Phase transitions in the coloring of random graphs

Lenka Zdeborová[1] and Florent Krząkała[2]

[1]*LPTMS, UMR 8626 CNRS et Université Paris-Sud, 91405 Orsay CEDEX, France*
[2]*PCT, UMR Gulliver 7083 CNRS-ESPCI, 10 rue Vauquelin, 75231 Paris, France*
(Received 11 April 2007; published 26 September 2007)

We consider the problem of coloring the vertices of a large sparse random graph with a given number of colors so that no adjacent vertices have the same color. Using the cavity method, we present a detailed and systematic analytical study of the space of proper colorings (solutions). We show that for a fixed number of colors and as the average vertex degree (number of constraints) increases, the set of solutions undergoes several phase transitions similar to those observed in the mean field theory of glasses. First, at the clustering transition, the entropically dominant part of the phase space decomposes into an exponential number of pure states so that beyond this transition a uniform sampling of solutions becomes hard. Afterward, the space of solutions condenses over a finite number of the largest states and consequently the total entropy of solutions becomes smaller than the annealed one. Another transition takes place when in all the entropically dominant states a finite fraction of nodes freezes so that each of these nodes is allowed a single color in all the solutions inside the state. Eventually, above the coloring threshold, no more solutions are available. We compute all the critical connectivities for Erdős-Rényi and regular random graphs and determine their asymptotic values for a large number of colors. Finally, we discuss the algorithmic consequences of our findings. We argue that the onset of computational hardness is not associated with the clustering transition and we suggest instead that the freezing transition might be the relevant phenomenon. We also discuss the performance of a simple local Walk-COL algorithm and of the belief propagation algorithm in the light of our results.

I. INTRODUCTION

Graph coloring is a famous yet basic problem in combinatorics. Given a graph and q-colors, the problem consists in coloring the vertices in such a way that no connected vertices have the same color [1]. The celebrated four-colors theorem assures that this is always possible for planar graphs using only four colors [2]. For general graphs, however, the problem can be extremely hard to solve and is known to be NP complete (NP states for nondeterministic polynomial time) [3], so that it is widely believed that no algorithm can decide in a polynomial time (with respect to the size of the graph) if a given arbitrary instance is colorable or not. Indeed, the problem is often taken as a benchmark for the evaluation of the performance of algorithms in computer science. It also has an important practical application as time tabling, scheduling, register allocation in compilers or frequency assignment in mobile radios.

In this paper, we study colorings of sparse random graphs [4,5]. Random graphs are one of the most fundamental sources of challenging problems in graph theory since the seminal work of Erdős and Rényi [6] in 1959. Concerning the coloring problem, a crucial observation was made by focusing on typical instances drawn from the ensemble of random graphs with a given average vertex connectivity c: as c increases a threshold phenomenon is observed. Below a critical value c_s a proper coloring of the graph with q colors exists with a probability going to one in the large size limit, while beyond c_s it does not exist in the same sense. This sharp transition also appears in other constraint satisfaction problems (CSPs) such as the satisfiability of Boolean formulas [1]. The existence of the sharp COLorable/UNCOLorable (COL/UNCOL) transition was partially [85] proven in [7],

and computing rigorously its precise location is a major open problem in graph theory. Many upper and lower bounds were established [8–15] for Erdős-Rényi and regular random graphs.

It was also observed empirically [16,17] that deciding colorability becomes on average much harder near to the coloring threshold c_s than far away from it. This onset of computational hardness cannot be explained only by the simple fact that near to the colorable threshold the number of proper colorings is small [18]. Some progress in the theoretical understanding has been done by the analysis of search algorithms [19,20]. For the coloring problem, it was proven [21] that a simple algorithm q-colors almost surely in linear time random graphs of average connectivity $c \leq q \ln q - 3q/2$ for all $q \geq 3$ (see [21] for references on previous works). For three-coloring the best algorithmic lower bound is $c = 4.03$ [10]. An important and interesting open question [22] is the existence of an $\epsilon > 0$ and of a polynomial algorithm which q-colors almost surely a random graph of connectivity $c = (1+\epsilon)q \ln q$ for arbitrary large q.

The sharp coloring threshold and the onset of hardness in its vicinity has also triggered a lot of interest within the statistical physics community following the discovery of a close relation between constraint satisfaction problems and spin glasses [23,24]. In physical terms, solving a CSP consists in finding ground states of zero energy. The limit of infinitely large graphs corresponds to the thermodynamic limit. In the case of the q-coloring problem, for instance, one studies the zero temperature behavior of the antiferromagnetic Potts model [25]. Using this correspondence, a powerful heuristic tool called the *cavity method* has been developed [24,26–28]; it allows an exact analytical study of the CSPs on sparse random graphs. Unfortunately, some pieces
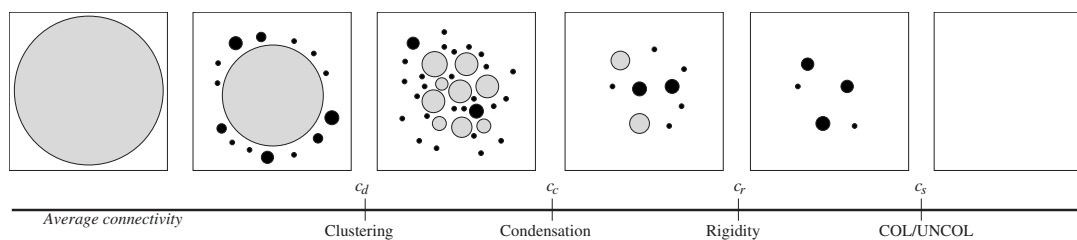
FIG. 1. A sketch of the set of solutions when the average connectivity (degree) is increased. At low connectivities (on the left), all solutions are in a single cluster. For larger c, clusters of solutions appear but the single giant cluster still exists and dominates the measure. At the dynamic/clustering transition c_d, the phase space slits in an exponential number of clusters. At the condensation/Kauzmann transition c_c, the measure condenses over the largest of them. Finally, no solutions exist above the COL/UNCOL transition c_s. The rigidity/freezing transition c_r (which might appear before or after the condensation transition) takes place when the dominating clusters (the minimal set of clusters that covers almost all proper colorings) start to contain frozen variables. The clusters containing frozen variables are colored in black and those which do not are colored in gray.

are still missing to make the cavity method a rigorous tool although many of its results were rigorously proven. The cavity method is equivalent to the famous replica method [29] (and unfortunately for clarity, it has also inherited some of its notations, as we shall see).

The cavity method and the statistical physics approach have been used to study the q-coloring of random graphs in [30–32]. The coloring threshold c_s was calculated [30], the self-consistency of the solution checked [31], and the large q asymptotics of the coloring threshold computed [31]. These results are believed to be *exact* although proving their validity rigorously remains a major subject in the field. Nevertheless—as the results obtained for the random satisfiability problem [28,33,34]—they agree perfectly with rigorous mathematical bounds [8–13,15], and with numerical simulations. The coloring threshold is thus fairly well understood, at least at the level of the cavity method.

A maybe more interesting outcome of the statistical physics analysis of the CSPs was the identification of a new transition, which concerns the structure of the set of solutions, and that appears before the coloring threshold [24,28,35]: while at low connectivities all solutions are in a single pure state (cluster) [86], the set of solutions splits in an exponential number of different states (clusters) at a connectivity strictly smaller than c_s. Roughly speaking, clusters are groups of nearby solutions that are in some sense disconnected from each other. Recently, the existence of the clustered phase was proven rigorously in some cases for the satisfiability problem [36,37]. A major step was made by applying the cavity equations on a single instance: this led to the development of a very efficient message-passing algorithm called *survey propagation* (SP) that was originally used for the satisfaction problem in [24] and later adapted for the coloring problem in [30]. Survey propagation allows one to find solutions of large random instances even in the clustered phase and very near to the coloring threshold.

Despite all this success, the cavity description of the clustered phase was not complete in many aspects, and a first improvement has been made with the introduction of a refined zero temperature cavity formalism that allows a more detailed description of the geometrical properties of the clusters [38,39]. We pursue in this direction and give for the first time a detailed characterization of the structure of the set of solutions. We observed, in particular, that the clustering threshold was not correctly computed, that other important transitions were overlooked, and the global picture was mixed up. The corrected picture that we describe in this paper is the following: when the connectivity is increased, the set of solutions undergoes several phase transitions similar to those observed in mean field structural glasses (we sketch these successive transitions in Fig. 1). First, the phase space of solutions decomposes into an exponential number of states, which are entropically negligible with respect to one large cluster. Then, at the clustering threshold c_d, even this large state decomposes into an exponential number of smaller states. Subsequently, above the condensation threshold c_c, most of the solutions are found in a finite number of the largest states. Eventually, the connectivity c_s is reached beyond which no more solutions exist. Another important transition, which we refer to as *rigidity*, takes place at c_r when a finite fraction of frozen variables appears inside the dominant pure states (those containing almost all the solutions). All those transitions are sharp and we computed the values of the corresponding critical connectivities.

A nontrivial ergodicity breaking takes place at the clustering transition; in consequence uniform sampling of solutions becomes hard. On the other hand, clustering itself is not responsible for the hardness of finding a solution. Moreover, until the condensation transition many results obtained by neglecting the clustering effect are correct. In particular, for all $c < c_c$: (i) the number of solutions is correctly given by the annealed entropy (and, for general CSP, by the replica symmetric entropy), and (ii) a simple message-passing algorithm such as belief propagation (BP) [40,41] converges to a set of exact marginals (i.e., the probability that a given node takes a given color). In consequence we can use BP plus a decimationlike strategy to find proper colorings on a given graph. Finally, we give some arguments to explain why the rigidity transition is a better candidate for the onset of computational hardness in finding solutions.

Our results are obtained within the one-step replica symmetry breaking approach, and we believe (and argue partially in Sec. IV), that our results would not be modified by considering further steps of RSB (as opposed to previous conclusions [31]).

A shorter and partial version of our results, together with a study of similar issues in the satisfiability problem, was

already published in [42]. We refer to [43] for a detailed discussion of the satisfiability problem. The paper is organized as follows: In Sec. II, we present the model. In Sec. III, we introduce the cavity formalism at the so-called replica symmetric level, and discuss in detail why and where this approach fails. In Sec. IV we take into account the existence of clusters of solutions and employ the so-called one-step replica symmetry breaking formalism to describe the properties of clusters. The results for several ensembles of random graphs are then presented in Sec. V. We finally discuss the algorithmic implications of our findings in Sec. VI and conclude by a general discussion. Some appendixes with detailed computations complete the paper.

II. MODEL

A. Definition of the model

For the statistical physics analysis of the q-coloring problem [30,32,44] we consider a Potts [25] spin model with antiferromagnetic interactions where each variable s_i (spin, node, vertex) is in one of the q different states (colors) $s = 1, \ldots, q$. Consider a graph $G = (\mathcal{V}, \mathcal{E})$ defined by its vertices $\mathcal{V} = \{1, \ldots, N\}$ and edges $(i, j) \in \mathcal{E}$, which connect pairs of vertices $i, j \in \mathcal{V}$; we write the Hamiltonian as

$$\mathcal{H}(\{s\}) = \sum_{(i,j) \in \mathcal{E}} \delta(s_i, s_j). \qquad (1)$$

With this choice there is no energy contribution for neighbors with different colors, but a positive contribution otherwise. The ground state energy is thus zero if and only if the graph is q-colorable. The Hamiltonian leads to a Gibbs measure [45] over configurations (where β is the inverse temperature) as follows:

$$\mu(\{s\}) = \frac{1}{Z_0} e^{-\beta \mathcal{H}(\{s\})}. \qquad (2)$$

In the zero temperature limit, where $\beta \to \infty$, this measure becomes uniform over all the proper colorings of the graph.

B. Ensembles of random graphs

We will consider ensembles of graphs that are given by a degree distribution $\mathcal{Q}(k)$. The required property of $\mathcal{Q}(k)$ is that its parameters should not depend on the size of the graph. All the analytical results will concern only very large sparse graphs ($N \to \infty$). Provided the second and higher moments of $\mathcal{Q}(k)$ do not diverge, such graphs are locally tree-like in this limit [4,5]. More precisely, call a d-neighborhood of a node i the set of nodes which are at distance at most d from i. For d arbitrary but finite the d-neighborhood is almost surely a tree graph when $N \to \infty$. This property is connected with the fact that the length ℓ of the shortest loops (up to a finite number of them) scales with the graph size as $\ell \sim \ln(N)$. We will consider the two following canonical degree distribution functions:

(i) Uniform degree distribution $\mathcal{Q}(k) = \delta(k - c)$ corresponding to the *c-regular random graphs*, where every vertex has exactly c neighbors.

(ii) Poissonian degree distribution $\mathcal{Q}(k) = e^{-c} c^k / k!$ corresponding to the Erdős and Rényi (ER) random graphs [6]. A simple way to generate graphs with N vertices from this ensemble is to consider that each link is present with probability $c/(N-1)$. The binomial degree distribution converges to the Poissonian in the large size limit.

It will turn out that the cavity technics simplify considerably for the regular graphs. However, ER graphs have the advantage that their average connectivity is a real number that can be continuously tuned, which is obviously very convenient when one wants to study phase transitions. It is thus useful to introduce a third ensemble, which still has the computational advantage of regular graphs, but that at the same time gives more freedom to vary the connectivity:

(iii) Biregular random graphs, where nodes with connectivity c_1 are all connected to nodes with connectivity c_2, and vice versa. There are thus two sets of nodes with degree distributions $\mathcal{Q}(k) = \delta(k - c_1)$ and $\mathcal{Q}(k) = \delta(k - c_2)$.

Notice that these graphs are bipartite by definition and therefore have always a trivial two-coloring, which we will have to dismiss in the following. This can be easily done within the cavity formalism (it is equivalent to neglecting the ordered "crystal" phase in glass models [46]).

In the first two cases, the parameter c plays the role of the average connectivity $c = \Sigma_k k \mathcal{Q}(k)$. In the cavity approach, a very important quantity is the *excess degree* distribution $\mathcal{Q}_1(k)$, i.e., the distribution of the number of neighbors, different from j, of a vertex i adjacent to a random edge (ij). It holds:

$$\mathcal{Q}_1(k) = \frac{(k+1)\mathcal{Q}(k+1)}{c}. \qquad (3)$$

This distribution remains Poissonian for Erdős-Rényi graphs, whereas the excess degree is equal to $c-1$ in the case of regular graphs. In the biregular case, there are two sets of nodes with excess degrees $c1-1$ and $c2-1$.

III. CAVITY FORMALISM AT THE REPLICA SYMMETRIC LEVEL

We start by reviewing the *replica symmetric* (RS) version of the cavity method [26,27] and its implications for the coloring problem. In the last part of the section we show when and why the RS approach fails.

A. Replica symmetric cavity equations

The coloring problem on a tree is solved exactly by an iterative method called the belief propagation algorithm [40] (note some boundary conditions have to be imposed, otherwise a tree is always two-colorable) that is equivalent to the replica symmetric cavity method [41]. At this level, the method is in fact a classical tool in statistical physics to deal with that tree structure that dates back to the original ideas of Bethe, Peierls, and Onsager [47]. It allows one to compute the marginal probabilities that a given node takes a given color and, in the language of statistical physics, observables like energy, entropy, average magnetization, etc. The applicability of the method goes, however, beyond tree graphs and
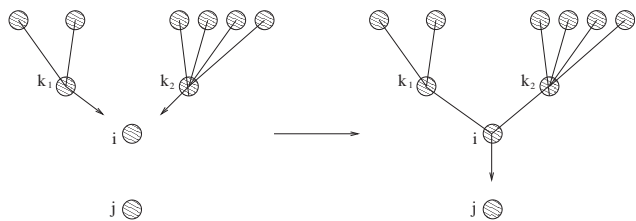
FIG. 2. Iterative construction of a tree by adding a new Potts spin.

we will discuss when it is correct for random locally treelike graphs in Sec. III C.

Let us now describe the RS cavity equations. Denote $\psi_{s_i}^{j \to j}$ the probability that the spin i has color s_i when the edge (ij) is not present and consider the iterative construction of a tree in Fig. 2. The probability follows the recursive relation

$$\psi_{s_i}^{j \to j} = \frac{1}{Z_0^{i \to j}} \prod_{k \in i-j} \sum_{s_k} e^{-\beta \delta_{s_i s_k}} \psi_{s_k}^{k \to i}$$

$$= \frac{1}{Z_0^{i \to j}} \prod_{k \in i-j} [1 - (1 - e^{-\beta}) \psi_{s_i}^{k \to i}], \quad (4)$$

where $Z_0^{i \to j}$ is a normalization constant (the cavity partition sum) and β the inverse temperature. The notation $k \in i-j$ means the set of neighbors of i except j. The normalization $Z_0^{i \to j}$ is related to the free energy shift after the addition of the node i and the edges around it, except (ij), as

$$Z_0^{i \to j} = e^{-\beta \Delta F^{i \to j}}. \quad (5)$$

In the same way, the free energy shift after the addition of the node i and all the edges around it is

$$e^{-\beta \Delta F^i} = Z_0^i = \sum_s \prod_{k \in i} [1 - (1 - e^{-\beta}) \psi_s^{k \to i}]. \quad (6)$$

The Z_0^i is at the same time the normalization of the total probability that a node i takes color s_i (the marginal of i), which reads

$$\psi_{s_i}^i = \frac{1}{Z_0^i} \prod_{k \in i} [1 - (1 - e^{-\beta}) \psi_{s_i}^{k \to i}]. \quad (7)$$

The free energy shift after the addition of an edge (ij) is

$$e^{-\beta \Delta F^{ij}} = Z_0^{ij} = 1 - (1 - e^{-\beta}) \sum_s \psi_s^{j \to i} \psi_s^{j \to j}. \quad (8)$$

The free energy density in the thermodynamic limit is then given by (see, for instance, [26])

$$f(\beta) = \frac{1}{N} \left(\sum_i \Delta F^i - \sum_{ij} \Delta F^{ij} \right). \quad (9)$$

Note that this relation for the free energy is variational, i.e., that if one differentiates with respect to β, then only the *explicit dependence* needs to be considered (see [26] for details). The energy (the number of contradictions) and the corresponding entropy (the logarithm of the number of solutions

with a given energy) densities can be then computed from the Legendre transform

$$-\beta f(\beta) = -\beta e + s(e), \quad (10)$$

where $f = F/N$, $e = E/B$, and $s = S/N$ are intensive variables.

The learned reader will notice that in some previous works using the cavity method [30,31], these equations were often written for a different object than the probabilities ψ. Instead, the so-called cavity magnetic fields h and biases u were considered. The two approaches are related via

$$e^{-\beta h_{s_i}^{i \to j}} \equiv Z_0^{i \to j} \psi_{s_i}^{i \to j} \equiv \prod_{k \in i-j} e^{-\beta u_{s_i}^{k \to i}}. \quad (11)$$

Strictly speaking the ψ are "cavity probabilities" while h and u are "cavity fields"; however, ψ are sometimes also referred to as cavity fields (or messages) in the literature, and the reader will thus forgive us if we do so.

Note that each of the two notations is suitable for performing the zero temperature limit in a different way: in the first one we fix the energy zero and we obtain the zero temperature BP recursion, which gives the marginal probabilities for each variable, while in the second case, we obtain a simpler recursion called *warning propagation* (49) that deals with the energetic contributions but neglects the entropic ones [28,30]. This is the origin of the discrepancy between the RS results of Refs. [32,30]. As we shall see, the differences between these two limits will be an important point in the paper.

B. Average over the ensemble of graphs and the RS solution

To compute (quenched) averages of observables over the considered ensemble of random graphs given by the degree distribution $\mathcal{Q}(k)$ we need to solve the self-consistent cavity functional equation

$$\mathcal{P}(\psi) = \sum_k \mathcal{Q}_1(k) \prod_{i=1}^k \int d\psi^i \mathcal{P}(\psi^i) \delta[\psi - \mathcal{F}(\{\psi^i\})], \quad (12)$$

where $\mathcal{Q}_1(k)$ is the distribution of the number of neighbors given that there is already one neighbor, and the function $\mathcal{F}(\{\psi^i\})$ is given by Eq. (4). Beware that ψ here is a q-components vector while we omit the vector notation to lighten the reading. This equation is quite complicated since the order parameter is nontrivial, but we can solve it numerically using the population dynamics method described in [26,27].

Throughout this paper we will search only for the color symmetric functions $\mathcal{P}(\psi)$, i.e., invariant under permutation of colors. Clearly with this assumption we might miss some solutions of Eq. (12). Consider, for example, $q > 2$ colors and the ensemble of random biregular graphs. Since every bipartite graph is two-colorable there are $q(q-1)/2$ corresponding color asymmetric solutions for $\mathcal{P}(\psi)$. For the ensemble of random graphs considered here, we later argue that this assumption is, however, justified.

