmed rigorously. Let us mention the work of Talagrand [11], who, using some of the tools developed by Guerra [12], proved the validity of the Parisi formula for the partition function of Sherrington–Kirkpatrick model (Parisi's original work [13] uses the replica method, but it can be reformulated in cavity terms [2]). Aldous [14] developed the local weak convergence method and proved the $\zeta(2)$-limit for the random assignment problem, initially found in [3]. In this same problem, Bayati *et al* [15] proved the convergence of a 'belief propagation' algorithm, which is basically the replica symmetric cavity method, for finding the lowest weight assignment in generic bipartite graphs. Recently, Bandyopadhyay and Gamarnik [16] have used this local weak convergence strategy to derive some results on the entropies in the problems of graph colouring and independent sets, in regions of parameters where the RS cavity solution is the correct one. The local weak convergence method was also used for weighted matchings in sparse random graphs in [17].

Because of these recent developments, and of the simple replica symmetric nature of the matching problem, we believe that it should be possible to turn all our results into rigorous statements. We hope that this work will also turn out to be useful in the opposite direction, i.e. that working on rigorous proofs of our results for matching will help to develop the rigorous version of the cavity method.

The matching problem on a graph is equivalent to a physical model of dimers. This was mostly studied on planar graphs (lattices), where there is a beautiful method by Kasteleyn [18], which shows how to count exactly dimer arrangements (perfect matchings). On non-planar regular graphs a Bethe mean field approximation, which is known to be exact on a Bethe lattice, has been developed in [19], and references therein. Our work generalizes these results and gives the solution of dimer models on sparse random graphs.

The paper is organized as follows. In section 2 we set up our notations and overview the main known results for the matching on sparse random graphs. In section 3 we introduce the cavity approach to the matching problem and derive the size of maximum matching and the entropy of matchings of a given size on a typical random graph. We also describe approximate polynomial algorithms for sampling and counting matchings on a given graph. In section 4 we give results for the size and the number of matchings for the ensemble of regular and Erdös–Rényi random graphs, and we show that the replica symmetric ansatz is stable for these two ensembles. In section 5 we discuss the alternative 1RSB solution at zero temperature which was obtained by Zhou and Ou-Yang [10]. The conclusion summarizes this work and gives some perspective on how it could be turned into rigorous results.

2. Background and notations

Consider a graph $G(V, E)$ with N vertices ($N = |V|$) and a set of edges E. A *matching* of G is a subset of edges $M \subset E$ such that each vertex is incident with at most one edge

J. Stat. Mech. (2006) P05003

in M. In other words, the edges in the matching M do not touch each other. The *size of the matching*, $|M|$, is the number of edges in M. We denote the size of the maximum possible matching by $|M^*|$. The trivial relation $|M^*| \leq N/2$ follows from the definition. If a maximum matching covers all the vertices, $|M^*| = N/2$, we call M^* a *perfect matching*.

Finding a maximum matching in a given graph G is a polynomial problem. For instance, the algorithm of [20] solves this problem with a computational complexity proportional to $\mathrm{O}(|E|\sqrt{|V|})$.

How many matchings of size $|M^*|$ can we actually find in G? No exact polynomial algorithm to answer this question is known. Counting the number of matchings of a given size was proven [21] to belong to the $\#P$-complete (sharp P-complete) class of problems. It means that if an exact polynomial algorithm for this problem existed we could also count solutions of all the other problems belonging to the NP class. It is generally believed that a polynomial procedure to solve $\#P$-complete problems does not exist. For this reason it is very useful to develop methods to count matchings fast (in polynomial time) but only approximately. Several works have been done in this direction [22]–[24].

In this paper we study not only properties of matchings on a given graph G but also on ensembles $\mathcal{G}$ of large sparse random graphs. When we claim that a property A is true for a typical random graph $G \in \mathcal{G}$ we mean that when G is chosen from the ensemble with its natural probability law, the probability that A is true goes to one as the size of G grows to infinity.

2.1. Rigorous results for matching on random graphs

In this section we give some known rigorous results for matchings on random graphs. From the point of view of matching, the simplest ensemble is the one of r-regular random graphs, i.e. all graphs where every vertex has degree (number of neighbours) r. In this ensemble the measure is uniform over all r-regular graphs.

Theorem 1 ([25]). *If $r \geq 3$ and the number of vertices N is even then almost every r-regular graph has a perfect matching. Denote by $\mathcal{N}_G$ the number of perfect matchings of an r-regular graph G. Then its first two moments are*

$$\mathbb{E}(\mathcal{N}_G) \approx \sqrt{2}e^{1/4}[(r-1)^{r-1}/r^{r-2}]^{N/2}, \tag{1}$$

$$\mathbb{E}[(\mathcal{N}_G)^2] \approx \sqrt{\frac{r-1}{r-2}}\mathrm{e}^{-(2r-1)/4(r-1)^2}\mathbb{E}(\mathcal{N}_G)^2. \tag{2}$$

In the statistical physics language we call the logarithm of the first moment $\log[\mathbb{E}(\mathcal{N}_G)]$ the *annealed entropy* of perfect matchings and the typical average $\mathbb{E}[\log \mathcal{N}_G]$[1] the *quenched entropy* of perfect matchings. Due to the concavity of logarithmic function the upper bound for the quenched entropy follows from (1)

$$\mathbb{E}[\log \mathcal{N}_G] \leq \log[\mathbb{E}(\mathcal{N}_G)] = N[(r-1)\log(r-1) - (r-2)\log r]/2 + \mathrm{O}(1). \tag{3}$$

[1] We should write $\mathbb{E}[\log(\mathcal{N}_G + 1)]$ for the quenched entropy to avoid $-\infty$ for graphs which do not have any perfect matching.

In section 4.1 we will show that for r-regular graphs the quenched entropy is in fact the same as the annealed one, i.e. in (3) the bound is tight. Note that the fact that $\mathbb{E}[(\mathcal{N}_G)^2] \sim [\mathbb{E}(\mathcal{N}_G)]^2$ to leading exponential order is not enough to prove that the quenched and annealed entropy are equal.

In this paper we will be interested in random graphs with a fixed degree distribution: we call $\mathcal{Q}(k)$ the probability that a randomly chosen vertex has degree k, in the asymptotic limit of large graphs. In particular, in Erdös–Rényi (ER) random graphs, where every edge is present with probability $p = c/(N-1)$, the degree sequence is Poissonian $\mathcal{Q}(k) = \mathrm{e}^{-c}c^k/k!$. Because of the existence of a fraction e^{-c} of isolated vertices perfect matchings almost surely do not exist in ER graphs. The size of maximum possible matching was computed in a seminal paper of Karp and Sipser [26].

Theorem 2 ([26]). *The maximum matching in an Erdös–Rényi random graph with N sites and mean degree c has on average size*

$$\mathbb{E}(|M^*|) = \frac{1 - p_1(c) + p_2(c) - cp_1(c) + cp_1(c)p_2(c)}{2} N, \tag{4}$$

where $p_1(c)$ is the smallest solution of equation $p = \exp[-c\exp(-cp)]$ and $p_2(c) = 1 - \exp[-cp_1(c)]$.

When $c < e$ there is only one solution for $p_1(c)$. When $c \geq e$ another pair of solutions for $p_1(c)$ appears.

2.2. Karp–Sipser leaf removal, the core

The Karp–Sipser theorem was originally proven by analysing a greedy leaf removal algorithm [26]. This algorithm consists of two steps.

(1) Given a graph G, if there are leaves choose randomly one of them i and its incident edge (ij). Put this edge to the matching and remove the two vertices i and j. Delete at the same time all the edges incident with j. Repeat until there are no leaves.

(2) If there are no leaves in G choose randomly an edge (ij) with uniform probability, add it to the matching and erase all the edges incident with i and j. Go to step (1).

We define as a *core of the graph* all the non-single vertices (and edges between them) which remain in the graph after the first step of the leaf removal procedure. The core does not depend on the history of the leaf removal [27]. Karp and Sipser proved that for $c \leq e$ the core is small (zero asymptotically), whereas for $c > e$ the core covers a finite fraction of all the vertices. They also proved that when a large (or order N) core exists, asymptotically all its nodes can be matched.

We call $v(c)$ the fraction of vertices in the core of a typical ER random graph of average degree c, $l(c)N$ the number of edges in the core, and $m(c)N$ the number of edges matched in the leaf removal procedure. It is known [27] that

$$v(c) = p_3(1 - cp_1), \qquad l(c) = \frac{c}{2}p_3^2, \qquad m(c) = p_2 - \frac{c}{2}p_1^2, \tag{5}$$

where p_1 and p_2 are the same parameters as in theorem 2 of Karp and Sipser, and $p_3 = 1 - p_1 - p_2$.

Properties of the core were also studied in [28] and [29]. From these results it follows that the degree distribution in the core is Poissonian-like

$$\mathcal{Q}(0) = \mathcal{Q}(1) = 0\,, \qquad \mathcal{Q}(k) = \frac{\mathrm{e}^{-cp_3}(cp_3)^k}{\mathcal{C}k!} \qquad \text{for } k > 1, \tag{6}$$

where $\mathcal{C}$ is a normalization constant. We will study connections between the Karp–Sipser leaf removal and our method in section 3.4.1.

