A theory explaining the limits and performances of algorithms based on simulated annealing in solving sparse hard inference problems

Maria Chiara Angelini[1,2] and Federico Ricci-Tersenghi[1,2,3]

[1]*Dipartimento di Fisica, Sapienza Università di Roma, 00185 Italy*
[2]*INFN - Sezione di Roma 1, 00185 Italy*
[3]*CNR - Nanotec, unità di Roma, 00185 Italy*

The planted coloring problem is a prototypical inference problem for which thresholds for Bayes optimal algorithms, like Belief Propagation (BP), can be computed analytically. In this paper, we analyze the limits and performances of the Simulated Annealing (SA), a Monte Carlo-based algorithm that is more general and robust than BP, and thus of broader applicability. We show that SA is sub-optimal in the recovery of the planted solution because it gets attracted by glassy states that, instead, do not influence the BP algorithm. At variance with previous conjectures, we propose an analytic estimation for the SA algorithmic threshold by comparing the spinodal point of the paramagnetic phase and the dynamical critical temperature. This is a fundamental connection between thermodynamical phase transitions and out of equilibrium behavior of Glauber dynamics. We also study an improved version of SA, called replicated SA (RSA), where several weakly coupled replicas are cooled down together. We show numerical evidence that the algorithmic threshold for the RSA coincides with the Bayes optimal one. Finally, we develop an approximated analytical theory explaining the optimal performances of RSA and predicting the location of the transition towards the planted solution in the limit of a very large number of replicas. Our results for RSA support the idea that mismatching the parameters in the prior with respect to those of the generative model may produce an algorithm that is optimal and very robust.

I. INTRODUCTION

Inference problems are widespread in all scientific disciplines and real-world applications based on data analysis. It is also well known that not all inference problems have the same difficulty, and the solving algorithms need possibly to be tested on the hardest problems. For this reason, focusing on classes of inference problems that are provably very hard is mandatory at the time of testing algorithms. Here we choose to work with inference problems defined on sparse random graphs that present the so-called *hard phase*, that is a range of the signal-to-noise ratio (SNR) where the solution is in principle achievable (by unbounded computational power), but any known algorithm running in a time growing polynomially with the problem size is not able to reach such a solution. Below we define more in detail this class of hard inference problems; for the moment it is enough to stress that the presence of the hard phase is related to the so-called *information to computational gap conjecture*. This conjecture, partially proved for some inference problems and some classes of algorithms, assumes that the minimal SNR needed by any polynomial-time algorithm to solve one of these hard inference problems is larger than the information theoretical lower bound [1–10].

FIG. 1. The schematic phase diagram for hard inference problems showing an information-to-computational gap. The SNR is here given by the graph mean degree c. While for $c < c_{\mathrm{IT}}$ it is impossible to detect the planted signal, for $c > c_{\mathrm{IT}}$ it is in principle possible, but any known algorithm requires a much larger SNR, $c > c_{\mathrm{alg}}$, to achieve signal detection. The value of c_{alg} depends on the algorithm and the present work aims at estimating it for simulated annealing algorithms.

In Fig. 1 we show a schematic phase diagram for a generic inference problem with a hard phase. The derivation of such a phase diagram has been possible working in the very commonly used setting of Bayesian inference known as the *teacher-student scenario*, firstly introduced in Ref. [11]. In this setting the teacher generates a random signal $\mathbf{x}$ with the prior probability distribution $P_p(\mathbf{x})$. The signal generated by the teacher is often called the *planted* configuration. Later, the teacher generates the data from the signal using a probabilistic model P_m. Without loss of generality, we assume that the data come into the form of a sparse graph G and that the SNR is directly related to the graph mean connectivity c. Applying a probabilistic model to a random signal, we get a random graph G generated with the likelihood $P_m(G|\mathbf{x})$. The teacher provides the student with the prior P_p, the model P_m and the data G, asking the student to infer the planted signal $\mathbf{x}$. Using the Bayes theorem, the student can easily write down the posterior probability distribution

$$P(\mathbf{x}|G) \propto P_m(G|\mathbf{x})P_p(\mathbf{x}) . \qquad (1)$$

The Bayes optimal estimator $\hat{\mathbf{x}}_{\mathrm{opt}}$ of the signal can be obtained by sampling the posterior $P(\mathbf{x}|G)$. A weak recovery of the signal can be achieved if the Bayes optimal estimator $\hat{\mathbf{x}}_{\mathrm{opt}}$ is better than a random guess: this happens for $c > c_{\mathrm{IT}}$, that is above the information theoretical threshold c_{IT}.

In hard inference problems, it is not enough to have an SNR larger than the c_{IT} threshold to recover the signal in a time growing polynomially with the problem size. A larger algorithmic threshold exists, c_{alg}, such that only for $c > c_{\mathrm{alg}}$ polynomial-time algorithms can actually detect the signal. The threshold c_{alg} clearly depends on the algorithm, and we are going to show how to estimate it for stochastic samplers based on the Simulated Annealing (SA) algorithm.

When the graph G representing the data is a random graph [12], the posterior can be efficiently sampled by using the

Bethe approximation and the Belief Propagation (BP) algorithm [13]. The algorithmic threshold for the BP algorithm c_{BP} is larger than c_{IT} for inference problems with a hard phase, and is nonetheless considered the optimal algorithmic threshold for algorithms running in linear time in the problem size N [14]. Other algorithms which are not Bayes optimal, i.e. not sampling exactly the Bayesian posterior, may have even larger algorithmic thresholds.

Unfortunately, the BP algorithm is known to be not very robust and of limited applicability when the graph G is not locally tree-like, as the presence of short loops may avoid convergence to a fixed point and, even at convergence, the estimated marginal probabilities may be far from the exact ones [15]. We are not saying BP can never be used for inference away from random graphs: indeed many works are showing the applicability of BP or inference methods strongly related to it on real-world data (see [16] just as an example). However, having at hand more robust algorithms that can work efficiently both on random graphs and on more general structured data is of primary importance.

To this aim, it is natural to consider samplers of the posterior probability distribution based on Monte Carlo Markov Chain (MCMC). By construction, an MCMC will eventually sample exactly from the posterior, but its limits and performances are mostly unknown. In some recent works, it has been proved how the algorithmic threshold of some MCMC algorithms is higher than the algorithmic thresholds of other Bayes-optimal algorithms in specific planted problems [17–19]. However, apart from the straightforward statement that MCMC methods are slowed down in a phase where ergodicity is broken in the large N limit [20], more precise general statements on when and why MCMC-based inference algorithms can find the solution are missing to the best of our knowledge.

MCMC-based inference algorithms could really be ideal if, on top of being robust, they would also be optimal in those cases where the optimal threshold can be analytically computed. To understand this last point, we believe that the study of inference problems defined on random sparse graphs is the perfect framework: thresholds can be computed analytically thanks to the Bethe approximation for the posterior distribution and MCMC simulations can be run efficiently on very large problem sizes since running times are linear in N for sparse models.

We study one of the simplest and most common inference problems, namely the planted coloring problem (to be defined in more detail below), using simple, but effective, MCMC-based algorithms: Simulated Annealing (SA) and Replicated Simulated Annealing (RSA). The main results that we achieve within this framework are the following.

Algorithmic threshold for Simulated Annealing:
We make a conjecture based on firm numerical evidence and strong analytical arguments for the determination of the algorithmic threshold of SA. Our conjecture is based on the comparison of the spinodal point and the dynamical phase transition. Although sub-optimal with respect to the Bayes-optimal threshold, we find that the SA algorithmic threshold is much better than what has been reported in previous studies.

Replicated Simulated Annealing is Bayes-optimal:
We provide strong numerical evidence that the algorithmic threshold for RSA is extremely close to the Bayes-optimal one of BP. RSA works very well even for a small number of coupled replicas, so the complexity overhead with respect to SA is minimal.

Nature of the phase transition:
We provide both numerical and analytical evidence that coupling replicas changes the nature of the glass transition in the random problem, from a discontinuous one to a continuous one. This observation may have important algorithmic consequences, as hard phases are known to exist only in models undergoing discontinuous phase transitions.

Analytical estimate for the RSA critical temperature:
We provide an analytical estimate of the transition temperature in the limit of a large number of coupled replicas for the RSA algorithm. Convergence to this limiting value is fast, so it provides a good estimate of the critical temperature also for a reasonably small number of replicas, those used in practical simulations.

Optimal temperature for MCMC is not Bayes optimal:
For the sake of simplicity, we study here an inference problem without additional noise, i.e. generated from a Gibbs distribution at $T = 0$. We provide analytical evidence that the optimal temperature for recovering the signal is strictly positive and not small at all, so far from the Bayes optimal one (this is in agreement with previous numerical results).

Analytical control of finite size effects:
Analytical predictions in the large N limit are of scarce practical utility if they are not complemented with the study of finite size effect, which may be very large. We provide an analytical explanation of finite size effects in the behavior of the SA algorithm, via the study of BP on finite graphs, where similar finite size effects appear.

Taken all together the above results provide a strong and coherent theory explaining the behavior of important algorithms like Simulated Annealing in solving hard inference problems defined on sparse graphs. We believe this theory has broad applicability, and preliminary results in other planted models support it (however we leave the study of other problems for future works, the planted coloring being already very interesting and of broad applicability *per se*).

The rest of the paper is organized as follows. In Sec. II we define the planted coloring problem and the associated BP algorithm. In Sec. III we study the performances of SA and put forward a conjecture for its algorithmic threshold. In Sec. IV we describe the RSA algorithm and we show that it can reach the threshold of Bayes-optimal algorithms. In Sec. V we describe the finite size effects in SA and explain them by studying the BP behavior on single instances of finite size. Finally, in Sec. VI we summarize the results found, the general picture emerging from this work and possible future implications. We leave to the appendices the more technical computations.

