Ground-state entropy of the random vertex-cover problem

Jie Zhou and Haijun Zhou

Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China
(Received 6 October 2008; published 20 February 2009)

Counting the number of ground states for a spin-glass or nondeterministic polynomial-complete combinatorial optimization problem is even more difficult than the already hard task of finding a single ground state. In this paper the entropy of minimum vertex covers of random graphs is estimated through a set of iterative equations based on the cavity method of statistical mechanics. During the iteration both the cavity entropy contributions and cavity magnetizations for each vertex are updated. This approach overcomes the difficulty of iterative divergence encountered in the zero-temperature first-step replica-symmetry-breaking (1RSB) spin-glass theory. It is still applicable when the 1RSB mean-field theory is no longer stable. The method can be extended to compute the entropies of ground states and metastable minimal-energy states for other random-graph spin-glass systems.

A combinatorial optimization (CO) problem is defined by an energy function $E(\vec{\sigma})$ on configurations $\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$ of an N-dimensional space, where each variable σ_i has only a finite number of states (e.g., $\sigma_i = \pm 1$). For CO problems in the nondeterministic polynomial-complete (NP-complete) class, searching for configurations of energy equal or very close to the lowest possible value is in general a very difficult task. Statistical physicists relate this computational hardness to the emergence of complex structures in the problem's configuration space and the proliferation of metastable macroscopic states (macrostates). The entropy spectrum (number of configurations at each minimal-energy level) gives a characterization of the energy landscape of a hard CO problem. Using the first-step replica-symmetry-breaking (1RSB) cavity method of spin-glass theory [1–3], the ground-state energy densities for several hard CO problems on random graphs have been calculated with high precision (see, e.g., Refs. [2,4,5]). But estimation of the ground-state entropy and the entropies of metastable states is still a challenging theoretical and computational issue.

In the 1RSB cavity approach, the ground-state energy E^0 of a hard CO problem is evaluated by first assuming that the configuration space of the system can be clustered into many macrostates. A variable i experiences in each macrostate an integer-valued field h_i. The distribution of this field among all the macrostates, $P_i(h_i)$, is then obtained through an iterative numerical scheme [2]. To calculate the ground-state entropy, a conventional technique is to introduce a temperature T and expand the field h_i to first order in T at the limit $T \rightarrow 0$ [6]:

$$h_i = m_i + T r_i, \qquad (1)$$

with m_i being an integer and r_i a finite real value. A set of iterative equations are derived to obtain the joint distribution $P_i(m_i, r_i)$ of the values m_i and r_i for each variable i. The ground-state entropy S^0 is then the first-order term in the free energy expansion $F(T) = E^0 - T S^0$. This approach works for some relatively simple problems [e.g., graph matching whose configuration space is ergodic [6], or random Q-coloring and K-satisfiability (K-SAT) which have zero

ground-state energy [7–9]] but it fails for many other NP-complete CO problems, for which the ground-state energy E^0 is positive and the $T \rightarrow 0$ 1RSB cavity equations are not stable. For these latter systems, at $T = 0$ the field h_i is not necessarily an integer and the correction r_i in Eq. (1) usually diverges at the $T \rightarrow 0$ limit. Consequently, although the $T \rightarrow 0$ 1RSB cavity method is able to estimate the ground-state energy of a hard CO problem with high accuracy (as only the value of m_i but not that of r_i is used), it often reports a negative or divergent ground-state entropy.

In this paper we use a different way to estimate the ground-state entropy of a CO problem or a finite-connectivity spin glass. We work directly at temperature zero and, within the 1RSB cavity framework, calculate both the cavity magnetizations and cavity entropies of each variable in each macrostate. A very small cutoff ϵ is naturally introduced in the iteration of cavity magnetizations to avoid divergence in the iteration of cavity entropies. This approach gives good results when tested on the random vertex-cover problem, the random 3-SAT problem, and the random $\pm J$ spin-glass model, even when the $T \rightarrow 0$ 1RSB mean-field spin-glass theory is no longer stable. It is also able to calculate the entropies for other minimal-energy levels. Here we focus on the vertex-cover problem (a prototypical NP-hard problem [10]) to demonstrate the main ideas of this method. Detailed calculations on other model systems will be reported elsewhere.