2.3. The annealed average

We denote by $\mathcal{N}_G(x)$ the number of matchings of size $|M| = xN/2$ in a graph G. Its expectation $\mathbb{E}[\mathcal{N}_G(x)]$ in the random r-regular graph ensemble can be computed as the number of all possible matchings of size $|M|$ times the leading order of the number of all r-regular graphs which contain a given matching M, divided by the leading order of the number of all r-regular graphs. We keep in mind that r is finite and $N \to \infty$. This gives the annealed entropy:

$$\frac{\log \mathbb{E}[\mathcal{N}_G(x)]}{N} = \left(x - \frac{r}{2}\right)\log r + \frac{r-x}{2}\log\left(r-x\right) - (1-x)\log\left(1-x\right) - \frac{x}{2}\log x. \tag{7}$$

Again, thanks to the concavity of the logarithm, the quenched entropy cannot be larger than the annealed one, $\mathbb{E}[\log(\mathcal{N})]/N \le \log \mathbb{E}(\mathcal{N})/N$. We will see in section 4.1 that for r-regular graphs this upper bound is actually tight.

In the ensemble of ER random graphs the expectation of the number of matchings of size $|M| = xN/2$ is computed in the very same way and reads

$$\mathbb{E}[\mathcal{N}_G(x)] \approx \exp\left\{\frac{Nx}{2}\left[\ln\frac{c}{x} - 1 - 2\left(\frac{1}{x} - 1\right)\ln\left(1-x\right)\right]\right\}. \tag{8}$$

If the exponent is negative (which happens for $c < e$ and x sufficiently large) then there is almost surely no matching of size $|M|$ in graph G. On the other hand, if the exponent is positive then equation (8) provides us with an upper bound on the quenched (typical) entropy

$$\frac{\mathbb{E}\{\log\left[\mathcal{N}_G(x)\right]\}}{N} \le \frac{\log \mathbb{E}(\mathcal{N})}{N} = \frac{x}{2}\left[\ln\frac{c}{x} - 1 - 2\left(\frac{1}{x} - 1\right)\ln\left(1-x\right)\right]. \tag{9}$$

For ER random graphs the bound is not tight. From equation (8) we see that for $c > e$ the average number of perfect matchings ($x = 1$) is exponentially large. But we know that for a typical ER graph no perfect matching exists (due to the presence of isolated vertices). The reason is that $\mathbb{E}[\mathcal{N}_G(x)]$ is dominated by a few exceptional graphs G which have a huge number of perfect matchings. The correct quenched average $\mathbb{E}\{\log\left[\mathcal{N}_G(x)\right]\}/N$ will be computed in section 4.2.

3. Cavity method: general formalism

3.1. Statistical physics description

We describe a matching by the variables $s_i = s_{(ab)} \in \{0, 1\}$ assigned to each edge $i = (ab)$ of G, with $s_i = 1$ if $i \in M$ and $s_i = 0$ otherwise. The hard constraints that two edges

doi:10.1088/1742-5468/2006/05/P05003

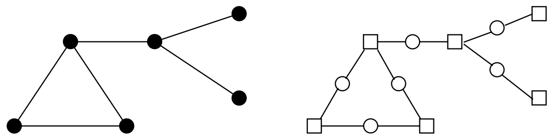

Figure 1. On the left, an example of a graph with six nodes and six edges. On the right, the corresponding factor graph with six functional nodes (squares) and six variable nodes (circles).

in a matching cannot touch impose that, on each vertex $a \in V$, $\sum_{b,(ab)\in E} s_{(ab)} \leq 1$. To complete our statistical physics description, we define for each given graph G an energy (or cost) function which gives, for each matching $M = \{s\}$, the number of unmatched vertices:

$$E_G(M = \{s\}) = \sum_a E_a(\{s\}) = N - 2|M|, \tag{10}$$

where $E_a = 1 - \sum_b s_{(ab)}$. The Boltzmann probability law in the space of matchings is defined by

$$P_G(M) = \frac{1}{Z_G(\beta)} \mathrm{e}^{-\beta E_G(M)}, \tag{11}$$

where β is the inverse temperature and $Z_G(\beta)$ is the partition function.

We use a factor graph representation [30] of the Boltzmann probability (11). With a graph G we associate a factor graph $\mathcal{F}(G)$ as follows (see figure 1): to each edge of G corresponds a variable node (circle) in $\mathcal{F}(G)$; to each vertex of G corresponds a function node (square) in $\mathcal{F}(G)$. We shall index the variable nodes by indices $i, j, k, \ldots$ and function nodes by $a, b, c, \ldots$. The variable i takes value $s_i = 1$ if the corresponding edge is in the matching, and $s_i = 0$ if it is not. For a given configuration $\mathbf{s} = \{s_1, \ldots, s_{|E|}\}$, the weight of function node a is

$$\psi_a(\mathbf{s}) = \mathbb{I}\left(\sum_{i\in V(a)} s_i \leq 1\right) \mathrm{e}^{-\beta(1-\sum_{i\in V(a)} s_i)}, \tag{12}$$

where $V(a)$ is the set of all the variable nodes which are neighbours of function node a, and the total Boltzmann weight of a configuration is $\prod_a \psi_a(\mathbf{s})/Z_G(\beta)$. Later on, when confusion cannot occur, we denote $V(a)$ just as a.

We want to compute the internal energy $E_G(\beta)$ (the expectation value of the number of unmatched vertices) and the entropy $S_G(\beta)$ (the logarithm of the number of matchings). For $\beta \to \infty$ (zero temperature limit) these two quantities give the ground state properties, i.e. respectively the size and entropy of the maximum size matchings.

We are interested in the 'thermodynamic' limit of large graphs ($N \to \infty$), and we shall compute expectations over ensembles of graphs of the densities of thermodynamical potentials $\epsilon(\beta) = \mathbb{E}[E_G(\beta)]/N$ and $s(\beta) = \mathbb{E}[S_G(\beta)]/N$, as well as the average free energy density

$$f(\beta) = \frac{-1}{\beta N}\mathbb{E}[\log Z_G(\beta)] = \frac{1}{N}\mathbb{E}[F_G(\beta)] = \epsilon(\beta) - \frac{1}{\beta}s(\beta). \tag{13}$$

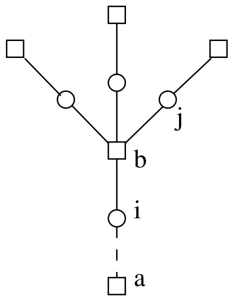

Figure 2. Part of the factor graph used to compute $P_{s_i}^{i\to a}$.

The reason for this interest is that one expects, for reasonable graph ensembles, $F_G(\beta)$ to be self-averaging. This means that the distribution of $F_G(\beta)/N$ becomes more and more sharply peaked around $f(\beta)$ when N increases.

3.2. The cavity method at finite temperature

In the following we use the cavity method at the replica symmetric (RS) level, as described in [9]. We introduce a 'cavity' in the factor graph by deleting the function node a and its incident edges, and we denote by $P_{s_i}^{i\to a}$ the probability that variable i takes value s_i (see figure 2). Because of the local tree-like structure of the (sparse) graphs that we study, it is reasonable to assume that the $P_{s_j}^{j\to b}$ for $j \in b-i$ are uncorrelated. This is the main assumption of the cavity method at the RS level (see [9]) and we will check later on its self-consistency. Using this assumption one gets

$$P_{s_i}^{i\to a} = \frac{1}{\mathcal{C}^{i\to a}} \sum_{\{s_j\}} \mathbb{I}\left(s_i + \sum_{j\in b-i} s_j \le 1\right) \mathrm{e}^{-\beta(1-s_i-\sum s_j)} \prod_{j\in b-i} P_{s_j}^{j\to b}, \tag{14}$$

where $\mathcal{C}^{i\to a}$ is a normalization constant.

For every edge between a variable i and a function node a, we define a cavity field $h^{i\to a}$ as

$$\mathrm{e}^{-\beta h^{i\to a}} \equiv \frac{P_0^{i\to a}}{P_1^{i\to a}}. \tag{15}$$

The recursion relation between cavity fields is then

$$h^{i\to a} = -\frac{1}{\beta} \log\left[\mathrm{e}^{-\beta} + \sum_{j\in b-i} \mathrm{e}^{\beta h^{j\to b}}\right]. \tag{16}$$

This is one form of the 'belief propagation' equations [30, 31]. The cavity fields can be interpreted as messages living on the edges of the factor graph, with some consistency rules on the function nodes, and one can try to solve them by an iterative 'message passing' procedure.