II. THE PLANTED COLORING PROBLEM

In the present work, we focus our attention on the planted coloring problem, which was the first planted problem studied in detail [21]. Let us start recalling briefly the coloring problem (both the random and the planted versions) and their phase transitions.

In the random q-coloring problem, a random graph of N vertices and mean degree c is given and we have to assign one among the q available colors to each vertex in a way to avoid monochromatic edges, i.e. adjacent vertices with the same color. In the thermodynamical limit $N \to \infty$, a typical random q-coloring problem undergoes several phase transitions [22, 23]. Before the so-called *dynamical* transition c_d, there are exponentially many (in N) solutions that all belong to a large single cluster. For $c_d < c < c_c$, with c_c being the *condensation* transition, the space of solutions splits in an exponential number of different clusters, whose number becomes sub-exponential for $c > c_c$. A typical random graph is q-colorable up to the *satisfiability* threshold c_s and for $c > c_s$ there is no solution to a typical random q-coloring problem. The actual values of the different transitions connectivities, $c_d \leq c_c \leq c_s$, depend on q and also the nature of the phase transition does: for $q < 4$ the dynamical and condensation transition coincide, leading to a continuous phase transition, while for $q \geq 4$ the transition is discontinuous. And this has consequences also on the corresponding planted model.

The planted coloring problem has been introduced in Ref. [21] and consists in first assigning a random q-coloring to the N vertices (this is the planted solution) and then building the random graphs of mean degree c by adding $M = cN/2$ edges connecting randomly chosen pairs of vertices of different colors. The inference problem corresponds to the recovery of the signal (the planted solution) given the random graph G built as explained above. So the mean degree c of the random graph can be seen as a sort of signal to noise ratio: for small c the graph G has exponentially many solutions (colorings) and it is impossible to identify the planted one, while for very large c the planted coloring is only compatible with the graph G and can potentially be identified.

To achieve a more detailed description of the phase transitions taking place in the planted coloring problem, we need to discuss the solutions to the BP equations associated to the problem. Calling $\psi_s^{i \to j}$ the probability that the vertex i takes the color s in absence of the link between i and j, the BP equations read [23]

$$\psi_s^{i \to j} = \frac{1}{Z^{i \to j}} \prod_{k \in \partial i \setminus j} \left(1 - \psi_s^{k \to i}\right) \qquad (2)$$

where $Z^{i \to j}$ is a normalization imposing $\sum_s \psi_s^{i \to j} = 1$ and $\partial i \setminus j$ indicates all the neighbours of i with the exclusion of j. The BP equations are usually solved iteratively and at convergence provide the marginal probability that variable i takes color s via

$$\psi_s^i = \frac{1}{Z^i} \prod_{k \in \partial i} \left(1 - \psi_s^{k \to i}\right), \qquad (3)$$

where Z^i is again a normalization constant.

The fixed point marginals are mainly of two kinds [24]: (i) uniform over the q colors ($\psi_s^i = \frac{1}{q}, \forall i, s$), identifying the paramagnetic (PM) phase; (ii) highly biased towards the planted solution, identifying the signal recovery or ferromagnetic (FM) phase. The stability analysis of the paramagnetic fixed point towards perturbations in the direction of the planted state reveals that the paramagnetic phase is locally stable for $c < c_{\mathrm{KS}} = (q-1)^2$ [21] (in that work the threshold was called c_l, but we prefer a name making evident the connection with the Kesten-Stigum bound in tree reconstruction). So for $c < c_{\mathrm{KS}}$, we expect that BP, if initialized with no information about the signal, will converge to the uninformative PM fixed point and recovery via BP is impossible. Vice versa, for $c > c_{\mathrm{KS}}$ any initial perturbation in the direction of the signal is amplified and BP will spontaneously converge to the informative FM fixed point, highly correlated with the signal. Thus c_{KS} is the algorithmic threshold for partial signal recovery via BP.

The procedure described above is called 'quite planting' as it adds the planted solution (and a cluster surrounding it) without changing the structure of solutions to the random problem for $c < c_c$. Since in this region the planted cluster is one among the exponentially many others, finding it is impossible. Instead, for $c > c_c$, the planted cluster of solutions dominates the total number of solutions and the planted solution can be detected in principle. For this reasons $c_c = c_{\mathrm{IT}}$ is called the *information theoretical* threshold.

Whether the planted solution can be actually found above c_{IT} depends on the nature of the phase transition [14]. In continuous phase transitions, we have $c_{\mathrm{IT}} = c_{\mathrm{KS}}$ and so, as soon as the planted cluster dominates the total number of solutions, BP can actually detect it. This happens in the planted q-coloring problem for $q < 4$. On the contrary, if the transition is discontinuous, we have a *hard* phase for $c_{\mathrm{IT}} < c < c_{\mathrm{KS}}$ where the planted solution is in principle detectable, but signal recovery via BP is impossible, due to the stability (and the strong attraction) of the PM fixed point. This happens for $q > 4$ in the planted q-coloring problem. The case $q = 4$ has remained elusive for a while and was finally solved in [25].

In the present work we focus on the case $q = 5$, that is an inference problem with a well defined hard phase. Threshold values for $q = 5$ are $c_d = 12.837(3)$, $c_c = c_{\mathrm{IT}} = 13.23(1)$ and $c_{\mathrm{KS}} = 16$ [23]. Despite a great effort to break this hard phase, no polynomial algorithm is known that can detect the planted solution for $c < c_{\mathrm{KS}}$.

Let us finally discuss a delicate point. By construction, the BP algorithm relies on the Replica Symmetric (RS) approximation, assuming that most of the solutions form a single and well-connected cluster [13]. Consequently, the BP fixed points can describe only RS phases, like the PM and FM phases discussed above. However, from the solution of the random model, we know that for $c > c_d$ the space of solutions splits into many different clusters: this is the so-called *glassy* phase, as it induces a very strong slowing down in many algorithms, and its description requires to assume Replica Symmetry Breaking (RSB) [26]. These glassy states, which are so important in the description of the random model, play no

role in the BP behavior when it is applied to the planted case and the performance of the BP algorithm is not improved by taking into account the glassy nature of the hard phase [20].

We have already mentioned that BP is Bayes-optimal: it can find optimal solutions for $c > c_{\mathrm{KS}}$ because it assumes the perfect knowledge of the model. Indeed, Eq. (3) assumes that we know that there is a hard constraint imposing the absence of links between nodes of the same color in the construction of the graph around the planted solution. It can be shown that, imposing the perfect knowledge of the model, there cannot exist an RSB phase, and the only possible transition is a usual ferromagnetic first order transition between PM and FM phases. In statistical mechanics language, assuming the perfect knowledge of the model corresponds to be on the so-called *Nishimori line* [27], where the RS assumption holds. So, in the planted coloring problem, thanks to the perfect knowledge of the generative model, only PM and FM phases can dominate the measure over the solutions. The first order transition takes place at c_{IT}, where the free-energies of the two phases are equal, with the FM free-energy being lower for $c > c_{\mathrm{IT}}$ (thus making the signal in principle detectable). The spinodal points are c_d and c_{KS}, as the FM state appears at c_d and the PM state becomes unstable at c_{KS}.

Leaving BP aside, and considering broader classes of algorithms is very likely that one has to consider situations where algorithms do not satisfy the Nishimori condition: either because they have no perfect knowledge about the model, or because, even knowing the model, they go out of equilibrium, thus violating the Nishimori condition. In all these cases, the glassy states — so important in the random model — should be taken again into account (even if they are irrelevant in the thermodynamic RS description of the planted model). We will see in the next sections that they will be particularly important to determine the behavior of MCMC-based solvers that do not assume the perfect knowledge of the model. This is an ideal example to show that moving away from equilibrium statistical mechanics to out of equilibrium processes one can find a much richer physics that requires a more complex description.

III. SIMULATED ANNEALING FOR THE PLANTED COLORING PROBLEM

In this section, we introduce the Simulated Annealing (SA) as a solver for the planted coloring problem, and we analyze its performance. SA is an MC-based algorithm introduced in Ref. [28] and widely used in many optimization problems, including the random coloring problem [29, 30]. In the context of the planted coloring problem, it was considered in Ref. [21], where the authors stated it can find the planted solution in the whole region $c > c_{\mathrm{KS}}$. We will show that this is not in agreement with our numerical results and propose a better conjecture.

Given a configuration $\boldsymbol{s}$ of colors of the nodes, in a statistical mechanics approach, we associate to it an energy counting the number of monochromatic edges

$$H(\boldsymbol{s}) = \sum_{(i,j) \in E} \delta_{s_i, s_j} \,, \tag{4}$$
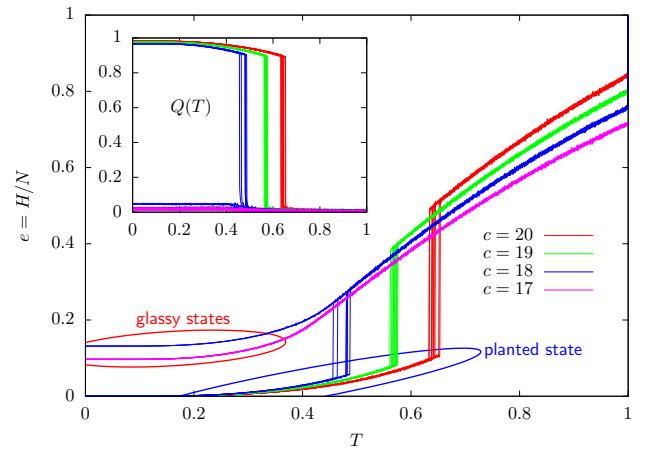

FIG. 2. The energy density as a function of the temperature in the Simulated Annealing for planted coloring with $q = 5$, $N = 10^5$ and $dT = 10^{-7}$. For each value of the mean degree c, we show the results for 10 different samples. For $c \lesssim 18$, SA does not find the planted solution, and get trapped in glassy states. Inset: overlap with the planted configuration as a function of the temperature (same parameters as in the main figure).

where E is the edge set of the graph. We then introduce a temperature $T \equiv 1/\beta$ in the model and a Gibbs-Boltzmann-like probability measure on the configurations

$$P_{GB}(\boldsymbol{s}) \propto e^{-\beta H(\boldsymbol{s})} \,. \tag{5}$$

In the $T = 0$ limit, this probability measure becomes the uniform distribution over the set of solutions (proper colorings), which is always non-empty thanks to the presence of the planted solution $\boldsymbol{s}^*$ with $H(\boldsymbol{s}^*) = 0$. Moreover the $T = 0$ distribution coincides with the posterior $P(\boldsymbol{s}|G)$ if the prior is uniform over all configurations. This indicates that the Bayes optimal temperature is $T_{\mathrm{Bayes}} = 0$. The aim of any inference algorithm should be the sampling of the $T = 0$ distribution to detect $\boldsymbol{s}^*$. We will see that the optimal temperature for MC-based algorithms is different from the Bayes optimal one $T_{\mathrm{Bayes}} = 0$.