Let us first briefly introduce the vertex-cover problem [11,12]. A graph $\mathcal{G}$ contains N vertices and M edges. Each edge (i, j) is between a pair of vertices i, j. The mean connectivity of the graph is $c = 2M/N$, which is the number of edges a vertex on average is connected to. A vertex cover C for graph $\mathcal{G}$ contains a subset of vertices of $\mathcal{G}$ such that for each edge $(i, j) \in \mathcal{G}$ at least one of its extremities is in C. If a vertex i is contained in a vertex cover C, we say it is covered in C. The energy of a vertex cover C is defined as its cardinality, $E(C) = |C|$. Graph $\mathcal{G}$ has many vertex covers, with energy ranging from the maximum value N to the minimum value $E_{\mathcal{G}}^0 = \min_C |C|$. We denote by $C_{\mathcal{G}}^0$ the set of minimum vertex covers (MVCs) of graph $\mathcal{G}$: $C_{\mathcal{G}}^0 = \{C : |C| = E_{\mathcal{G}}^0\}$. To determine exactly the energy E^0 of MVCs for a graph in general is a very hard computational problem; to count the num-

ber of MVCs is even harder. Using the cavity method of statistical mechanics, the mean energy density of MVCs for random graphs has been evaluated in Refs. [4,11,13]. The entropy of MVCs for a random graph was also estimated by computer simulations [12]. This work complements Ref. [12] by giving an analytical estimation of the ground-state entropy $S_{\mathcal{G}}^0 \equiv \ln |\mathcal{C}_{\mathcal{G}}^0|$.

We denote by π_i the probability of a vertex i being contained in the MVCs,

$$\pi_i = \frac{1}{|\mathcal{C}_{\mathcal{G}}^0|} \sum_{C \in \mathcal{C}_{\mathcal{G}}^0} \mathbb{I}(i \in C), \tag{2}$$

where $\mathbb{I}$ is the indicator function. Similarly to π_i, for a vertex $j \in \partial i$ (the set of vertices that are connected to i by an edge), we denote by $\pi_{j|i}$ its probability of being in the MVCs of the cavity graph $\mathcal{G} \backslash i$ (which is obtained from $\mathcal{G}$ by removing vertex i and all its edges).

We consider here the ensemble of random graphs with mean connectivity c. When the graph size N is sufficiently large the length of a loop in the graph is of order $\ln_c(N)$, and a random graph $\mathcal{G}$ is locally treelike. Consider the set ∂i of vertices in the neighborhood of a randomly chosen vertex i. If the edges between i and these vertices are deleted, in the cavity graph $\mathcal{G} \backslash i$ the shortest length between any two vertices $j, k \in \partial i$ diverges logarithmically with N. As a first step, it is therefore assumed that these vertices are uncorrelated in the cavity graph $\mathcal{G} \backslash i$, and the probability of finding these vertices in a MVC of $\mathcal{G} \backslash i$ can be expressed in a factorized form:

$$\mathbb{P}(\partial i \subset C : C \in \mathcal{C}_{\mathcal{G} \backslash i}^0) = \prod_{j \in i} \pi_{j|i}. \tag{3}$$

Equation (3) is called the Bethe-Peierls approximation and the corresponding cavity method is referred to as replica symmetric (RS). Under this assumption, the energy and entropy of MVCs for a random graph $\mathcal{G}$ can be expressed as [4]

$$E_{\mathcal{G}}^0 = \sum_i \Delta E_i - \sum_{(i,j)} \Delta E_{(i,j)}, \tag{4}$$

$$S_{\mathcal{G}}^0 = \sum_i \Delta S_i - \sum_{(i,j)} \Delta S_{(i,j)}, \tag{5}$$

where ΔE_x and ΔS_x are, respectively, the contribution to the system's ground-state energy and entropy from a vertex $(x = i)$ or an edge $[x = (i, j)]$. For example, $\Delta E_i = E_{\mathcal{G}}^0 - E_{\mathcal{G} \backslash i}^0$ and $\Delta S_i = \ln |\mathcal{C}_{\mathcal{G}}^0| - \ln |\mathcal{C}_{\mathcal{G} \backslash i}^0|$.