Assuming that one has found the cavity fields, one can deduce from them the various marginal probabilities and the free energy. For instance, the expectation value (with respect to Boltzmann's distribution) of the occupation number s_i of a given edge $i = (ab)$

is given by

$$\langle s_i \rangle = \frac{1}{1 + \mathrm{e}^{-\beta(h^{i\to a} + h^{i\to b})}}. \tag{17}$$

To compute the free energy we first define the free energy shift $\Delta F_{a+i\in V(a)}$ after addition of a function node a and all the edges i around it, and the free energy shift ΔF_i after addition of an edge $i = (ab)$. These are given by

$$\mathrm{e}^{-\beta \Delta F_{a+i\in V(a)}} = \mathrm{e}^{-\beta} + \sum_{i\in a} \mathrm{e}^{\beta h^{i\to a}}, \tag{18}$$

$$\mathrm{e}^{-\beta \Delta F_i} = 1 + \mathrm{e}^{\beta(h^{i\to a} + h^{i\to b})}. \tag{19}$$

The total free energy is then [9, 32]

$$F_G(\beta) = \sum_a \Delta F_{a+i\in V(a)} - \sum_i \Delta F_i. \tag{20}$$

This form of free energy is variational, i.e. the derivative $\partial(\beta F_G(\beta))/\partial h^{i\to a}$ vanishes if and only if the fields $h^{i\to a}$ satisfy (16). This allows us to compute easily the internal energy

$$\begin{aligned} E_G(\beta) &= \sum_a \frac{\mathrm{e}^{-\beta} - \sum_{i\in a} h^{i\to a} \mathrm{e}^{\beta h^{i\to a}}}{\mathrm{e}^{-\beta} + \sum_{i\in a} \mathrm{e}^{\beta h^{i\to a}}} + \sum_i \frac{(h^{i\to a} + h^{i\to b}) \mathrm{e}^{\beta(h^{i\to a} + h^{i\to b})}}{1 + \mathrm{e}^{\beta(h^{i\to a} + h^{i\to b})}} \\ &= N - 2\sum_i \langle s_i \rangle = \sum_a \frac{1}{1 + \sum_{i\in a} \mathrm{e}^{\beta(1 + h^{i\to a})}}. \end{aligned} \tag{21}$$

The second and third equalities have been derived using equation (16). In the last term we can recognize the probability that a node a is not matched. The entropy is then obtained as

$$S_G(\beta) = \beta[E_G(\beta) - F_G(\beta)]. \tag{22}$$

All the equations (14)–(22) hold on a single large sparse graph G. In section 3.4 we will describe how to use them to build algorithms for counting and sampling the matchings on a given graph.

We now study the typical instances in an ensemble of graphs. We denote the average over the ensemble by $\mathbb{E}(\cdot)$. We assume that the random graph ensemble is given by a prescribed degree distribution $\mathcal{Q}(k)$. Let us call $\mathcal{P}_\beta(h)$ the distribution of cavity fields over all the edges of a large typical graph from the graph ensemble. It satisfies the following self-consistent equation:

$$\mathcal{P}_\beta(h) = \sum_{k=1}^{\infty} \frac{k}{c} \mathcal{Q}(k) \int \prod_{i=1}^{k-1} [\mathrm{d}h^i \mathcal{P}_\beta(h^i)]\, \delta\left[h + \frac{1}{\beta} \log\left(\mathrm{e}^{-\beta} + \sum_i \mathrm{e}^{\beta h^i}\right)\right]. \tag{23}$$

The term $k\mathcal{Q}(k)/c$ is the normalized degree distribution of the function node a when one picks up uniformly at random an edge $a - i$ from the factor graph; $c = \sum_k k\mathcal{Q}(k)$ is the mean degree. This equation for distribution $\mathcal{P}_\beta(h)$ can be solved numerically by a technique of population dynamics [9].

J. Stat. Mech. (2006) P05003

The average of the free energy density is then

$$f(\beta) = \frac{\mathbb{E}[F_G(\beta)]}{N} = -\frac{1}{\beta}\sum_{k=0}^{\infty}\mathcal{Q}(k)\int\prod_{i=1}^{k}\left[\mathrm{d}h^i\mathcal{P}_\beta(h^i)\right]\log\left(\mathrm{e}^{-\beta}+\sum_i \mathrm{e}^{\beta h^i}\right)$$
$$+\frac{c}{2\beta}\int \mathrm{d}h^1\,\mathrm{d}h^2\,\mathcal{P}_\beta(h^1)\,\mathcal{P}_\beta(h^2)\log\left(1+\mathrm{e}^{\beta(h^1+h^2)}\right). \quad (24)$$

This expression for the free energy is in its variational form (see [9]), i.e. the functional derivative $\delta f(\beta)/\delta\mathcal{P}_\beta(h)$ vanishes if and only if $\mathcal{P}_\beta$ satisfies (23). The average energy density is then equal to

$$\epsilon(\beta) = \sum_{k=0}^{\infty}\mathcal{Q}(k)\int\prod_{i=1}^{k}\left[\mathrm{d}h^i\mathcal{P}_\beta(h^i)\right]\frac{\mathrm{e}^{-\beta}-\sum_i h^i\mathrm{e}^{\beta h^i}}{\mathrm{e}^{-\beta}+\sum_i \mathrm{e}^{\beta h^i}}$$
$$+\frac{c}{2}\int \mathrm{d}h^1\mathrm{d}h^2\mathcal{P}_\beta(h^1)\mathcal{P}_\beta(h^2)\frac{(h^1+h^2)\mathrm{e}^{\beta(h^1+h^2)}}{1+\mathrm{e}^{\beta(h^1+h^2)}}. \quad (25)$$

The average entropy density is

$$s(\beta) = \beta[\epsilon(\beta) - f(\beta)]. \quad (26)$$

All our computations up to now rely on the only assumption (the ‘RS cavity assumption’) that the neighbours of a node in a cavity are uncorrelated. A necessary condition for the validity of this assumption is that the following nonlinear (spin-glass) susceptibility be finite [33, 34]:

$$\chi_{\mathrm{SG}} = \sum_i \mathbb{E}(\langle s_0 s_i\rangle_c^2) = \sum_{d=0}^{\infty}\alpha^d\,\mathbb{E}(\langle s_0 s_d\rangle_c^2). \quad (27)$$

Here $\langle s_0 s_i\rangle_c$ is the connected correlation function between reference edge 0 and edge i, α^d is the average number of vertices at distance d from the edge 0, for general degree distribution $\alpha = \sum_{k=0}^{\infty} k(k+1)\mathcal{Q}(k+1)/c$. The susceptibility is finite if and only if

$$\overline{\lambda}_T = \lim_{d\to\infty}\alpha\left[\mathbb{E}(\langle s_0 s_d\rangle_c^2)\right]^{1/d} < 1. \quad (28)$$

We will call $\overline{\lambda}_T$ the finite temperature stability parameter. A necessary condition for the RS cavity assumption to hold is that $\overline{\lambda}_T < 1$.

Using the fluctuation–dissipation relation when edge i is at distance d from edge 0 we have for the correlation function

$$\mathbb{E}(\langle s_0 s_d\rangle_c^2) = \mathcal{C}\mathbb{E}\left[\left(\frac{\partial h_d}{\partial h_0}\right)^2\right] = \mathcal{C}\mathbb{E}\left[\prod_{i=1}^{d}\left(\frac{\partial h_i}{\partial h_{i-1}}\right)^2\right], \quad (29)$$

where $\mathcal{C}$ is a d-independent constant. The field h_i is according to (16) a function of h_{i-1} and other fields $h_{i-1}^{(k)}$ incoming to h_i; see figure 3.

$$h_i = -\frac{1}{\beta}\log\left[\mathrm{e}^{-\beta}+\mathrm{e}^{\beta h_{i-1}}+\sum_{k=1}^{p_{i-1}}\mathrm{e}^{\beta h_{i-1}^{(k)}}\right]. \quad (30)$$

The number p_{i-1} of the incoming fields is chosen according to the probability distribution $\mathcal{Q}_2(p_{i-1})$ of the number of neighbours of a node given this node already has two other neighbours, $\mathcal{Q}_2(k) = (k+2)(k+1)\mathcal{Q}(k+2)/(\alpha c)$; in particular for Poisson distributions $\mathcal{Q}_2 = \mathcal{Q}$. The values of the fields $h_{i-1}^{(k)}$ are chosen randomly from the distribution (23).

J. Stat. Mech. (2006) P05003

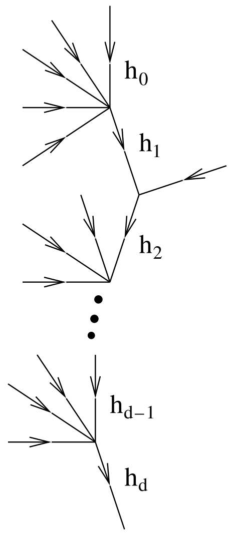

Figure 3. Chain of the cavity fields used to compute the finite temperature stability parameter.

3.3. Zero temperature limit

The zero temperature limit ($\beta \to \infty$) corresponds to the ground state (maximum matching) of our system. Let us investigate the explicit behaviour of that limit.