The meaning of the distribution $P_{GB}(\boldsymbol{s})$ with $T > 0$ is to relax the hard constraints into soft ones: configurations with monochromatic edges are admitted, but with a probability that is more suppressed the lower is the temperature. We start our SA from a random configuration of colors and at sufficiently high T. At each step of the SA algorithm we decrease the temperature by dT and we do a Monte Carlo Sweep (MCS), that corresponds to the attempt to update the color of each of the N variables following the usual Metropolis rule: in practice, we change the color s_i into s_i', chosen at random between the q possible colors, with probability 1 if in this way the number of monochromatic edges decreases, while we update it with probability $e^{-\beta n}$ if the number of monochromatic edges increases by n after the update.

In Fig. 2 we show the intensive energy, $e = H/N$, of the configurations visited by SA with parameter $dT = 10^{-7}$ in 10 samples of size $N = 10^5$ of the planted coloring problem

with $q = 5$. Differences between samples are negligible (but at $c = 18$), while the behavior of SA changes drastically with the mean degree c.

For $c > 18$, SA finds a configuration highly correlated with the planted solution s^* on each sample. This is confirmed also by the data in the inset, showing the overlap between the configuration s reached by SA and the planted solution s^* defined as

$$Q = \frac{\max_{\pi \in S_q} \sum_i \delta_{s_i^*, \pi(s_i)} - 1/q}{1 - 1/q} \,, \qquad (6)$$

where S_q is the group of permutations of q elements. The overlap defined in this way is null for a random guess, while $Q = 1$ if there is a perfect recovery of the planted solution.

On the contrary, for $c < 18$, SA is unable to reach the planted solution in any sample, and it gets trapped in glassy states, which are metastable states (local minima of higher energy), orthogonal to the planted state ($Q \simeq 0$). So the numerical data strongly suggest that $c_{\rm SA} \simeq 18$ is the algorithmic threshold for SA. Indeed, exactly at $c = 18$, we see in Fig. 2 that the behavior of SA is sample dependent: a typical behavior that is often observed at a spinodal point.

The strict inequality $c_{\rm KS} < c_{\rm SA}$ implies SA is not Bayes optimal and there exists a region, $c \in (c_{\rm KS}, c_{\rm SA})$, where BP performs better than SA.

Let us comment briefly on the choice of N and dT. We have studied several values of N and dT, all providing data consistent with the claim above. We decided to show data for $N = 10^5$ since this value is large enough that we can consider the system as in the thermodynamic limit. Moreover the choice $dT = 10^{-7}$ allows the SA to run for a large enough time, but not so large to allow for activated processes of barrier crossing, that should take a time $O(\exp(N))$. In this way, we are analyzing the regime in which SA is a linear algorithm.

Data in Fig. 2 clearly show that a dynamical first order transition is taking place at the temperature where SA makes a jump to reach the planted cluster of solutions. Moreover when such a dynamical first order transition does not take place (i.e. for $c < c_{\rm SA}$) the overlap remains null and SA stays very far from (actually orthogonal to) the planted solution.

To explain why SA is not able to find the planted solution for $c < c_{\rm SA}$, we need to look to the extension to the finite temperature of the phase transitions at $T = 0$. How the thresholds c_d and $c_{\rm KS}$ change in temperature was already computed in Ref. [21]: we report them with the names $T_d(c)$ and $T_{\rm KS}(c)$ in Fig. 3. For $T < T_d$ glassy states appear while for $T < T_{\rm KS}$ the paramagnetic solution is no more stable with respect to perturbations toward the planted solution.

The SA protocol starts from very high temperatures and keeps cooling very slowly (ideally in an adiabatic way). Which critical line, either T_d or $T_{\rm KS}$, it will find first depends on the c value. Defining the SA algorithmic threshold $c_{\rm SA}$ as the crossing point of these two critical lines

$$T_d(c_{\rm SA}) = T_{\rm KS}(c_{\rm SA}) \,, \qquad (7)$$

we conjecture the following scenario. If $c > c_{\rm SA}$, decreasing T the paramagnetic phase becomes unstable with respect to
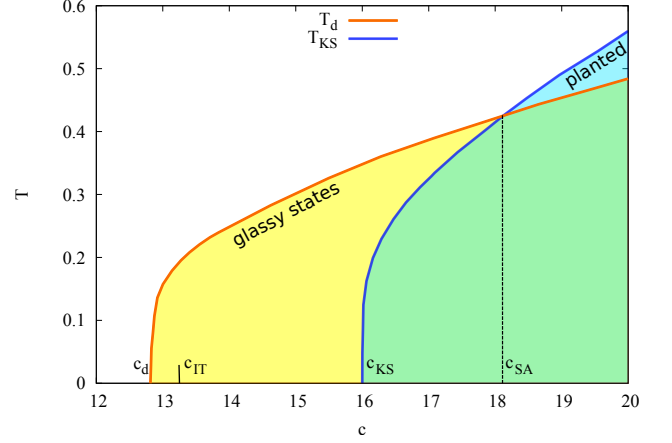

FIG. 3. Phase diagram in the $c - T$ plane for the planted-coloring problem with $q = 5$. $T_d(c)$ is the critical line below which glassy states appear, while below $T_{\rm KS}(c)$ the paramagnetic solution becomes unstable toward the planted solution. The Information-Theoretical threshold c_{IT} corresponds to c_c, while we claim that SA is not able to find the planted solution for $c < c_{SA}$. We can extract c_{SA} as the connectivity at which $T_d(c) = T_{\rm KS}(c)$. Both in the yellow and the green region, an MCMC algorithm gets trapped in glassy states, and the only region in which it can nucleate the planted solution is the blue one.

the planted state and SA jumps into the planted cluster, thus achieving perfect recovery in the $T = 0$ limit (the planted state is very stable at low temperatures and SA cannot leave it when glassy states appear below T_d). Conversely, if $c < c_{\rm SA}$, decreasing T the paramagnet first becomes unstable toward the glassy states at T_d. Remind that for $T > 0$ the Nishimori condition is not satisfied and so glassy states are not only possible but likely to play a central role. And indeed we observe SA falling inside glassy states for $c < c_{\rm SA}$. The barriers to escaping these glassy states are extensive, and SA cannot jump anymore to the planted solution, even if the temperature is decreased below $T_{\rm KS}$: once the dynamics is trapped by a glassy state, it will stay there for a time that is exponential in the size of the system.

Based on the arguments above the conjecture that $c_{\rm SA}$ is the algorithmic threshold for SA is very plausible, on top of being fully supported by numerical observations.

IV. REPLICATED SIMULATED ANNEALING

Having understood that the SA algorithmic threshold $c_{\rm SA}$ is larger than the BP one $c_{\rm KS}$ because of the existence of glassy states than can trap the SA algorithm, a natural question is whether another MC-based algorithm could do better than SA. In this section, we study the Replicated Simulated Annealing (RSA) introduced in Ref. [31] and provide strong numerical evidence that its algorithmic threshold $c_{\rm RSA}$ matches $c_{\rm KS}$, thus being as good as the Bayes optimal BP.

The idea beyond RSA is to simulate y replicas (copies) of

the system coupled together with ferromagnetic coupling (i.e. replicas prefer to be in similar configurations). The corresponding energy function is the following

$$H(\boldsymbol{s}) = \sum_{a=1}^{y} \sum_{(i,j)\in E} \delta_{s_i^a, s_j^a} - \frac{\gamma}{2y} \sum_{a\neq b} \sum_{i=1}^{N} \delta_{s_i^a, s_i^b} \,, \quad (8)$$

where γ controls the intensity of the ferromagnetic coupling between replicas, that we decided to scale by a factor y in order to get a well defined limit when $y \to \infty$.

RSA works exactly as SA: we start at a high enough temperature from random independent configurations for each replica. Each MCS consists in the attempt to change the color of each variable in each replica, according to the energy in Eq. (8). After each MCS the temperature is decreased by dT. In all our numerical experiments we set $\gamma = 1$ (a complete study of the optimal value of γ could be performed, but it is beyond the scope of the present work).

The lowering of the temperature forces the system to look for lower energy configurations, both reducing the number of monochromatic edges in each replica and increasing the similarity between replicas. In this way, RSA should converge towards low energy states of large entropy, as larger clusters of solutions can more easily accommodate $y \gg 1$ replicas. The reason why this should favor the planted cluster of solutions against the glassy states is that beyond c_c the planted cluster has the largest entropy: in fact, glassy states are many, but the entropy of each glassy state is relatively small, such that their total entropy is not larger than the one of the states dominating the thermodynamics. This picture is not limited to the present problem and generalizes to other inference problems where the signal competes with many more random "glassy" configurations.