Another important observation is that for regular graphs Eq. (12) crucially simplifies: the solution *factorizes* [48] in the sense that the order parameter ψ is the same for each

edge. This is due to the fact that, locally, every edge in such a regular graph has the same environment. All edges are therefore equivalent and thus the distribution $\mathcal{P}(\psi)$ has to be a delta function. For the biregular graphs, the solution of Eq. (12) also factorizes, but the two sets of nodes of connectivity c_1 and c_2 (each of them being connected to the other) have to be considered separately.

It is immediate to observe that $\mathcal{P}(\psi) = \delta(\psi - 1/q)$ (i.e., each of the q components of each cavity field ψ equals $1/q$), is always a solution of Eq. (12). By analogy with magnetic systems we shall call this solution *paramagnetic*. Numerically, we do not find any other solution in the colorable phase. For regular random graphs the paramagnetic solution is actually the only factorized one. The number of proper colorings predicted by the RS approach is thus easy to compute. Since all messages are of the type $\mathcal{P}(\psi) = \delta(\psi - 1/q)$, the free energy density simply reads

$$-\beta f_{\mathrm{RS}} = \ln q + \frac{c}{2} \ln\left(1 - \frac{1 - e^{-\beta}}{q}\right).\tag{13}$$

The entropy density at zero temperature thus follows:

$$s_{\mathrm{RS}} = \ln q + \frac{c}{2} \ln\left(1 - \frac{1}{q}\right).\tag{14}$$

It coincides precisely with the annealed (first moment) entropy. We will see in the following that, surprisingly, the validity of this formula goes well beyond the RS phase (actually until the so-called condensation transition).

C. Validity conditions of the replica symmetric solution

We used the main assumption of the replica symmetric approach when we wrote Eq. (4): we supposed that the cavity probabilities $\psi_{s_i}^{k \to i}$ for the neighbors k of the node i are "sufficiently" independent in the absence of the node i, because only then the joint probability factorizes. This assumption would be true if the lattice were a tree with noncorrelated boundary conditions, but loops, or correlations in the boundaries, may create correlations between the neighbors of node i (in the absence of i) and the RS cavity assumption might thus cease to be valid in a general graph. The aim here is to be more precise and quantify this statement both from a rigorous and heuristic point of view.

1. Gibbs measure uniqueness condition

Proving rigorously the correctness of the RS cavity assumption for random graphs is a crucial step that has not yet been successfully overcome in most cases. The only success so far was obtained by proving a far too strong condition: *the Gibbs measure uniqueness* [49–52]. Roughly speaking, the Gibbs measure (2) is unique if the behavior of a spin i is totally independent from the boundary conditions (i.e., very distant spins) for *any* possible boundary conditions. More precisely, let us define $\{s_l\}$ colors of all the spins at a distance at least l from the spin i. The Gibbs measure μ is unique if and only if the following condition holds for every i (and in the limit $N \to \infty$):

$$\mathbb{E}\left[\sup_{\{s_l\},\{s_l'\}} \sum_{s_i=1}^{q} |\mu(s_i|\{s_l\}) - \mu(s_i|\{s_l'\})|\right] \xrightarrow[l \to \infty]{} 0,\tag{15}$$

where the average is over the ensemble of graphs. In [50,51], it was proven that the Gibbs measure in the coloring problem on random regular graphs is unique only for graphs of degree $c < q$. To the best of our knowledge, this has not been computed for Erdős-Rényi graphs (later in this section, we argue on the basis of a physical argument that it should be $c < q - 1$ in this case).

2. Gibbs measure extremality condition

In many cases, the RS approach is observed to be correct beyond the uniqueness threshold. It was thus suggested in [53] (see also [42]) that the Gibbs measure extremality provides a proper criterion for the correctness of the replica symmetric assumption. Roughly speaking, the difference between uniqueness and extremality of a Gibbs measure is that although there may exist boundary conditions for which the spin i is behaving differently than for others, such boundary conditions have a null measure if the extremality condition is fulfilled. Formally (and keeping the notations from the previous section), the extremality corresponds to

$$\mathbb{E}\left[\sum_{\{s_l\}} \mu(\{s_l\}) \sum_{s_i=1}^{q} |\mu(s_i|\{s_l\}) - \mu(s_i)|\right] \xrightarrow[l \to \infty]{} 0.\tag{16}$$

In mathematics the "extremal Gibbs measure" is often used as a synonym for a "pure state." Recently, the authors of [53] provided rigorous bounds for the Gibbs measure extremality of the coloring problem on trees.

There exist two heuristic equivalent approaches to check the extremality condition. In the first one, one studies the divergence of the so-called "point-to-set" correlation length [54,55]. The second one is directly related to the cavity formalism: one should check for the existence of a nontrivial solution of the one-step replica symmetry breaking equations (1RSB) at $m=1$ (see Sec. V). Both these analogies were remarked in [53] and exploited in [42]. We will show in Sec. V that the extremality condition ceases to be valid at the clustering threshold c_d, beyond which the 1RSB formalism will be needed.

3. Local stability: A simple self-consistency check

A necessary, simple to compute but not sufficient, validity condition for the RS assumption is the nondivergence of the spin glass susceptibility (see, for instance, [56]). If it diverges, a spin glass transition occurs, and the replica symmetry has to be broken [29]. The local stability analysis thus gives an upper bound to the Gibbs extremality condition, which remarkably coincides with the rigorous upper bound of [57]. This susceptibility is defined as

$$\chi_{\mathrm{SG}} = \frac{1}{N} \sum_{i,j} \langle s_i s_j \rangle_c^2.\tag{17}$$

The connectivities above which it diverges at zero temperature can be computed exactly within the cavity formalism

(we refer to Appendix A for the derivation). It follows for regular and Erdős-Rényi graphs,

$$c_{RS}^{reg} = q^2 - 2q + 2, \quad c_{RS}^{ER} = q^2 - 2q + 1, \quad (18)$$

while the stability of the biregular graphs of connectivities c_1, c_2 is equivalent to regular graphs with $c = 1 + \sqrt{(c_1-1)(c_2-1)}$.

Note that for regular and ER graphs the RS instability threshold is in the colorable phase only for $q=3$. Indeed the five-regular graphs are three-colorable [31] (and rigorous results in [13,14]) and exactly critical since $c_{RS}^{reg}(3)=5$, and for ER graphs the COL/UNCOL transition appears at $c_s \approx 4.69$ [30] while $c_{RS}^{ER}(3)=4$. This means that the replica symmetry breaking transition appears continuously at the point c_{RS} so that above it the RS approach is not valid anymore. For all $q \geq 4$, however, the local stability point is found beyond the best upper bound on the coloring threshold for both regular and ER graphs. In this case, the extremality condition will not be violated by the continuous mechanism, but we will see that, instead, a discontinuous phase transition, as happens in mean field structural glasses, will take place.

Interestingly enough, a similar computation can be made for the ferromagnetic susceptibility $\chi_F = \frac{1}{N}\sum_{i,j}\langle s_i s_j \rangle_c$ (see again Appendix A). It diverges at $c=q$ for regular graph and $c=q-1$ for Erdős-Rényi graphs. This divergence (called *modulation* instability in [56]) would announce the transition towards an antiferromagnetic ordering on a tree, which is, however, incompatible with the frustrating loops in a random graph (although such order might exist on the biregular graphs). This is precisely the solution we dismiss when considering only the color symmetric solution of Eq. (12). Note, however, that the presence of this instability shows that the problem ceases to a have a unique Gibbs state (although it is still extremal) as for some specific (and well-chosen) boundary conditions, an antiferromagnetic solution may appear. Indeed it coincides perfectly with the rigorous uniqueness condition $c=q$ for regular graphs, and suggests strongly that the uniqueness threshold (or at least an upper bound) for Erdős-Rényi graphs is $c=q-1$.

IV. ONE-STEP REPLICA SYMMETRY BREAKING FRAMEWORK

So far we described the RS cavity method for coloring random graphs and explained that the extremality of the Gibbs measure gives a validity criterion. We now describe the one-step replica symmetry breaking cavity solution [26,27]. In this approach, the nonextremality of the Gibbs measure is cured, by decomposing it into several parts (pure states, clusters) in such a way that within each of the states the Gibbs measure becomes again extremal.

This decomposition has many elements or states, not just a finite number like the q states of the usual ferromagnetic Potts models. It is actually found that the number of pure states is growing exponentially with the size of the system. Let us define the state-entropy function $\Sigma(f)$—called the *complexity*—which is just the logarithm of the number of

states with internal free energy density f, i.e., $\mathcal{N}(f) = \exp[N\Sigma(f)]$. In the glass transition formalism, this complexity is usually referred to as the *configurational entropy*. Dealing with exponentially many pure states is obviously a nightmare for all known rigorous approaches to the thermodynamic limit. The heuristic cavity method overcomes this problem elegantly, as was shown originally in the seminal work of [26,27].

Another very useful intuition about the 1RSB cavity method comes from the identification of states α with the fixed points $\{\psi\}$ of the belief propagation equations (4). The goal is thus to compute the statistical properties of these fixed points. Each of the states is weighted by the corresponding free energy (9) to the power m, where m is just a parameter analogous to the inverse temperature β (in the literature m is often called the Parisi replica symmetry breaking parameter [29,58]). The probability measure over states $\{\psi\}$ is then

$$\tilde{\mu}(\{\psi\}) = \frac{Z_0(\{\psi\})^m}{Z_1} = \frac{1}{Z_1} e^{-\beta m N f(\{\psi\})}, \quad (19)$$

where Z_1 is just the normalization constant. To write the analog of the belief propagation equations we need to define the probability (distribution) $P^{i \to j}(\psi^{i \to j})$ of the fields $\psi^{i \to j}$. This can be obtained from those of incoming fields as

$$P^{i \to j}(\psi_{s_i}^{i \to j}) = \frac{1}{Z_1^{i \to j}} \prod_{k \in i-j} \int d\psi_{s_i}^{k \to i} P^{k \to i}(\psi_{s_i}^{k \to i}) \delta[\psi_{s_i}^{i \to j}$$
$$- \mathcal{F}(\{\psi_{s_i}^{k \to i}\})](Z_0^{i \to j})^m$$
$$\equiv \frac{1}{Z_1^{i \to j}} \int_{POP} \delta(\psi - \mathcal{F})(Z_0^{i \to j})^m. \quad (20)$$

The function $\mathcal{F}$ is given by Eq. (4) and the delta function ensures that the set of fields $\psi^{i \to j}$ is a fixed point of the belief propagation (4). The reweighting term $(Z_0^{i \to j})^m$ takes into account the change of the free energy of a state after the addition of a cavity spin i and its adjacent edges except (ij), as defined in Eq. (5). This term appears for the same reason as a Boltzmann factor $e^{-\beta \delta_{s_i, s_k}}$ in Eq. (4): it ensures that the state α is weighted by $(Z_\alpha)^m$ in the same way a configuration $\{s\}$ is weighted by $e^{-\beta \mathcal{H}(\{s\})}$ in Eq. (2). Finally, $Z_1^{i \to j}$ is a normalization constant. In the second line of Eq. (20) we introduced an abbreviation that will be used from now on to make the equations more easily readable. The notation POP comes from "population dynamics," which refers to the numerical method we use to solve Eq. (20). The probability distribution $P(\psi)$ can be represented numerically by a set of fields taken from $P(\psi)$, and then the probability measure $P(\psi)d\psi$ becomes uniform sampling from this set; for more details see Appendix D.

We define the "replicated free energy" and compute it in analogy with Eq. (9) as

$$\Phi(\beta,m) \equiv -\frac{1}{\beta m N}\ln(Z_1) = \frac{1}{N}\left(\sum_i \Delta\Phi^i - \sum_{ij}\Delta\Phi^{ij}\right),$$
$$(21)$$

where

$$e^{-\beta m \Delta\Phi^i} = \int_{\mathrm{POP}} e^{-\beta m \Delta F^i}, \quad e^{-\beta m \Delta\Phi^{ij}} = \int_{\mathrm{POP}} e^{-\beta m \Delta F^{ij}}.$$
$$(22)$$

Putting together Eqs. (19) and (21) we have

$$Z_1 = e^{-\beta m N \Phi(\beta,m)} = \sum_{\{\psi\}} e^{-\beta m N f(\{\psi\})} = \int_f df e^{-N[\beta m f(\beta) - \Sigma(f)]},$$
$$(23)$$

where the sum over $\{\psi\}$ is over all states (or BP fixed points). In the interpretation of [59] m is the number of replicas of the system, thus the name "replicated free energy" for $\Phi(\beta,m)$. Note that we are using the word "replica" only to refer to the established terminology as no replicas are needed within the cavity formalism. From the saddle point method, it follows that the Legendre transform of complexity function $\Sigma(f)$ gives the replicated free energy $\Phi(m)$,

$$-\beta m \Phi(\beta,m) = -\beta m f(\beta) + \Sigma(f). \quad (24)$$

Notice that this equation is correct only in the highest order in the system size N, i.e., in densities and at the thermodynamic limit. From the properties of the Legendre transform we have

$$\Sigma = \beta m^2 \partial_m \Phi(\beta,m), \quad f = \partial_m[m\Phi(\beta,m)], \quad \beta m = \partial_f \Sigma(f). \quad (25)$$

Thus, from Eq. (21), the free energy reads

$$f(\beta) = \sum_i \frac{\int_{\mathrm{POP}} \Delta F^i e^{-\beta m \Delta F^i}}{\int_{\mathrm{POP}} e^{-\beta m \Delta F^i}} - \sum_{ij} \frac{\int_{\mathrm{POP}} \Delta F^{ij} e^{-\beta m \Delta F^{ij}}}{\int_{\mathrm{POP}} e^{-\beta m \Delta F^{ij}}}.$$
$$(26)$$

When the parameter m is equal to one (the number of replicas is actually one in the approach of [59]), then $-\beta\Phi(\beta,1) = -\beta f(\beta) + \Sigma(f)$ reduces to the usual free energy function considered in the RS approximation

$$\Phi(\beta,1) = e - \frac{\Sigma + s}{\beta} = e - T s_{\mathrm{tot}}, \quad (27)$$

where s is the internal entropy density of the corresponding clusters and s_{tot} is the total entropy density of the system.

A. Analyzing the 1RSB equations

Combining Eqs. (21) and (26) we can compute Σ and f for each value m, which gives us implicitly $\Sigma(f)$. To compute the thermodynamic observables in the model we have to minimize the total free energy $f_{\mathrm{tot}} = f(\beta) - \Sigma/\beta$ over such val-

ues of f where the complexity $\Sigma(f)$ is non-negative (so that the states exist in the thermodynamic limit). The minimum of the total free energy corresponds to a value of parameter $m = m^*$ and states with the free energy f^* dominate the thermodynamics. Three different cases are then observed:

(a) If there is only the trivial (replica symmetric) solution at $m=1$, then the Gibbs measure (2) is extremal and the replica symmetric approach is correct. If at the same time a nontrivial solution exists for some $m \neq 1$, then the clusters corresponding to this solution have no influence on the thermodynamics.

(b) If there is a nontrivial solution at $m=1$ with a positive complexity, then $m^*=1$ minimizes the total free energy. The system is in a *clustered phase* with an exponential number of dominating states.

(c) If, however, the complexity is negative at $m=1$, then the corresponding states are absent with probability one in the thermodynamic limit. Instead the total entropy is dominated by clusters corresponding to m^* such that $\Sigma(m^*)=0$: the system is in a *condensed phase*. Note that the condition $\Sigma(m^*)=0$ corresponds to the maximum of the replicated free energy (21).

The transitions between these cases are well known in structural glass phenomenology where they appear when the temperature is lowered [60,61]. The transition from the paramagnetic phase to the clustered one is usually referred to as the *dynamical transition* [62] or the *clustering transition*. It is not a true thermodynamic transition as the total free energy of the system at $m^*=1$ is still equal to the replica symmetric one [Eq. (9)] (see Appendix C) and thus is an analytical function of connectivity. However, the phase space is broken into exponentially many components and, as a consequence, the dynamics fall out of equilibrium beyond this transition.

The second transition from the clustered to the condensed phase is, however, a genuine thermodynamic transition (the free energy has a discontinuity in the second derivative at c_c) and is called the *replica symmetry breaking transition*, or the static *glass transition*. At this point the measure condenses into few clusters, and we shall call it the *condensation transition*. In structural glasses, it corresponds to the well known *Kauzmann transition* [63]. The sizes of the clusters in the condensed phase follow the so-called Poisson-Dirichlet process, which is discussed shortly in Appendix B.

The procedure to compute the replicated free energy (21) and the related observables was described above for a single large random graph. To compute the averages over the ensemble of random graphs, we need to solve an equation analogous to Eq. (12),

$$\mathcal{P}[P(\psi)] = \sum_k \mathcal{Q}_1(k)\prod_{i=1}^k \int dP^i(\psi)\mathcal{P}[P^i(\psi)]\delta[P(\psi)$$
$$- \mathcal{F}_2(\{P^i(\psi)\})], \quad (28)$$

where the functional $\mathcal{F}_2$ is given by Eq. (20). Solving this equation for a general ensemble of random graphs and a general parameter m is a numerically quite tedious problem. In the population dynamics algorithm [26,27] we need to deal with a population of populations of q-components fields.

It is much more convenient to look at the ensemble of random regular graphs where a factorized solution $\mathcal{P}[P(\psi)] = \delta[P(\psi) - P_0(\psi)]$ must exist. Then we are left with only one functional equation, Eq. (20).

Before discussing the zero temperature limit, we would like to point out that there exists another very important case in which Eq. (28) simplifies. For $m=1$, as first remarked and proved in [53], when dealing with the problem of reconstruction on trees, the equations can be written (and numerically solved) in a much simpler way. We again refer to Appendix C for details. Especially for the Poissonian random graphs this simplification is very useful.

B. Zero temperature limit

We now consider the zero temperature limit $\beta \to \infty$ of the 1RSB equations (24)–(28) to study the coloring problem. In most of the previous works [27,28,30] the *energetic zero temperature limit* was employed. The $\beta \to \infty$ limit of Eq. (24) was taken in such a way that $m\beta = y$ remains constant. The replicated free energy (24) then becomes

$$-y\Phi_e(y) = -ye + \Sigma(e).$$ (29)

It is within this approach that the survey propagation (SP) algorithm was derived. The connectivity at which the complexity function $\Sigma(e=0)$ becomes negative is the coloring threshold. Above this connectivity $\Sigma(e)$ was used to compute the minimal number of violated constraints (the ground state energy). The reweighting in Eq. (20) becomes $e^{-y\Delta E^{i \to j}}$, and when $y \to \infty$ all the configurations with positive energy are forbidden.

In this paper we adopt the *entropic zero temperature limit*, suggested originally in [38,39]. The difference in the two approaches was already underlined in Sec. III A. We want to study the structure of proper colorings, i.e., the configurations of zero energy, and we thus fix the energy to zero. Then we obtain the entropy by considering $-\beta f = s$ and introduce a *free entropy*—or in replica term a "replicated entropy"—as $\Phi_s(m) = -\beta m \Phi(\beta, m)|_{\beta \to \infty}$. Equation (24) then becomes

$$\Phi_s(m) = ms + \Sigma(s).$$ (30)

The belief propagation update (4) becomes

$$\psi_{s_i}^{j \to j} = \frac{1}{Z_0^{i \to j}} \prod_{k \in i-j} (1 - \psi_{s_i}^{k \to i}),$$ (31)

while the 1RSB equation (20) keeps the same expression (and thus the same computational complexity).

The partition sum Z_0 in Eq. (2) becomes in this limit the number of proper colorings or solutions. The clusters are now sets of such solutions, and are weighted by their size to the power m. The free entropy $\Phi_s(m)$ is then computed as

$$\Phi_s(m) = \frac{1}{N}\left(\sum_i \Delta\Phi_s^i - \sum_{ij} \Delta\Phi_s^{ij} \right)$$
$$= \frac{1}{N}\left(\sum_i \ln \int_{\text{POP}} (\Delta Z^i)^m - \sum_{ij} \ln \int_{\text{POP}} (\Delta Z^{ij})^m \right),$$ (32)

where ΔZ^i and ΔZ^{ij} are given by Eqs. (5) and (6). The analysis from the previous section is valid also for the entropic zero temperature limit. The information extracted from the number of clusters of a given size $\Sigma(s)$ is one of the main results of this paper and will be discussed and interpreted in Sec. V.