Our numerical studies of (23) show that for large β the cavity field distribution $\mathcal{P}_\beta(h)$ peaks around three different values $h \in \{0, \pm 1\}$.

$$\mathcal{P}(h) = p_1\delta(h-1) + p_2\delta(h+1) + p_3\delta(h), \tag{31}$$

where p_1, p_2 and p_3 are the weights (probabilities) of $h = 1$, -1 and 0. The cavity field update (16) becomes

$$h^{i\to a} = -\max_{j\in b-i}(-1, h^{j\to b}). \tag{32}$$

These equations may also be derived by working directly at zero temperature as in [35]. We define the cavity energy $E^{i\to a}_{s_i}$ as the ground state energy of the subgraph containing edge i when constraint a is absent (figure 2) and edge i takes value s_i. The analogue of (14) is

$$E^{i\to a}_{s_i} = \min_{\{s_j\}} \mathbb{I}\left(s_i + \sum_{j\in b-i} s_j \le 1\right)\left[\sum_{j\in b-i} E^{j\to b}_{s_j} + (1 - s_i - \sum_{j\in b-i} s_j)\right]. \tag{33}$$

If one defines the cavity fields as

$$h^{i\to a} = E^{i\to a}_0 - E^{i\to a}_1 \tag{34}$$

then (33) gives back the cavity field update (32). The difference between cavity energies when i is (is not) matched may be ± 1 or 0, and these are the three possible values of cavity fields.

Equation (23), taken in the $\beta \to \infty$ limit, shows that

$$p_1 = \frac{1}{c}\sum_{k=0}^{\infty}(k+1)\mathcal{Q}(k+1)p_2^k, \tag{35}$$

$$p_2 = \frac{1}{c}\sum_{k=0}^{\infty}(k+1)\mathcal{Q}(k+1)[1-(1-p_1)^k], \tag{36}$$

$$p_3 = \frac{1}{c}\sum_{k=0}^{\infty}(k+1)\mathcal{Q}(k+1)[(1-p_1)^k - p_2^k]. \tag{37}$$

The possible solutions to these equations depend on the distribution $\mathcal{Q}(k)$.

(a) There always exists a solution with $p_3 = 0$, $p_1 = 1 - p_2$.

(b) For graphs without leaves, $\mathcal{Q}(1) = 0$, there exists a solution with $p_3 = 1$, $p_1 = p_2 = 0$.

(c) For graphs with leaves $\mathcal{Q}(1) > 0$ a solution with $0 < p_1, p_2, p_3 < 1$ exists if the mean degree is sufficiently large.

Let us stress at this point that our numerical solution of (23) for the cavity field distribution at very small but nonzero temperatures corresponds to case $p_3 > 0$, (b) or (c). In other words, whereas at zero temperature there exist two mathematically possible solutions of (35)–(37), at arbitrary small temperature only the one with $p_3 > 0$ exists. In the rest of this section we describe this 'small temperature' solution. Case (a), which exists only at strictly zero temperature, and which forbids the cavity fields $h = 0$, will be discussed in section 5.

Using (24) the ground state energy, related to the size of the maximum matching, is

$$\epsilon_0 = \mathcal{Q}(0) + \sum_{k=1}^{\infty}\mathcal{Q}(k)[p_2^k + (1-p_1)^k - 1] + cp_1(1-p_2). \tag{38}$$

If we consider solution (b) for p_1, p_2, p_3 for graphs with no leaves, $Q(1) = 0$, then the ground state energy is $\epsilon_0 = Q(0)$, i.e. asymptotically all the non-isolated vertices are matched. In other words, in an ensemble of random graphs with minimal degree two (e.g. regular graphs) almost every graph has an almost perfect matching. This is in agreement with the result of Karp and Sipser [26] and also with a stronger result of Frieze and Pittel [36], who also found the (small) number of vertices which cannot be matched.

To compute the average ground state entropy we need to expand the free energy at low temperatures $f(\beta \to \infty) = e_0 - s_0/\beta + \mathrm{O}(1/\beta^2)$. This requires us to study the 'evanescent' parts of the cavity fields [37], i.e. the leading corrections to their value at $\beta = \infty$. Numerically we have observed that at $\beta \gg 1$ the three delta peaks (31) keep their weights (p_1, p_2, p_3) and spread as

$$h = 1 + \frac{\log \nu}{\beta} \qquad \text{for the peak around } h = 1, \tag{39}$$

$$h = -1 + \frac{\log \mu}{\beta} \qquad \text{for the peak around } h = -1, \tag{40}$$

$$h = \frac{\log \gamma}{\beta} \qquad \text{for the peak around } h = 0. \tag{41}$$

From (23) we derive self-consistently the distributions $\mathcal{A}$ of the evanescent cavity fields ν, μ, γ

$$\mathcal{A}_1(\nu) = \sum_{k=0}^{\infty} \mathcal{C}_1(k) \int \prod_{i=1}^{k} [\mathrm{d}\mu_i \mathcal{A}_2(\mu_i)]\, \delta\left(\nu - \frac{1}{1 + \sum_i \mu_i}\right), \tag{42}$$

$$\mathcal{A}_2(\mu) = \sum_{k=1}^{\infty} \mathcal{C}_2(k) \int \prod_{i=1}^{k} [\mathrm{d}\nu_i \mathcal{A}_1(\nu_i)]\, \delta\left(\mu - \frac{1}{\sum_i \nu_i}\right), \tag{43}$$

$$\mathcal{A}_3(\gamma) = \sum_{k=1}^{\infty} \mathcal{C}_3(k) \int \prod_{i=1}^{k} [\mathrm{d}\gamma_i \mathcal{A}_3(\gamma_i)]\, \delta\left(\gamma - \frac{1}{\sum_i \gamma_i}\right), \tag{44}$$

where the combinatorial factors $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ are given by $\mathcal{C}_1(k) = (k+1) p_2^k \mathcal{Q}(k+1)/p_1 c$, $\mathcal{C}_2(k) = (p_1^k/k!) \sum_{m=k}^{\infty} ((1-p_1)^{m-k} \mathcal{Q}(m+1)(m+1)!/(m-k)! p_2 c)$, $\mathcal{C}_3(k) = (p_3^k/k!) \sum_{m=k}^{\infty} (p_2^{m-k} \mathcal{Q}(m+1)(m+1)!/(m-k)! p_3 c)$. Using equations (39)–(44) we expand the free energy to order $1/\beta$ and get the ground state entropy of maximum matchings

$$\begin{aligned} s_0 = {} & \sum_{k=1}^{\infty} \frac{p_3^k}{k!} \sum_{m=k}^{\infty} \frac{p_2^{m-k} \mathcal{Q}(m) m!}{(m-k)!} \overline{\log\left(\sum_{i=1}^{k} \gamma_i\right)} - c p_1 p_3 \overline{\log \gamma} - \frac{c p_3^2}{2} \overline{\log(1 + \gamma_1 \gamma_2)} \\ & - c p_1 (p_1 + p_3) \overline{\log \nu} + \sum_{k=1}^{\infty} \frac{p_1^k}{k!} \sum_{m=k}^{\infty} \frac{(1-p_1)^{m-k} \mathcal{Q}(m) m!}{(m-k)!} \overline{\log\left(\sum_{i=1}^{k} \nu_i\right)} \\ & + \sum_{k=0}^{\infty} \mathcal{Q}(k) p_2^k \overline{\log\left(1 + \sum_{i=1}^{k} \mu_i\right)} - c p_1 p_2 \overline{\log(1 + \mu\nu)}, \end{aligned} \tag{45}$$

where the overlines denote expectations over independent random variables with distribution $\mathcal{A}_1$ (for ν-variables), $\mathcal{A}_2$ (for μ-variables), and $\mathcal{A}_3$ (for γ-variables).

To conclude this section we describe the zero temperature version of the stability analysis for the cavity assumption. What follows is equivalent to the stability analysis of the replica symmetric assumption with respect to replica symmetry breaking for discrete sets of cavity fields [38]. Here we will describe this stability analysis in an intuitive way as a spread of changes in the cavity field update that is analogous to what is called *bug proliferation* in the context of the stability of the one step replica symmetry breaking ansatz [33, 34, 39].