In Fig. 4 we show the behaviour of RSA for $N = 10^5$ and $c = 17$. The curves for $y = 1$ are the same shown in Fig. 2 reporting the inability of SA in finding the planted solution. Instead, RSA can find the planted solution as long as $y > 1$. Surprising enough, already the smallest value $y = 2$ seems very effective in the task of detecting the planted signal, with 9 over the 10 samples shown in the figure reaching a perfect recovery (although the temperature where detection takes place, that corresponds to a jump in the energy, fluctuates a little bit among samples). For larger values of y, the behavior of RSA is practically sample-independent and seems to follow an almost deterministic law. A possible explanation for this extremely regular path is that the PM state in the replicated model is becoming unstable at the temperature where the jump takes place: if this happens the RSA dynamics has no other option than flowing towards the planted solution, thus eventually detecting the signal in the $T = 0$ limit.

To test the above hypothesis we compute $T_{\text{KS}}(c)$ in the replicated planted model. Moreover, to test the conjecture of the previous section locating the algorithmic threshold c_{RSA} for RSA, we compute also $T_d(c)$ in the replicated random model. The details of these computations are presented in the appendices and the main results are summarized in Fig. 5 for $y = 2, 4, 8$. It is immediately evident that the critical lines for $y > 1$ are located at much higher temperatures than those for
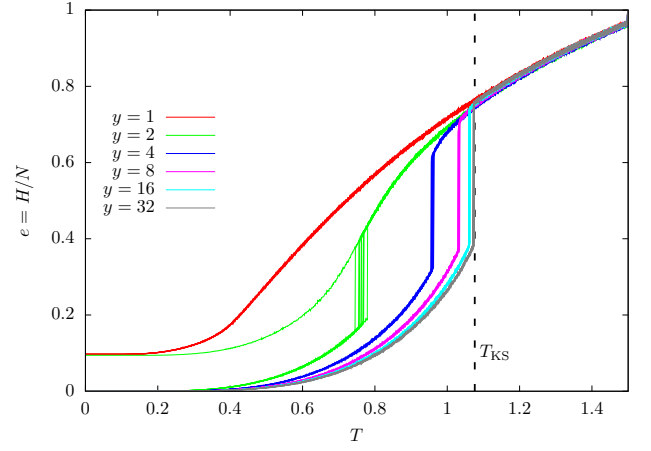

FIG. 4. The energy density as a function of the temperature in the Replicated Simulated Annealing for planted coloring with $q = 5$, $N = 10^5$, $dT = 10^{-7}$ and $c = 17$, changing the number y of replicas. For each value of y, 10 different samples are simulated. While for $y = 1$ SA does not find the planted solution, as long as $y > 1$ RSA finds it. The vertical line represents T_{KS} as found by the linearization of BP equations around the PM solution in the limit $y \to \infty$.
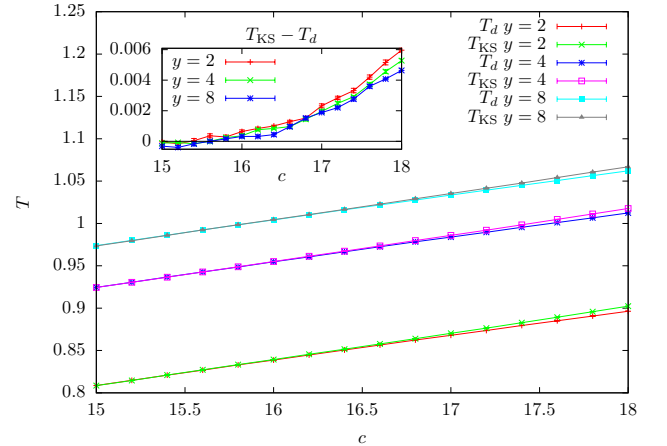

FIG. 5. T_{KS} and T_d as a function of c in the replicated model with $y = 2, 4, 8$. Data are averaged over 10 samples of size $N = 10^6$ (which is large enough to avoid any significant finite size effects). Inset: the difference $T_{\text{KS}} - T_d$ is plotted to help locating the crossing point that defines the algorithmic threshold c_{RSA}, perfectly compatible with the BP threshold $c_{\text{KS}} = 16$.

the original model ($y = 1$) shown in Fig. 3. And these critical temperatures are indeed very close to the temperature where the jump takes place. In particular, the vertical dashed line in Fig. 4 marks the value of T_{KS} in the $y \gg 1$ limit, which turns out to be a very accurate analytical prediction for the jump location (i.e. the signal detection) in RSA run with $y \gg 1$ and a very good approximation also for finite values of y.

Moreover the critical lines $T_{\text{KS}}(c)$ and $T_d(c)$ shown in Fig. 5
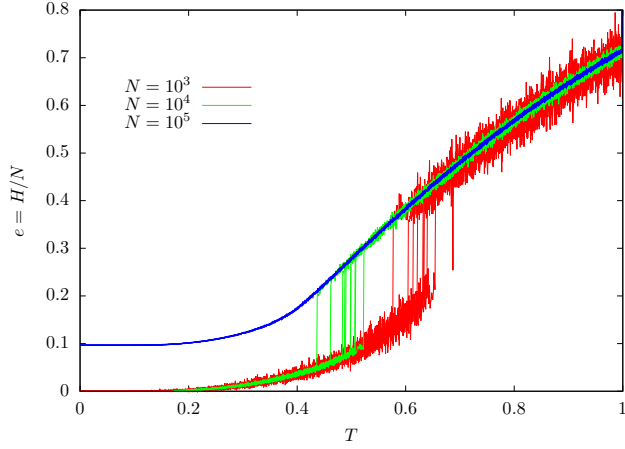
FIG. 6. The energy density as a function of the temperature during Simulated Annealing for planted coloring with $q = 5$, $c = 17$, $dT = 10^{-7}$, and different problem sizes. For each value of N, 10 different samples are simulated. For $N = 10^5$ SA does not find the planted solution, while it always does for smaller sizes.
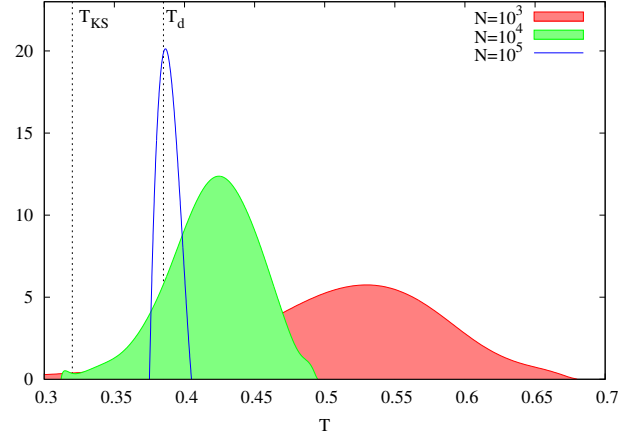
FIG. 7. Frequency of the $\max(T_{\text{KS}}, T_d)$, extracted by running linearized BP equations on given instances of different sizes at $c = 17$ (more details on the computation in the text). Filled curves (for $N = 10^3, 10^4$) are the ones for which $\max(T_{\text{KS}}, T_d) = T_{\text{KS}}$ for each pair of temperatures, while the empty curve indicates that $\max(T_{\text{KS}}, T_d) = T_d$ (and this is always true for $N = 10^5$ when SA is not able to reach the planted state). Vertical dashed lines represent T_{KS} and T_d in the thermodynamic limit.

for $y > 1$ are extremely close. Indeed we report in the inset of Fig. 5 their difference to help the reader in locating their crossing-point that defines the algorithmic threshold c_{RSA} (according to the conjecture put forward in the previous section). Within the statistical uncertainties, the difference $T_{\text{KS}} - T_d$ does not depend on y and becomes null at $c_{\text{RSA}} \simeq 16$, perfectly compatible with the algorithmic threshold $c_{\text{KS}} = 16$ of the Bayes optimal BP. So RSA is as good as the Bayes optimal BP algorithm in detecting the planted signal.

As long as $y > 1$, changing the number of replicas y does not seem to lead to improvements, as we also see directly from the RSA simulation (see Fig. 4). The best choice is thus to use a small y value, even $y = 2$, because the time for running RSA scales as $O(y \cdot N)$.

Let us finally discuss an important difference between the original model and the replicated one. For $y > 1$, in the random model, the transition towards the glassy states becomes continuous. So, while in the original model ($y = 1$) glassy states at T_d are already well-formed states and become dominant with a discontinuous transition, in the replicated model ($y > 1$) the glassy states arise exactly at T_d with a continuous transition. This observation allows us to compute T_d very easily for $y > 1$ via the linearization of the BP equations in the random model, similarly to the computation of T_{KS} in the planted model (more details in the Appendix).

V. FINITE SIZE EFFECTS

Finite size effects are sizeable only close to the critical point. As a typical example, we show in Fig. 6 the behaviour of SA at connectivity $c = 17 < c_{\text{SA}}$ for sizes $N = 10^3, 10^4, 10^5$. At first sight, the behavior of SA seems strange enough: it can find the planted solution for $N \leq 10^4$, even if the PM state should be stable given that we are running SA

with $c < c_{\text{SA}}$. As long as the PM is locally stable, the barrier one has to cross to leave it is $O(N)$ and thus the corresponding timescale should be $O(\exp(N))$, definitely a huge number for the sizes $N = 10^3, 10^4$ shown in Fig. 6.