C. Role of frozen variables

In this section we discuss the presence and the role of the frozen variables and explain the connection between the energetic and the entropic zero temperature limits. This allows us to revisit (and extend) the survey propagation equations. Remember that the components of the cavity field $\psi_{s_i}^{i \to j}$ are the probabilities that the node i takes the color s_i when the constraint on the edge ij is not present. In the zero temperature limit we can classify them in two categories:

(i) A *hard field* corresponds to the case when all components of $\psi^{i \to j}$ are zero except one, s. Then only that color is allowed for the spin i, in the absence of edge (ij).

(ii) A *soft field* corresponds to the case when more than one component of $\psi_{s_i}^{i \to j}$ is nonzero. The variable i is thus not frozen in the absence of edge (ij), and the colors of all the nonzero components are allowed.

This distinction is also meaningful for the full probabilities $\psi_{s_i}^i$ [Eq. (4)], if $\psi_{s_i}^i$ is a hard field then the variable i is frozen. In the colorable region there cannot exist a finite fraction of frozen variables (even if we consider properly the permutational symmetry of colors) since by adding a link the connectivity changes by $1/N$ but the probability of becoming uncolorable would be finite. On the contrary, in the 1RSB picture, we observe that a finite fraction of variables can be frozen within a single cluster. In other words, in all the solutions that belong to this given cluster a finite fraction of variables can take one color only. By adding a link into the graph, the connectivity grows by $1/N$, and there is a finite probability that a cluster with frozen variables disappears. The distinction between hard and soft fields is useful not only for the intuition about clusters, but also for the analysis of the cavity equations and it also leads to the survey propagation algorithm.

The distribution of fields over states $P^{i \to j}(\psi^{j \to j})$ [Eq. (20)] can be decomposed into the hard- and soft-field parts

$$P^{i \to j}(\psi^{j \to j}) = \sum_{s=1}^{q} \eta_s^{i \to j} \delta(\psi^{j \to j} - r_s) + \eta_0^{i \to j} \widetilde{P}^{i \to j}(\psi^{j \to j}),$$
(33)

where $\widetilde{P}^{i \to j}$ is the distribution of the soft fields and the normalization is $\sum_{s=0}^{q} \eta_s^{i \to j} = 1$.

Interestingly, the presence of frozen variables in the entropically dominating clusters is connected to the divergence of the size of average minimal rearrangement [55,64]. Precisely, choose a random proper coloring $\{s\}$ and a random node i in the graph. The average minimal rearrangement is the Hamming distance to the nearest solution in which node i has a color different from s_i averaged over the nodes i, the proper colorings, and graphs in the ensemble.

Another interesting role of the frozen variables arises within the *whitening procedure*, introduced in [65] and studied, between others, for the satisfiability problem in [66,67]. This procedure is equivalent to the *warning propagation* (WP) update [Eq. (49)], which we outlined in Sec. III A. Whitening is able to identify if a solution belongs to a cluster with frozen variables or not. Particularly, the result of the whitening is a set of hard cavity fields.

Since the survey propagation algorithm is computing statistics over the states that contain hard fields, then the solution found after decimating the survey propagation result should *a priori* also contain hard fields. However, recent works show that if one applies the whitening procedure starting from solutions found by SP on large graphs, whitening converges every time to the trivial fixed point (see detailed studies for K-SAT in [66,67]). A possible solution to this apparent paradox is discussed in Sec. VI.

1. Hard fields in the simplest case, m=0

Let us now consider the survey propagation equations originally derived in [30] from the energetic zero temperature limit (29) when $y \to \infty$. For simplicity we will write them only for the three-coloring. We consider the 1RSB cavity equation (20) for $m \to 0$, then the reweighting factor $(Z_0^{i \to j})^m$ is equal to zero when the arriving hard fields are contradictory, and equal to one otherwise. The update of probability η_s that a field is frozen in direction s is then written from Eq. (20),

$$\eta_s^{i \to j} = \frac{\prod_{k \in i-j} (1 - \eta_s^{k \to i}) - \sum_{p \neq s} \prod_{k \in i-j} (\eta_0^{k \to i} + \eta_p^{k \to i}) + \prod_{k \in i-j} \eta_0^{k \to i}}{\sum_p \prod_{k \in i-j} (1 - \eta_p^{k \to i}) - \sum_p \prod_{k \in i-j} (\eta_0^{k \to i} + \eta_p^{k \to i}) + \prod_{k \in i-j} \eta_0^{k \to i}}. \tag{34}$$

In the numerator there is a telescopic sum counting the probability that color s and only color s is not forbidden by the incoming fields. In the denominator the telescopic sum is counting the probability that there is at least one color which is not forbidden. If we do not want to actually find a proper coloring on a single graph but just to compute the replicated free energy or entropy, we can further simplify Eq. (34) by imposing the color symmetry. Indeed, the probability that in a given state a field is hard in direction of a color s has to be independent of s (except $s=0$, which corresponds to a soft field). Then Eq. (34) becomes, now for a general number of colors q,

$$\eta^{i \to j} = w(\{\eta^{k \to i}\}) = \frac{\sum_{l=0}^{q-1} (-1)^l \binom{q-1}{l} \prod_{k \in i-j} [1 - (l+1)\eta^{k \to i}]}{\sum_{l=0}^{q-1} (-1)^l \binom{q}{l+1} \prod_{k \in i-j} [1 - (l+1)\eta^{k \to i}]}. \tag{35}$$

Note that since $\partial \Sigma(s)/\partial s = -m$, the value $m=0$ corresponds to the point where the function $\Sigma(s)$ has a zero slope. If a nontrivial solution of Eq. (35) exists, then $\Sigma(s)|_{m=0}$ is the maximum of the curve $\Sigma(s)$ and is counting the logarithm of the number of clusters of size s, which is due to the expo-

nential dependence, also the logarithm of the number of all clusters, regardless of their size. There are two points that we want to emphasize:

(1) Suppose that a nontrivial solution of Eq. (35) exists, i.e., many clusters exist and their number can be computed with the energetic zero temperature limit calculations. Then the clusters might be very small and contain very few solutions in comparison to bigger, less numerous clusters, or in comparison to a giant single cluster which might still exist. This situation cannot be decided by the energetic formalism that weights clusters equally independently of their size.

(2) Suppose, on the contrary, that a nontrivial solution of Eq. (35) does not exist. It might still well be that many clusters exist, but the $\Sigma(s)$ curve has no part with zero slope.

We will see that these two cases are actually observed. The energetic method, which can locate the coloring threshold and from which the survey propagation can be derived, is therefore not a good tool to study the clustering transition.

2. Generalized survey propagation recursion

Let us compute how the fraction of hard fields η evolves after one iteration of Eq. (20) at general m. There are two steps in each iteration of Eq. (20). In the first step, η iterates via Eq. (35). In the second step we reweight the fields. Writing $P_m^{\text{hard}}(Z)$ the—unknown—distribution of the reweightings Z^m for the hard fields, one gets

$$\eta^{i \to j} = \frac{1}{\mathcal{N}} \int dZ P_m^{\text{hard}}(Z) Z^m w(\{\eta^{k \to i}\}) = \frac{w(\{\eta^{k \to i}\})}{\mathcal{N}} \int dZ P_m^{\text{hard}}(Z) Z^m = \frac{w(\{\eta^{k \to i}\})}{\mathcal{N}} \overline{Z_{\text{hard}}^m}. \tag{36}$$

A similar equation can formally be written for the soft fields

$$1 - q\eta^{i \to j} = \frac{1 - qw(\{\eta^{k \to i}\})}{\mathcal{N}} \overline{Z_{\text{soft}}^m}. \tag{37}$$

Writing explicitly the normalization $\mathcal{N}$, we finally obtain the generalized survey propagation equations

$$\eta^{i \to j} = \frac{w(\{\eta^{k \to i}\})}{qw(\{\eta^{k \to i}\}) + [1 - qw(\{\eta^{k \to i}\})]r(\{\eta^{k \to i}\})}, \quad \text{with} \quad r(\{\eta^{k \to i}\}) = \frac{\overline{Z_{\text{soft}}^m}}{\overline{Z_{\text{hard}}^m}}. \tag{38}$$

In order to do this recursion, the only information needed is the ratio r between soft- and hard-field reweightings, which is in general difficult to compute since it depends on the full distribution of soft fields.

There are two cases where Eq. (38) simplifies so that the hard-field recursion become *independent* from the soft-field distribution. The first case is, of course, $m=0$ then $r=1$ independently of the edge (ij), and the equation reduces to the original SP. The second case arises for $m=1$, where one can use the so-called reconstruction formalism and obtain again a closed set of equations. The computation is done in Appendix C, and the SP equations at $m=1$ read

$$\eta_s^{i \to j} = \frac{1}{q} \sum_{l=0}^{q-1} \left[(-1)^l \sum_{s_1, \dots, s_l \neq s} \prod_{k \in i-j} \left(1 - \frac{q}{q-1} \sum_{\alpha=1}^{l} \eta_{s_\alpha}^{k \to i} \right) \right]. \tag{39}$$

It would be interesting in the future to use Eq. (38) in an algorithm to find proper graph colorings, as it has been done with the original SP equation [30]. As an approximation one might also use a value r independent of the edge (ij), but different from one.

For the purpose of the present work, it is important to notice that it is also possible to use Eq. (38) in the population dynamics to simplify the numerical evaluation of the 1RSB solution by separating the hard-field and the soft-field contributions. Indeed, it gives the exact density of hard fields provided the ratio r is calculated, which is doable numerically (see Appendix D). This allows us to monitor precisely the hard-field density and only the soft-field part is given by the population dynamics. This turns out to greatly improve the precision of the numerical solution of the cavity equations and to considerably fasten the code.

3. Presence of frozen variables

A natural question is "When are the hard fields present?" or more precisely, "When does Eq. (38) have a nontrivial solution $\eta > 0$?" First notice that in order to constrain a node into one color, one needs at least $q-1$ incoming fields that forbid all the other colors. It means that function $w(\{\eta^{k \to i}\})$ defined in Eq. (35) is identically zero for $k < q-1$ and might be nonzero only for $k \geq q-1$, where k is the number of incoming fields.

In the limit $r \to 0$ (which corresponds to $m \to -\infty$) Eq. (38) gives $\eta = 1/q$ if $w(\{\eta^{k \to i}\})$ is positive, and $\eta = 0$ if $w(\{\eta^{k \to i}\})$ is zero. Updating Eq. (38) on a given graph, from initial conditions $\eta = 1/q$ everywhere, is equivalent to recursive removing of all the nodes of connectivity smaller than q. This

shows that the first nontrivial solution with hard fields exists if and only if the q-core [68] of the graph is extensive. For regular graphs it is simply at connectivity $c = q$ while for Erdős-Rényi graphs these critical connectivities can be computed exactly and read, for small q, $c_3 = 3.35$, $c_4 = 5.14$, $c_5 = 6.81$ [68]. Indeed we see that the first nontrivial solution to the 1RSB equation appears much before those of the original SP equation at $m = 0$.

On a regular graph, the equations further simplify as η factorizes (is edge independent) and follows a simple self-consistent equation

$$\eta = w(\eta) \frac{1}{qw(\eta) + [1 - qw(\eta)]r}. \tag{40}$$

This equation can be solved for every possible ratio r so that for all $c \geq q$, we can compute and plot the curve $\eta(r)$. We show the results in Fig. 3 for different numbers of colors q. On this plot, we observe that $\eta = 1/q$, as predicted, for $r = 0$. It then gets smaller for a larger value of the ratio and, at a critical value r_{crit}, the solution disappears discontinuously and only the (trivial) solution $\eta = 0$ exists. The values r_{crit} correspond to a critical value of m_r. For all $m > m_r$ no solution containing frozen variables can exist.

D. Validity conditions of the 1RSB solution

Now that we have discussed in detail the 1RSB formalism the next question is "Is this approach correct?" To answer this question, one has to test if the Gibbs measure is extremal *within* the thermodynamically dominating pure states. This is equivalent to checking if the two-step replica symmetry breaking (2RSB) solution is nontrivial. Computing explicitly the 2RSB solution is, however, very complicated numerically, especially for Erdős-Rényi graphs. Instead, the local stability of the 1RSB solution towards 2RSB should be checked, in analogy with the RS stability in Sec. III C 3. It is indeed a usual feature in spin glass physics to observe that the 1RSB glass phase become unstable at low temperatures towards a more complex RSB phase and this phenomenon is called the Gardner transition [69].

To perform the stability analysis [31,34,56,70], one first writes the 2RSB recursion—where the order parameter is a distribution of distributions of fields on every edge $P_1[P_2(\psi)]$—and then two types of 1RSB instabilities have to be considered depending on the way the 2RSB arises from the 1RSB solution. The first type of instability—called *states aggregation*—corresponds to $\delta[P(\psi)] \to P_1[P_2(\psi)]$ while the second type—called *states splitting*—corresponds to

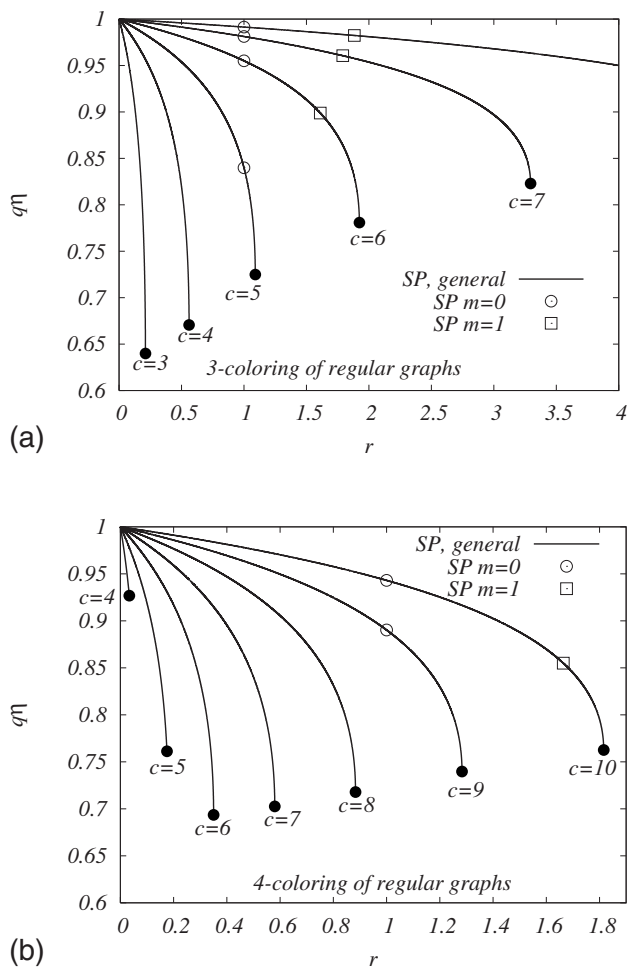FIG. 3. The lines are solutions of Eq. (40) and give the total fraction $q\eta$ of hard fields for a given value of ratio $r = \overline{Z}_{\mathrm{soft}}^m / \overline{Z}_{\mathrm{hard}}^m$ for $q = 3$ (a) and $q = 4$ (b) in regular random graphs. There is a critical value of the ratio (full point) beyond which only the trivial solution $\eta = 0$ exists. Note that the solutions at $m = 0$ and $m = 1$ only exist for a connectivity large enough.

$P[\delta(\psi)] \to P_1[P_2(\psi)]$. A complete stability analysis is left for future works, but it is worth discussing the relevance of the results derived over the last few years [31,34,70].

The 1RSB stability was studied for the coloring problem in [31] but only for the energetic zero temperature limit (29). In this case the parameter $y = \beta m$ is conjugated to the energy. The results derived in [31]—as well as those previously derived for other problems [34,70]—thus concern *only the clusters of sizes corresponding to $m = 0$ at zero (for $y = \infty$) or at positive (for finite y) energy*. The main result of [31,34,70] was that the 1RSB approach was stable in the vicinity of the coloring threshold c_s. As we shall see the clusters corresponding to $m = 0$ are those dominating the total entropy at the coloring threshold, and as a consequence its location is thus *exact* within the cavity approach. The states of the lowest energy (the ground states) in the uncolorable phase also correspond to $m = 0$, and thus the conclusions of [31,34,70] concerning the uncolorable phase are also correct. In particular, a Gardner transition towards further steps of RSB ap-

pears in the uncolorable phase beyond a connectivity denoted c_G in [31].

On the other hand, in the colorable phase the stability of the entropically dominating clusters that correspond to $m > 0$ should be investigated. Some more relevant information can be, however, already drawn from known results. It was indeed found that the 1RSB approach at $m = 0$ is type I stable for all y, and type II stable for all $y > y_I$ in the vicinity of the coloring threshold. These results concern the states of positive energy, but keeping in mind the interpretation of y as a slope in the T, m diagram, we see that the clusters of zero energy corresponding to small but nonzero positive m and zero temperature are also stable with respect to both types of stabilities. Near the colorable threshold, the value of m^*, which describes the dominating clusters, is close to zero and as a consequence *all the dominating clusters are 1RSB stable in the vicinity of the coloring threshold*. Far from the coloring threshold, the stability analysis of [31,34] is irrelevant. In particular, the predictions of a full-RSB colorable phase made in [31,34,70] is *not correct*. Quite the contrary; our preliminary results indicate that *all the dominating clusters are 1RSB stable for $q > 3$*.

The three-coloring is, however, a special case as the clustering transition is continuous. Although the type II instability seems irrelevant in this case as well, we cannot at the moment dismiss a type I instability close to the clustering transition. Indeed the entropically relevant clusters correspond to values of m^* close to one in this case, and it is simple to show that the clusters at $m = 1$ are type I unstable in the case where there is a continuous transition: this is because the type I stability is equivalent to the convergence of the 1RSB update on a single graph. Since for $m = 1$ the averages of the 1RSB fields satisfy the RS belief propagation equations, and since we know from the RS stability analysis in Sec. III C 3 that those equations do not converge in the RS unstable region (i.e., for $c > c_{\mathrm{RS}} = 4$ in three-coloring of Erdős-Rényi graphs), it then follows that the three-coloring is unstable against state aggregation at $m = 1$ for all connectivities $c > 4$. Therefore, it is possible that the 1RSB result for three-coloring is only approximate for what concerns the number and the structure of solutions close to the clustering transition (note that the critical values for the phase transition are, however, correct and do not depend on that). This, and related issues [71], will be hopefully clarified in future works.

To conclude, we believe that all the transition points we discuss in this paper (and those computed in the K-SAT problem in [42]) as well as the overall picture, are exact and would not be modified by considering further steps of replica symmetry breaking.

V. COLORING OF RANDOM GRAPHS: CAVITY RESULTS

We now solve the 1RSB equations, and discuss and interpret the results. We solve Eq. (28) by the population dynamics technique; the technical difficulties and the precision of the method are discussed in Appendix D. Let us stress at this point that the correctness of Eq. (28) is guaranteed only in the limit of large graphs ($N \to \infty$), unfortunately, the cavity

method does not give us any direct hint about the finite graph-size corrections. We start by the results for the regular random graphs, then consider the ensemble of biregular graphs and after that we turn towards Erdős-Rényi graphs. Finally, we discuss the limit of a large number of colors.

A. Regular random graphs

Let us fix the number of colors q, vary the connectivity, and identify successively all the transitions that we shall encounter. For the sake of the discussion, we choose as a typical example the six-coloring and we discuss later in details the cases, for a different number of colors, where some transitions are missing or are arriving in a different order. We solved the 1RSB equation (20) for regular graphs, where the distribution $P^{i \to j}(\psi)$ is the same for every edge (ij) (see Appendix D) and plot the curves for $\Sigma(s)$ we obtained doing so in Fig. 4. We now describe the phase space of solutions when the connectivity is increased:

(1) At very low connectivities $c < q$, only the paramagnetic RS solution is found at all m, $P(\psi) = \delta(\psi - 1/q)$. The phase space is made of a single RS cluster.

(2) For larger connectivities $c \geq q$, we saw in Sec. IV C 3 that the 1RSB equations start to have nontrivial solutions with hard fields in an interval $[-\infty, m_r]$. Interestingly, another nontrivial solution, without hard fields, can be found numerically in an interval $[m_s, \infty]$, and we shall call this one the soft-field solution. As the connectivity increases, we find that m_r increases while m_s decreases, so that the gap $[m_r, m_s]$, where no nontrivial solution exists, is getting continuously smaller. However, there is no nontrivial solution at $m = 1$ for connectivities smaller than c_d (see Fig. 4 for the example of the six-coloring at $c = 17$). This means that the Gibbs measure (2) is still extremal. In other words the large RS state still exists and is entropically dominant [its entropy (14) is noted by a circle in Fig. 4]. Despite the fact that an exponential number of clusters of solutions exist and that the SP equations converge to a nontrivial result, a random proper coloring will almost surely belong to the large RS cluster.