Consider a node b with $k+1$ neighbours, figure 2. Choose one incoming cavity field $h^{j \to b}$ and one outgoing field $h^{i \to a}$. Now consider probability $P_k(\alpha_{\mathrm{o}} \to \gamma_{\mathrm{o}} | \alpha_{\mathrm{i}} \to \gamma_{\mathrm{i}})$ that the value of the outgoing field changes from $\alpha_{\mathrm{o}} \in \{\pm 1, 0\}$ to $\gamma_{\mathrm{o}} \in \{\pm 1, 0\}$ providing the incoming one has been changed from $\alpha_{\mathrm{i}} \in \{\pm 1, 0\}$ to $\gamma_{\mathrm{i}} \in \{\pm 1, 0\}$. More precisely, P is the probability of having a set of $k-1$ other incoming fields such that it causes the change $\alpha_{\rm o} \to \gamma_{\rm o}$ given that the change $\alpha_{\rm i} \to \gamma_{\rm i}$ happened. There are eight different combinations of cavity fields such that $P_k(\alpha_{\rm o} \to \gamma_{\rm o}|\alpha_{\rm i} \to \gamma_{\rm i})$ is nonzero

$$P_k(1 \to -1|-1 \to 1) = P_k(-1 \to 1|1 \to -1) = p_2^{k-1}, \tag{46}$$

$$P_k(1 \to 0|-1 \to 0) = P_k(0 \to 1|0 \to -1) = p_2^{k-1}, \tag{47}$$

$$P_k(-1 \to 0|1 \to 0) = P_k(0 \to -1|0 \to 1) = (p_2+p_3)^{k-1}, \tag{48}$$

$$P_k(-1 \to 0|1 \to -1) = P_k(0 \to -1|-1 \to 1) = (p_2+p_3)^{k-1} - p_2^{k-1}. \tag{49}$$

In the first step we change a cavity field from $\alpha_{\rm i}$ to $\gamma_{\rm i}$. The average probability of change $\alpha_{\rm o}$ to $\gamma_{\rm o}$ that follows is

$$\overline{P}(\alpha_{\rm o} \to \gamma_{\rm o}|\alpha_{\rm i} \to \gamma_{\rm i}) = \sum_{k=0}^{\infty} \mathcal{Q}_2(k) P_{k+1}(\alpha_{\rm o} \to \gamma_{\rm o}|\alpha_{\rm i} \to \gamma_{\rm i}). \tag{50}$$

The most important change is given by the largest eigenvalue $\overline{\lambda}_{\max}$ of the matrix $\overline{P}$. In analogy with (27) we are interested in the stability parameter $\overline{\lambda}_0 = \alpha \overline{\lambda}_{\max}$

$$\overline{\lambda}_0 = \alpha \sqrt{\sum_{k=0}^{\infty} \mathcal{Q}_2(k) p_2^k} \sqrt{\sum_{k=0}^{\infty} \mathcal{Q}_2(k)(p_2+p_3)^k}. \tag{51}$$

If $\overline{\lambda}_0 > 1$, the total number of changes after many steps will diverge and we cannot hope the cavity assumption will be valid. On the other hand, if $\overline{\lambda}_0 < 1$, then the first change will not spread very far and the RS assumption is locally stable.

Note also that $\overline{\lambda}_0$ and $\overline{\lambda}_{T\to 0}$ are not equal, because they count different quantities. But we expect that their positions relative to the threshold value of unity are the same. In other words, both of them correctly describe the stability at zero temperature. The advantage of $\overline{\lambda}_0$ is that it is far easier to compute than the $d \to \infty$ limit of $\overline{\lambda}_{T,d}$ (28)–(29).

3.4. Algorithms following from the cavity method

3.4.1. Zero temperature message passing and leaf removal. The zero temperature limit of the cavity field update (32) can be seen as a message passing (warning propagation) algorithm. The interpretation of the three different cavity fields is the following: $h = 1$ means *'I want you to match me'*, $h = -1$ means *'I want you not to match me'*, $h = 0$ means *'No preferences, do what you want'*. The interpretation of the cavity field update (32) is the following.

- If one or more of my neighbours wants me to match them, I match one of them, and I send *do not match me.*
- If none of my neighbours wants me to match it, and at least one of them does not have any preferences I send *no preferences, do what you want.*
- If all of my neighbours are saying do not match me, or if I have no neighbours, I send *match me.*

J. Stat. Mech. (2006) P05003

This message passing procedure starting from all the $h = 0$ is equivalent to the step (1) of the Karp and Sipser leaf removal procedure in the following sense: run the message passing until you find a fixed point. Then the edges where a message $h = 1$ or $h = -1$ is sent from at least one side are exactly those edges which have been matched or removed in the leaf removal procedure. Consequently the edges in the core are those which have zero sent from both sides. Using equation (16) at a very small temperature we can use arbitrary initial conditions.

3.4.2. Uniform sampling. Solving the BP equations (16) on a given graph G by iterations allows us to sample typical matchings from Boltzmann's distribution (11) at inverse temperature β (i.e. matchings of size $[1 - E_G(\beta)]/2$).

Such a sampling can be done as follows: one chooses a variable node i, computes $\langle s_{\rm i}\rangle$ from (17), and generates the value of $s_{\rm i}$ as $s_{\rm i} = 1$ with probability $\langle s_{\rm i}\rangle$, and $s_{\rm i} = 0$ with probability $1 - \langle s_{\rm i}\rangle$. Once $s_{\rm i}$ has been fixed, this imposes that all the fields $h^{i\to a}$ (for all function nodes a connected to i) are equal either to $+\infty$ (if $s_{\rm i} = 1$) or to $-\infty$ (if $s_{\rm i} = 0$). One runs again the BP equations, with these extra constraints, and iterates this procedure.

3.4.3. Counting matchings on a single graph. Our results may be also used to estimate the size (25) and number of matchings (26) on arbitrary spare large graph G. The size of the maximum matching is obtained from the zero temperature limit of (20) or (21); this is not very interesting since the solution to this problem is well known; see, e.g., [20]. An algorithm to compute approximately the number of matchings of given size is more interesting.

`INPUT: The graph, the inverse temperature` β`, a maximum number of iterations` t_{max}.
`OUTPUT: The entropy of matchings` $S_G = \log \mathcal{N}_G$ `of size` $(1 - E_G)/2$. `If at the end` $E_G = -1$ `the procedure failed to converge.`

(1) `Initialize all the cavity fields` $h^{i\to a}$ `to some random value.`
(2) `Iterate belief propagation equations (16) until they converge, i.e. the values of the cavity fields do not change anymore, or until the number of iterations exceeds` t_{max}.
(3) $E_G = -1$. `If the number of steps is` $> t_{max}$ `STOP. Else: compute the energy` E_G `and the free energy` F_G `of matchings corresponding to` β `according to (21) and (20), compute the entropy` $S_G = \beta[E_G - F_G]$.

In order to compute the total number of matchings one needs to take the $\beta \to 0$ limit. To compute the number of maximal matchings one takes $\beta \to \infty$. In both cases the algorithm can be rewritten and simplified.

Note also that the complexity of this algorithm is only linear in number of edges. The convergence and correctness in the highest order ($(\log \mathcal{N}_G)/N$) for large sparse graphs or trees is expected for the same reasons as the correctness of results of cavity method for the ensembles of r-regular and ER random graphs.

J. Stat. Mech. (2006) P05003

On small or loopy graphs the cavity field update does not have to converge or its fixed point may depend on the initial conditions. However, it would be interesting to apply it to 'real world' graphs as in [40], or to compare the results with those of existing methods [22]–[24].

4. Application to random graph ensembles

We compute and discuss the results for two random graph ensembles, the r-regular and ER graphs.

4.1. Random regular graphs

For r-regular regular graphs ($\mathcal{Q}(k) = \delta_{kr}$) the results are particularly simple. All the vertices are equivalent in the cavity method. This means that the solution of (23) is $\mathcal{P}_\beta(h) = \delta(h - h_r)$, where h_r is the solution of

$$h_r = -\frac{1}{\beta} \log\,[\mathrm{e}^{-\beta} + (r-1)\mathrm{e}^{\beta h_r}], \tag{52}$$

given by

$$h_r = \frac{1}{\beta} \log \left[\frac{\sqrt{4(r-1) + \mathrm{e}^{-2\beta}} - \mathrm{e}^{-\beta}}{2(r-1)} \right]. \tag{53}$$

The free energy density (24) simplifies to

$$f = -\frac{1}{\beta} \log\,[\mathrm{e}^{-\beta} + r\mathrm{e}^{\beta h_r}] + \frac{r}{2\beta} \log\,[1 + \mathrm{e}^{2\beta h_r}]. \tag{54}$$

The energy density (25) reads

$$\epsilon = \frac{\mathrm{e}^{-\beta} - r h_r \mathrm{e}^{\beta h_r}}{[\mathrm{e}^{-\beta} + r\mathrm{e}^{\beta h_r}]} + \frac{r h_r \mathrm{e}^{2\beta h_r}}{[1 + \mathrm{e}^{2\beta h_r}]}. \tag{55}$$

The entropy, related to the number of matchings, is computed using (26). With a bit of algebra one finds that the quenched entropy as a function of energy is equal to the annealed result of equation (7), where $\epsilon = 1 - x$. This result is compatible with, but slightly stronger than, theorem 1 of Bollobás and McKay [25].

The matching problem is equivalent to the physical problem of dimers on a graph. There it is natural to compute the density of dimers rp as a function of β (which is half of the chemical potential in the context of dimers). For r-regular graphs we find that this 'equation of state' is

$$\mathrm{e}^{2\beta} = \frac{p(1-p)}{(1-rp)^2}. \tag{56}$$

This result has already been obtained in [19] for the dimer problem on the Bethe lattice.

For $r \geq 2$ in the zero temperature limit one has $h_r = 0$, which corresponds to solution (b) of (35)–(37). The ground state energy density is then $\epsilon_0 = 0$; this means that asymptotically almost all the vertices may be matched for almost every r-regular graph. This agrees with the stronger mathematical result that for $r \geq 3$ there exists a perfect

J. Stat. Mech. (2006) P05003

matching almost surely in graphs with even number of vertices. From (45) one finds that the ground state entropy density $\lim_{\beta\to\infty} s(\beta)$ is

$$s_0 = [(r-1)\log(r-1) - (r-2)\log r]/2, \tag{57}$$

in agreement with the annealed result (3) of Bollobás and McKay.