We shall distinguish three situations when writing down the expressions for ΔE_i and ΔS_i. The total number of MVCs of cavity graph $\mathcal{G} \backslash i$ which contain the set ∂i is $|\mathcal{C}_{\mathcal{G} \backslash i}^0| \Pi_{j \in \partial i} \pi_{j|i}$. This number is positive if all the cavity probabilities $\pi_{j|i} > 0$. In this case, when vertex i is added, it will not be in any MVC of $\mathcal{G}$ but the set ∂i will be contained in every MVC of $\mathcal{G}$. Then we have $\pi_i = 0$ (vertex i is always noncovered), $\Delta E_i = 0$, and $|\mathcal{C}_{\mathcal{G}}^0| = |\mathcal{C}_{\mathcal{G} \backslash i}^0| \Pi_{j \in \partial i} \pi_{j|i}$, and the change in entropy is $\Delta S_i = \Sigma_{j \in i} \ln \pi_{j|i}$. On the other hand, if in the set ∂i there are at least two vertices that are not in any MVCs of cavity graph $\mathcal{G} \backslash i$, then vertex i will be present in all the MVCs of graph $\mathcal{G}$.

Adding vertex i to a MVC of cavity graph $\mathcal{G} \backslash i$ results in a MVC of graph $\mathcal{G}$ ($|\mathcal{C}_{\mathcal{G}}^0| = |\mathcal{C}_{\mathcal{G} \backslash i}^0|$). In this case $\pi_i = 1$ (i is always covered), $\Delta E_i = 1$, and $\Delta S_i = 0$.

Now we consider the remaining case, namely, in the set ∂i there is only one vertex (say j) which is not in any MVC of graph $\mathcal{G} \backslash i$. In this case, each MVC of graph $\mathcal{G}$ contains either vertex i or vertex j but not both. The energy increase is $\Delta E_i = 1$, and the total number of MVCs for graph $\mathcal{G}$ is

$$|\mathcal{C}_{\mathcal{G}}^0| = |\mathcal{C}_{\mathcal{G} \backslash i}^0| + |\mathcal{C}_{\mathcal{G} \backslash i,j}^0| \prod_{k \in \partial i \backslash j} \pi_{k|i}, \tag{6}$$

where $\mathcal{C}_{\mathcal{G} \backslash i,j}^0$ is the complete set of MVCs for the cavity graph $\mathcal{G} \backslash i, j$ (the remaining graph after further removing vertex j from $\mathcal{G} \backslash i$), and $\partial i \backslash j$ is the set of direct neighbors except j of vertex i. The first term on the right-hand side of the equality in Eq. (6) is the number of MVCs that contain vertex i, while the second term is the number of MVCs that contain j. In this case vertex i is said to be unfrozen. The entropy change $\Delta S_i = \ln(1 + e^{-\Delta S_{j|i}} \Pi_{k \in \partial i \backslash j} \pi_{k|i})$, and

$$\pi_i = 1/(1 + e^{-\Delta S_{j|i}} \prod_{k \in \partial i \backslash j} \pi_{k|i}),$$

where $\Delta S_{j|i} = \ln |\mathcal{C}_{\mathcal{G} \backslash i}^0| - \ln |\mathcal{C}_{\mathcal{G} \backslash i,j}^0|$ is the entropy change due to adding vertex j to the cavity graph $\mathcal{G} \backslash i, j$. We refer to $\Delta S_{j|i}$ as the cavity entropy of vertex j. Notice that, in this case, to properly calculate ΔS_i, one needs to consider not only all the MVCs of cavity graph $\mathcal{G} \backslash i$ but also vertex covers with a higher energy $E_{\mathcal{G} \backslash i}^0 + 1$.

The energy and entropy contribution of an edge (i, j) can also be obtained similarly. If at least one of the cavity probabilities $\pi_{j|i}$ and $\pi_{i|j}$ is positive, then $\Delta E_{(i,j)} = 0$ and $\Delta S_{(i,j)} = \ln[1 - (1 - \pi_{i|j})(1 - \pi_{j|i})]$. On the other hand, if both $\pi_{i|j}$ and $\pi_{j|i}$ are zero, then $\Delta E_{(i,j)} = 1$ and $\Delta S_{(i,j)} = \ln(e^{-\Delta S_{i|j}} + e^{-\Delta S_{j|i}})$.