How can we explain this observation? Can we achieve an analytical prediction, at least qualitatively, of the above finite size effects? A clear and simple answer comes from the study of the critical values T_{KS} and T_d on *samples of a given finite size*. Indeed both critical values can be computed by running BP on a given sample: T_{KS} is found by analyzing the stability of the PM state in the planted model via the linearization of BP equations around the paramagnetic fixed point; T_d is found by running BP on the random model initialized in a planted glassy state at a given temperature and looking at the temperature at which the glassy state becomes unstable towards the paramagnet. More details are given in App. C.

We cannot compare directly T_{KS} and T_d for a given sample, because, as explained exhaustively in the Appendix, we are looking at the ensemble of graphs planted at $T = 0$ to compute T_{KS}, and to the ensemble of graphs planted at $T > 0$ to find T_d. However, to compare the two, we can reasonably assume that the finite size is effectively changing the mean degree c. T_d and T_{KS} being both monotonic functions of c, we can assume that the sample that has the highest T_d will also have the highest T_{KS}. Following the same reasoning, we order values of T_{KS} and T_d and pair them following their order. For each pair, we can now look to the $\max(T_{\text{KS}}, T_d)$, because we conjecture that the highest of the two is the one determining the fate of SA in the search for the planted signal. In Fig. 7 we show the frequency of the $\max(T_{\text{KS}}, T_d)$ for different sizes at $c = 17$. Filled curves (for $N = 10^3, 10^4$) indicate that $\max(T_{\text{KS}}, T_d) = T_{\text{KS}}$ for each pair, while the empty curve (for

$N = 10^5$) indicates that $\max(T_{\text{KS}}, T_d) = T_d$ for each pair. The data are thus in perfect agreement with our claim that SA can find the planted state only if $\max(T_{\text{KS}}, T_d) = T_{\text{KS}}$ and this happens with a high probability for $N = 10^3$ and 10^4, while we never observed it for $N = 10^5$.

The reason beyond this unexpected behavior is due to the very different finite size effects in T_{KS} and T_d: while the latter deviates slightly from the value it takes in the thermodynamic limit, we observe in T_{KS} substantial deviations, very biased towards larger temperatures. This in turn destabilizes the PM state at high temperatures in samples of finite size allowing for signal detection (a very favorable effect of finite size effects!).

VI. DISCUSSION AND PERSPECTIVES

In this work we have analyzed the limits and the performances of simple, but efficient, algorithms based on MCMC (i.e. SA and RSA) in solving the planted coloring problem on sparse random graphs. This problem is a well-understood inference problem (in the teacher-student scenario), showing an information-to-computational gap: the planted solution can in principle be detected for $c > c_{\text{IT}}$, but any known polynomial algorithm can find it only for $c > c_{\text{alg}} \geq c_{\text{KS}} > c_{\text{IT}}$.

For planted problems defined on random graphs and for which the generative model is known — two conditions not always easy to meet — the BP algorithm is Bayes-optimal and can find the planted solution as soon as $c > c_{\text{KS}}$. The BP algorithm works at the same temperature used in the planting process ($T = 0$ in the present model) and for this reason, it satisfies the Nishimori condition that keeps the algorithm away from glassy states: indeed in glassy states, the replica symmetry is spontaneously broken, while the Nishimori condition implies the replica symmetry must be preserved. This lucky coincidence makes BP the optimal one among a large class of message passing algorithms: as shown in Ref. [20] generalizing BP to include RSB effects leads to worst algorithmic performances (probably due to the presence of the uninformative glassy states that we have seen trapping the MCMC algorithms).

Willing to understand the more general situation where algorithms are run with a temperature T different from the planting temperature used in the problem generation ($T > T_p = 0$ in our case), we need to abandon the Nishimori condition. In this situation, the glassy states, which are uninformative from the point of view of the inference problem, can play a key role, influencing negatively the performances of the algorithms. We have shown in the present work that algorithms based on MCMC, like SA, can indeed be trapped by glassy states and fail to detect the signal, even in the region where the FM state has lower energy with respect to the PM state.

Decreasing the temperature in a SA (or RSA) simulation the system faces a critical choice, which is schematically depicted in Fig. 8: when it leaves the PM state it can either flow to the FM state (and thus detect the planted signal) or get trapped in glassy states (which are uncorrelated to the planted signal). In the latter case, the FM state becomes unreachable, even if it has lower energy because the dynamics in the glassy states is
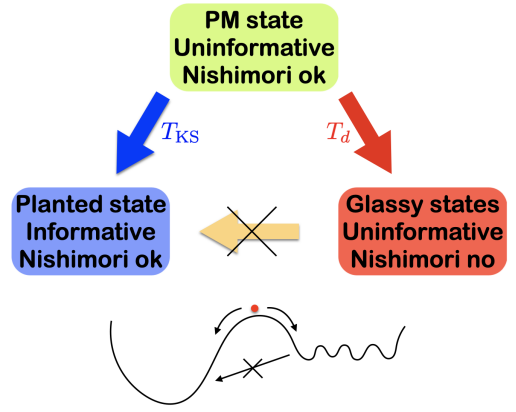

FIG. 8. A schematic picture of the general scenario faced by an annealing process looking for the planted signal. Decreasing the temperature the fate of the annealing algorithm is determined by which critical temperature is met first between T_{KS} and T_d (as explained in the text). The lower diagram should remind us that once the system gets trapped in glassy states it can not reach anymore the planted signal due to the large barriers.

extremely slow and the barriers (both energetic and entropic [32]) make paths leaving the glassy states extremely improbable.

The above observations allow us to make a conjecture on the location of the algorithmic threshold for MCMC-based algorithms. In the literature, there is not a clear indication on which is the correct threshold for these algorithms: while some works claim that MCMC solvers can reach the BP threshold [21], some other references suggest that indeed glassy states do influence negatively the performances of the non-Bayes optimal algorithms [20, 33–36], but do not provide any estimate of the actual algorithmic threshold.

In this work, we conjecture that the fate of an MCMC-based annealing algorithm depends on which critical line it meets before. If the PM spinodal line $T_{\text{KS}}(c)$ is met first, the MCMC evolution spontaneously leaves the PM state heading towards the FM states (the only existing before the appearance of glassy states) and detecting the planted signal. On the contrary, if the dynamical critical line $T_d(c)$ is met first, glassy states become the most attractive for the MCMC evolution and the system gets trapped in glassy states forever. We have seen that the above conjecture predicts with high accuracy the algorithmic thresholds measured in numerical simulations for both SA and RSA. For SA we find that the algorithmic threshold is suboptimal ($c_{\text{SA}} > c_{\text{KS}}$), thus correcting previous claims of optimality [21]. For RSA, instead, we find an optimal algorithmic threshold ($c_{\text{RSA}} \simeq c_{\text{KS}}$).

The optimality of RSA is a very good news: being an MCMC-based algorithm, it is extremely robust and can be used with confidence also in more structured problems, like those coming from realistic applications, where message passing algorithms are likely to fail. The RSA algorithm has been introduced in Ref. [31] as a smart way to sample configurations with large local entropy. To the best of our knowledge, it was never used before in inference problems. We have

adopted it with the following rationale: according to the scenario depicted above, the inference process can be seen as a competition between the single FM planted state and the exponentially many glassy states, the two competitors having similar chances of attracting the dynamics close to the algorithmic threshold. Given that the glassy states are many, we expect the basin of attraction of a *single* glassy state to be much smaller than the one of the FM planted state. For this reason, in the replicated system the y copies are more likely to be attracted by the FM state having a larger basin than by any single glassy state (while being distributed over several glassy states is not convenient because of the coupling among replicas).

In other words, the improvement of RSA over SA is due to the planted state being different from glassy states (and this is often the case in inference problems), but it can not be extended straightforwardly to other types of problems. For example, in a recent paper we compared the performances of SA and RSA for an optimization problem, the largest independent set problem on sparse random graphs [37], and in that case, we did not find a great improvement in the replicated version. However, in hard optimization problems, one is often trying to find a solution that lies inside a glassy state (e.g. in the clustered phase of constraint satisfaction problems [23, 38, 39]). Thus, if the competition is taking place between glassy states of similar internal entropy, the replicated version of the algorithms does not lead to any great improvement (as we saw in the largest independent set problem [37]). So, it is worth stressing that the very good performances of RSA in inference problems [40] is mainly due to the internal entropy of the planted state being larger than the one of each glassy state, and this difference is general, being the planted state the one with the lowest energy and being the basin of attraction larger for lower energy states (see e.g. the discussion in Appendix A of Ref. [32]).

We have studied annealing algorithms as they represent the safer way to test a broad range of temperatures: any good practitioner would start running SA if she knows anything about the model to be optimized, and the output of SA would already provide a rough but useful indication of the critical temperature region to be better explored. However, once one gets a more detailed picture of the energy landscape (as we did here for the planted coloring problem), the natural question is "What is the optimal temperature to run MCMC, eventually the replicated one?". The answer is clear from Figs. 2 and 4: the optimal temperature for MCMC is the one where the jump to the planted state takes place (let us call it T_{jump}). Indeed for $T > T_{jump}$ the PM state is stable, while decreasing the temperature the dynamics becomes slower and the time for doing the jump increases (and even eventually diverges if glassy states with extensive barriers appear).

Noticing that in all cases where the signal can be detected by MCMC-based algorithms we have $T_{jump} > 0$, we can straightforwardly conclude the optimal MCMC temperature is not the Bayes optimal one. A similar conclusion has been achieved recently for MCMC-based algorithms solving the planted clique problem [18, 41] and in the rank-one matrix estimation problem [42]. So, it seems a general statement that, once we consider out of equilibrium processes solving an inference problem, the Bayes optimal temperature plays a marginal or null role and the optimal temperature is typically higher (larger thermal fluctuations enhance the exploration of the configuration space). This is reminiscent of what has been found for Langevin dynamics in another inference problem, the spiked matrix-tensor model: in that case, the Langevin algorithm is sub-optimal with respect to message passing algorithms, due to the presence of glassy states, and it is computationally advantageous for it to mismatch the parameters to find the planted solution of the problem [34–36].