(3) If the connectivity is increased at and above the clustering threshold c_d, a nontrivial solution with positive complexity Σ is found at $m = 1$. In Fig. 4, we see that this happens at $c_d = 18$ for the regular six-coloring. At this point, the RS Gibbs measure (2) ceases to be extremal and the single large RS cluster splits into exponentially numerous components. To cover almost all proper colorings we need to consider exponentially many clusters $\mathcal{N} \sim e^{N\Sigma(m^*=1)}$. The probability that two random proper colorings belong to the same cluster is going exponentially to zero with the system size. The connectivity c_d is thus the true clustering (dynamic) transition. This is not, however, a thermodynamic phase transition because the 1RSB total entropy reduces to the RS entropy (14) at $m = 1$, which is analytical in c. Thus the RS approach gives a correct number of solution and correct marginals as long as the complexity function at $m = 1$ is non-negative.

(4) For even larger connectivities $c \geq c_c$, the complexity at $\Sigma(m = 1)$ becomes negative, e.g., $c_c = 19$ for six-coloring. It means that the clusters corresponding to $m = 1$ are absent with probability one. The total entropy is then smaller than
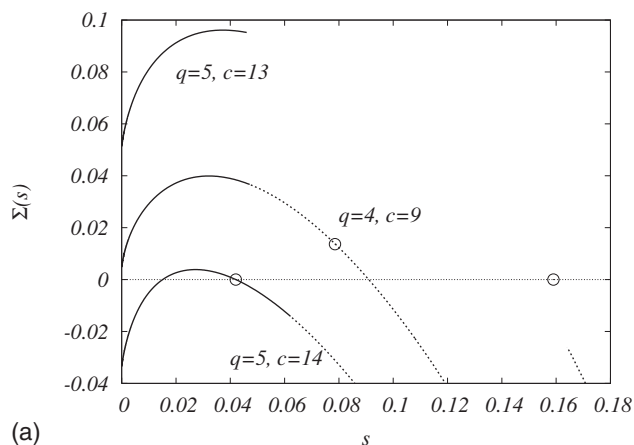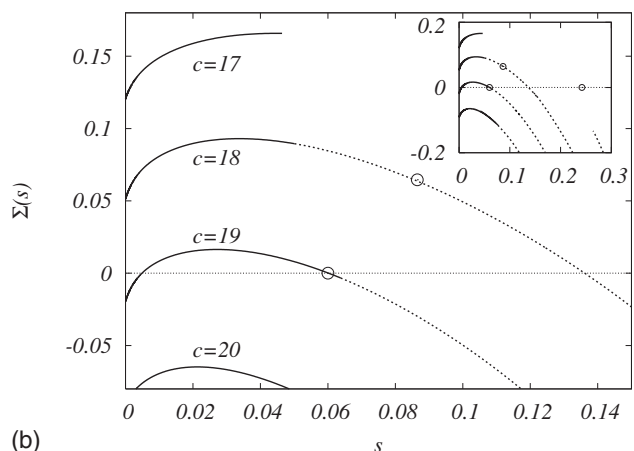
FIG. 4. Complexity as a function of the internal state entropy for the q-coloring problem on random regular graphs of connectivity c. The full line corresponds to the clusters where a finite fraction of hard fields (frozen variables) is present and the dotted line to the clusters without hard fields. The circle signs the entropically dominating clusters. (a) ($q = 4$, $c = 9$) is in the clustered phase; ($q = 5$, $c = 13$) is in a simple replica symmetric phase; and ($q = 5$, $c = 14$) is in the condensed clustered phase. (b) Results for six-coloring for connectivities 17 (RS), 18 (clustered), 19 (condensed), and 20 (uncolorable). For 4-, 5-, and six-coloring all the smaller connectivities are in the RS phase while all the larger one are uncolorable.

the RS or annealed one and is dominated by clusters corresponding to $m^* < 1$ such that $\Sigma(m^*) = 0$. The ordered weights of the entropically dominating clusters follow the Poisson-Dirichlet process (explained in Appendix B). As a consequence, the probability that two random proper colorings belong to the same cluster is finite in the thermodynamic limit. Another way to describe the situation is that the entropy condenses into a finite number of clusters. This condensation is a true thermodynamic transition, since the total entropy is non-analytical at c_c (there is a discontinuity in its second derivative with respect to connectivity). The condensation is analogous to the static (Kauzmann) glass transition observed in mean field models of glasses [60,61].

(5) For connectivities $c \geq c_s$ ($c_s = 20$ for six-coloring) even the maximum of the complexity $\Sigma(m = 0)$ becomes negative. In this case proper colorings are absent with probability go-

ing to one exponentially fast with the size of the graph, and we are in the uncolorable phase.

It is useful to think of the growing connectivity as additions of the constraints into a fixed set of nodes. From this point of view the set of solutions, which exists at connectivity c, gets smaller when new edges are introduced and the connectivity is increased. This translates into the cartoon in the Introduction (Fig. 1) where all the successive transitions are represented. Finally, another important transition has to be considered:

(6) There is a connectivity c_r beyond which the measure is dominated by clusters that contain a finite fraction of *frozen variables*. For the regular six-coloring, $c_r = 19$. We refer to this as the *rigidity transition*, by analogy with [72,73].

The presence or the absence of hard fields inside a given cluster is crucial: if a cluster contains only soft fields, then after the addition of a small but finite fraction of new constraints, its size will get smaller (or it will split). If, however, a cluster contains a finite fraction of frozen variable, then after adding a small but finite fraction of links the cluster will almost surely disappear.

Since the connectivities of regular graphs are integer numbers, we define the dynamical threshold c_d as the smallest connectivity where a nontrivial 1RSB solution exists at $m=1$, the condensation transition c_c as the smallest connectivity where complexity at $m=1$ is negative, c_r the smallest connectivity where hard fields are present at m^*, and the coloring threshold c_s as the first uncolorable case. The scenario described here is observed for all cases of the regular ensemble, although, since connectivities are integers, the transitions are not very well separated at small q. We summarize the results in Table I.

Note that for $q > 3$, the local RS stability discussed in Sec. III C 3 is irrelevant in the colorable regime. The only subtle case being for three-coloring of five-regular graphs where the RS solution is only marginally stable, i.e., the spin glass correlation function goes to zero only algebraically instead of exponentially (from this point of view $c = 5$ would correspond to the critical point well known in the second order phase transitions). More interesting cases will arise in the other ensembles of random graphs.

B. Results for the biregular ensemble

The biregular ensemble allows us to fine-tune the connectivity while preserving the factorization of the 1RSB solution, which is crucial for the numerical precision. It is actually more correct to say that the solution is "bi-factorized," as all the messages going from the nodes with connectivity c_1 to c_2 are the same and the other way around. The biregular ensemble allows us to describe with large precision two interesting cases, which reappear in the Erdős-Rényi ensemble and which are not present in the regular ensemble (again, due to the discrete nature of the connectivity). Let us remind here that bipartite graphs are always two-colorable, but we consider only the color symmetric cavity solutions and that is why we get a nontrivial result from this ensemble.

In Fig. 5(a) is the result for the complexity as a function of entropy $\Sigma(s)$ for four-coloring of 5-21-biregular graphs.

TABLE I. Top: The transition thresholds for regular random graphs: c_{SP} is the smallest connectivity with a nontrivial solution at $m=0$; the clustering threshold c_d is the smallest connectivity with a nontrivial solution at $m=1$; the rigidity threshold c_r is the smallest connectivity at which hard fields are present in the dominant states, the condensation c_c is the smallest connectivity for which the complexity at $m=1$ is negative, and c_s is the smallest UNCOL connectivity. Note that three-coloring of five-regular graphs is exactly critical for that $c_d = 5^+$. The rigidity transition may not exist due to the discreteness of the connectivities. Bottom: Values of m^* (corresponding to the dominating clusters), and in the range of $[-\infty, m_r]$ the hard-field solution exists, and in the range $[m_s, \infty]$ the soft-field solution exists.

q	c_{SP} [31]	c_d [53]	c_r	c_c	c_s [31]
3	5	5^+		6	6
4	9	9		10	10
5	13	14	14	14	15
6	17	18	19	19	20
7	21	23		25	25
8	26	29	30	31	31
9	31	34	36	37	37
10	36	39	42	43	44
20	91	101	105	116	117

q	c	m^*	m_r	m_s
5	3	RS^+	0.12	1.2(1)
4	8	RS	−0.03	3.4(1)
4	9	1	0.41	0.41
5	12	RS	−0.02	3.7(1)
5	13	RS	0.20	2.0(1)
5	14	0.50	0.90	0.90
6	16	RS	−0.02	4.3(1)
6	17	RS	0.05	3.2(1)
6	18	1	0.40	0.40
6	19	0.92	0.96	0.96
7	21	RS	0.01	4.7(1)
7	22	RS	0.17	3.2(1)
7	23	1	0.60	0.60
7	24	1	0.95	0.95

The replica symmetric solution on this case is locally stable. We see clearly the gap between the hard-field and the soft-field solution, and yet we are already beyond the clustering transition c_d; actually the system is in the condensed phase. This example is similar to what happens for the four-coloring of Erdős-Rényi graphs.

The second interesting case, Fig. 5(b), is given by the results for $\Sigma(s)$ for the four-coloring of 4-c-biregular graphs, which are RS unstable for $c > 28$. Both the clustering and the condensation transitions coincide with the RS instability $c_d = c_c = 28$. The survey propagation equations have a nontrivial solution starting from $c_{SP} = 37$. The rigidity transition is at $c_r = 49$. Finally the coloring threshold is $c_s = 57$. Qualitatively, the results for this 4-c-biregular ensemble are the same as
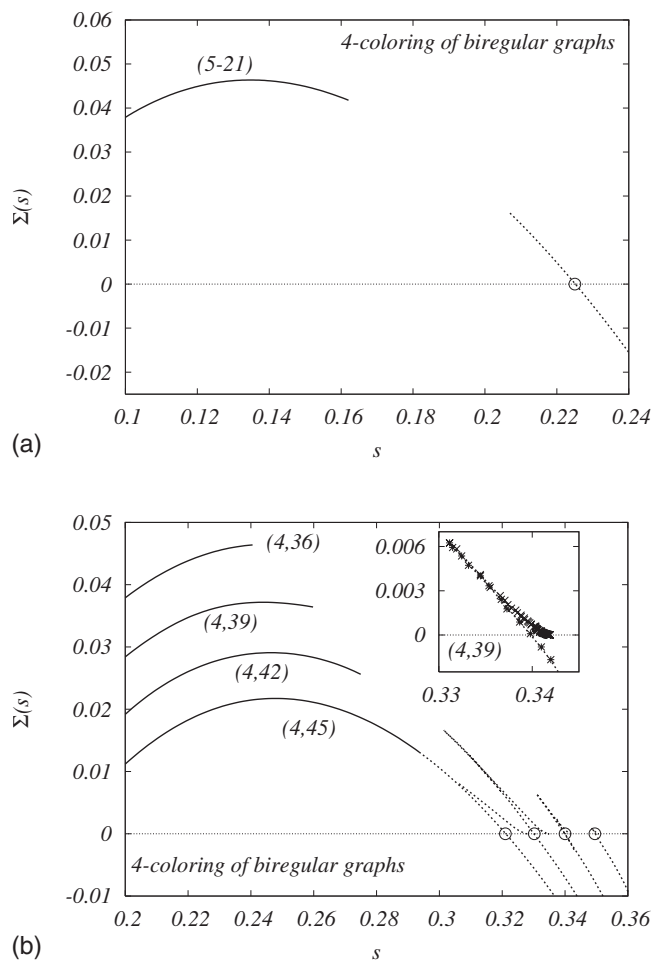
FIG. 5. The complexity as a function of entropy for four-coloring of biregular graphs. (a) 5-21-biregular graph, an example where the entropy is dominated by clusters with soft fields while the gap in the curve $\Sigma(s)$ still exists. (b) 4-c-biregular graphs for c =36,39,42,45. In all these cases the replica symmetric solution is locally unstable. In the dependence $\Sigma(s)$ we see an unphysical branch of the complexity, which is zoomed in the inset for $c=39$.

those for the three-coloring of Erdős-Rényi random graphs.

We see that for $c \leq 42$ the gap between the hard-field (full line) and soft-field (dotted line) solution exists. For $m > m_s$ there is a nonphysical nontrivial soft-field solution, the convex part of the line in the figure, zoomed in the inset. It means that for $m < m_r$ we actually can find two solutions depending on if we start or not with a population containing enough hard fields. The unphysical branch survives even when the gap $[m_r, m_s]$ closes (see the example of $c=45$ in the figure).

We would like to stress at this point the enormous similarity of the soft-field part of the curve $\Sigma(s)$ to the one in Fig. 4 in Ref. [35]. Actually the variational results of [35] should be very precise and relevant near to the continuous clustering transition (this is also the case for the three-coloring of Erdős-Rényi graphs or for 3-SAT).

C. Results for Erdős-Rényi random graphs

For Erdős-Rényi random graphs obtaining the solution of Eq. (28) is computationally more involved as the solution is

no longer factorized. In the population dynamics a population of populations has to be updated, which is numerically possible only for small populations, and so one has to be careful that the finite population-size corrections are small enough (see details in Appendix D). However, all the computations can be done with the same computational complexity as for the regular graphs for $m=0$, the energetic zero temperature limit (Sec. IV C 1), and for $m=1$ (Appendix C). That is enough to obtain the SP, clustering, condensation, and COL/UNCOL transitions (from which the first and last one was computed in [30]). We can also compute exactly when hard fields appear for $m=1$, eq. (C9), this transition is further studied in [64]. Finally, using the generalized survey propagation equation introduced in Sec. IV C 2, the rigidity transition can be computed quite precisely.

1. General case for $q > 3$, discontinuous clustering transition

The phase transitions in q-coloring of random Erdős-Rényi graphs for $q > 3$ are qualitatively identical to those discussed in the case of random regular graphs. We plot the results for the total entropy (number of solutions) and complexity (number of clusters which dominate the entropy) in the four- and five-coloring in Fig. 6.

At the clustering transition c_d the complexity becomes discontinuously positive, and the large RS cluster suddenly splits in an exponential number of smaller ones. The total entropy $\Sigma^* + s^*$ is given by the RS formula (14) up to the condensation transition c_c. At the condensation transition the complexity of the dominating clusters becomes zero, the total entropy $s^{\text{tot}} = s^* < s_{\text{RS}}$ is given by the point where $\Sigma(s^*)$ $=0$. The function $s^{\text{tot}}(c)$ is nonanalytical at the point c_c; it has a discontinuity in the second derivative. At the coloring threshold c_s all the clusters of solutions disappear; note, however, that the total entropy of the last existing clusters is strictly positive (about a half of the total entropy at the condensation transition). That means that the COL/UNCOL transition is not only sharp but also discontinuous in terms of the entropy of solutions. Note that the positive entropy has two contributions: the trivial and smaller one coming from the presence of leaves and other small subgraphs, and the nontrivial and more important one connected with the fact that the ground state entropy is positive even in the uncolorable phase or for the random regular graphs.

Finally, we located the rigidity transition, when frozen variables appear in the dominating clusters. For $3 \leq q \leq 8$ this transition appears in the condensed phase. As the number of colors grows it approaches the clustering transition. All four critical values c_d, c_r, c_c, and c_s are summarized in Table II; values of c_{SP} and $c_r(m=1)$ are given for comparison.

2. Special case of three-coloring, continuous clustering transition

The only case which is left to be discussed is the three-coloring of Erdős-Rényi graphs. It is different from $q > 3$ because the replica symmetric solution is locally unstable in the colorable phase (see Sec. III C 3). The extremality condition underlying the RS assumption ceases to be valid because of the mechanism discussed in Sec. III C 3, with a divergence of the spin glass correlation length: the main dif-
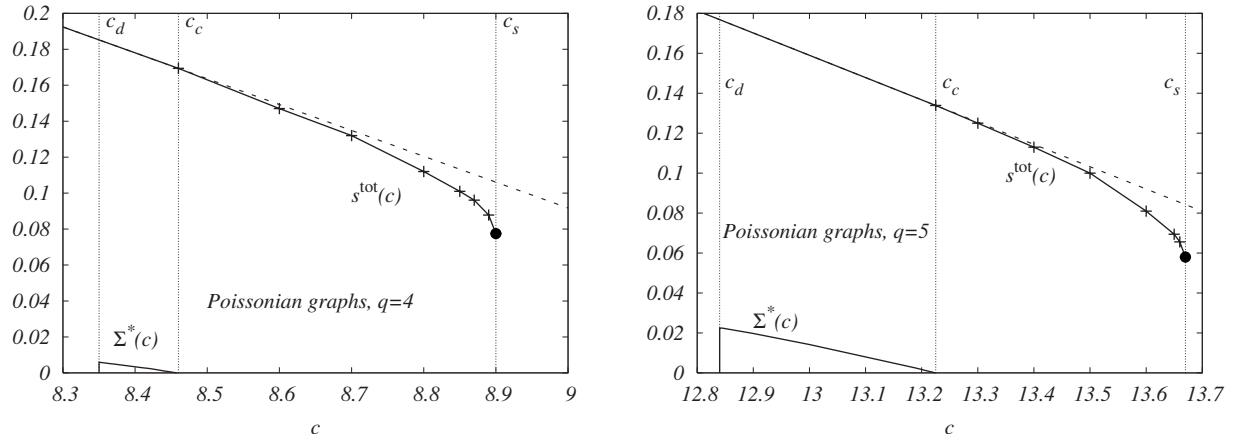
FIG. 6. The 1RSB total entropy and complexity of the dominating clusters for four- and five-coloring of Erdős-Rényi random graphs. The complexity jumps discontinuously at the clustering transition c_d while the total entropy stays analytical. The complexity disappears at the condensation transition c_c causing a nonanalyticity in the total entropy. Finally, the total entropy discontinuously disappears at the coloring threshold. Dashed is the RS entropy left for comparison.

ference with the previous cases is therefore that the clustering transition is continuous and coincides with the condensation transition.

However, the phenomenology does not differ too much from the other cases: $c_{RS}=c_d=c_c=4$; the phase where the entropy is dominated by an exponential number of states is thus missing and the complexity corresponding to $m=1$ is always negative (see Fig. 7 left together with the dependence of the total entropy on the connectivity). Note that the curves $\Sigma(s)$ for the three-coloring have been already studied in [38,39] where the authors considered, however, only the range of connectivities $c=[4.42,4.69]=[c_{SP},c_s]$.

All the results derived for the four-coloring of 4-c-biregular bipartite graphs are quantitatively valid also here. We are thus not surprised by the fact that in interval $c=[4,4.42]$ the survey propagation algorithm gives us a trivial result: simply the maximum of the curve $\Sigma(s)$ does not exist yet there is no nontrivial solution at $m=0$. Yet, the entropy is dominated by a finite number of largest clusters which do not contain hard fields. The two solutions (hard-field and soft-field) join at a connectivity around 4.55. Finally, at $c_r=4.66$

the hard fields arrive to the dominating states (and in consequence to all others).

3. Overlap structure

We now give some results about the overlap structure in the random coloring to elaborate the intuition about clusters. First, consider marginal probabilities $\psi_{s_i}^{j,\alpha}$ within a cluster α. Note that due to the color symmetry there exist another $q!-1$ clusters different only in the permutation of colors. We define the intracluster overlap of two solutions (averaged over states) as

$$\delta = \frac{1}{N}\sum_i \sum_{s_i} \langle (\psi_{s_i}^{j,\alpha})^2 \rangle_\alpha. \tag{41}$$

In the paramagnetic phase $\delta=1/q$, otherwise we have to compute it from the fixed point of Eq. (20). The overlap between two solutions which lie in two clusters, which differ just by permutation π of colors is

TABLE II. Critical connectivities c_d (dynamical, clustering), c_r (rigidity, rearrangements), c_c (condensation, Kauzmann), and c_s (COL/UNCOL) for the phase transitions in the coloring problem on Erdős-Rényi graphs. The connectivities c_{SP} (where the first nontrivial solution of SP appears) and $c_{r(m=1)}$ (where hard fields appear at $m=1$) are also given. The error bars consist of the numerical precision on the evaluation of the critical connectivities by the population dynamics technique (details are given in Appendix D).