The stability parameter $\overline{\lambda}_T$ (28) for r-regular graphs is

$$\overline{\lambda}_T = (r-1)\left(\left.\frac{\partial h_1}{\partial h_0}\right|_{h_1=h_0=h_r}\right)^2 = (r-1)\left[\frac{\sqrt{4(r-1)+\mathrm{e}^{-2\beta}}-\mathrm{e}^{-\beta}}{2(r-1)}\right]^4. \tag{58}$$

We see that $\overline{\lambda}_T < 1$ for all finite temperatures and $r \geq 2$. In the zero temperature limit $\overline{\lambda}_{T\to 0} = 1/(r-1)$. The zero temperature stability parameter (51) for r-regular graphs is $\overline{\lambda}_0 = 0$ for $r > 2$, and $\overline{\lambda}_0 = 1$ for $r = 2$. This agrees qualitatively with the behaviour of $\overline{\lambda}_{T\to 0}$. It is worth noticing that the ferromagnetic stability parameter, defined as $(r-1)\left((\partial h_1/\partial h_0)|_{h_1=h_0=h_r}\right)$, is also at finite temperature smaller than unity when $r \geq 2$. It should be possible to use this result in order to show that the matching properties on the root of a large tree are completely independent of boundary conditions, which could then allow for a rigorous proof of our results following the lines of Bandyopadhyay and Gamarnik [16].

In order to exclude the possibility of a discontinuous transition towards a phase with broken replica symmetry, which we cannot see by analysing the stability, we wrote the 1RSB [9] equations for the r-regular graph. We have seen clearly numerically that their solution reduces to the replica symmetric one. So all the evidence suggests that the RS cavity assumption should be valid for r-regular graphs and so we expect our result for the quenched entropy to be exact.

4.2. ER graphs

For the ER random graphs, with degree distribution $\mathcal{Q}(k) = \mathrm{e}^{-c}c^k/k!$, we have solved numerically the equation (23) by the population dynamics method. Using (24)–(26) we have then computed the energy density $\epsilon(\beta)$ (related to the size of the matching as $|M| = (1-\epsilon)N/2$) and the entropy density $s(\beta)$ for values of $\beta \in (-\infty, \infty)$. In figure 4 we show the entropy versus the size of matching for mean degrees $c = 1, 2, 3$ and 6.

The maxima of the curves in figure 4 give the entropy of all the possible matchings, regardless of their sizes. The lower right ends of the curves give the ground state energy (the size of the maximal matching) and the ground state entropy (number of maximal matchings). They are computed using by the following direct zero temperature method.

The zero temperature equations (35)–(37) for the Poissonian distribution become

$$p_1 = \mathrm{e}^{-c(1-p_2)}, \qquad p_2 = 1 - \mathrm{e}^{-cp_1}, \qquad p_3 = 1 - p_1 - p_2. \tag{59}$$

Analysing these equations we can see that for $c \leq e$ there exists only solution (a), with $p_1 + p_2 = 1$ and $p_3 = 0$. For $c > e$ there exists a second solution (c) with $p_1 + p_2 < 1$ and $1 > p_3 > 0$. From the population dynamics solution of equation (23) at very small temperatures for $c > e$ we found that solution (c) with $p_1 + p_2 < 1$ is the proper zero temperature limit for $c > e$.

J. Stat. Mech. (2006) P05003

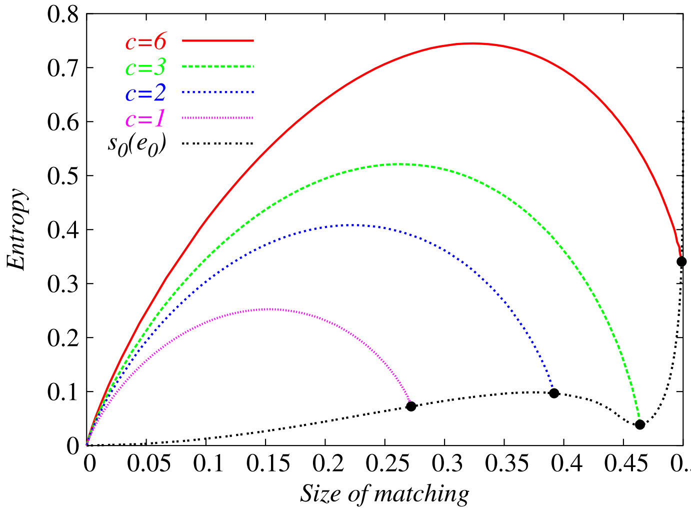

Figure 4. Entropy density $s(m) = \log \mathcal{N}(m)/N$ as a function of relative size of the matching $m = |M|/N = (1-\epsilon)/2N$ for ER random graphs with mean degrees $c = 1, 2, 3, 6$. The lower curve is the ground state entropy density $s_0(\epsilon_0)$ for all mean degrees. The curves are obtained by solving equations (23)–(26) with a population dynamics, using a population of sizes $N = 2 \times 10^4$ to 2×10^5 and the number of iterations $t_m = 10\,000$.

The ground state energy (38) then reads

$$\epsilon_0 = 1 - 2\frac{|M^*|}{N} = -p_2 + p_1 + cp_1 - cp_1p_2. \tag{60}$$

This is the exact result of Karp and Sipser [26], theorem 2.

The ground state entropy for ER graphs is computed using population dynamics equations (42)–(44) with combinatorial factors

$$\mathcal{C}_1(k) = \frac{\mathrm{e}^{-cp_2}(cp_2)^k}{k!}, \qquad \mathcal{C}_2(k) = \frac{\mathrm{e}^{-cp_1}(cp_1)^k}{(1-\mathrm{e}^{-cp_1})k!}, \qquad \mathcal{C}_3(k) = \frac{\mathrm{e}^{-cp_3}(cp_3)^k}{(1-\mathrm{e}^{-cp_3})k!}. \tag{61}$$

Factors $\mathcal{C}_{\mathrm{i}}(k)$ are a properly normalized Poissonian distribution with mean equal to the concentration of corresponding cavity fields. The ground state entropy (45) finally simplifies to

$$s_0 = -(1+cp_1)p_3\overline{\log\gamma} - \frac{cp_3^2}{2}\overline{\log(1+\gamma_1\gamma_2)} \\ - p_2\overline{\log\mu} - p_1(1+cp_1+cp_3)\overline{\log\nu} - cp_1p_2\overline{\log(1+\mu\nu)}. \tag{62}$$

We call the first two terms in equation (62) the *core entropy* s_{c}, the averages (denoted by overlines) are over the distribution (44). The rest (last three terms) we call the *non-core entropy* s_{nc}; the averages are over the distributions (42) and (43). The reason for these names is the following. Since we know the size and degree distribution on the core (6) we can use equation (45) directly only for the core, and indeed we will obtain the first two terms of equation (62), the core entropy. The rest is the entropy corresponding to the choice of the matching in the non-core part of the graph.

J. Stat. Mech. (2006) P05003

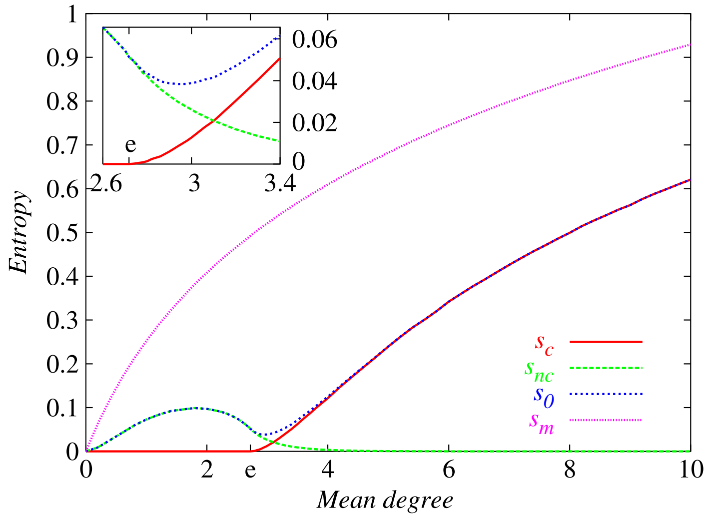

Figure 5. The ground state entropy density s_0 (giving leading exponential behaviour of the number of maximum matchings) and the full entropy s_m (giving leading exponential behaviour of the number of all possible matchings) as a function of mean degree c in ER random graphs. The detail is in the inset. The ground state entropy is the sum of s_{c}, the contribution of the core, and s_{nc}, the contribution of the parts of the graph removed in the leaf removal procedure. We see that $s_{\mathrm{c}} > 0$ only for $c > e$, because the core covers a finite fraction of vertices only when $c > e$.

Figure 5 shows the core and non-core entropies of the maximum matchings and their sum as a function of the average degree c. The fourth (upper) line in figure 5 is the total entropy of all matchings.