The mismatch between the optimal MCMC temperature and the Bayes-optimal temperature is a good news for all those situations when the generative model is not known and matching the planting temperature would be simply impossible by ignorance. So, having robust algorithms that can work even without knowing exactly the properties of the generative model is extremely useful, and the only requirement is to have an idea of the optimal temperature at which the algorithm should be run (the jump temperature in the present case).

We have attempted to get an analytical estimation of the jump temperature T_{jump}, by checking the stability of the PM state in the replicated model (the details of the computation are in App. B). However, the replicated model has many short loops and the Bethe approximation is no longer exact (one should consider super-variables involving the y replicas and taking q^y values, but we leave this computation for future work). Luckily enough in the large y limit, the interaction among replicas becomes weak (we scaled it by $1/y$ to get a well-defined limit for $y \to \infty$) and the Bethe approximation is again valid. In the $y \gg 1$ limit we can thus obtain an estimate of T_{jump} (it is reported in Fig. 4 with a vertical dashed line) that compares extremely well with numerics even for smaller values of y. This is, in our opinion, a very relevant achievement as predicting the dynamical behavior of out of equilibrium algorithms is typically very challenging and the cases where this can be put in connection to the thermodynamical phase diagrams (as those shown in Figs. 3 and 5) are very few, especially for sparse models.

A final comment is about the nature of the transitions in the replicated system. We already pointed out that the transition from the PM state towards the glassy states changes its nature adding replicas, passing from a discontinuous transition for $y = 1$ to a continuous transition for $y > 1$. This could sound strange and unexpected, however, we believe there is a simple explanation. Let us consider the formulation of the problem where the y coupled variables on a single site form a super-variable taking q^y values. The super-variable is subject to a local field that favors configurations where the y replicas are similar. It is well known that disordered models undergoing a discontinuous Replica-Symmetry Breaking transition (e.g. the famous p-spin model [43]) can change their nature to a continuous transition in presence of an external field. This change in the type of the transition should lead us to rethink the whole concept of the hard phase in inference problems, which is fundamentally grounded on the discontinuous nature of the phase transition.

In conclusion, what is the most robust, eventually optimal,

and easy-to-use algorithm to solve hard inference problems? According to the theory and the numerical data presented here, we believe RSA with few replicas, say $y = O(10)$, is the best option. It is more robust than BP and it can be applied essentially to any problem. Its running time is $O(yN)$, so only a factor y slower than the simpler SA, which is the first option for any practitioner. Maybe parallel tempering (PT) could be better than RSA, but it is more difficult to use and requires a pre-processing operation where the temperature schedule has to be optimized. Moreover, RSA can be restricted to a small temperature range around T_{KS}, thus reducing the running times, while PT requires many more temperatures to be run efficiently.

Let us also comment that RSA can be viewed as a standard annealing process, that is a dynamical process trying to minimize the energy function, but using an effective Hamiltonian or cost function where configurations are *reweighted*. In the original work introducing RSA the motivation was to sample configurations of large local entropy [31]. However the idea can be generalized and we believe it is really worth exploring further the possibility of increasing the chances of reaching the planted signal by modifying the energy landscape. This has been shown useful in random constraint satisfaction problems, where the reweighting of solutions can modify the critical lines and enhance the chance of finding a solution [44, 45]. Also in the tensor PCA problem, the modification of the energy landscape achieved via the use of many replicas can lead to sensible improvement in signal detection [46].

Having found an MCMC-based algorithm (RSA) that has the same algorithmic threshold c_{KS} as the Bayes-optimal BP, some questions naturally arise (that we leave for future works).

Does the RSA optimality hold in general? And if yes, why? Can we get an improved RSA by optimizing the coupling among replicas? We believe the answer to these questions will come from studying numerically more inference problems and analytically the $y \to \infty$ limit.

There exists an even better MCMC-based algorithm that can work below c_{KS}, i.e. breaking the hard phase? The most immediate answer would be negative, given that BP (a Bayes-optimal algorithm) stops working at c_{KS}. Moreover recent works connect the performance of a broad class of algorithms, called *stable algorithms*, to the structure of the configurational space, suggesting that the clustering of solutions in the hard phase is an insurmountable obstacle for stable algorithms [47].

However, we notice that MCMC-based algorithms do not belong to the class of stable algorithms (at least if run for a long enough time), so there is a hope they can work efficiently even in a clustered hard phase, where message-passing algorithms are deem to fail. Moreover smarter MCMC-based algorithm, like RSA or PT, actually work in an extended space (in replicas the former and in temperature the latter), and in this extended space the clustering of the measure may be less drastic. In support of this, we have shown in the present work how the nature of the phase transition changes in the replicated problem.

It is also worth noticing that algorithms working above c_{KS} scale linearly with the problem size N. This is obvious for message-passing algorithms run for a finite number of iterations, and we have shown in this work that the same linear scaling holds for successful MCMC-based algorithms. While running message-passing algorithms for times super-linear in N has no advantage at all, running MCMC-based algorithms for longer times is beneficial. Obviously exponential times are not interesting, but the possibility of having polynomial algorithms based on MCMC working below c_{KS} in the hard phase is to be seriously taken into account.
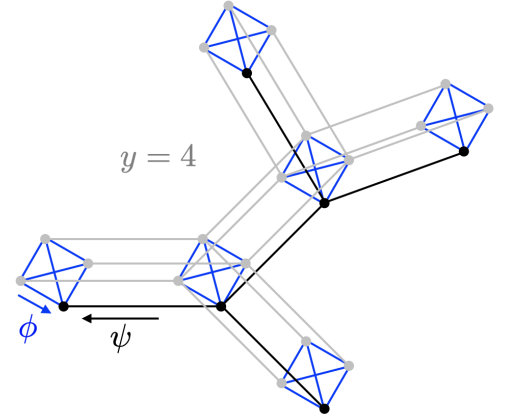
VII. ACKNOWLEDGMENTS

FIG. 9. A schematic view of the replicated graph for $y = 4$. The black graph is the original one, while the gray graphs are the $y - 1$ replicas. Blue ferromagnetic couplings are present among every pair of replicated vertices with the same site index. Replicated BP uses two kind of messages, ψ and ϕ, represented by black and blue arrows.

Appendix A: Replicated Belief Propagation

In this appendix, we derive the Belief Propagation (BP) equations for the coloring problem with y coupled replicas at inverse temperature β, which we call Replicated BP (RBP). For this model, the probability distribution of the variables follows the Gibbs-Boltzmann probability associated with the Hamiltonian in Eq. (8). As for the original model, one would like to estimate the marginal probabilities through the fixed point of some BP-like equations. However, in the replicated case, the graph is no more locally tree-like, because of the presence of small loops between the y coupled replicas and the BP equations are not guaranteed to provide the correct solution to the problem. We will see that in the limit $y \gg 1$, the interaction among replicas is weak enough and the critical temperatures predicted by RBP will describe perfectly the

transitions observed in the Replicated Simulated Annealing (RSA) in the limit of large y.

Making reference to Fig. 9, where every site of the replicated graph is identified by a pair (i, a), with i being the vertex of the original graph, taking values in $[1, N]$, and a is the replica index running in $[1, y]$, the RBP equations for the coloring model with y coupled replicas at inverse temperature β can be written straightforwardly

$$
\begin{aligned}
\psi_s^{(i,a)\to(j,a)} &= \frac{1}{Z^{(i,a)\to(j,a)}} \prod_{b(\neq a)} \left(1 - (1 - e^{\gamma\beta/y})\phi_s^{(i,b)\to(i,a)}\right) \prod_{k\in\partial i\setminus j} \left(1 - (1 - e^{-\beta})\psi_s^{(k,a)\to(i,a)}\right) , \\
\phi_s^{(i,a)\to(i,b)} &= \frac{1}{Z_\phi^{(i,a)\to(i,b)}} \prod_{c(\neq a,b)} \left(1 - (1 - e^{\gamma\beta/y})\phi_s^{(i,c)\to(i,a)}\right) \prod_{k\in\partial i} \left(1 - (1 - e^{-\beta})\psi_s^{(k,a)\to(i,a)}\right) ,
\end{aligned}
\tag{A1}
$$

with Z and Z_ϕ being the normalization factors enforcing $\sum_s \psi_s^{(i,a)\to(j,a)} = \sum_s \phi_s^{(i,a)\to(i,b)} = 1$ for every edge of the replicated graph. We call ψ the cavity messages passed between nodes of the same replica, and ϕ the cavity messages passed between different replicas on the same vertex of the original graph. In the limit $y = 1$, the ϕ messages are irrelevant and the above equations reduce to the standard BP equations for the original (unreplicated) model.

To simplify Eqs. (A1) we assume a Replica Symmetric (RS) ansatz, where cavity messages do not depend on the replica indices, that is $\psi_s^{(i,a)\to(j,a)} = \psi_s^{i\to j}$ and $\phi_s^{(i,a)\to(i,b)} = \phi_s^i$, and then we get

$$
\begin{aligned}
\psi_s^{i\to j} &= \frac{1}{Z^{i\to j}} \left(1 - (1 - e^{\gamma\beta/y})\phi_s^i\right)^{y-1} \prod_{k\in\partial i\setminus j} \left(1 - (1 - e^{-\beta})\psi_s^{k\to i}\right) , \\
\phi_s^i &= \frac{1}{Z_\phi^i} \left(1 - (1 - e^{\gamma\beta/y})\phi_s^i\right)^{y-2} \prod_{k\in\partial i} \left(1 - (1 - e^{-\beta})\psi_s^{k\to i}\right) ,
\end{aligned}
\tag{A2}
$$

We have verified numerically that, fixing $\gamma = 1$, and solving Eqs. (A1) and (A2) on any given graph, we get exactly the same result, both in the high-temperature PM state and in the low-temperature FM state. So, the RS assumption is valid in the present case and we will focus only on Eqs. (A2) hereafter.