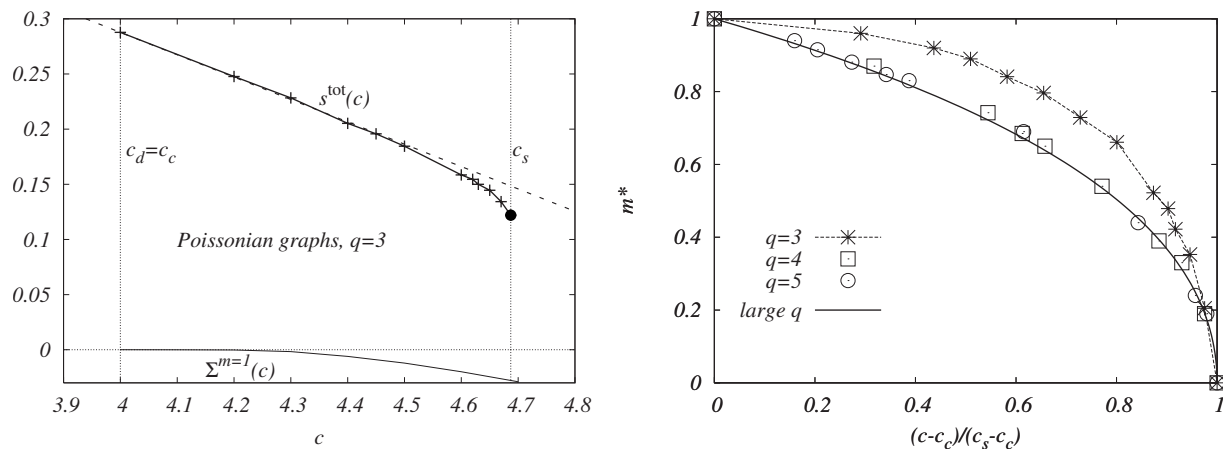| q	c_d	c_r	c_c	c_s	c_{SP}	$c_{r(m=1)}$
3	4	4.66(1)	4	4.687(2)	4.42(1)	4.911
4	8.353(3)	8.83(2)	8.46(1)	8.901(2)	8.09(1)	9.267
5	12.837(3)	13.55(2)	13.23(1)	13.669(2)	12.11(2)	14.036
6	17.645(5)	18.68(2)	18.44(1)	18.880(2)	16.42(2)	19.112
7	22.705(5)	24.16(2)	24.01(1)	24.455(5)	20.97(2)	24.435
8	27.95(5)	29.93(3)	29.90(1)	30.335(5)	25.71(2)	29.960
9	33.45(5)	35.658	36.08(5)	36.490(5)	30.62(2)	35.658
10	39.0(1)	41.508	42.50(5)	42.93(1)	35.69(3)	41.508

FIG. 7. Left: The total entropy for three-coloring of Erdős-Rényi random graphs. The dashed line is the replica symmetric (also the annealed) entropy, left for comparison. The complexity at $m=1$ is shown; it is negative for $c>4$, however, for connectivity near to four it is very near to zero. Right: The values of parameter m^*, $\Sigma(m^*)=0$, as a function of connectivity for $q=3,4,5$ and in the large q limit. The connectivity c is rescaled as $(c-c_c)/(c_s-c_c)$. It is striking that for $q>3$ the curves are so well fitted by the large q limit one. We are even not able to see the difference due to the error bars which are roughly of the point size.

$$\delta_j = \delta\frac{j-1}{q-1} + \frac{q-j}{q(q-1)}, \qquad (42)$$

where j is the number of fixed positions in the permutation π (in particular, $\delta_q = \delta$, and $\delta_1 = 1/q$). In Fig. 8 we show the overlap structure for three- and four-coloring. The probabilities that two random solutions have one of the overlaps can be computed from the Poisson-Dirichlet process described in Appendix B. In fact this is not a self-averaging quantity [29].

D. Large q asymptotics

We give here the exact analytical large q expansion of the previous results. In the asymptotic computations the regular and Erdős-Rényi ensembles are equivalent (the corrections are of smaller order in q than the orders we give). We refer to Appendix E for the explicit derivation of the formulas.

At large q a first set of transitions arises for connectivities scaling as $q \ln q$,

$$c_{SP} = c_r(m=0) = q[\ln q + \ln \ln q + 1 - \ln 2 + o(1)], \qquad (43)$$

$$c_r = c_r(m=1) = q[\ln q + \ln \ln q + 1 + o(1)]. \qquad (44)$$
$${\scriptstyle q\to\infty}$$

c_{SP} was already computed in [31] and c_r is the rigidity transition. The clustering transition has to appear before the rigidity one $c_d < c_r$. For all the finite q cases we looked at, c_d was between c_{SP} and c_r.

A second set of transitions arises for connectivities scaling as $2q \ln q$. We found

$$c_c = 2q \ln q - \ln q - 2\ln 2 + o(1), \qquad (45)$$

$$c_s = 2q \ln q - \ln q - 1 + o(1). \qquad (46)$$

The condensation thus appears very close to the COL/UNCOL transition and both are very far from the clustering
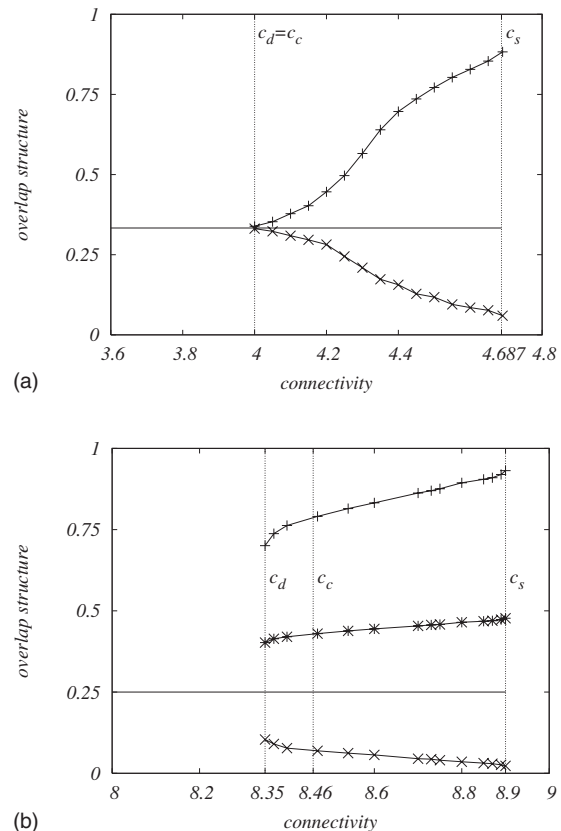

(a)

(b)

FIG. 8. (a) Overlaps structure in three-coloring of random graphs as a function of connectivity. The intracluster overlap (upper curve) grows continuously from 1/3 at the clustering transition $c=4$. In the figure from up there are $\delta=\delta_3$, δ_1, and δ_0. (b) Overlaps structure in four-coloring of random graphs as a function of connectivity. The intracluster overlap (upper curve) jumps discontinuously from 1/4 at the clustering transition $c=8.35$. The probability that two random solutions belong to the same cluster, however, is zero between the clustering and condensation transition [8.35,8.46]. In the figure from up there are $\delta=\delta_4$, δ_2, δ_1, and δ_0.
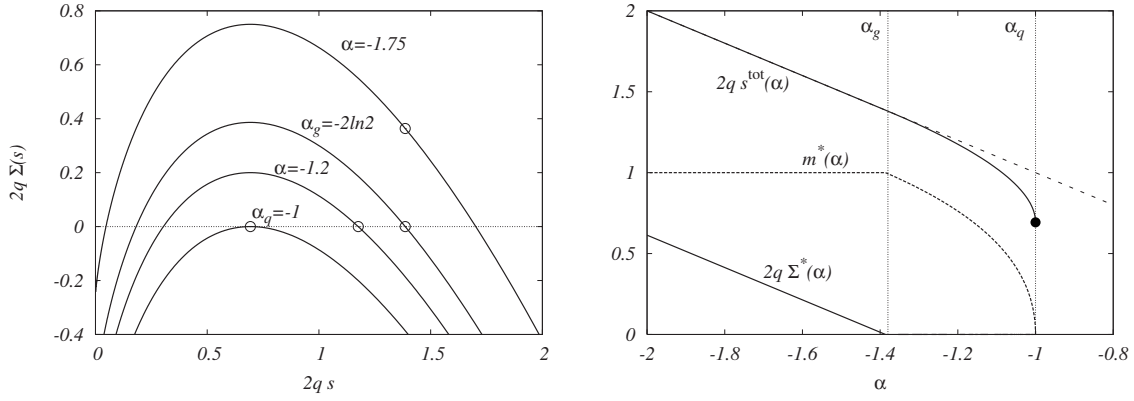
FIG. 9. Analytical result for the large q asymptotics close to the COL/UNCOL transition for $c=2q \ln q - \ln q + \alpha$. Left: (rescaled) complexity versus (rescaled) internal entropy for different connectivities. The condensation transition appears for $\alpha=-2 \ln 2$. The maximum of the complexity becomes zero at c_s for $\alpha=-1$. Right: Total entropy (s^{tot}), complexity (Σ^*), and the parameter m^* versus α. Notice how the values for the total number of solutions are already very close to those for finite low q in Figs. 6 and 7.

and rigidity transitions (those are on a half way in the phase diagram).

We show also in Appendix E that for connectivity $c=2q \ln q - \ln q + \alpha$, one has

$$2qs(m) \simeq 2^m \ln 2, \qquad (47)$$

$$2q\Sigma(m) \simeq 2^m - 2 - m2^m \ln 2 - \alpha. \qquad (48)$$

Since the RS free energy is correct until c_c, which differs just by constant from c_s, that means that for all connectivities below c_c the number of solutions is correctly given by the replica symmetric entropy (14). Indeed, the value $s(m=1)$ can be obtained by a large expansion of Eq. (14).

In Fig. 9 we plot the complexity of dominating clusters $\Sigma^* = \Sigma(m^*)$, the total entropy $s^{\mathrm{tot}} = \Sigma^* + s^*$, and the physical value of m^* as a function of connectivity $c=2q \ln q - \ln q + \alpha$. Note that the properly scaled values of the total number of solutions at c_c and c_s, and the values c_c, c_s themselves, are already very close to those at $q=3,4,5$ [see Figs. 6 and 7 (left)]. The closeness is particularly striking for the values m^* for $q=4$ and $q=5$ [see Fig. 7 (right)].

These formulas show that in the large q limit, near to the coloring threshold, it is the number of clusters which change with connectivity (i.e., α), and not their internal entropy (size). In the leading order, adding constraints near to the COL/UNCOL transition thus destroys clusters of solutions, but does not make them smaller: this is due to the fact that these clusters are dominated by frozen variables so that adding a link kills them most of the time. We also computed the entropy value at the condensation transition, and found $s(m=1)=\ln 2/q$. The entropy of the last cluster (exactly at the COL/UNCOL transition) is $s=\ln 2/2q$.

VI. ALGORITHMIC CONSEQUENCES

In this section we give some algorithmic consequences of our findings. First, we discuss the whitening procedure. We then introduce a random walk algorithm adapted from the Walk-SAT strategy and study its performance. We show, in particular, that the clustering or dynamical transition *does not*

correspond to the onset of hardness in the problem and argue that it is instead the rigidity transition. Finally, we discuss the performance of the belief propagation algorithm in counting and finding solutions, and show that it works much better than previously anticipated.

A. Whitening procedure

The whitening procedure as introduced in [65] can distinguish between solutions which belong to a cluster containing hard fields and those which do not. Generally, whitening is equivalent to the warning propagation (version of belief propagation which distinguish only if a field is hard or not). Warning propagation for coloring was derived in [30]. Let us call $u^{i \to j} = (1,0,0,\ldots,0)$ the hard field in the direction of the first color, i.e., in the absence of node j the node i takes only the first color in all the colorings belonging to the cluster in consideration, and similarly for other colors. Denote $u^{i \to j} = (0,0,0,\ldots,0)$ if $\psi^{j \to j}$ is not frozen in the cluster, we say that the oriented edge $i \to j$ is then "white." The update for u's follows from Eq. (4),

$$u_s^{i \to j} = \min_r \left(\sum_{k \in i - j} u_r^{k \to i} + \delta_{r,s} \right) - \min_r \left(\sum_{k \in i - j} u_r^{k \to i} \right). \quad (49)$$

To see if a solution $\{s_i\}$ belongs to a cluster with frozen variables or not we initialize a warning propagation with $u_s^{i \to j} = \delta_{s,s_i}$, and update iteratively according to Eq. (49) until a fixed point is reached (the updates every time converge, because starting from a solution we are only adding white edges). In the fixed point if all edges are white, then the solution $\{s_i\}$ does not belong to a frozen cluster, or if some of the edges stay colored (nonwhite), then the solution $\{s_i\}$ belongs to a frozen cluster. Note that in the K-SAT problem (but not in general), whitening is equivalent to a more intuitive procedure, where the directed edges are not considered [66,67].

We wish to offer here an explanation of a paradox observed in [66,67]. The SP algorithm gives information on the frozen variables in the most numerous clusters ($m=0$). Yet, the solutions which are found by the standard implementa-

tion (decimation and SP plus Walk-SAT) do not belong to clusters with frozen variables, since they always give a trivial whitening result (all directed edges are white) [66,67]. We suggest that the decimation strategy drives the system towards a solution belonging to a large cluster, which does not contain frozen variables. In this case, it is logical that the result of the whitening is trivial, as it is observed. We believe this is reason why no nontrivial whitenings are observed so far in the study of the K-SAT problem on large graphs.

Note that beyond the rigidity transition this argument does not work anymore, since there all the dominant clusters contain frozen variables. More precisely, for $q \geq 9$ we could, in principle, end up in soft clusters even beyond the rigidity transition (since that one concerns only the dominant states); future works should investigate further if this is possible. Interestingly, in the coloring problem we have not been able to find solutions beyond the rigidity transition even with survey propagation algorithm (compare c_r with the performance of SP in [30]). Further, more systematic investigations have to be done about these issues, employing other strategies for the use of the survey propagation equations (for example, the reinforcement [74]).

B. Walk-COL algorithm to color random graph

In this paper, we have computed the correct clustering transition c_d for the random coloring problem. Beyond this transition, Monte Carlo algorithms are proven not to reach equilibrium as their time of equilibration diverges [55,62]. It was often claimed, or assumed, that this point corresponds to the onset of hardness of the problem. However, the fact that the physical dynamics does not equilibrate just means that the complete set of solutions will not be correctly sampled—indeed Monte-Carlo experiments clearly display slow relaxation [75]—but not that no solutions can be eventually found. This simple fact explains the results of [32] where a simple annealing procedure was shown to three-color an ER graph beyond $c_d = 4$.

In this section, we use a local search strategy, which does not satisfy the detailed balance condition. Therefore, we do not expect to be able to find *typical* solutions, however, it might be possible to find *some* solutions to the problem. The Walk-COL algorithm [87] is a simple adaptation of the celebrated Walk-SAT [76]. More precisely, we adapted the method designed for satisfiability in [77]. Given a graph, and starting from an initial random configuration, we recursively apply the following procedures:

(1) Choose at random a spin which is not satisfied (i.e., at least one of its neighbors has the same color).

(2) Change randomly its color. Accept this change with probability one if the number of unsatisfied spins has been lowered, otherwise accept it with probability p.

(3) If there still are unsatisfied nodes, go to step (1) unless the maximum running time is reached.

The probability p has to be tuned in each different case for a better efficiency of the algorithm. Typically, values between 0.01 and 0.05 give good results. We shall now briefly discuss the performance of the algorithm, to illustrate the two following points: (a) When the phase space is RS, we
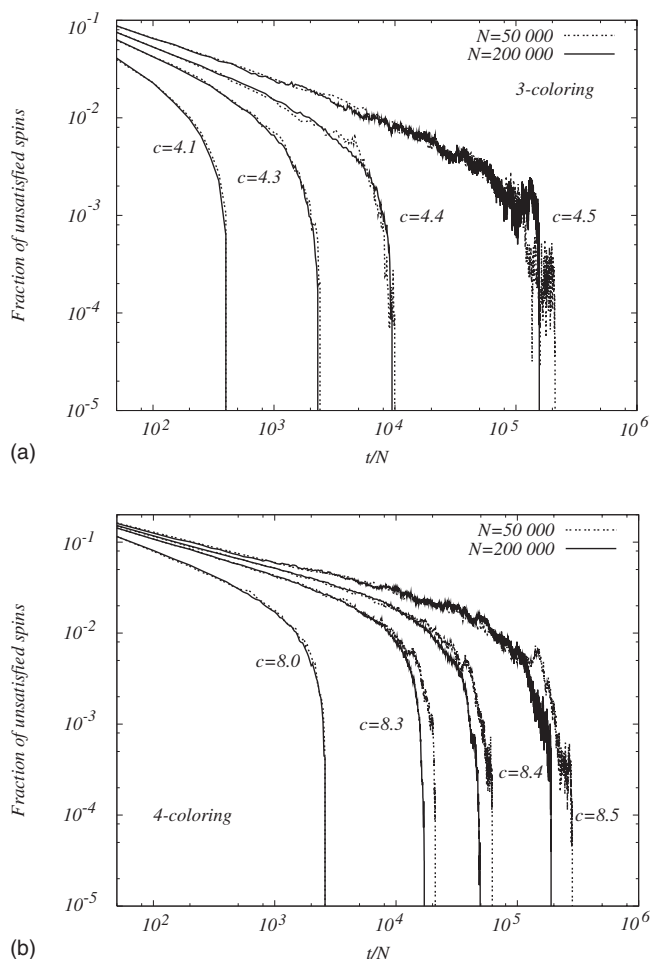

FIG. 10. Performance of the Walk-COL algorithm in coloring random graphs for three-coloring (a) and four-coloring (b). We plot the rescaled time (averaged over five instances) needed to color a graph of connectivity c. The strategy allows one to go beyond the clustering transition ($c_d = 4$ for three-coloring and $c_d = 8.35$ for four-coloring) in linear time with respect to the size of the graph.

observe that Walk-COL finds a solution in linear time. (b) Even in the "complex" phase for $c > c_d$, the algorithm can find in some cases solutions in linear time.

Concerning the first point, we tested the algorithm in the RS phase of regular random graphs for $q = 3, 4, 5, 6, 7$. In all these cases, we were able to color in linear time all the graphs of connectivities that correspond to a replica symmetric solution. In particular, the cases $(q=3, c=5)$, $(q=5, c=13)$, $(q=6, c=17)$, $(q=7, c=21)$, $(q=7, c=22)$ are found to be colorable with the Walk-COL algorithm even if a nontrivial solution to the SP equations exists.

Concerning the second point, we considered the three- and four-coloring of Erdős-Rényi random graphs. The results are shown in Fig. 10 where the percentage of unsatisfied spins versus the number of attempted flips (averaged over five different realizations) divided by N is plotted. We observe that the curves corresponding to different values of N superpose quite well (and that actually the results for $N=2 \times 10^5$ are systematically lower than those for $N=5 \times 10^4$) so that an estimation of the time needed to color a graph can be obtained. The connectivities of these graphs are beyond the

dynamical transition ($c_d=4$ for three-coloring and $c_d=8.35$ for four-coloring). It would be interesting to systematically test Walk-COL, as it has been done for Walk-SAT in [77], to derive the precise connectivity at which it ceases to be linear.

Already these results show that the dynamical transition is not a problem for the algorithms. This can also be observed in a number of numerical experiments for the satisfiability [77,78] and the coloring [73,79] problems.

We believe, however, that the rigidity transition plays a fundamental role for the average computational complexity. A first argument for this is that, for large graphs, it seems that all the known algorithms are only able to find solutions with a trivial whitening, i.e., solutions that belong to clusters without hard fields. Beyond the rigidity transition however, the clusters without hard fields become very rare (in the sense that all the dominating clusters and almost all the smaller, more numerous ones, contain hard fields). For $q \geq 9$ maybe the connectivity where hard fields appear in clusters corresponding to $\Sigma(m)=0$, $m>1$ should be considered. This suggests that the known algorithms will not be able to find a solution beyond this point.

A second argument is the following: local search algorithms are either attracted into a solution or stuck in a metastable state. These metastable states, in order to be able to trap the dynamics, have to contain a finite fraction of hard fields. Given an algorithm, determining which of these two situations happens is not only a question of existence of states, but also a question of basins of attraction and a theoretical analysis of such basins is a very difficult task so that the precise analysis of the behavior of local algorithms remains a hard problem. However, the metastable states are known, from the cavity formalism, to be much more numerous than the zero-energy states. Moreover, the basin of attraction of a zero-energy state that contains hard fields does not probably differ much from those of the metastable state (while, on the other hand, the basin of attraction of a zero-energy state, which does *not* contain hard fields, might be slightly different and arguably relatively larger). It thus seems to us reasonable that local algorithms will get trapped by the metastable states beyond the rigidity transition.

A similar conclusion was reached recently in [73] where the recursive implementation of the Walk-COL algorithm was studied and found to be somehow simpler to analyze. Again the strategy was found to be efficient (with linear time with respect to the size of the graph) beyond c_d *but below* c_r. The precise algorithmic implications of the rigidity transition thus require further investigation, maybe in the lines of [73,78].