On each of the $2M$ directed edges $j \to i$ of graph $\mathcal{G}$ there is a cavity probability $\pi_{j|i}$ and a cavity entropy $\Delta S_{j|i}$. After all these $4M$ values are determined, the total ground-state energy and entropy can then be evaluated using Eqs. (4) and (5). Starting from a random initial condition, the values of $\pi_{j|i}$, $\Delta S_{j|i}$, and the cavity energy increase $\Delta E_{j|i}$ can be determined by the following iterative algorithm. Consider vertices k in the neighborhood $\partial j \backslash i$ of a vertex j:

(1) If all $\pi_{k|j} > 0$, then $\pi_{j|i} = 0$, $\Delta E_{j|i} = 0$, $\Delta S_{j|i} = \Sigma_{k \in \partial j \backslash i} \ln \pi_{k|j}$.

(2) If two or more $\pi_{k|j} = 0$, then $\pi_{j|i} = 1$, $\Delta E_{j|i} = 1$, $\Delta S_{j|i} = 0$.

(3) If only one $\pi_{k|j} = 0$, then $\Delta E_{j|i} = 1$, and

$$\pi_{j|i} = \left(1 + e^{-\Delta S_{k|j}} \prod_{l \in \partial j \backslash i, k} \pi_{l|j}\right)^{-1}, \tag{7}$$

$$\Delta S_{j|i} = \ln\left(1 + e^{-\Delta S_{k|j}} \prod_{l \in \partial j \backslash i, k} \pi_{l|j}\right). \tag{8}$$

When Eq. (7) is used to update the cavity probability $\pi_{j|i}$, sometimes $\pi_{j|i}$ becomes extremely close to zero. A cutoff ϵ is therefore introduced in the numerical scheme: If the obtained value of $\pi_{j|i}$ is less than ϵ we set $\pi_{j|i} = 0$. The theoretical entropy values shown in Fig. 1 are obtained by setting a cutoff value of $\epsilon = 10^{-10}$ (more discussion of the choice of ϵ is given below).

The above discussion is concerned with one large graph
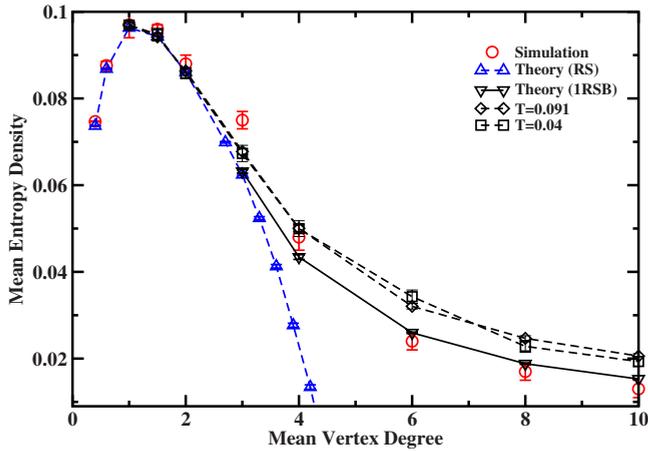
FIG. 1. (Color online) Ground-state entropy density of the vertex-cover problem as a function of the mean vertex degree c of the random graph. Circles are simulation data [11]; up and down triangles are, respectively, theoretical results obtained by the replica-symmetric and the 1RSB cavity methods. The finite-temperature entropy density values at $T=0.091$ and 0.04 are also shown as a comparison.

$\mathcal{G}$. As we are interested in the graph-averaged values for the ground-state energy and entropy, a population dynamics simulation can be performed similarly by storing a large array of $\pi_{j|i}$ and $\Delta S_{j|i}$ and then updating this array [1,4]. The ground-state entropy density $s^0 = \lim_{N \to \infty} S_{\mathcal{G}}^0/N$ as a function of mean connectivity c of the random graph $\mathcal{G}$ is evaluated by this numerical scheme (Fig. 1, up triangles). The ground-state entropy first increases with c and reaches a maximum at $c \approx 1$; then it decreases with increasing c. For $c < 2.7183$ the RS cavity method is known to be valid [4,13] and it correctly predicts the ground-state entropy and energy density for the system. For $c > 2.7183$, however, the RS prediction is systematically lower than the simulation result of Hartmann and Weigt (circles in Fig. 1) and even becomes negative for $c > 4.4$ (such an entropy crisis was also observed in the hitting set problem [5]). For $c > 2.7183$ the entropy predicted by the RS cavity method depends strongly on the cutoff ϵ. If a smaller ϵ value is used, the entropy decreases even faster with increasing mean connectivity c.