Appendix B: Computation of the Kesten-Stigum transition temperature in the replicated model

The Kesten-Stigum transition temperature corresponds to the spinodal point of the PM state, that is to the lowest temperature such that the PM state is locally stable. If the trivial

fixed point of the RBP equations ($\psi_s = \phi_s = \frac{1}{q}$) is correctly describing the PM state of the replicated model, we can compute $T_{KS}(y, \gamma)$ via the study of the local stability of the RBP trivial fixed point with respect to fluctuations leading to the planted FM state.

We perform a linear expansion around the PM trivial solution in the variables

$$
\epsilon_s^{i\to j} \equiv \psi_s^{i\to j} - \frac{1}{q} , \qquad \delta_s^i \equiv \phi_s^i - \frac{1}{q} .
\tag{B1}
$$

At leading order in this expansion, Eqs. (A2) become

$$
\begin{aligned}
\epsilon_s^{i\to j} &= \sum_{k\in\partial i\setminus j}\sum_t \left.\frac{\partial\psi_s^{i\to j}}{\partial\psi_t^{k\to i}}\right|_{\text{PM}} \epsilon_t^{k\to i} + \sum_t \left.\frac{\partial\psi_s^{i\to j}}{\partial\phi_t^i}\right|_{\text{PM}} \delta_t^i = A\sum_{k\in\partial i\setminus j}\left[\sum_t \epsilon_t^{k\to i} - q\epsilon_s^{k\to i}\right] + B(y-1)\left[\sum_t \delta_t^i - q\delta_s^i\right] , \\
\delta_s^i &= \sum_{k\in\partial i}\sum_t \left.\frac{\partial\phi_s^i}{\partial\psi_t^{k\to i}}\right|_{\text{PM}} \epsilon_t^{k\to i} + \sum_t \left.\frac{\partial\phi_s^i}{\partial\phi_t^i}\right|_{\text{PM}} \delta_t^i = A\sum_{k\in\partial i}\left[\sum_t \epsilon_t^{k\to i} - q\epsilon_s^{k\to i}\right] + B(y-2)\left[\sum_t \delta_t^i - q\delta_s^i\right] ,
\end{aligned}
\tag{B2}
$$

where the subscript PM indicates that the function should be computed at the PM trivial solution $\psi_s = \phi_s = \frac{1}{q}$ and

$$
A = \frac{1 - e^{-\beta}}{q(q - 1 + e^{-\beta})} , \qquad B = \frac{1 - e^{\gamma\beta/y}}{q(q - 1 + e^{\gamma\beta/y})} .
$$

Eqs. (B2)) define a $q(N + M) \times q(N + M)$ matrix associated to a given graph of N vertices and M edges. To study the stability of the paramagnetic solution one should look at its maximum eigenvalue $\lambda_{\max}$. When $\lambda_{\max} = 1$ the PM solution looses its local stability. In this way we can locate $T_{\text{ks}}(y, \gamma)$,

that we have shown (fixing $\gamma = 1$) in Fig. 3 for $y = 1$, in Fig. 5 for $y > 1$ and in Fig. 4 with a vertical dashed line in the $y \gg 1$ limit. The latter value is a very good estimate of the jump temperature for large y, where the planted signal can be detected.

At this point, one could think to obtain an analytic estimation for $T_{\text{KS}}(y, \gamma)$ computing the eigenvalues of the linear transformation in Eqs. (B2), assuming that in the large N limit the original graph becomes locally a tree, and so the cavity messages on the right-hand sides are uncorrelated. This is a standard argument, that indeed works for $y = 1$ and pro-

vides the analytical value $T_{\text{KS}}(y = 1) = -1/\log\left[\frac{c-(q-1)^2}{q-1+c}\right]$ of Ref. [21].

In the following, we show that this kind of computation for $y > 1$ does not give the correct result compared to the large N limit estimation of T_{KS} from the single graphs illustrated before. Let us introduce $P_{s,t}(\epsilon^{i \to j})$ and $Q_{s,t}(\delta^{(i,a) \to (i,b)})$ as the probability distributions of the two kind of perturbations (between the nodes and between the replicas) in the direction of color s if the spin i was planted in the color t. In the large N limit, assuming a locally tree-like replicated graph, we can write closed equations for these distributions

$$P_{s,t}(\epsilon) = \sum_{d=0}^{\infty} \frac{d}{c} r_d \int \sum_{\{\sigma_j\},\{\tau_j\},\{\sigma_a\}} \prod_{j=1}^{d-1} d\epsilon_j P_{\sigma_j,\tau_j}(\epsilon_j) p(\tau_j|t) \prod_{a=1}^{y-1} d\delta_a Q_{\sigma_a,t}(\delta_a) \, \delta(\epsilon - f_1(\{\epsilon_j\}, \{\delta_a\})) \,,$$

$$Q_{s,t}(\delta) = \sum_{d=0}^{\infty} r_d \int \sum_{\{\sigma_j\},\{\tau_j\},\{\sigma_a\}} \prod_{j=1}^{d} d\epsilon_j P_{\sigma_j,\tau_j}(\epsilon_j) p(\tau_j|t) \prod_{a=1}^{y-2} d\delta_a Q_{\sigma_a,t}(\delta_a) \, \delta(\delta - f_2(\{\epsilon_j\}, \{\delta_a\})) \,,$$

(B3)

where r_d is the probability of having a vertex of degree d in the original graph (a Poisson distribution on mean c in the model we have studied in this work), $\frac{d}{c} r_d$ is the probability that, taken a link $i \to j$ at random, i has degree d, $p(\tau|t) = \frac{\mathbb{I}(\tau \neq t)}{q-1}$ is the probability that a neighbour of site i was planted in the state τ given that i was planted in the state t, and finally the functions f_1 and f_2 are defined respectively by the right hand

sides of the first and second equation in (B2). Please note that we are allowing δ_a to take different values for each replica. We are thus analyzing the linear expansion of Eq. (A1) rather than Eq. (B2) which is the linear expansion of Eq. (A2).

At this point, given that we are looking for a transition towards a "ferromagnetic" state, we can restrict our attention to the first moments of the above distributions

$$\langle \epsilon_{s,t} \rangle \equiv \int d\epsilon \, P_{s,t}(\epsilon) \, \epsilon = \sum_{d=0}^{\infty} \frac{d}{c} r_d \left\{ \sum_{\{\tau_j\} \neq t} \frac{A}{q-1} \sum_{j=1}^{d-1} \left[\sum_{\sigma_j} \langle \epsilon_{\sigma_j,\tau_j} \rangle - q \langle \epsilon_{s,\tau_j} \rangle \right] + B \sum_{a=1}^{y-1} \left[\sum_{\sigma_a} \langle \delta_{\sigma_a,t} \rangle - q \langle \delta_{s,t} \rangle \right] \right\} \,,$$

$$\langle \delta_{s,t} \rangle \equiv \int d\delta \, Q_{s,t}(\delta) \, \delta = \sum_{d=0}^{\infty} r_d \left\{ \sum_{\{\tau_j\} \neq t} \frac{A}{q-1} \sum_{j=1}^{d} \left[\sum_{\sigma_j} \langle \epsilon_{\sigma_j,\tau_j} \rangle - q \langle \epsilon_{s,\tau_j} \rangle \right] + B \sum_{a=1}^{y-2} \left[\sum_{\sigma_a} \langle \delta_{\sigma_a,t} \rangle - q \langle \delta_{s,t} \rangle \right] \right\} \,.$$

(B4)

Exploiting the symmetry under permutation of the colors, we can identify the only four relevant observables: $\epsilon^e \equiv \langle \epsilon_{s,s} \rangle$, $\epsilon^d \equiv \langle \epsilon_{s,t} \rangle$ with $s \neq t$, $\delta^e \equiv \langle \delta_{s,s} \rangle$ and $\delta^d \equiv \langle \delta_{s,t} \rangle$ with $s \neq t$. Our goal is to look at fluctuations towards the planted state, thus it is convenient to look at the differences $\epsilon^d - \epsilon^e$ and $\delta^d - \delta^e$, that satisfy the following closed equations

$$\epsilon^d - \epsilon^e = (\epsilon^d - \epsilon^e) Ac \frac{q}{q-1} - qB(y-1)(\delta^d - \delta^e) \,,$$

$$\delta^d - \delta^e = (\epsilon^d - \epsilon^e) Ac \frac{q}{q-1} - qB(y-2)(\delta^d - \delta^e) \,.$$

(B5)

The analysis of the stability of the PM solution in the tree-limit of the replicated graph is thus reduced to the computation of

the largest eigenvalues of the matrix

$$\begin{bmatrix} Ac\frac{q}{q-1} & -qB(y-1) \\ Ac\frac{q}{q-1} & -qB(y-2) \end{bmatrix}$$

(B6)