C. Belief propagation algorithm to color random graph

Another consequence of our results, that we already discussed shortly in [42], is that the standard belief propagation (BP) algorithm gives correct marginals until the condensation transition. It is actually a simple algebraic fact that the 1RSB approach at $m=1$ gives the same results for the marginals (average probabilities over all solutions) as the simple RS approach (see Appendix C). Moreover, the logarithm of the number of solutions in clusters corresponding to $m=1$ is

also equal to the RS one. This suggests one use the BP marginals (as was already suggested in [40]) and a decimation procedure to find proper colorings. Compared with the SP algorithm, which has computational complexity proportional to $q!$ (factorial), the BP is only q. We have seen, moreover, that for large numbers of colors, the condensation point is very close to the COL/UNCOL transition, so that BP could be used in a large interval of connectivities.

As a simple application, we tested how the straightforward implementation of the BP algorithm plus decimation allows one to find solutions of three- and four-coloring of random Erdős-Rényi graphs. Note that in the three-coloring the clustered but not condensed phase is missing, so the argumentation above does not concern this case. The algorithm works by iterating the following procedures:

(i) Run BP on the graph for a given number l of iterations.

(ii) Consider the most biased variable, and color it with its most probable color.

Two problems have to be mentioned. The first one is rather trivial: since at the beginning all colors are symmetric, the first color had to be put at random. The second one is more serious and concerns the convergence of BP. Indeed, we saw that there is local instability in the BP (replica symmetric) equations at connectivity $c=4$ for the three-coloring of random graphs, so that the BP equations *do not converge*. This seems to be a problem restricted to the three-coloring, but even in the case of four or more colors, the BP equations do not converge on the decimated graph when a finite fraction (typically few percent) of variables is fixed. The reason for that is to be understood.

Nevertheless, since we merely want to design an effective tool to solve the coloring problem, we choose to avoid this problem by fixing the number of iteration l at each step and thus ignore the nonconvergence. We tried the method on both the three- and four-coloring and obtained unexpectedly good results. We used the following protocol in the code: We first try to find a solution with $l=10$. If we do not succeed, we restart with $l=20$ and once more with $l=40$. We tried that on ten different samples for different connectivities. The probability to find a proper coloring with these conditions is shown in Fig. 11.

We thus observe that the BP strategy is able to find solutions, even beyond the condensation transition. This shows clearly that the decimation procedure is a nontrivial one, and that the problem is not really hard in that region of connectivities. Note that the SP algorithm plus decimation has been shown to work in the three-coloring very well until about 4.60 [30]: our results are thus very close to those obtained using SP. This raises again the question of the rigidity transition $c_r=4.66$, which might also be problematic for the decimated survey propagation solver.

VII. CONCLUSIONS

Let us summarize the results. They are perhaps best illustrated looking back to the cartoon in Fig. 1, where the importance of the size of clusters is evidenced. We find that the set of solutions of the q-coloring problem undergoes the following transitions as the connectivity is increased:
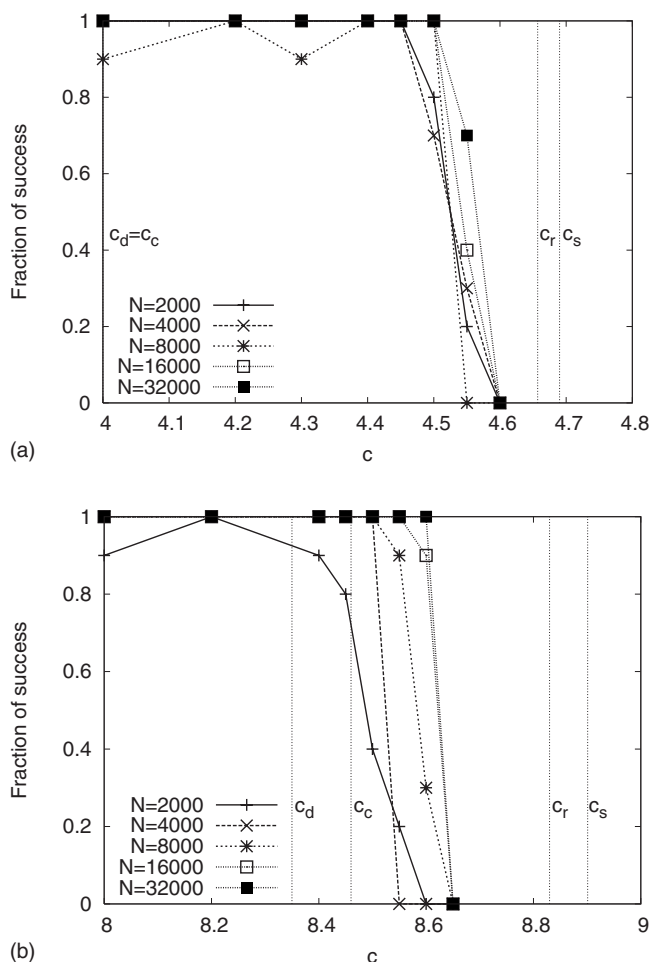
FIG. 11. Performance of the BP algorithm plus decimation in coloring random graphs for three-coloring (a) and four-coloring (b). The strategy described in the text allows one to color random graphs beyond the clustering and even the condensation transitions.

(i) At low connectivity, $c < c_d$, many clusters might exist but they are very small and the measure over the set of solutions is dominated by the single giant cluster described by the replica symmetric approach.

(ii) Only at the *dynamical* transition c_d the giant cluster decomposes abruptly into an exponentially large number of clusters (pure states). For connectivities $c_d < c < c_c$, the measure is dominated by an exponential number of clusters. Yet, the total number of solutions is given by the replica symmetric entropy (14), and the marginals are given by the fixed point of the replica symmetric equations (belief propagation) (4). Starting from this transition the uniform sampling of solution becomes hard.

(iii) At connectivity c_c the set of solutions undergoes a *condensation* transition, similar to the one appearing in mean field spin glasses. In the condensed phase the measure is dominated by a finite number of the largest clusters. The total entropy is strictly smaller than the replica symmetric one and has a discontinuity in the second derivative at c_c.

(iv) When connectivity c_s is reached, no more clusters exist: this is the COL/UNCOL transition. Note that the entropy of the last existing clusters is strictly positive, and not given only by the contribution of the isolated nodes, leaves, and other small subgraphs, the COL/UNCOL transition is thus discontinuous in entropy.

This picture is very similar to the well-known scenario of the glass transition in temperature, with the dynamical and glass (Kauzmann) transition [61]. In some cases, the main one being the three-coloring of Erdős-Rényi graphs, the clustering and the condensation transition merge and a continuous transition take place at $c_d = c_c$, which then coincide with the local instability of the replica symmetric solution. Interestingly, the variational approach of [35] is very precise near to the continuous clustering transition. Since the 3-SAT problem behaves in the same manner, this solves the apparent contradictions between the results of [35] and [24].

In addition to the transitions describing the organization of clusters, another important phenomenon concerning the internal structure of clusters takes place. A finite fraction of frozen variables can appear in the clusters (a frozen variable takes the same color in all the solutions that belong to the cluster). We found that the fraction of such variables in each cluster undergoes a first order transition and jumps from zero to a finite fraction at a connectivity that depends on the size of the cluster. In particular:

(v) There exists a critical connectivity c_r (rigidity or freezing) at which the thermodynamically relevant clusters—those that dominate the Gibbs measure—start to contain a finite fraction of frozen variables.

The results above were obtained within the 1RSB scheme, but should not change when considering further steps of RSB (an exception might be the three-coloring near to the clustering transition).

We discussed some algorithmic consequences of these transitions. First, the belief propagation algorithm is efficient in counting solution and estimating marginals until the condensation transition. More interestingly, it can also be used, just like survey propagation, together with a decimation procedure in order to find solutions as we numerically demonstrated. Second, the dynamical transition is not the one at which simple algorithms fail as we illustrated using the Walk-COL strategy. For the three-coloring of an ER graph, there is even a rigorous proof of algorithmic performance beyond $c_d = 4$ and until $c = 4.03$ [10]. We argued that, instead, the rigidity phenomenon is responsible for the onset of computational hardness. This is a major point that we hope to see more investigated in the future.

Our study opens a way to many new and promising investigations and developments. For instance, we wrote the equivalent of the survey propagation equations for a general value of m, which has a particularly simple form for $m = 1$ [Eq. (39)].

It would be interesting to use these equations to find solutions. The behavior at finite temperature and the performance of the annealing procedure are also of interest. It would furthermore be interesting to rediscuss other finite connectivity spin glass models such as for instance the lattice glass models [56] in the light of our findings. The stability towards more steps of replica symmetry breaking, or the supersymmetric approach [71], should be further investigated. Finally, it would be interesting to combine the entropic and energetic approach to investigate the frozen variables in the

metastable states. We hope that our results will stimulate activity in these lines of thought.

APPENDIX A: STABILITY OF THE PARAMAGNETIC SOLUTION

In this appendix, we show how to compute the stability of the paramagnetic solution towards the continuous appearance of a 1RSB solution. This happens, as usual for continuous transition, when the spin glass correlation length, or equivalently, the spin glass susceptibility, diverges. Obviously, the presence of the diverging correlation length invalidates the premise of the RS cavity method. Recall that the spin glass susceptibility is defined as

$$\chi_{SG} = \frac{1}{N} \sum_{i,j} \langle \vec{s}_i \cdot \vec{s}_j \rangle_c^2, \tag{A1}$$

and can be rewritten for the present purpose as

$$\chi_{SG} = \sum_{d=0}^{\infty} \gamma^d \mathbb{E}(\langle \vec{s}_0 \cdot \vec{s}_d \rangle_c^2), \tag{A2}$$

where we consider the average over graphs, in the thermodynamic limit, where spins s_0 and s_d are at distance d. The factor γ^d stands for the average number of neighbors at distance d, when $d \leqslant \ln N$. Assuming that the limit for large d of the summands in Eq. (A2) exists (with the limit $N \to \infty$ performed first), we relate it to the *stability parameter* as follows:

$$\bar{\lambda} = \lim_{d \to \infty} \gamma [\mathbb{E}(\langle \vec{s}_0 \cdot \vec{s}_d \rangle_c^2)]^{1/d}. \tag{A3}$$

Then the series in Eq. (A2) is essentially geometric, and converges if and only if $\bar{\lambda} < 1$.

Using the fluctuation-dissipation theorem we relate the correlation $\langle s_0 s_d \rangle_c$ to the variation of magnetization in s_0, caused by an infinitesimal field in s_d. Finally, using the fact that we perform the large-N limit first, the variation above is dominated by the direct influence through the length-d path between the two nodes, and this induces a "chain" relation: if the path involves the nodes $(d, d-1, \ldots, 0)$ we have

$$\mathbb{E}(\langle \vec{s}_0 \cdot \vec{s}_d \rangle_c^2) = \mathcal{C} \sum_{\substack{b \in \partial d \\ a \in \partial 0}} \mathrm{Tr}\mathbb{E}\left[\left(\frac{\partial \psi^{a \to 0}}{\partial \psi^{b \to d}}\right)^2\right]$$

$$= \mathcal{C}\mathrm{Tr}\mathbb{E}\left[\prod_{\ell=1}^{d} \left(\frac{\partial \psi^{l \to l-1}}{\partial \psi^{l+1 \to l}}\right)^2\right]. \tag{A4}$$

The stability parameter of the paramagnetic solution of the cavity equations towards small perturbations can be computed from the following Jacobian:

$$T^{\tau\sigma} = \left. \frac{\partial \psi_\tau^{1 \to 0}}{\partial \psi_\sigma^{2 \to 1}} \right|_{RS}, \tag{A5}$$

which gives the infinitesimal probability that a change in the input probability $\psi_\sigma^{2 \to 1}$ will change the output probability $\psi_\tau^{1 \to 0}$. The index RS says that the expression has to be evaluated at the RS paramagnetic solution.

This matrix has only two different entries, all the diagonal elements are equal, and all the nondiagonal elements are also equal. As an immediate consequence all Jacobians commute and are thus simultaneously diagonalizable so that it will be sufficient to study the effect after one cavity iteration (one step in the chain). The matrix T has only two distinct eigenvalues,

$$\lambda_1 = \left. \left(\frac{\partial \psi_1^{1 \to 0}}{\partial \psi_1^{2 \to 1}} - \frac{\partial \psi_1^{1 \to 0}}{\partial \psi_2^{2 \to 1}} \right) \right|_{RS},$$

$$\lambda_2 = \left. \left(\frac{\partial \psi_1^{1 \to 0}}{\partial \psi_1^{2 \to 1}} + (q-1) \frac{\partial \psi_1^{1 \to 0}}{\partial \psi_2^{2 \to 1}} \right) \right|_{RS}. \tag{A6}$$

The second eigenvalue corresponds to the homogeneous eigenvector $(1, 1, \ldots, 1)$ and describes a fluctuation changing all $\psi_\tau^{2 \to 1}, \tau = 1, \ldots, q$, by the same amount, and maintains the color symmetry. It is thus not likely to be the relevant one and we will see that indeed $\lambda_2 = 0$. The first eigenvalue, however, is $(q-1)$-fold degenerate and its eigenvectors are spanned by $(1, -1, 0, \ldots, 0)$, $(0, 1, -1, 0, \ldots, 0), \ldots$, $(0, \ldots, 0, 1, -1)$. The corresponding fluctuations explicitly break the color symmetry, and are in fact the critical ones. Using the cavity recursion (4), the two derivatives simply read

$$\frac{\partial \psi_1^{1 \to 0}}{\partial \psi_2^{2 \to 1}} = (1 - e^{-\beta}) \frac{(\psi_1^{1 \to 0})^2}{1 - (1 - e^{-\beta}) \psi_1^2},$$

$$\frac{\partial \psi_1^{1 \to 0}}{\partial \psi_1^{2 \to 1}} = (1 - e^{-\beta}) \left(\frac{(\psi_1^{1 \to 0})^2}{1 - (1 - e^{-\beta}) \psi_1^{2 \to 1}} - \frac{\psi_1^{1 \to 0}}{1 - (1 - e^{-\beta}) \psi_1^{2 \to 1}} \right), \tag{A7}$$

so that the values of the two eigenvalues evaluated at the RS solution, where all ψ are equal to $1/q$, are

$$\lambda_1 = \frac{1}{1 - \dfrac{q}{1 - e^{-\beta}}}, \quad \lambda_2 = 0. \tag{A8}$$

The stability parameter (A3) is thus $\bar{\lambda} = \gamma \lambda_1^2$ and the critical temperatures below which the instability sets in are

$$T_c^{\mathrm{reg}}(q,c) = -1/\ln\left(1 - \frac{q}{\sqrt{c-1}+1}\right),$$

$$T_c^{ER}(q,c) = -1/\ln\left(1 - \frac{q}{\sqrt{c}+1}\right) \qquad (A9)$$

for regular and Erdős-Rényi graphs, respectively. Thus at zero temperature the critical connectivities read

$$c_{RS\ stab.}^{reg} = q^2 - 2q + 2, \quad c_{RS\ stab.}^{ER} = q^2 - 2q + 1. \quad (A10)$$

These results coincide perfectly with the numerical simulations of the cavity recursion of [32]. The analytical expressions equivalent to Eq. (A9) were in fact first obtained in [57] in the context of the reconstruction problem on trees as an upper bound for the Gibbs measure extremality, and its connection with the statistical physics approach was explained in [53]. The case of biregular random graphs can be easily understood by noticing that two recursions should be considered, one with $\gamma = c_1 - 1$, and one with $\gamma = c_2 - 1$. As a consequence, the stability point is equivalent in this case to the one of a regular random graph with an effective connectivity equal to $c = 1 + \sqrt{(c_1-1)(c_2-1)}$.

Another instability appears when $\gamma|\lambda_1| > 1$. This has been referred to as the *modulation* instability in [56]. Actually, this is the continuous instability towards the appearance of the antiferromagnetic order. Since at zero temperature $\lambda_1 = (1-q)^{-1}$, then for connectivities larger than $c_{mod} = q$ for random regular graphs (and $c_{mod} = q-1$ for Erdős-Rényi) the paramagnetic solution becomes unstable towards the antiferromagnetic order. However, this is correct if we study a tree with some given (and well chosen) boundary condition, but as noted in [56], the antiferromagnetic solution is impossible on random graphs because of the existence of frustrating loops of arbitrary length. The cavity equations (4) can actually never converge towards an antiferromagnetic solution of a random graph. Instead, when iterating, the fields oscillate between different solutions (thus the name *modulation*).

In other words, although on a random tree with special boundary conditions there exists for $c > c_{mod}$ a nontrivial solution to the cavity recursion (for the Gibbs state is no longer unique [Eq. (15)]), this solution does not exist on a random graph (and the Gibbs state is still extremal [Eq. (16)]). Note that this instability could anyway be a source of numerical problems that can be overcome considering that the distribution of cavity fields $\mathcal{P}(\psi)$ over the ensemble of random graphs has to be symmetric in the color permutation. Another possibility is to randomly mix the new and old populations in the population dynamics so that the antiferromagnetic oscillations are destroyed.

APPENDIX B: THE RELATIVE SIZES OF CLUSTERS IN THE CONDENSED PHASE

In this section, we introduce the Poisson-Dirichlet point process and we shortly review some of its important properties. We also sketch its deep connection with the size of clusters in the condensed phase. Poisson-Dirichlet (PD) point process is a set of points $\{x_i\}$, $i = 1, \ldots, \infty$ such that $x_1 > x_2 > x_3 > \cdots$ and $\Sigma_{i=1}^{\infty} x_i = 1$. To construct these points we consider a Poisson process $\{y_i\}$, $i = 1, \ldots, \infty$ of intensity measure y^{-1-m^*}, $0 < m^* < 1$ (note that this measure is not a probability
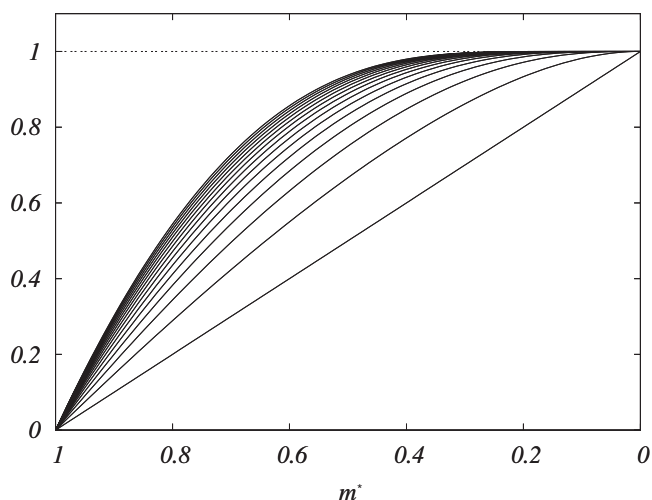

FIG. 12. The sketch of size of the largest clusters for a given value of parameter m^*. The lower curve is related to the average size of the largest clusters as $1/\mathbb{E}[1/x_1] = 1 - m^*$. The following curves are related to the size of i largest clusters, their distances are $\mathbb{E}[R_i]\mathbb{E}[R_{i-1}]\cdots\mathbb{E}[R_1](1-m^*)$.

measure). We order the sequence $\{y_i\}$ in such a way that $y_1 > y_2 > y_3 > \cdots$ and define the PD point process as

$$x_j = \frac{y_j}{\sum_{i=1}^{\infty} y_i}. \qquad (B1)$$

If we identify the parameter m^* with the value for which the complexity is zero, $\Sigma(m^*) = 0$, then y_i is proportional to the number of solutions in cluster i (or to $e^{-\beta F}$ for nonzero temperature), and x_i is the size on that cluster relative to the total number of solutions. This connection was (on a nonrigorous level) understood in [80] (for more mathematical review see [81]). Note that due to the permutation symmetry in graph coloring there are every time $q!$ copies of one clusters (different in the color permutation).

To get a feeling about the PD statistics let us answer in Fig. 12 the following question: Given the value m^* how many clusters do we need to cover fraction r of solutions, in other words, what is the smallest k such that $\Sigma_{i=1}^k x_i > r$? The mathematical properties of the PD process are very clearly reviewed in [82]. To avoid confusion, note at this point that the PD process we are interested in is the PD$(m^*, 0)$ in the notation of [82]. In the mathematical literature, it is often referred to as the PD$(0, \theta)$ without indexing by the two parameters. Let us be reminded of two useful results. Any moment of any x_i can be computed from the generating function

$$\mathbb{E}[\exp(-\lambda/x_i)] = e^{-\lambda} \phi_{m^*}(\lambda)^{i-1} \psi_{m^*}(\lambda)^{-i}, \qquad (B2)$$

where $\lambda \geq 0$ and the functions ϕ_{m^*} and ψ_{m^*} are defined as

$$\phi_{m^*}(\lambda) = m^* \int_1^{\infty} e^{-\lambda x} x^{-1-m^*} dx, \qquad (B3)$$

$$\psi_{m^*}(\lambda) = 1 + m^* \int_0^1 (1 - e^{-\lambda x}) x^{-1-m^*} dx. \quad \text{(B4)}$$

Another relation is that the ratio of two consequent points $R_i = x_{i+1}/x_i$, $i = 1, 2, \dots$ is distributed as $im^* R_i^{im^*-1}$. In particular, its expectation is $\mathbb{E}[R_i] = im^*/(1+im^*)$ and the random variables R_i are mutually independent. We used these relations to obtain data in Fig. 12.