When the mean connectivity c of the random graph is larger than 2.7183, the Bethe-Peierls approximation Eq. (3) is no longer a good assumption, as there are strong long-range correlations among distant vertices [13]. For example, consider two vertices i and j whose shortest-path length is of order $\ln_c N$ and suppose these two vertices both are unfrozen ($\pi_i > 0$, $\pi_j > 0$) among the space of MVCs of graph $\mathcal{G}$. The Bethe-Peierls approximation assumes that the probability of finding a MVC which contains both i and j is equal to $\pi_i \pi_j > 0$. However, it may be the case that there is not a single MVC in which both i and j are present [13]. To take into account such long-range correlations, in the 1RSB cavity theory the MVC set $\mathcal{C}_{\mathcal{G}}^0$ of a graph $\mathcal{G}$ is clustered into many subsets $\mathcal{C}_{\mathcal{G}}^{0,\alpha}$ which are indexed by an index α. In each such subset it is assumed that the Bethe-Peierls approximation still holds, and Eq. (3) is replaced by

$$\mathbb{P}(\partial i \subset C : C \in \mathcal{C}_{\mathcal{G} \backslash i}^{0,\alpha}) = \prod_{j \in i} \pi_{j|i}^{\alpha}, \qquad (9)$$

where $\pi_{j|i}^{\alpha}$ is the probability of vertex j being in the MVCs of the αth subset $\mathcal{C}_{\mathcal{G} \backslash i}^{0,\alpha}$ of the cavity graph $\mathcal{G} \backslash i$. Because of Eq. (9), the iterative equations for $\pi_{j|i}$ and $\Delta S_{j|i}$ as mentioned before are still valid in each subset of MVCs. To characterize the property of different clusters, a probability distribution $\mathcal{P}_{j|i}(\pi, s)$ is introduced on each directed edge of the graph, which is equal to the fraction of MVC clusters with $\pi_{j|i} = \pi$ and $\Delta S_{j|i} = s$. For a single graph $\mathcal{G}$ these $2M$ probability distributions $\mathcal{P}_{j|i}$ can again be obtained by iterations, and the corresponding distribution of $\mathcal{P}_{j|i}(\pi, s)$ among all the edges of the random graph can be obtained by mean-field population dynamics.

The iteration equation for $\mathcal{P}_{j|i}(\pi, s)$ reads

$$\mathcal{P}_{j|i}(\pi, s) \propto \int \prod_{k \in \partial j \backslash i} \int d\pi_{k|j} ds_{k|j} \mathcal{P}_{k|j}(\pi_{k|j}, s_{k|j})$$
$$\times e^{-y \Delta E_{j|i}} \delta(\pi - \pi_{j|i}) \delta(s - \Delta S_{j|i}). \qquad (10)$$

The value of the reweighting parameter y in the above equation is chosen such that the complexity parameter is equal to zero [2]. In writing down Eq. (10), it is further assumed that the joint probability of observing the cavity values $(\pi_{k|j}, \Delta S_{k|j}), (\pi_{l|j}, \Delta S_{l|j}), \ldots$ for vertices $k, l, \ldots \in \partial j \backslash i$ can be written in a factorized form:

$$\mathbb{P}(\pi_{k|j}, \Delta S_{k|j}; \ldots) = \prod_{m \in \partial j \backslash i} \mathcal{P}_{m|j}(\pi_{m|j}, \Delta S_{m|j}). \qquad (11)$$

Details of the 1RSB numerical iteration scheme for the vertex-cover problem can be found in Ref. [4].

The mean entropy density of MVCs for random graphs of mean connectivity c as obtained by this 1RSB cavity method is shown in Fig. 1 (down triangles). The theoretical predictions are in good agreement with simulation results [12]. The mean ground-state energy density as obtained by the present method is also in good agreement with the simulation and theoretical results of Ref. [4]. In the calculation we have used a cutoff value $\epsilon = 10^{-10}$. Such a cutoff is necessary for the random vertex-cover problem, as the 1RSB cavity approach is not stable to further steps of replica-symmetry breaking [14]. Figure 2 shows that the predicted value of the mean ground-state entropy density is not sensitive to the cutoff parameter ϵ when $\epsilon \leq 10^{-5}$.