The finite temperature stability parameter $\overline{\lambda}_T$ (28) for ER random graphs is

$$\overline{\lambda}_T = c\left[\mathbb{E}(\langle n_0 n_d \rangle_c^2)\right]^{1/d}. \tag{63}$$

We have to compute it numerically as described in figure 3 and equations (28)–(29). As for the r-regular graphs we find that $\overline{\lambda}_T$ grows as temperature decreases, see on the left on figure 6. So we may analyse only the zero temperature limit, and if that is stable, then also the finite temperature is stable. On the right in figure 6 we can also see the dependence of $\overline{\lambda}_T$ on the distance d. Although we are not able to compute precisely its $d \to \infty$ limit, all the evidence speaks for the fact that even for $d \to \infty$ the stability parameter $\overline{\lambda}_T$ never exceeds unity.

To check this, we look directly at the zero temperature stability parameter $\overline{\lambda}_0$ (51), which for the ER graph reads

$$\overline{\lambda}_0 = cp_1\sqrt{1 + \frac{p_3}{p_1}}. \tag{64}$$

Its value is also depicted in figure 6. We can see that $\overline{\lambda}_0 < 1$ (stable) for all mean degrees except $c = e$ where $\overline{\lambda}_0 = 1$ (marginally stable). Supported by the numerical data in figure 6 we expect that in the $d \to \infty$ limit the $\overline{\lambda}_T$ would behave in the qualitatively same way.

J. Stat. Mech. (2006) P05003

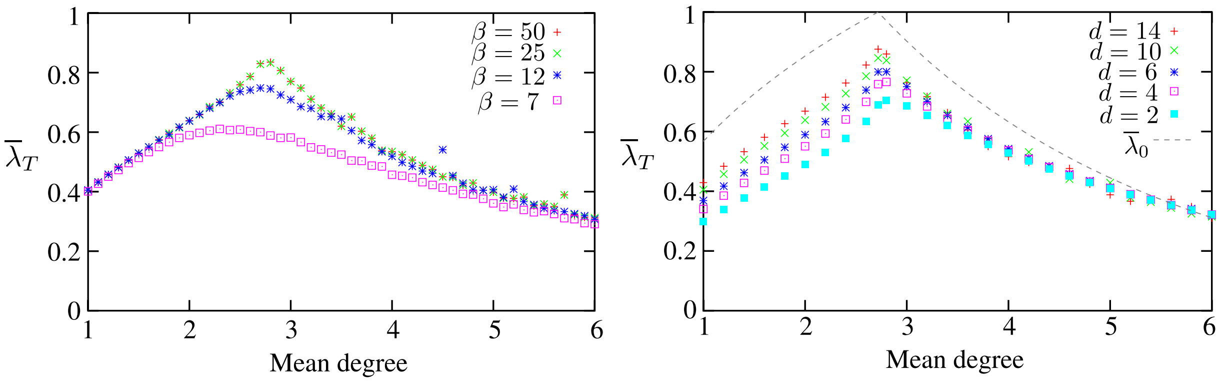

Figure 6. On the left the finite temperature stability parameter $\overline{\lambda}_T$ (63) for distance $d = 10$ as a function of mean degree. The upper curve corresponds to the smallest temperature ($\beta = 50$). Data were obtained for size of population $N = 40\,000$ and time $t = 40\,000$. On the right the dependence of $\overline{\lambda}_T$ for temperature $\beta = 50$ on the mean degree and distance d. We can see that $\overline{\lambda}_T$ is growing slightly with d in the regime $c < e$. For larger d we would need very big populations to obtain reliable data. The continuous line depicts the zero temperature stability parameter $\overline{\lambda}_0$, equation (64).

From this analysis it is reasonable to conjecture that in ER random graphs the replica symmetric cavity assumption is correct and all our results, in particular for the entropy, are exact. Another strong argument in favour of the validity of replica symmetry at any finite temperature will be given in section 5.

5. Ergodicity breaking at zero temperature

The size of the maximum matchings in ER graphs was studied recently by Zhou and Ou-Yang (Z–O) [10], using the cavity method directly at zero temperature [35], with a one step RSB solution. In this section we discuss the difference between their approach and ours, in particular as far as RSB effects are concerned. We keep to ER random graphs.

One should first emphasize that both approaches give the same result for the size of the largest matching in ER graphs, and this result also agrees with the rigorous value of Karp and Sipser. Our formalism is more general in two respects. (1) We can work at finite temperature, which gives access to the full distribution of the number of matchings versus their size. (2) We study the limit of zero temperature ($\beta \to \infty$) keeping the leading corrections of order $1/\beta$ in the fields (see (39)–(41)); this allows us to study the entropy of maximal matchings.

The issue of RSB at zero temperature, which is present in the Z–O approach, and absent in ours, is a somewhat subtle one. We shall just present a few arguments of explanation.

First one should note that one does not expect ergodicity to be broken at any finite temperature in this problem. We have not tried to write a formal proof of this statement, but a first strong argument comes from the fact that the energy barriers are finite. Let us define as *a step* the fact of removing (adding) an edge from (to) a matching so that the new configuration is still a matching. By adding an edge to a matching we lower the energy by 2, whereas by removing an edge we increase the energy by 2. Using these steps one may go from any matching M to any other matching M'. Furthermore, if $|M'| \geq |M|$, one can choose the steps in such a way that at every step the energy is not higher than $E_M + 4$. In other words, energy barriers in the matching are at most 4. This argument suggests that there should not be ergodicity breaking at finite temperature, provided there are no diverging entropic barriers. Another indication of ergodicity comes from the rapid mixing results of the Monte Carlo procedure in the related problem of sampling perfect matchings in bipartite graphs [41].

Let us now focus on the zero temperature approach. The finite energy barriers between almost perfect matchings on the core become effectively infinite at zero temperature so the breaking of ergodicity cannot be excluded. The RS cavity method gives the equations (32) for the cavity fields, easily derived from (33). The more subtle issue is the support of the distribution of cavity fields. Because a field $h^{i\to a}$ is defined as the difference (34) of the ground state energies conditioned to i being absent/present in the matching, it is clear that $h^{i\to a} \in \{-1, 0, 1\}$, and in ER graphs with $c > e$ one should thus choose between solutions (a) and (c) of equations (35)–(37). If one considers equations (32) on a graph which is a tree, one finds that actually on all edges $h_{i\to a} \in \{-1, 1\}$. Because ER graphs are locally tree-like (seen from a randomly chosen point, the subgraph of its environment up to a fixed distance d is a tree with probability one in the large N limit), it is tempting to restrict the cavity fields to ± 1 values. This is what is done in Z–O. Then the RS solution, for any value of the average degree c, is necessarily solution (a). This solution is unstable towards 1RSB at $c > e$, which forces one to study the 1RSB solution in this regime, as was done in Z–O.

The 1RSB solution for the maximal matchings is able to nicely reconstruct the information that is contained in the $h = 0$ fields of our RS solution with support on $\{-1, 0, 1\}$ as follows. Let us consider an edge $i \to a$ which should pass $h^{i\to a} = 0$ in our formalism. In the 1RSB formalism it passes a message which is a probability distribution on the space of cavity fields, with support $\{-1, 1\}$, of the form $\alpha\delta_{h,-1} + (1-\alpha)\delta_{h,1}$; the distribution of α is related to our distribution $\mathcal{A}_3$ (44). Consequently, the complexity computed by Z–O is equal to the complexity of the core. This means that different almost perfect matchings on the core form the different states, each state containing only one of them (similarly as in the XOR-SAT problem [42, 43], or in the multi-index matching [44]). It is interesting to notice that, through the restriction of cavity fields to $h \in \{-1, 1\}$, the Z–O method at the RS level completely neglects loops, the effect of which is recovered only at the 1RSB level. Conversely, the inclusion of the value zero in our cavity fields allows us to take into account loops directly, in which case RSB is not needed.

This physical interpretation of the Z–O 1RSB solution is confirmed by its stability analysis. Using notations of [33] one should compute the type I instability (interpreted as state aggregation) of the 1RSB solution. Type II instability (interpreted as division of states) is irrelevant here, because each state corresponds to a single almost perfect matching on the core and cannot divide further. We have found that the 1RSB solution is stable, but only if one considers maximal matchings ($y \to \infty$ in the Z–O notation): any departure from this limit mixes the various configurations and restores ergodicity, as expected.

J. Stat. Mech. (2006) P05003

6. Conclusion and discussion

We have argued that the replica symmetric cavity solution is exact for counting matchings on random graphs. We have computed the size and quenched (typical) entropy in two random graph ensembles. For r-regular graphs we have shown that the quenched entropy of matchings of a given size agrees with the annealed one. For the Erdös–Rényi random graphs we have shown how our method reproduces the result of Karp and Sipser for the size of maximum matching, and we computed the quenched entropy of matchings of a given size (figure 4).

Our method provides an algorithm for counting and uniform sampling of matchings on a given sparse graph, which should give the exact entropy for graphs with a girth that diverges in the large size limit. It would be very interesting to apply it to 'real world' graphs, e.g. the internet, as in [40]. Also its systematic study on graphs with smaller girth and comparison with existing approximative methods [22]–[24] could reveal some interesting properties.