APPENDIX C: THE 1RSB FORMALISM AT $m=1$ AND THE RECONSTRUCTION EQUATIONS

In this appendix we discuss the considerable simplification of Eq. (28) that is obtained by working directly at $m = 1$. This was first remarked and proved in [53] when dealing with the tree reconstruction problem (for a discussion of a case where the RS solution is not paramagnetic, see [43]). We first introduce the probability distribution of fields (20) averaged over the graph

$$\bar{P}(\psi) \equiv \int dP \mathcal{P}[P(\psi)] P(\psi)$$

$$= \sum_k \mathcal{Q}_1(k) \frac{1}{Z_1} \int \prod_{i=1}^k d\psi^i \bar{P}^i(\psi^i) \delta[\psi - \mathcal{F}(\{\psi^i\})] Z_0,$$

$$\text{(C1)}$$

where Z_1 is computed from Eq. (20) as

$$Z_1 \equiv \int \prod_{i=1}^k d\psi^i P^i(\psi^i) Z_0 = \sum_{j=1}^q \prod_{i=1}^k (1 - \bar{\psi}^i), \quad \text{(C2)}$$

where $\bar{\psi} = \int d\psi P(\psi) \psi$. Generally, $\bar{\psi}$ is a solution of the RS equation (12), which is easily seen from Eqs. (20) and (28). Since the RS solution is the paramagnetic one $\bar{\psi} = 1/q$ the form of Z_1 is particularly simple.

In the next step, we want to get rid of the term Z_0 in Eq. (C1). We thus introduce q distribution functions $\bar{P}_s$,

$$\bar{P}_s(\psi) = q \psi_s \bar{P}(\psi). \quad \text{(C3)}$$

It is then easy to show that if $\bar{\psi} = 1/q$ then $\bar{P}_s(\psi)$ satisfies

$$\bar{P}_s(\psi) = \sum_k \mathcal{Q}_1(k) \sum_{s_1 \dots s_k} \prod_{i=1}^k \pi(s_i|s) \int \delta[\psi$$

$$- \mathcal{F}(\{\psi^j\})] \prod_{i=1}^k d\psi^j \bar{P}_{s_i}(\psi^j), \quad \text{(C4)}$$

where

$$\pi(s_i|s) = \frac{1 - (1 - e^{-\beta}) \delta(s_i, s)}{q - (1 - e^{-\beta})}. \quad \text{(C5)}$$

We solve Eq. (C4) by population dynamics. In order to do this, one needs to deal with q populations of q-component fields, and to update them according to Eq. (C4). It is only a functional equation and not a double functional as the general 1RSB equation (28). Moreover, the absence of the re-weighting term Z_0^m simplifies the population dynamics algorithm significantly. Finally, it is important to note that the computational complexity here is the same as the one for regular and ER random graphs.

A crucial theorem is also proven in [53]: the population dynamics of Eq. (C4) has a nontrivial solution if and only if it converges to a nontrivial solution starting from initial conditions

$$\bar{P}_r^{\text{init}}(\psi_s) = \delta(r, s). \quad \text{(C6)}$$

This shows that when a paramagnetic solution is found, then no other solutions exist.

Similar manipulations allow us to obtain the replicated free energy (21), which is equal in this case to the replica symmetric free energy (9), and the free energy (26) inside the corresponding states as

$$-\beta f(\beta) = \sum_k \mathcal{Q}(k) \sum_{s_1, \dots, s_k} \sum_s \frac{1}{q} \prod_{i=1}^k \pi(s_i|s) \int (\ln Z_0^i) \prod_{i=1}^k d\psi^i \bar{P}_{s_i}(\psi^i) - \frac{c}{2} \sum_{s_1, s_2} \frac{\pi(s_1|s_2)}{q} \int (\ln Z_0^{12}) d\psi^1 d\psi^2 \bar{P}_{s_1}(\psi^1) \bar{P}_{s_2}(\psi^2), \quad \text{(C7)}$$

where the normalization factors Z_0^i, Z_0^{ij} are defined by Eqs. (6) and (8). The complexity follows from Eq. (24). Since the replicated free energy $\Phi(\beta, 1)$ is equal, according to Eq. (27), to the total free energy, we showed the statement used several times in the paper, i.e., the total free energy (entropy) at $m = 1$ is equal to the replica symmetric free energy (9).

Another important point is that one can write the recursion separating the hard and soft fields. In general, at zero temperature, we can write the distribution $\bar{P}_q(\psi_s)$ in Eq. (C4) as

$$\bar{P}_r(\psi) = \sum_s \mu_{r,s} \delta(\psi_s - 1) + \left(1 - \sum_s \mu_{r,s}\right) \tilde{P}_r(\psi). \quad \text{(C8)}$$

Plugging this to Eq. (C4) and taking into account the initial condition (C6) and color symmetry, we see that $\mu_{q,s} = q \eta \delta(q, s)$, where η satisfies

$$q \eta = \sum_k \mathcal{Q}_1(k) \sum_{m=0}^{q-1} (-1)^m \binom{q-1}{m} \left(1 - \frac{mq}{q-1} \eta\right)^k. \quad \text{(C9)}$$

On ER graphs, the sum can be performed analytically and one finds

$$q\eta = \left(1 - e^{-[q\eta c/(q-1)]}\right)^{q-1}. \quad \text{(C10)}$$

This equation can be solved iteratively starting from $\eta^{\text{init}} = 1/q$. It is a very simple equation, as the one obtained for $m=0$, which gives us a very efficient way to compute the fraction of hard fields at $m=1$ for both regular and ER graphs. Indeed η is larger than zero only above a certain average connectivity $c_r(m=1)$.

APPENDIX D: NUMERICAL METHODS

In this section, we detail the numerical methods we used to solve the 1RSB equations (20) and (28), and the procedures used to generate the data. We use a population dynamics method, as introduced in [26,27], and model the distribution $P^{i \to j}(\psi^{i \to j})$ by a population of N vectors $\psi^{j \to j}$. To compute $P^{i \to j}(\psi^{i \to j})$ knowing the $P^{k \to i}(\psi^{k \to i})$ for all incoming k we perform the 1RSB recursion in Eq. (20) in two steps: (i) first we compute the new vectors $\psi^{j \to j}$ using the simple RS recursion in Eq. (31) (this is the iterative step) and (ii) we take into account the weight $(Z_0^{i \to j})^m$ for each of the vectors (this is the reweighting step). For the reweighting we tried different strategies; two of them perform very well.

(a) For every field ψ in the population, we keep its weight Z_0. We then compute the cumulative distribution of weights Z_0 and sample uniformly the incoming fields. Using dichotomy we generate a random field with its proper weight in $O[\ln(N)]$ steps. A complete iteration thus takes $O[N \ln N]$ steps.

(b) We compute N new vectors and then we make a new population when we clone some of them while erasing others so that in this new population each field is present according to its weight (in principle, one can even change the size of the population, although we have not implemented this strategy). This second approach can be implemented in linear time (generating an ordered list of random numbers is a linear problem; see [83]), but is a bit less precise as we introduce redundancy in the population.

We finally choose to use the second strategy, as we observed that it performs almost as good as the first one (for a given size of population) while it was much faster, so that, for given computer time, it allows a better representation of the population. We also force the population to be color symmetric by adding a random shift of colors in the incoming messages. This is needed in order to avoid the antiferromagnetic solution. The learned reader will notice that this is equivalent to solving a *disordered* Potts glass instead of an antiferromagnet model. Indeed the fact that an Ising antiferromagnet on a random graph is equivalent to an Ising spin glass was already noticed [84].

Another important issue is the presence of hard fields. In Fig. 13 are histograms of the first component of the vectors in the population for three- and four-coloring of five- and nine-regular random graphs, respectively. It is interesting to see how they peak around fractional values due to the presence of hard fields (see the three upper ones). Maybe even more interesting are the lower ones where *no hard fields are*
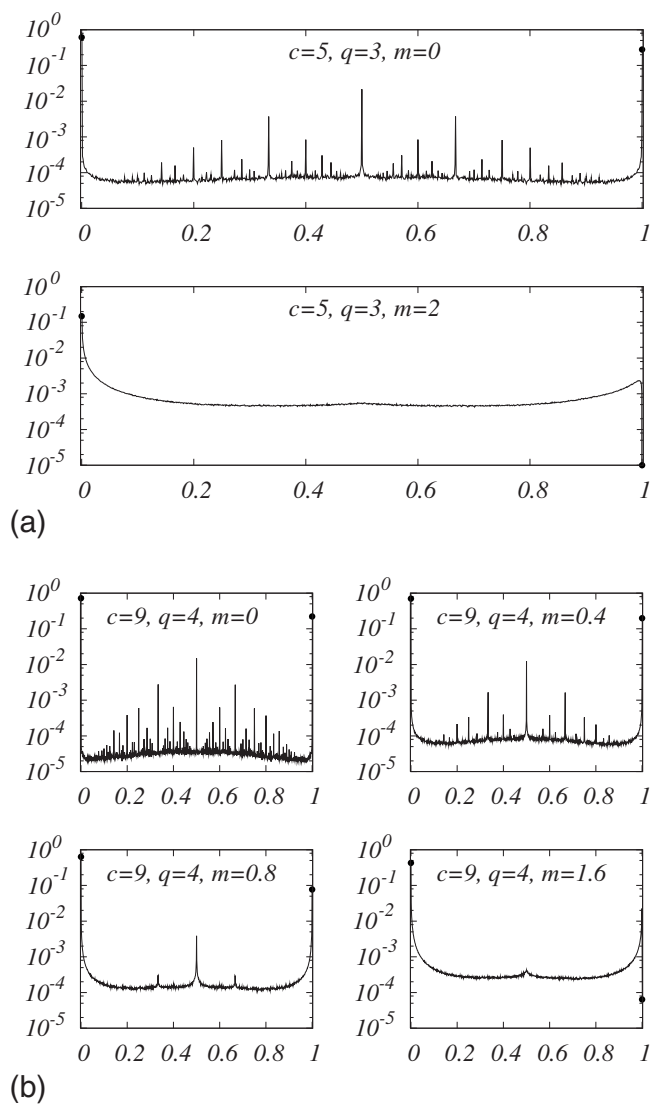

FIG. 13. Histograms of the first component of the cavity field ψ_1, i.e., the probability that a node takes color one. Notice the logarithmic y scale. Frozen fields ($\psi_1 = 1$) are present in the solution for the three upper cases; there are delta peaks on 0, 1, 1/2, and other simple fractions depending on q. Notice that even when frozen fields are not present, there are many almost frozen fields (the distributions only concentrate around 0, 1, 1/2, and other fractions).

present. However, since there are soft fields with values $1 - \epsilon$, where ϵ can be almost arbitrarily small, one cannot see from these pictures the absence of frozen variables. For the case $c=9$, $q=4$, $m=0.8$ for instance, the presence of the quasihard fields makes the distribution clearly concentrate on values around one, zero, and half (note, however, that the amplitude—on a logarithmic scale—is far less important).

The quasihard fields are therefore very hard to distinguish numerically from the true hard ones. This is evidenced in Fig. 14 where we plot the fraction of hard fields computed using the expression (38) together with a numerical estimate made by population dynamics without the separate hard and soft implementation. We show that the fraction of fields of value $1 - \epsilon$ is not zero in regions where we know that there are no hard fields even for $\epsilon = 10^{-20}$. This demonstrates the
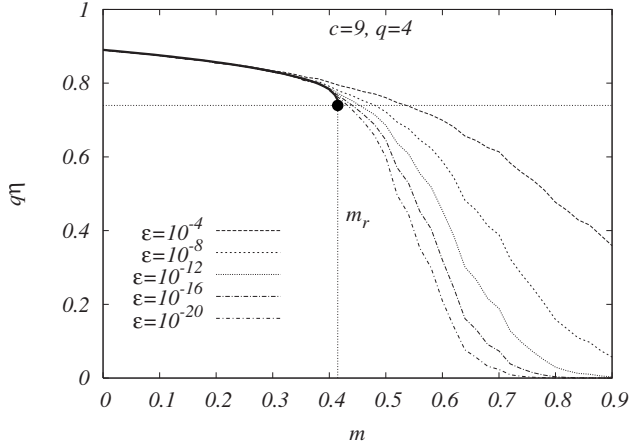
FIG. 14. Fraction of the hard and the quasihard cavity fields $q\eta$ (a field is quasihard if $\psi > 1 - \epsilon$) in the four-coloring of nine-regular graphs. The bold line is obtained with the analytical computation of the fraction of hard fields and the dot corresponds to the threshold m_r.

presence of quasihard fields, with ϵ going to zero as the critical m is approached. This transition is further studied in [64].

An important simplification of the 1RSB update (20) arises when we consider the soft and the hard fields separately. The fraction of hard fields can be computed using the generalized SP equation (38), provided the ratio Z_{soft}^m/Z_{hard}^m is computed. This considerably reduced the size of the population as only soft field has to be kept in memory. Another way to further speed up the code is to generate directly soft fields with a uniform measure instead of waiting for them to come out from the 1RSB recursion. Indeed they might be quite rare in the region of small m and one can spend a considerable amount of time before being able to sample them correctly. Generating the soft fields with a uniform weight turns out to be rather easy using the following method: (i) Choose two random colors q_1 and q_2. (ii) Perform the usual recursion (31) in order to have a new vector but forbid incoming hard fields to q_1 and q_2. (iii) To obtain a uniform soft-field generator, the resulting field should be weighted by $1/\binom{s}{2}$ where s is the number of non-null components in the vector. This is specially useful in the case of Erdős-Rényi graphs.

The formula for the free entropy $\Phi_s(m)$ (32) also simplifies in this case. Consider a given site i; the site-free entropy term can be split into three parts when (i) the field is hard, (ii) the total field is soft, and (iii) the field is contradictory. Then

$$\Phi_s^i = \ln[p_{hard}\overline{(Z_{hard})^m} + p_{soft}\overline{(Z_{soft})^m}], \quad (D1)$$

where p_{hard} or p_{soft} are the probabilities that the total field is frozen or soft, and are given by the SP recursion. Indeed the probability that the total field is not contradictory ($p_{hard} + p_{soft}$) is the denominator in Eq. (35) while p_{hard} is the numerator of Eq. (35). The link part can also be simplified using the fact that contradictions arise when two incoming frozen messages of the same color are chosen, so that

$$\Phi_s^{ij} = \ln[p_{no\ contr}\overline{(Z_{no\ contr})^m}], \quad (D2)$$

where $p_{no\ contr}$ is simply $(1 - q\eta^{i \to j}\eta^{j \to i})$.

For $m=0$ the formula further simplifies as $Z_{hard}^m = Z_{soft}^m = Z_{no\ contr}^m = 1$ so that

$$\Phi_s^i = \ln\left(\sum_{l=0}^{q-1}(-1)^l \binom{q}{l+1}\prod_k[1 - (l+1)\eta^{k \to i}]\right), \quad (D3)$$

$$\Phi_s^{ij} = \ln(1 - q\eta^{i \to j}\eta^{j \to i}). \quad (D4)$$

This is precisely what was obtained within the energetic cavity approach in [30]. The numerical population dynamics implementation with mixed hard and soft strategy is therefore as precise as it could be since we obtain the *exact* evaluation in the $m=0$ case. This simple computation also demonstrates how one can recover the energetic zero temperature limit from the generic formalism.

Finally, we obtain the function $\Phi_s(m)$. We fit this function using an ansatz $\Phi_s(m) = a + b2^m + c3^m + \cdots$ and then perform the Legendre transform to obtain the entropy and complexity. It is also possible to compute directly the complexity from the population data using the expression of the derivative of the potential directly in the code. Both methods lead to very good results. We show an example of the raw data and their fit in Fig. 15, where the data have been obtained with relatively small population ($N=5000$) but where the mixed strategy separating the hard and soft fields have been used. For the purely soft-field branch, we used $N=50\,000$. It took a few hours up to a few days to generate these curves on present Intel PCs.

In the case of biregular random graphs, one needs two different populations: one for the fields going from nodes with connectivity c_1 and one for the fields going from nodes with connectivity c_2. Then each iteration for population 1 (respectively, 2) should be performed using as incoming messages the vectors of population 2 (respectively, 1.). Again, one can separately perform the recursion for the hard-field fractions in both populations.

The case of Erdős-Rényi random graphs is more involved, as one needs a large number N_{pop} of populations, each of them of size N. In this case, using the separate hard and soft fields implementation and the formulas (D1) and (D2) for complexity is crucial, as it allows a good precision even for smaller population sizes. We used typically $2N_{pop}/c \approx (1-3) \times 10^3$ and $N \approx (1-3) \times 10^2$. The error bars in Table II are computed from several independent runs of the population dynamics. In each case we were able to make the equilibration times and the population sizes large enough such that by doubling the time or the population size we did not observe any significant systematic changes in the average results.

APPENDIX E: HIGH-q ASYMPTOTICS

The quenched averages in the large q limit are the same for the regular and Erdős-Rényi graphs and we thus consider directly the regular ensemble of connectivity $c=k+1$. The appearance of a nontrivial 1RSB solution for $m=0$, which corresponds to c_{SP}, was already computed in [31] and reads
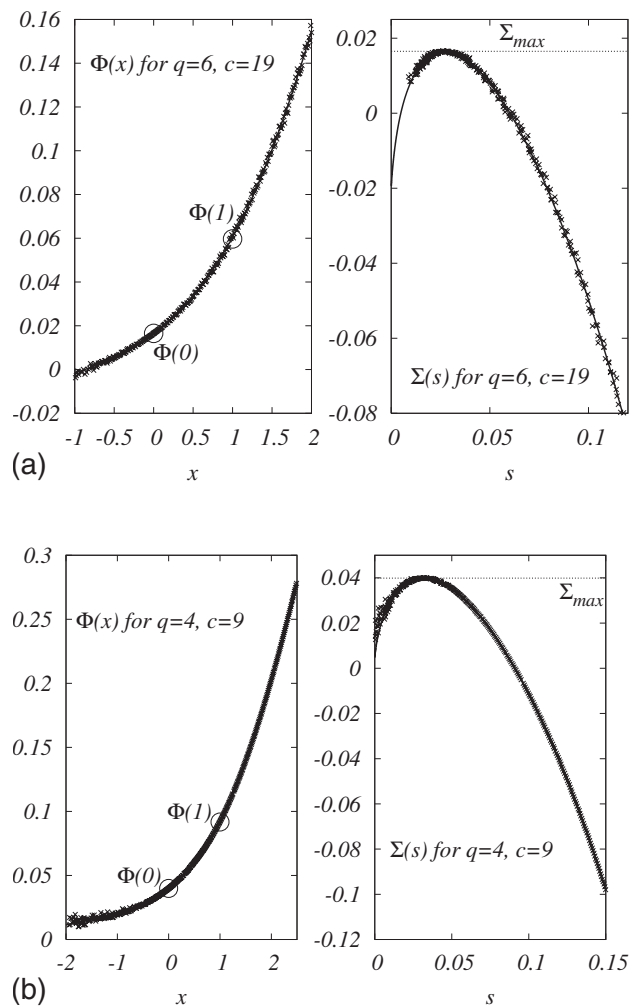
FIG. 15. The numerical results for the free entropy (30) and its fit with a function $a+b2^x+c3^x+\cdots$ for the six-coloring of 19-regular graphs and the four-coloring of nine-regular graphs. Circles give the analytical results at $m=0$ and $m=1$. On the right parts, we present the complexity versus internal entropy with the numerical points and the Legendre transform of the fit of the free entropy. The analytical result for Σ_{max} is also shown.

$$c_r(m=0) = c_{SP} = q[\ln q + \ln \ln q + 1 - \ln 2 + o(1)], \tag{E1}$$

$$\eta_d(m=0) = \frac{1}{q}\left[1 - \frac{1}{\ln q} + o\left(\frac{1}{\ln q}\right)\right], \tag{E2}$$

while the coloring threshold is [31]

$$c_s = 2q \ln q - \ln q - 1 + o(1). \tag{E3}$$

We now show how the connectivity where a solution with hard fields at $m=1$ first appears, and how the complete free entropy $\Phi_s(m)$ [Eq. (30)] can be computed close to the COL/UNCOL transition.