We also carry out lengthy population dynamics simulations at finite temperatures using different protocols. The entropy values obtained at $T=0.091$ and 0.04 are shown in Fig. 1. At these low temperatures, although the obtained energy density values are almost indiscernible from the ground-state values, there is still a gap between the finite-temperature and the ground-state entropy densities when $c > 2.7183$. If the temperature is lowered further, the quality of the simulation results deteriorates, possibly because of insufficient population size and insufficient equilibrium and sampling times. We were also unable to remove this gap by using instead the expansion Eq. (1), because the population dynamics di-
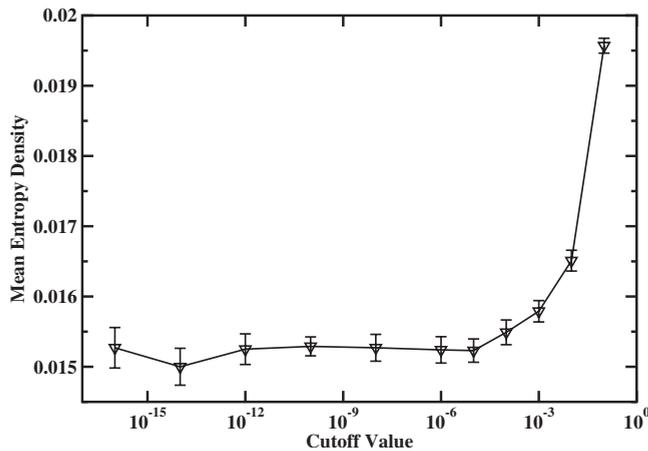
FIG. 2. Predicted mean ground-state entropy of the random vertex-cover problem at connectivity $c=10$ as a function of the cutoff parameter ϵ.

verges. In comparison with these approaches, the zero-temperature direct method is computationally much more efficient and also easier to implement.

In summary, we have calculated the ground-state entropy for the random vertex-cover problem using the 1RSB cavity approach of spin-glass theory. In our method, both the cavity probabilities $\pi_{j|i}$ and cavity entropies $\Delta S_{j|i}$ of each vertex j in a cluster of MVC solutions are recorded. We have paid special attention to unfrozen vertices (each of which belongs to some but not all MVCs of the graph). As demonstrated by Eq. (6), the entropy contribution of an unfrozen i comes not only from the MVCs of the graph $\mathcal{G}\backslash i$ but also from other higher-energy configurations of $\mathcal{G}\backslash i$. Similarly, the cavity entropy $\Delta S_{j|i}$ of a vertex j also has two sources of contribution. Equation (6) is rather simple for the vertex-cover problem, while for some other NP-hard CO problems and spin-glass models (e.g., the random-graph $\pm J$ spin glass) counting the entropy contribution of an unfrozen vertex can be more complicated. A cutoff parameter ϵ is introduced so that, if $\pi_{j|i} < \epsilon$ in one cluster, it is set to be zero. With this cutoff parameter, the present cavity method can still give good estimations for the ground-state entropy of a hard CO problem or spin-glass system even if more steps of replica-symmetry breaking are needed to fully describe the system.

[1] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).
[2] M. Mézard and G. Parisi, J. Stat. Phys. **111**, 1 (2003).
[3] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[4] M. Weigt and H. Zhou, Phys. Rev. E **74**, 046110 (2006).
[5] M. Mézard and M. Tarzia, Phys. Rev. E **76**, 041124 (2007).
[6] L. Zdeborová and M. Mézard, J. Stat. Mech.: Theory Exp. (2006) P05003.
[7] M. Mézard, M. Palassini, and O. Rivoire, Phys. Rev. Lett. **95**, 200202 (2005).
[8] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborova, Proc. Natl. Acad. Sci. U.S.A. **104**, 10318 (2007).
[9] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, J. Stat. Mech.: Theory Exp. (2008) P04004.
[10] A. K. Hartmann and M. Weigt, J. Phys. A **36**, 11069 (2003).
[11] M. Weigt and A. K. Hartmann, Phys. Rev. Lett. **84**, 6118 (2000).
[12] M. Weigt and A. K. Hartmann, Phys. Rev. E **63**, 056127 (2001).
[13] H. Zhou, Phys. Rev. Lett. **94**, 217203 (2005).
[14] J. Zhou, H. Ma, and H. Zhou, J. Stat. Mech.: Theory Exp. (2007) L06001.