There are two obvious generalizations of the matching problem where we expect that our method could be used straightforwardly. One is the matching with weights on edges (preferences to be matched) which is a dimer model on random graphs with quenched disorder. Another generalization is that instead of allowing a vertex to have no or one ($k = 1$) matched edge around itself, we could allow no or $k > 1$ edges around a vertex to be matched. Then $k = 2$ would mean we are interested in sets of loops, a model that has been studied recently in [40, 45]. The case $k > 2$, corresponding to k-regular subgraphs, is being studied by [46].

We hope that the replica symmetric nature of matching on random graphs should allow us to turn all our results into rigorous theorems. In this respect there are two directions which look particularly promising. One is to generalize the local weak convergence approach of [16] in order to turn our results into rigorous theorems when β is small enough and/or c is far enough from e (for ER graphs). The second one is to use Guerra's interpolation method [12, 47, 48] in order to turn our results into rigorous upper bounds for the entropy. More ambitiously, one can hope that the study of this matching problem will help to turn the cavity method into a rigorous tool.

Acknowledgments

We thank Florent Krzakala, Thierry Mora, Martin Weigt, Guilhem Semerjian and Federico Ricci-Tersenghi for very helpful discussions and comments. This work has been supported by the EC through the network MTR 2002-00319 STIPCO and the FP6 IST consortium EVERGROW.

References

[1] Lovasz L and Plummer M D, 1986 *Matching Theory* (Amsterdam, New York: North-Holland)
[2] Mézard M, Parisi G and Virasoro M A, 1987 *Spin-Glass Theory and Beyond* (*Lecture notes in Physics* vol 9) (Singapore: World Scientific)
[3] Mézard M and Parisi G, *Replicas and optimization*, 1985 *J. Phys. Lett.* **46** L771
[4] Mézard M and Parisi G, *A replica analysis of the travelling salesman problem*, 1986 *J. Physique* **47** 1285
[5] Weigt M and Hartmann A K, *Number of guards needed by a museum: a phase transition in vertex covering of random graphs*, 2000 *Phys. Rev. Lett.* **84** 6118

[6] Mézard M, Parisi G and Zecchina R, *Analytic and algorithmic solution of random satisfiability problems*, 2002 *Science* **297** 812

[7] Mézard M and Zecchina R, *The random K-satisfiability problem: from an analytic solution to an efficient algorithm*, 2002 *Phys. Rev.* E **66** 056126

[8] Mulet R, Pagnani A, Weigt M and Zecchina R, *Coloring random graphs*, 2002 *Phys. Rev. Lett.* **89** 268701

[9] Mézard M and Parisi G, *The Bethe lattice spin glass revisited*, 2001 *Eur. Phys. J.* B **20** 217

[10] Zhou H and Ou-Yang Z, *Maximum matching on random graphs*, 2003 *Preprint* cond-mat/0309348

[11] Talagrand M, 2003 *Spin Glasses: a Challenge for Mathematicians* (Berlin: Springer)

[12] Guerra F, 2003 *Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model* (*Communications in Mathematical Physics* vol 233) pp 1–12

[13] Parisi G, *A sequence of approximated solution to the S–K model for spin glasses*, 1980 *J. Phys. A: Math. Gen.* **13** L115

[14] Aldous D J, *The* $\zeta(2)$ *limit in the random assignment problem*, 2001 *Random Struct. Algorithms* **18** 381

[15] Bayati M, Shah D and Sharma M, *Maximum weight matching via max-product belief propagation*, 2005 *Information Theory (Proc. ISIT)* p 1763 [cs.IT/0508101]

[16] Bandyopadhyay A and Gamarnik D, *Counting without sampling. New algorithms for enumeration problems using statistical physics*, 2006 *ACM–SIAM Symp. on Discrete Algorithms* p 890 [math.PR/0510471]

[17] Gamarnik D, Nowicki R and Swirscsz G, *Maximum weight independent sets and matchings in sparse random graphs. Exact results using the local weak convergence method*, 2004 *Lecture Notes Comput. Sci.* **3122** 357

[18] Kasteleyn P W, *The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice*, 1961 *Physica* **27** 1209

[19] Harris A B and Cohen M, *Dimer statistics on a Bethe lattice*, 2004 *Preprint* cond-mat/0408618

[20] Micali S and Vazirani V V, *An* $O(|E|\sqrt{|V|})$ *algorithm for finding maximum matching in general graphs*, 1980 *Proc. 21st IEEE Symp. on Foundations of Computer Science* pp 17–27

[21] Valiant L G, *The complexity of computing the permanent*, 1979 *Theor. Comput. Sci.* **8** 189

[22] Jerrum M and Sinclair A, *The Markov chain Monte Carlo method: an approach to approximate counting and integration*, 1996 *Approximation Algorithms for NP-Hard Problems* ed D Hochbaum (Boston, MA: PWS Publishing Company) p 482

[23] Fürer M and Kasiviswanathan S P, *Approximately counting perfect matchings in general graphs*, 2005 *SIAM Proc. ALENEX/ANALCO 2005*

[24] Chien S, *A determinant-based algorithm for counting matchings in a general graph*, 2004 *Proc. 15th Annual ACM-SIAM Symp. on Discrete Algorithms* pp 728–35

[25] Bollobás B and McKay B D, *The number of matching in random regular graphs and bipartite graphs*, 1986 *J. Combin. Theory* B **41** 80

[26] Karp R M and Sipser M, *Maximum matchings in sparse random graphs*, 1981 *Proc. 22nd Annual IEEE Symp. on Foundations of Computer Science* pp 364–75

[27] Bauer M and Golinelli O, *Core percolation in random graphs: a critical phenomena analysis*, 2001 *Eur. Phys. J.* B **24** 339

[28] Aronson J, Frieze A and Pittel B, *Maximum matchings in sparse random graphs: Karp–Sipser revisited*, 1998 *Random Struct. Algorithms* **12** 111

[29] Hartmann A K and Weigt M, 2005 *Phase Transitions in Combinatorial Optimization Problems* (New York: Wiley–VCH)

[30] Kschischang F R, Frey B J and Loelinger H-A, *Factor graphs and the sum–product algorithm*, 2001 *IEEE Trans. Inf. Theory* **47** 498

[31] Pearl J, 1988 *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* 2nd edn (San Francisco, CA: Morgan Kaufmann Publishers)

[32] Yedidia J S, Freeman W T and Weiss Y, *Understanding belief propagation and its generalizations*, 2003 *Exploring Artificial Intelligence in the New Millennium* chapter 8, pp 239–69

[33] Montanari A and Ricci-Tersenghi F, *On the nature of the low-temperature phase in discontinuous mean-field spin glasses*, 2003 *Eur. Phys. J.* B **33** 339

[34] Rivoire O, Biroli G, Martin O C and Mézard M, *Glass models on Bethe lattices*, 2004 *Eur. Phys. J.* B **37** 55

[35] Mézard M and Parisi G, *The cavity method at zero temperature*, 2003 *J. Stat. Phys.* **111** 1

[36] Frieze A and Pittel B, *Perfect matchings in random graphs with prescribed minimal degree*, 2003 *ACM–SIAM Symp. on Discrete Algorithms* pp 148–57

[37] Biroli G, Monasson R and Weigt M, *A variational description of the ground state structure in random satisfiability problems*, 2000 *Eur. Phys. J.* B **14** 551

J. Stat. Mech. (2006) P05003

[38] Castellani T, Krzakala F and Ricci-Tersenghi F, *Spin glass models with ferromagnetically biased couplings on the Bethe lattice: analytic solutions and numerical simulations*, 2005 *Eur. Phys. J.* B **47** 99

[39] Mertens S, Mézard M and Zecchina R, *Threshold values of random K-SAT from the cavity method*, 2005 *Random Struct. Algorithms* **28** 340–73

[40] Marinari E, Monasson R and Semerjian G, *An algorithm for counting circuits: application to real-world and random graphs*, 2006 *Europhys. Lett.* **73** 8

[41] Jerrum M, Sinclair A and Vigoda E, *A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries*, 2004 *J. ACM* **51** 671

[42] Mézard M, Ricci-Tersenghi F and Zecchina R, *Alternative solutions to diluted p-spin models and XORSAT problems*, 2003 *J. Stat. Phys.* **111** 505

[43] Cocco S, Dubois O, Mandler J and Monasson R, *Rigorous decimation-based construction of ground pure states for spin glass models on random lattices*, 2003 *Phys. Rev. Lett.* **90** 047205

[44] Martin O C, Mézard M and Rivoire O, *Random multi-index matching problems*, 2005 *J. Stat. Mech.* P09006

[45] Marinari E and Semerjian G, *On the number of circuits in random graphs*, 2006 *Preprint* cond-mat/0603657

[46] Pretti M and Weigt M, *Sudden emergence of q-regular subgraphs in random graphs*, 2006 *Preprint* cond-mat/0603819

[47] Franz S and Leone M, *Replica bounds for optimization problems and diluted spin systems*, 2003 *J. Stat. Phys.* **111** 535

[48] Franz S, Leone M and Toninelli F L, *Replica bounds for diluted non-Poissonian spin systems*, 2003 *J. Phys. A: Math. Gen.* **36** 10967