1. Appearance of hard fields at $m=1$

We first show that the correct scaling for the appearance of hard fields at $m=1$ is

$$k = q[\ln q + \ln \ln q + \alpha], \tag{E4}$$

and compute the value of α. In the order $\mathcal{O}(q)$ we can write also $k=(q-1)[\ln(q-1)+\ln\ln(q-1)+\alpha]$. The starting point is Eq. (C9), with $\mathcal{Q}_1(x)=\delta(x-k)$. In the large q limit the fraction of hard fields is $\mu(q,k)=q\eta(q,k)=1-\theta(q,k)$, where $\theta(q,k)=o(1)$ is the fraction of soft fields. We check self-consistently at the end of the computation that only the two first terms of Eq. (C9) are important. Then we have

$$\mu(q,k) = 1 - (q-1)\left(1 - \frac{1}{q-1}\mu(q,k)\right)^k$$

$$= 1 - (q-1)e^{-[k\mu(q,k)/(q-1)]}. \tag{E5}$$

A self-consistent equation for $\theta(q,k)$ follows:

$$\ln(q-1)\theta(q,k) = (q-1)^{\theta(q,k)}e^{-\alpha}, \tag{E6}$$

which is solved by $\theta(q,k)=\gamma(\alpha)/\ln(q-1)$ where

$$\gamma(\alpha)e^{-\gamma(\alpha)} = e^{-\alpha}. \tag{E7}$$

The maximum of the left-hand side is $1/e$ for $\gamma(\alpha)=1$. It means that a solution of Eq. (E7) exists for $\alpha>1$. Finally, the hard fields appear in the 1RSB solution for $m=1$ at connectivity

$$c_r(m=1) = q[\ln q + \ln \ln q + 1 + o(1)]. \tag{E8}$$

The clustering transition c_d should be between c_{SP} and $c_r(m=1)$ as this is what we observed for finite q. We see that c_{SP} and $c_r(m=1)$ differs only in the third order and both are very far from the coloring threshold and also from the condensation transition as we show in Appendix E 2. It would be interesting to compute a large q expansion of the connectivity at which the hard fields appear in all the clusters (for all m such that $\Sigma(m)>0$). Together with our conjecture about rigidity being responsible for the computational hardness, that might give a hint about the answer on the long-standing question [22]: "Is there a polynomial algorithm and ϵ such that the algorithm would color random graphs of average connectivity $(1+\epsilon)q \ln q$ for all large q?"

2. Condensation transition

To compute the large-q asymptotic of the condensation transition, we first need to derive the large-q expansion of the free entropy (30) in the connectivity regime $c=2q \ln q$. Let us show self-consistently that the following scaling is relevant for the condensation transition in the large q limit:

$$c_s = 2q \ln q - \gamma \ln q + \alpha, \tag{E9}$$

$$\eta = \frac{1}{q} - \frac{B}{q^2}, \tag{E10}$$

and compute the constants γ, α, B. Using the above scaling, the function $w(\eta)$ (35) is dominated by the first two terms in

the numerator and denominator, and reads in the first two leading orders $1 - qw(\eta) \approx \frac{qe^{-k\eta}}{2}$ so that

$$w(\eta) = \frac{1}{q} - \frac{1}{2q^2} + \mathcal{O}\left(\frac{\ln q}{q^3}\right) \qquad (E11)$$

independently of γ, α, and B. To take into account the re-weighting we expand Eq. (38) in the two leading orders

$$\eta = \frac{1}{q} - \frac{1}{2q^2}\frac{\overline{Z_s^m}}{\overline{Z_h^m}} + \mathcal{O}\left(\frac{\ln q}{q^3}\right). \qquad (E12)$$

Note that almost all the incoming fields are hard, i.e., have one component of value 1. Since there are on average only $2B \ln q$ incoming soft fields, the leading order of the hard-field reweighting [the normalization in Eq. (4)] is different from 1 with a probability only $\mathcal{O}(\ln q/q)$. Similarly, almost all the soft fields have two nonzero and equal components. The normalization in Eq. (4) is thus almost surely 2, thus the average reweighting factor of the soft fields is

$$\overline{Z_s^m} = 2^m + \mathcal{O}\left(\frac{\ln q}{q}\right). \qquad (E13)$$

Finally,

$$\eta = \frac{1}{q} - \frac{2^m}{2q^2} + \mathcal{O}\left(\frac{\ln q}{q^3}\right). \qquad (E14)$$

Therefore the constant B in Eq. (E10) is $B = 2^m/2$, independent of γ and α.

The computation of the complexity requires the next order in the hard-field reweighting. Indeed the normalization in Eq. (4) might not be 1 but 1/2; and this happens when there is a soft field arriving of the color corresponding to the hard field in consideration. The probability of this event is $\frac{2c(1-q\eta)}{q} = \mathcal{O}\left(\frac{q}{\ln q}\right)$. The hard-field reweighting is thus

$$\overline{Z_h^m} = 1\left[1 - \frac{2c(1-q\eta)}{q}\right] + \frac{2c(1-q\eta)}{q}\frac{1}{2^m} + o\left(\frac{\ln q}{q}\right). \qquad (E15)$$

We now expand the replicated free energy (30) in the large q limit and regime (E9). We remind one that from Eqs. (6) and (8),

$$\Phi_s(m) = \ln\overline{(Z_0^i)^m} - \frac{c}{2}\ln\overline{(Z_0^{ij})^m}. \qquad (E16)$$

The averages are over the population in the sense of Eq. (21).

The site free energy is the logarithm of the average of the total field normalization. This average can be split into three parts when (i) the total field is a hard field, (ii) the total field is a soft field, and (iii) the total field is contradictory (and its normalization zero). The probability that the total field is not contradictory is the denominator in Eq. (35),

$$g(\eta) = \sum_{l=0}^{q-1} (-1)^l \binom{q}{l+1}[1 - (l+1)\eta]^c, \qquad (E17)$$

where again only the first two terms are relevant in the expansion. The site free energy is then

$$\ln\overline{(Z_0^i)^m} = \ln g(\eta) + \ln[qw(\eta)\overline{Z_h^m} + \{1 - qw(\eta)\}\overline{Z_s^m}]$$

$$\simeq \ln\left[q(1-\eta)^c - \frac{q(q-1)}{2}(1-2\eta)^c\right]$$

$$+ \ln\left[1 - \frac{1}{2q} - \frac{2c(1-q\eta)}{q}\left(1 - \frac{1}{2^m}\right) + \frac{2^m}{2q}\right] + o\left(\frac{1}{q}\right), \qquad (E18)$$

where

$$\ln\left[q[1-\eta]^c + \frac{q(q-1)}{2}[1-2\eta]^c\right]$$

$$= \ln q + c\ln[1-\eta] + \ln\left(1 - \frac{q-1}{2}\left[\frac{1-2\eta}{1-\eta}\right]^c\right)$$

$$\approx \ln q + c\ln\left[1 - \frac{1}{q} + \frac{2^m}{2q^2}\right] + \ln\left[1 - \frac{1}{2q} + o(1/q)\right]. \qquad (E19)$$

To compute the link contribution in Eq. (E16) we need to consider two fields $\psi_s^{j \to j}$ and $\psi_s^{j \to i}$ and to compute

$$Z_0^{ij} = 1 - \sum_{s=1}^{q} \psi_s^{j \to j}\psi_s^{j \to i}. \qquad (E20)$$

There are three different cases:

(1) Two hard fields are chosen, then $Z_0^{ij} = 0$ with probability $q\eta^2$ (this is of order $1/q$) and $Z_0^{ij} = 1$ with probability $q(q-1)\eta^2$ (this is of order 1).

(2) Two soft fields are chosen, then $Z_0^{ij} = 1$ with probability $(1-q\eta)^2$ (this is of order $1/q^2$), all other situations being $\mathcal{O}(1/q^3)$. Let us remind one that the dominant soft fields are two-component of type $(1/2, 1/2, 0, 0, \dots)$.

(3) One hard and one soft field is chosen, then $Z_0^{ij} = 1$ with probability $2\eta(1-q\eta)(q-2)/q$, and $Z_0^{ij} = 1/2$ with probability $4\eta(1-q\eta)$ (this is of order $1/q^2$).

On average, one thus obtains for the link contribution,

$$\ln\overline{(Z_0^{ij})^m} = \ln\left[[1 - q\eta^2 - 4\eta(1-q\eta)]1^m + 4\eta(1-q\eta)\frac{1}{2^m}\right]$$

$$+ o\left(\frac{1}{q^2}\right). \qquad (E21)$$

Putting together the two pieces (E18) and (E21), expanding η according to Eq. (E14), and considering only the highest order in c, we can finally write the free energy as

$$\Phi_s(m) = \ln q - \frac{c}{2q} + \frac{2^m - 2}{2q} - \frac{c}{4q^2} + o\left(\frac{1}{q}\right). \qquad (E22)$$

The internal entropy $s(m)$ and the complexity $\Sigma = \Phi_s(m) - ms(m)$ are then

$$s(m) = \frac{\partial\Phi_s(m)}{\partial m} = \frac{2^m \ln 2}{2q}, \qquad (E23)$$

$$\Sigma(m) = \ln q - \frac{c}{2q} + \frac{2^m - 2 - m2^m \ln 2}{2q} - \frac{c}{4q^2}, \quad \text{(E24)}$$

and the complexity is thus zero for $c_{\Sigma=0} = 2q \ln q - \ln q - 2 + 2^m[1 - m \ln 2] + o(1)$. In particular, one has for the coloring and the condensation thresholds,

$$c_{\Sigma=0}(m = 0) = 2q \ln q - \ln q - 1 + o(1), \quad \text{(E25)}$$

$$c_{\Sigma=0}(m = 1) = 2q \ln q - \ln q - 2 \ln 2 + o(1). \quad \text{(E26)}$$

For connectivity $c = 2q \ln q - \ln q + \alpha$, one gets

$$2qs(m) \simeq 2^m \ln 2, \quad \text{(E27)}$$

$$2q\Sigma(m) \simeq 2^m - 2 - m2^m \ln 2 - \alpha. \quad \text{(E28)}$$

[1] S. Cook, *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing* (ACM, New York, 1971), pp. 151–158; M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (Freeman, New York, 1979); C. H. Papadimitrious, *Computational Complexity* (Addison-Wesley, Reading, MA, 1994).

[2] K. Appel and W. Haken, Ill. J. Math. **21**, 421 (1977); **21**, 491 (1977).

[3] R. M. Karp, *Complexity of Computer Computations*, edited by R. E. Miller and J. W. Thatcher (Plenum, New York, 1972), p. 85.

[4] B. Bollobás, *Random Graphs* (CUP, Cambridge, 2001).

[5] S. Janson, T. Luczak, and A. Rucinski, *Random Graphs* (Wiley, New York, 2000).

[6] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[7] D. Achlioptas and E. Friedgut, Random Struct. Algorithms **14**, 63 (1999).

[8] D. Achlioptas, A. Naor, and Y. Peres, Nature (London) **435**, 759 (2005).

[9] S. Dubois and O. Mandler (unpublished).

[10] D. Achlioptas and C. Moore, J. Comput. Syst. Sci. **67**, 441 (2003).

[11] D. Achlioptas and C. Moore, *Proceedings of RANDOM* (Springer, Berlin, 2004), pp. 219–228.

[12] L. Shi and N. C. Wormald, Combinatorics, Probab. Comput. **16**, 309 (2007).

[13] G. Kemkes and N. C. Wormald, Research Report No. CORR 2005-22, Faculty of Mathematics, University of Waterloo, 2005 (unpublished).

[14] J. Díaz, G. Grammatikopoulos, A. C. Kaporis, L. M. Kirousis, X. Perez, and D. G. Sotiropoulos, *Proceedings of the 13th Annual European Symposium on Algorithms* (Springer, Berlin, 2005), pp. 215–225.

[15] T. Luczak, Combinatorica **11**, 45 (1991).

[16] P. Cheeseman, B. Kanefsky, and W. M. Taylor, *Proceedings of the 12th International Joint Conference on Artificial Intelligence* (IJCAI-91, Sydney, Australia, 1991).

[17] B. Selman and S. Kirkpatrick, Artif. Intell. **81**, 273 (1996); B. Selman, D. G. Mitchell, and H. J. Levesque, *ibid.* **81**, 17 (1996).

[18] D. L. Mammen and T. Hogg, Am. Econ. Rev. **7**, 47 (1997).

[19] C. P. Williams and T. Hogg, Artif. Intell. **70**, 73 (1994); T. Hogg, B. A. Huberman, and C. P. Williams, *ibid.* **81**, 1 (1996).

[20] S. Cocco and R. Monasson, Phys. Rev. Lett. **86**, 1654 (2001).

[21] D. Achlioptas and M. Molloy, *Proceedings of FOCS-97* (IEEE, New York, 1997).

[22] A. Frieze and C. J. H. McDiarmid, Random Struct. Algorithms **10**, 5 (1996).

[23] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, Nature (London) **400**, 133 (1999).

[24] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).

[25] F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[26] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).

[27] M. Mézard and G. Parisi, J. Stat. Phys. **111**, 1 (2003).

[28] M. Mézard and R. Zecchina, Phys. Rev. E **66**, 056126 (2002).

[29] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond*, Lecture Notes in Physics, Vol. 9 (World Scientific, Singapore, 1987).

[30] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. Lett. **89**, 268701 (2002); A. Braunstein, R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Phys. Rev. E **68**, 036702 (2003).

[31] F. Krzakala, A. Pagnani, and M. Weigt, Phys. Rev. E **70**, 046705 (2004).

[32] J. van Mourik and D. Saad, Phys. Rev. E **66**, 056120 (2002).

[33] A. Braunstein, M. Mézard, and R. Zecchina, Random Struct. Algorithms **27**, 201 (2005).

[34] A. Montanari, G. Parisi, and F. Ricci-Tersenghi, J. Phys. A **37**, 2073 (2004).

[35] G. Biroli, R. Monasson, and M. Weigt, Eur. Phys. J. B **14**, 551 (2000).

[36] M. Mézard, T. Mora, and R. Zecchina, Phys. Rev. Lett. **94**, 197205 (2005).

[37] D. Achlioptas and F. Ricci-Tersenghi, *Proceedings of the 38th Annual ACM Symposium on Theory of Computing* (ACM, New York, 2006), pp. 130–139.

[38] M. Mézard, M. Palassini, and O. Rivoire, Phys. Rev. Lett. **95**, 200202 (2005).

[39] O. Rivoire, These de doctorat, 2005 (unpublished); CNRS http://tel.ccsd.cnrs.fr/tel-00009956/en/.

[40] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, 2nd ed. (Morgan Kaufmann, San Francisco, 1988).

[41] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Exploring Artificial Intelligence in the New Millennium* (Morgan Kaufmann, San Francisco, 2003), Chap. 8, pp. 239–236.

[42] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Proc. Natl. Acad. Sci. U.S.A. **104**, 10318 (2007).

[43] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian (unpublished).

[44] D. J. Gross, I. Kanter, and H. Sompolinsky, Phys. Rev. Lett. **55**, 304 (1985).

[45] H. O. Georgii, *Gibbs Measures and Phase Transitions* (Walter

De Gruyer, Berlin, 1988).

[46] S. Franz, M. Mézard, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, Europhys. Lett. **55**, 465 (2001).

[47] H. A. Bethe, Proc. R. Soc. London, Ser. A **150**, 552 (1935); R. Peierls, *ibid.* **154**, 207 (1936); L. Onsager J. Am. Chem. Soc. **58**, 1486 (1936); P. R. Weiss, Phys. Rev. **74**, 1493 (1948).

[48] K. Y. Wong and D. Sherrington, J. Phys. A **21**, L459 (1988); Y. Y. Goldschmidt and P.-Y. Lai, J. Phys. A **23**, L775 (1990).

[49] D. Aldous, Random Struct. Algorithms **18**, 381 (2001).

[50] A. Bandyopadhyay and D. Gamarnik, *ACM-SIAM Symposium on Discrete Algorithms* (ACM, New York, 2006), p. 890.

[51] J. Jonasson, Stat. Probab. Lett. **57**, 243 (2002).

[52] A. Montanari and Devavrat Shah, *Proceedings of Symposium of Discrete Algorithms (SODA), New Orleans, September, 2007* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2007).

[53] M. Mézard and A. Montanari, J. Stat. Phys. **124**, 1317 (2007).

[54] J. P. Bouchaud and G. Biroli, J. Chem. Phys. **121**, 7347 (2004).

[55] A. Montanari and G. Semerjian, J. Stat. Phys. **124**, 103 (2006); Phys. Rev. Lett. **94**, 247201 (2005).

[56] O. Rivoire, G. Biroli, O. C. Martin, and M. Mézard, Eur. Phys. J. B **37**, 55 (2004).

[57] H. Kesten and B. P. Stigum, Ann. Math. Stat. **37**, 1463 (1966); J. Math. Anal. Appl. **17**, 309 (1966).

[58] G. Parisi, J. Phys. A **A13**, L115 (1980).

[59] R. Monasson, Phys. Rev. Lett. **75**, 2847 (1995).

[60] B. Derrida, Phys. Rev. Lett. **45**, 79 (1980); Phys. Rev. B **24**, 2613 (1981).

[61] T. R. Kirkpatrick and D. Thirumalai, Phys. Rev. Lett. **58**, 2091 (1987); T. R. Kirkpatrick, D. Thirumalai, and P. G. Wolynes, Phys. Rev. A **40**, 1045 (1989); M. Mézard and G. Parisi, Phys. Rev. Lett. **82**, 747 (1999); M. Mézard, *More is Different*, edited by N. Ong and R. Bhatt (Princeton University Press, Princeton, 2002).

[62] A. Montanari and G. Semerjian, J. Stat. Phys. **125**, 23 (2006).

[63] W. Kauzmann, Chem. Rev. (Washington, D.C.) **43**, 219 (1948).

[64] G. Semerjian, e-print arXiv:0705.2147v1.

[65] G. Parisi, J. Stat. Phys. (to be published).

[66] E. N. Maneva, E. Mossel, and M. J. Wainwright, *Proceedings of SODA* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).

[67] A. Braunstein and R. Zecchina, J. Stat. Mech.: Theory Exp. (2004), P06007.

[68] B. Pittel, J. Spencer, and N. Wormald, J. Comb. Theory, Ser. B **67**, 111 (1996).

[69] E. Gardner, Nucl. Phys. B: Field Theory Stat. Syst. **74**, 297 (1985).

[70] A. Montanari and F. Ricci-Tersenghi, Eur. Phys. J. B **33**, 339 (2003).

[71] G. Parisi and T. Rizzo, Phys. Rev. B **72**, 184431 (2005); M. Müller, L. Leuzzi, and A. Crisanti, *ibid.* **74**, 134431 (2006).

[72] C. S. O'Hern, L. E. Silbert, A. J. Liu, and S. R. Nagel, Phys. Rev. E **68**, 011306 (2003).

[73] F. Krzakala and J. Kurchan, Phys. Rev. E **76**, 021122 (2007).

[74] J. Chavas, C. Furtlehner, M. Mézard, and R. Zecchina, J. Stat. Mech.: Theory Exp. (2005), P11016.

[75] P. Svenson and G. Nordahl, Phys. Rev. E **59**, 3983 (1999).

[76] B. Selman, H. A. Kautz, and B. Cohen, *Proceedings of AAAI-94, Seattle* (American Association for Artificial Intelligence, Menlo Park, CA, 1994).

[77] J. Ardelius and E. Aurell, Phys. Rev. E **74**, 037702 (2006).

[78] S. Seitz, M. Alava, and P. Orponen, J. Stat. Mech.: Theory Exp. (2005), P06006.

[79] S. Boettcher and A. G. Percus, Phys. Rev. E **64**, 026114 (2001).

[80] M. Mézard, G. Parisi, and M. A. Virasoro, J. Phys. (France) Lett. **46**, L-217 (1985).

[81] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians* (Springer, New York, 2000).

[82] J. Pitman and M. Yor, Ann. Probab. **25**, 855 (1997).

[83] L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, New York, 1986).

[84] T. Castellani, F. Krzakala, and F. Ricci-Tersenghi, Eur. Phys. J. B **47**, 99 (2005).

[85] More precisely, [7] proves the existence of a sharp threshold for a possibly size dependent sequence of thresholds, whose convergence is not proven.

[86] In this paper we use the words "cluster" and "state" as synonyms.

[87] A slightly different adaptation, closer to [76], was performed a few years ago by Andrea Pagnani and Martin Weigt (private communication).