The value of T_{KS} resulting from this computation for $y = 64$ is shown in Fig. 10 with the name 'RPopDyn', as we have checked it coincides with the result found by running the population dynamic algorithm to solve Eqs. (B2). The comparison with the values of T_{KS} obtained from the solution of the RBP equations on several graph realizations for different sizes N reveals that in the large N limit the two computations coincide only for $\gamma = 0$, that is when replicas do not interact. As long as $\gamma > 0$ and the replicas are coupled, the large N limit of the single graph computation does not coincide with the analytical estimation obtained under the assumption of a
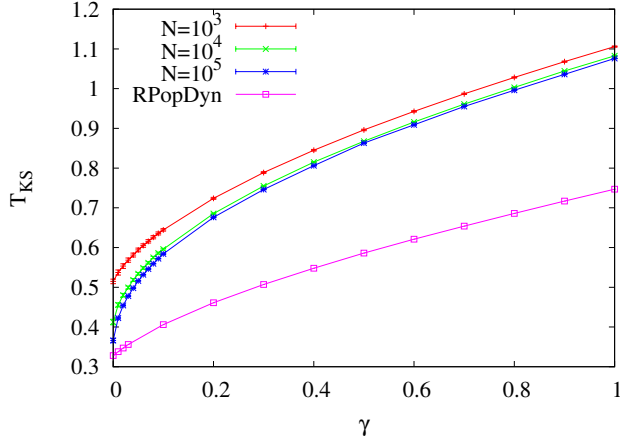
FIG. 10. $T_{\text{KS}}(y = 64, \gamma)$ at $c = 17$ as a function of the coupling γ among replicas. The upper curves have been obtained running RBP on several graph of a given size N, while the lower curve has been obtained running population dynamics to solve Eqs. (B2) or equivalently computing the largest eigenvalue of matrix in (B6). The limit $N \to \infty$ of the former computation coincides with the latter only at $\gamma = 0$. As soon as $\gamma > 0$, the assumption of the replicated graph being locally tree-like is wrong and the latter computation fails, while running RBP on a given graph still provides a reliable value for T_{KS}.

locally tree-like replicated graph.

The reason for the mismatch between the two computations is that for $y > 1$ and $\gamma > 0$ the replicated graph is not locally tree-like and cavity messages are correlated. While running RBP on a given graph these correlations can be taken into account, the population dynamics algorithm, by construction, ignores any correlation among cavity messages. The correlations among messages coming from different replicas are positive, thanks to the ferromagnetic coupling among replicas, and this, in turn, produces a much larger value for T_{KS} than under the assumption of uncorrelated messages. Willing to improve the analytical computation (or the solution via the population dynamics algorithm) one should consider a *super-variable* at every node of the original graph, taking q^y values. In this case, the graph connecting the super-variables is indeed locally tree-like and a population dynamical computation would give the correct estimates for the thermodynamical thresholds (we leave this computation for future work).

Appendix C: Computation of the dynamical transition temperature

The dynamical temperature T_d is the temperature below which glassy states are locally stable (and also attractive for the out of equilibrium dynamics). T_d is not modified by the addition of the planted state, which is roughly orthogonal to the glassy states. For this reason, the usual procedure to identify T_d is as follows: one plants a configuration typical of the Gibbs measure at temperature T, then constructs a graph compatible with it, initializes BP near enough to this planted glassy configuration, and runs the BP equations at temperature T. If the fixed point reached by BP is the PM one then glassy states are locally unstable and thus $T > T_d$, while if the fixed point is correlated with the planted glassy state, then $T < T_d$. In this way we have located T_d for $y = 1$ (see data in Fig. 3). [48]

In principle one could use the same planting technique to locate T_d in the $y > 1$ case, paying attention to plant a state according to the replicated Hamiltonian in Eq. (8). To do this, one really needs to run the BP algorithm with y different replicas, according to Eqs. (A1), because the different replicas will not be planted necessarily with the same color (being the temperature positive $T > 0$ each replica will have a non zero probability to differ from the planted configuration). Computing T_d in this way, we have discovered that the transition becomes a continuous one as soon as $y > 1$ (at least for $\gamma = 1$). This is signaled by the fact that the overlap between the fixed point reached by BP initialized near the planted glassy state at T tends to zero continuously in the limit $T \to T_d^-$. This fact implies that we can compute T_d in the $y > 1$ case just by looking at the instability of the paramagnetic solution through the computation of the maximal eigenvalue of Eqs. (B2) on a given graph *without* the $T = 0$ planted solution (exactly as we did to compute T_{KS} on a graph with the $T = 0$ planted solution). We checked that this computation gives exactly the same result for T_d as the previously described planting procedure. However the study of the linearized equations is much faster, the convergence of BP initialized around the planted glassy state being very slow in the vicinity of T_d. With this linearization method we have computed the T_d values for $y > 1$ shown in Fig. 5.

[1] P. W. Holland, K. B. Laskey, and S. Leinhardt, Social networks **5**, 109 (1983).

[2] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Physical Review Letters **107**, 065701 (2011).

[3] E. Richard and A. Montanari, Advances in Neural Information Processing Systems **27** (2014).

[4] E. Abbe and C. Sandon, arXiv preprint arXiv:1512.09080 (2015).

[5] Y. Deshpande and A. Montanari, Foundations of Computational Mathematics **15**, 1069 (2015).

[6] T. Lesieur, F. Krzakala, and L. Zdeborová, in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (IEEE, 2015), pp. 680–687.

[7] Y. Chen and J. Xu, The Journal of Machine Learning Research **17**, 882 (2016).

[8] S. B. Hopkins, T. Schramm, J. Shi, and D. Steurer, in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing* (2016), pp. 178–191.

[9] A. S. Bandeira, A. Perry, and A. S. Wein, Portugaliae Mathematica **75**, 159 (2018).

[10] B. Barak, S. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, SIAM Journal on Computing **48**, 687 (2019).

[11] E. Gardner and B. Derrida, Journal of Physics A: Mathematical and General **22**, 1983 (1989).

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

[13] M. Mezard and A. Montanari, *Information, physics, and computation* (Oxford University Press, 2009).

[14] L. Zdeborová and F. Krzakala, Advances in Physics **65**, 453 (2016).

[15] J. S. Yedidia, W. T. Freeman, Y. Weiss, et al., Exploring artificial intelligence in the new millennium **8**, 0018 (2003).

[16] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Physical Review E **84**, 066106 (2011).

[17] D. Gamarnik and I. Zadik, arXiv preprint arXiv:1904.07174 (2019).

[18] M. C. Angelini, P. Fachin, and S. de Feo, Journal of Statistical Mechanics: Theory and Experiment **2021**, 113406 (2021).

[19] Z. Chen, E. Mossel, and I. Zadik, arXiv preprint arXiv:2204.01911 (2022).

[20] F. Antenucci, S. Franz, P. Urbani, and L. Zdeborová, Physical Review X **9**, 011020 (2019).

[21] F. Krzakala and L. Zdeborová, Physical review letters **102**, 238701 (2009).

[22] R. Mulet, A. Pagnani, M. Weigt, and R. Zecchina, Physical review letters **89**, 268701 (2002).

[23] L. Zdeborova and F. Krzakala, Physical Review E **76**, 031131 (2007).

[24] Note1, for more complicated scenarios we address the reader to the thorough survey of phases in Ref. [25].

[25] F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Physical Review E **99**, 042109 (2019).

[26] M. Mézard and G. Parisi, The European Physical Journal B-Condensed Matter and Complex Systems **20**, 217 (2001).

[27] H. Nishimori, *Statistical physics of spin glasses and information processing: an introduction*, 111 (Clarendon Press, 2001).

[28] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, science **220**, 671 (1983).

[29] M. Chams, A. Hertz, and D. De Werra, European Journal of Operational Research **32**, 260 (1987).

[30] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, Operations research **39**, 378 (1991).

[31] C. Baldassi, C. Borgs, J. T. Chayes, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina, Proceedings of the National Academy of Sciences **113**, E7655 (2016).

[32] M. Bellitti, F. Ricci-Tersenghi, and A. Scardicchio, Physical

[33] H. Sompolinsky, N. Tishby, and H. S. Seung, Physical Review Letters **65**, 1683 (1990).

[34] S. S. Mannelli, G. Biroli, C. Cammarota, F. Krzakala, P. Urbani, and L. Zdeborová, Physical Review X **10**, 011057 (2020).

[35] S. S. Mannelli, F. Krzakala, P. Urbani, and L. Zdeborova, in *international conference on machine learning* (PMLR, 2019), pp. 4333–4342.

[36] S. Sarao Mannelli, G. Biroli, C. Cammarota, F. Krzakala, and L. Zdeborová, Advances in Neural Information Processing Systems **32** (2019).

[37] M. C. Angelini and F. Ricci-Tersenghi, Physical Review E **100**, 013302 (2019).

[38] F. Krzakała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Proceedings of the National Academy of Sciences **104**, 10318 (2007).

[39] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, Journal of Statistical Mechanics: Theory and Experiment **2008**, P04004 (2008).

[40] Note2, in addition to the results shown in the present work about the planted random graph coloring, we have preliminary results also for the planted bicoloring of random hyper-graphs supporting our statements.

[41] M. C. Angelini, Journal of Statistical Mechanics: Theory and Experiment **2018**, 073404 (2018).

[42] F. Pourkamali and N. Macris, arXiv preprint arXiv:2107.08927 (2021).

[43] A. Crisanti and H.-J. Sommers, Zeitschrift für Physik B Condensed Matter **87**, 341 (1992).

[44] L. Budzynski, F. Ricci-Tersenghi, and G. Semerjian, Journal of Statistical Mechanics: Theory and Experiment **2019**, 023302 (2019).

[45] A. G. Cavaliere, T. Lesieur, and F. Ricci-Tersenghi, Journal of Statistical Mechanics: Theory and Experiment **2021**, 113302 (2021).

[46] G. Biroli, C. Cammarota, and F. Ricci-Tersenghi, Journal of Physics A: Mathematical and Theoretical **53**, 174003 (2020).

[47] D. Gamarnik, Proceedings of the National Academy of Sciences **118**, e2108492118 (2021).

[48] Note3, one could locate T_d also performing a standard 1RSB computation, looking when a 1RSB solution is stable and different from the PM one, however, this procedure is much more expensive than the planting procedure we used.